



UNIVERSITAT DE  
BARCELONA

FACULTAD DE BIOLOGÍA  
DEPARTAMENTO DE GENÉTICA

TESIS DOCTORAL

# **Caracterización del Neoblasto como Modelo de Célula Madre**

Gustavo Rodríguez Esteban  
2015





UNIVERSITAT DE  
BARCELONA

FACULTAD DE BIOLOGÍA  
DEPARTAMENTO DE GENÉTICA  
PROGRAMA DE DOCTORADO EN GENÉTICA  
AÑO 2008

TESIS DOCTORAL

# Caracterización del Neoblasto como Modelo de Célula Madre

Memoria presentada por  
**Gustavo Rodríguez Esteban**

para optar al grado de  
**Doctor por la Universidad de Barcelona**

Tesis doctoral realizada bajo la dirección del Dr. Emili Saló Boix y el  
Dr. Josep Francesc Abril Ferrando en el Departamento de Genética  
de la Facultad de Biología de la Universidad de Barcelona.

Los directores

El autor

**Dr. Emili Saló Boix**

**Dr. Josep F. Abril Ferrando**

**Gustavo Rguez. Esteban**

Barcelona, 2 de noviembre de 2015



# Índice general

<b>Lista de figuras</b>	<b>9</b>
<b>Lista de tablas</b>	<b>11</b>
<b>Introducción</b>	<b>1</b>
<b>Introducción</b>	<b>3</b>
<b>1. Células madre: regeneración, cáncer y envejecimiento</b>	<b>5</b>
1.1. Regeneración en metazoos . . . . .	5
1.1.1. Origen evolutivo de la regeneración . . . . .	6
1.1.2. Tipos de regeneración . . . . .	6
1.1.3. Mecanismos de regeneración . . . . .	7
1.1.4. Orígenes celulares durante la regeneración . . . . .	7
1.1.5. Organismos modelo en el estudio de la regeneración . . . . .	8
1.1.6. Presente y futuro en el estudio de la regeneración . . . . .	10
1.2. Cáncer . . . . .	11
1.3. Envejecimiento . . . . .	13
<b>2. La planaria como organismo modelo para el estudio de las células madre</b>	<b>15</b>
2.1. Filogenia y morfología de la planaria . . . . .	15
2.2. La plasticidad de las planarias . . . . .	16
2.3. La planaria como organismo modelo . . . . .	17
2.4. Regeneración en planaria . . . . .	19
2.4.1. Cicatrización de la herida . . . . .	19
2.4.2. Formación del blastema . . . . .	20
2.4.3. Identidad posicional y especificación de los ejes corporales . . . . .	20
<b>3. El neoblasto</b>	<b>23</b>
3.1. Citología del neoblasto . . . . .	23
3.2. El neoblasto como célula madre . . . . .	24
3.3. Los neoblastos y la línea germinal . . . . .	26
3.3.1. Cuerpos cromatoides . . . . .	28
3.4. Dinámica de los neoblastos . . . . .	29
3.4.1. Respuesta proliferativa de los neoblastos . . . . .	29

## Índice general

3.5. La población de neoblastos es heterogénea . . . . .	30
3.6. Diferenciación de los neoblastos . . . . .	34
3.6.1. Los modelos de neoblasto nativo y especializado en la regeneración . . . . .	34
<b>4. Proteómica</b>	<b>39</b>
4.1. Separación de las proteínas de la muestra . . . . .	42
4.1.1. 2D SDS-PAGE . . . . .	42
4.1.2. <i>Difference gel electrophoresis</i> . . . . .	43
4.2. Espectrometría de masas . . . . .	44
4.3. Análisis computacional . . . . .	45
4.4. Aproximación proteómica . . . . .	46
<b>5. Secuenciación de RNA y análisis de la expresión génica</b>	<b>49</b>
5.1. <i>Microarrays</i> . . . . .	49
5.2. <i>Next-generation sequencing</i> . . . . .	50
5.2.1. 454 GenomeSequencer FLX instrument (Roche Applied Science) . . . . .	52
5.2.2. Illumina (Solexa) Genome Analyzer . . . . .	52
5.2.3. Applied Biosystems ABI SOLiD system . . . . .	54
5.2.4. Estado actual . . . . .	55
5.3. <i>Digital Gene Expression</i> en planaria . . . . .	55
<b>Objetivos</b>	<b>59</b>
<b>Resultados</b>	<b>63</b>
<b>Informe de los directores sobre la participación del doctorando y el índice de impacto de las publicaciones</b>	<b>65</b>
<b>Artículo 1</b>	<b>71</b>
<i>Heterogeneity of planarian stem cells in the S/G2/M phase</i> . . . . .	71
<b>Artículo 2</b>	<b>85</b>
<i>A proteomics approach to decipher the molecular nature of planarian stem cells</i> . . . . .	85
<b>Artículo 3</b>	<b>103</b>
<i>Smed454 dataset: unravelling the transcriptome of Schmidtea mediterranea</i>	103
<b>Artículo 4</b>	<b>125</b>
<i>Digital Gene Expression approach over multiple RNA-Seq data sets to detect neoblast transcriptional changes in Schmidtea mediterranea.</i>	125

<b>Discusión general</b>	<b>153</b>
<b>Discusión</b>	<b>155</b>
La población de neoblastos es heterogénea . . . . .	155
Proteómica y transcriptómica . . . . .	155
Proteómica . . . . .	156
Transcriptómica . . . . .	158
Correlación entre expresión de proteína y mRNA . . . . .	160
Perspectivas . . . . .	163
<i>Single-cell sequencing</i> en planaria . . . . .	164
La planaria como organismo modelo en biomedicina . . . . .	165
La planaria como organismo modelo en cáncer . . . . .	168
<b>Conclusiones</b>	<b>171</b>
<b>Bibliografía</b>	<b>175</b>
<b>Anexos</b>	<b>199</b>
<b>Artículo 5</b>	<b>201</b>
<i>egr-4, a target of EGFR signaling, is required for the formation of the brain     primordia and head regeneration in planarians</i> . . . . .	201
<b>Agradecimientos</b>	<b>219</b>





# Índice de figuras

2.1. Posición filogenética de las planarias en el reino animal . . . . .	16
2.2. Crecimiento y decrecimiento en la planaria <i>S. mediterranea</i> . . . . .	17
2.3. Diferentes estadios del proceso de regeneración de la planaria <i>S. mediterranea</i> . . . . .	18
2.4. Morfología de una planaria . . . . .	21
3.1. Distribución de los neoblastos en la planaria <i>S. mediterranea</i> . . . . .	23
3.2. Imagen de un neoblasto en el microscopio electrónico de transmisión	24
3.3. Experimento clásico de trasplante de neoblastos . . . . .	26
3.4. Modelo de diferenciación en distintos linajes celulares . . . . .	37
4.1. Esquema de la aproximación proteómica empleada en este proyecto	42
5.1. Pirosecuenciación en 454 GS FLX . . . . .	53
5.2. Secuenciación en Illumina Genome Analyzer . . . . .	54
5.3. Protocolo experimental de DGE . . . . .	56



# Índice de tablas

3.1. Algunos genes propios de neoblasto . . . . .	33
---	----

# Introducción



# Introducción

Esta es una historia que, en cierta manera, se ha escrito empezando por el final. El dogma central de la biología molecular establece que la información génica contenida en el DNA se transcribe a RNA y de este se traduce en forma de proteína (Crick, 1970). En nuestro caso recorreremos el camino en sentido inverso: de las proteínas a los genes.

Un sistema biológico debe explicarse en términos de actividad, regulación y modificación de sus proteínas, los principales efectores en último término de cualquier fenotipo observado. El conjunto de las proteínas expresadas en cada momento dependiendo del tipo celular y de los estímulos recibidos se conoce como proteoma y constituye el objeto de estudio de la proteómica. El primero de los dos grandes proyectos que se tratarán en esta tesis aborda la caracterización del proteoma de los neoblastos, las células madre de la planaria, un invertebrado del filo de los platelmintos (literalmente “gusanos planos”) con unas particularidades que lo hacen un modelo de gran valor para el estudio *in vivo* de las células madre.

Sin embargo, las técnicas que brinda la proteómica en la actualidad, si bien han demostrado su utilidad y su potencial, no son aún lo bastante resolutivas como para dar respuesta a muchas de las preguntas que nos plantea la compleja biología del neoblasto. Este término engloba a una población de células que se está revelando mucho más heterogénea de lo que se había presumido hasta hace pocos años. Por ello, ha sido necesario descender al nivel inmediatamente anterior a la traducción en la cadena de eventos moleculares y sacar partido de la ventaja tecnológica que proporciona actualmente el estudio del RNA con la nueva generación de plataformas secuenciadoras para intentar dilucidar los misterios de la biología de estas células madre.

El estudio de cualquier población celular requiere identificarla no solo a través de una descripción morfológica y funcional de las células que la componen sino que es imprescindible encontrar marcadores que permitan aislarla y, llegado el caso, subdividirla en grupos de células con características aún más concretas. Este es el caso de la población de neoblastos, dada la heterogeneidad celular por la que necesariamente está compuesta.

Para poder llevar a cabo estos análisis *high-throughput* ha sido necesario desarrollar técnicas experimentales y computacionales. A nivel experimental cabe destacar la puesta a punto del protocolo de *cell sorting* para el aislamiento de los neoblastos. En el campo computacional, se han desarrollado *pipelines* de ensamblado transcriptómico y mapeado sobre las secuencias de referencia, y se han implementado los métodos estadísticos necesarios para determinar aquellos genes diferencialmente expresados en neoblastos. Estos genes han sido validados en el laboratorio para corroborar su

## *Introducción*

expresión en neoblastos y observar el fenotipo tras su inhibición. Algunos, además, han sido caracterizados con mayor detalle (*Smed-SmB*, *Smed-meis-like*, *Smed-nf-YA/YB-2/YC*), lo que ha servido para profundizar en el conocimiento de la biología de los neoblastos.

# 1 Células madre: regeneración, cáncer y envejecimiento

*This brief condescension to evil  
finally destroyed the balance of my soul.*

“Strange Case of Dr Jekyll and Mr Hyde” (1886)  
Robert Louis Stevenson

Regeneración y cáncer pueden considerarse las dos caras de una misma moneda. La hiperproliferación celular es útil para llevar a cabo la reparación de lesiones en muchos animales y, en algunos, incluso la regeneración de partes perdidas. Sin embargo, esta hiperproliferación celular es al mismo tiempo la principal condición patológica responsable del cáncer en humanos. Estudios recientes atribuyen un papel de ciertas proteínas generalmente asociadas a cáncer también en la regeneración (Oviedo *et al.*, 2008; Pearson & Alvarado, 2009). Descubrir porqué a pesar de poseer un mismo acervo génico, organismos como planarias, *Drosophila*, nemátodos y muchas otras especies, parecen no presentar la misma frecuencia o evidencia de cáncer (Pearson & Sánchez Alvarado, 2008), es de gran importancia para entender las causas, orígenes y mecanismos de esta enfermedad.

## 1.1. Regeneración en metazoos

La regeneración en el reino animal es un fenómeno más extendido de lo que puede parecer. Comprender la diversidad de los mecanismos regenerativos en la extensa filogenia de los metazoos representa uno de los retos fundamentales en biología. Descifrar las diferentes estrategias de regeneración supone un desafío a la vez que representa una oportunidad para la todavía joven biomedicina.



### 1.1.1. Origen evolutivo de la regeneración

Desde un punto de vista evolutivo, la capacidad de regenerar es adaptable y mantenida por selección, lo cual implica que si el daño o la pérdida de tejido en una cierta especie es un evento frecuente, y ecológicamente relevante, esa especie es más propensa a regenerar. Aún así, no hay que asumir la regeneración como necesariamente adaptativa puesto que no hay evidencia de su relevancia ecológica en muchas especies altamente regenerativas y la pérdida de esta capacidad ha sido frecuente en la filogenia del reino animal (Bely & Nyberg, 2010). Además, la capacidad de regenerar entre diferentes tejidos de un mismo organismo puede variar drásticamente. Mientras que algunos animales como cnidarios, platelmintos o equinodermos pueden regenerar un nuevo individuo a partir de tan solo un pequeño fragmento de su cuerpo, muchas otras especies únicamente pueden regenerar partes de tejidos específicos y fallan a la hora de regenerar órganos vitales, como en el caso del sistema nervioso central en mamíferos (Franco *et al.*, 2013).

### 1.1.2. Tipos de regeneración

Hay diferentes variantes del proceso de regeneración, aunque no existe todavía consenso sobre si todas ellas pueden ser consideradas como tal (Franco *et al.*, 2013):

- **Regeneración fisiológica**

Se refiere a la sustitución natural de partes corporales desechadas o desgastadas, como glóbulos rojos, la mucosa intestinal de mamíferos o la cornamenta en los cérvidos.

- **Hipertrofia**

Se conoce como la capacidad que poseen algunos órganos internos para incrementar su masa ante una pérdida de tejido y el aumento de la demanda funcional. No se produce, sin embargo, una recuperación de la anatomía original del órgano. Un ejemplo de esto es el hígado, en el cual las células mitóticamente activas no se encuentran cerca de la superficie de la herida sino repartidas por el tejido restante.

- **Regeneración reparativa**

Implica el reemplazo de una parte dañada o perdida del organismo, que puede ir desde una única célula hasta partes completas del cuerpo, como en la regeneración de las extremidades en tritones, o la reconstrucción de un organismo entero en planaria a partir de un fragmento centesimal de la masa original.

- **Reproducción asexual**

Se caracteriza por la subdivisión natural de un único organismo en una o más partes y la reorganización de cada una de las partes en un individuo nuevo y completo.

### 1.1.3. Mecanismos de regeneración

Tradicionalmente se han distinguido dos mecanismos alternativos por los cuales puede llevarse a cabo la regeneración: la morfálaxis, que se basa en una reutilización y redistribución de células ya existentes sin necesidad de proliferación celular activa, y la epimorfosis, que sí requiere proliferación y la formación de un blastema típico (Tanaka & Reddien, 2011).

- **Epimorfosis**

El concepto de regeneración epimórfica fue introducido por Thomas Hunt Morgan en 1901 (Morgan, 1901). Se caracteriza por la formación de un blastema a partir de la multiplicación de una masa de células indiferenciadas que surge a raíz de la interacción entre el epitelio y el mesénquima, que contiene y expresa información morfogenética intrínseca

- **Morfálaxis**

Al contrario de lo que ocurre en la epimorfosis, en la morfálaxis no hay un grupo local definido de células indiferenciadas que proliferan. En lugar de eso, los tejidos perdidos surgen a partir de una extensiva reorganización de las células preexistentes, que en la fase inicial ocurre sin proliferación detectable, donde las células entran en transdiferenciación

Sin embargo, como se verá, estas dos estrategias no tienen por qué ser excluyentes.

### 1.1.4. Orígenes celulares durante la regeneración

La perspectiva desde la que observamos la regeneración está cambiando al ritmo que marcan los nuevos hallazgos sobre la plasticidad de las células madre y la capacidad de desdiferenciación de células que hasta ahora se presumían comprometidas de forma irreversible. Las estrategias regenerativas incluyen el reordenamiento de tejidos preexistentes, el uso de células madre somáticas adultas y la desdiferenciación. Esta última implica el proceso por el cual células diferenciadas pierden su identidad tisular, volviéndose indiferenciadas, pudiendo así rediferenciarse en células de su mismo tipo original, o bien, rediferenciarse en células de un linaje diferente, proceso conocido como transdiferenciación.

Los precursores celulares de los tejidos regenerantes son también objeto de una intensa investigación y pueden incluir (Franco *et al.*, 2013):

1. Desdiferenciación de células maduras/adultas que participarán en el proceso de regeneración. En algunos casos estas células pueden dar lugar a células de fenotipo diferente a aquel al cual pertenecían (transdiferenciación).
2. Proliferación del resto de células parenquimáticas sin diferenciación.
3. Proliferación de células madre (progenitoras) presentes en el tejido lesionado.

4. Influjo de células provenientes de fuera del tejido dañado.

El origen de las células que participan en el proceso de regeneración ha sido siempre una de las cuestiones más esquivas en el estudio de la regeneración. Por ejemplo, las células del blastema responsables de la regeneración de las extremidades en anfibios tienen un origen local cercano al plano de amputación. En planaria, en cambio, las células migradoras que participan en el blastema pueden tener un origen más distante. Durante la regeneración de la cabeza en *Hydra* no hay blastema como centro de proliferación celular. En su lugar, la regeneración deriva de la reorganización directa de células preexistentes en un proceso morfoláctico típico (Franco *et al.*, 2013).

División celular, activación de células madre, desdiferenciación y transdiferenciación son procesos que contribuyen a la regeneración en diferentes contextos. La extensa variedad regenerativa en los metazoos no puede explicarse a través de una única estrategia. Con el árbol filogenético del reino Metazoa en mente, parece lógico pensar que la facultad de regenerar es un atributo ancestral que se ha ido perdiendo a lo largo del curso de la evolución, a medida que los organismos aumentaban en complejidad. Asimismo, el paralelismo existente entre regeneración y desarrollo, especialmente postembrionario, sugiere que la primera es probablemente un producto derivado del segundo.

### 1.1.5. Organismos modelo en el estudio de la regeneración

Ciertos invertebrados como algunas planarias (filo Platyhelminthes o Plathelminthes), *Hydra* (filo Cnidaria) o esponjas (filo Porifera) son capaces de regenerar un individuo completo a partir de un pequeño fragmento de su cuerpo. Esos organismos, junto con algún modelo de vertebrado como el pez cebra (*Danio rerio*), están aportando una masa crítica de conocimiento acerca de los fundamentos celulares y moleculares que subyacen en los diferentes tipos de regeneración. En general, todas las especies poseen, en mayor o menor medida, la capacidad necesaria de regenerar tejidos u órganos dañados. Sin embargo, los métodos empleados para ello son variados entre especies o incluso entre tejidos. El descubrimiento de los mecanismos de reparación en estos organismos promete ayudar a revelar también esos mismos secretos en mamíferos, abriendo así las puertas a la medicina regenerativa (Li *et al.*, 2015).

#### Modelos invertebrados

Metazoos basales como las esponjas regeneran utilizando células madre pluripotentes, llamadas arqueocitos. *Hydra*, en cambio, cuenta con tres tipos diferentes de células madre restringidas a linajes concretos para regenerar todos los tejidos: las células epiteliales endodérmicas y ectodérmicas son células unipotentes que se renuevan continuamente, mientras que las células intersticiales, situadas entre las células epiteliales del ectodermo, son células madre multipotentes capaces de diferenciarse en neuronas, células glandulares, nematocistos y gametos. El tipo de regeneración

que presenta *Hydra* es de tipo morfolático, es decir, por morfogénesis tisular a partir de células preexistentes sin necesidad de proliferación celular. Por ejemplo, tras la amputación de la región anterior, las células de la columna gástrica entran en determinación y diferenciación para reemplazar la parte perdida. Aunque la regeneración *Hydra* puede ocurrir sin una división celular significativa, sí que se producen nuevas células de manera continua para el mantenimiento del pólipo (Tanaka & Reddien, 2011). También se ha descrito que *Hydra* puede regenerar mediante epimorfosis, tras la inducción de apoptosis en la estructura dañada, dependiendo de las condiciones (Chera *et al.*, 2009).

### Modelos vertebrados

El pez cebra constituye uno de los principales organismos modelo en los estudios de regeneración en vertebrados. Si bien sus capacidades regenerativas no son tan amplias como las de los modelos invertebrados, son capaces de regenerar aletas, corazón y sistema nervioso central. La formación del blastema tras la amputación de parte de la aleta caudal o del corazón implica el reclutamiento de células progenitoras por desdiferenciación y proliferación de células próximas a la región lesionada, que se convierten en células progenitoras de su propio linaje, aunque no pueden descartarse fenómenos de transdiferenciación.

Otros vertebrados como los anfibios urodelos (tritones, salamandras y axolotes) pueden regenerar extremidades, cola, mandíbula inferior, cerebro, médula espinal y el ventrículo del corazón. El blastema formado en la regeneración de sus extremidades está compuesto por una masa de células mesenquimáticas indiferenciadas cuyo origen no está claro pero parece que tanto células multinucleadas desdiferenciadas que dan lugar a una progenie mononucleada, como células progenitoras locales, participan en la regeneración del miembro perdido. La cuestión sobre si las células del blastema son pluripotentes, multipotentes o restringidas a un linaje continúa abierta. Lo que sí parece más o menos claro es que no se da una transdiferenciación entre tipos celulares y cada célula permanece fiel a su identidad durante todo el proceso de regeneración. Por contra, en la regeneración de la lente ocular en ciertos anfibios sí es bien conocido que se produce una transdiferenciación de células epiteliales pigmentarias, que se desdiferencian en células precursoras para posteriormente dar lugar a los linajes necesarios para restaurar la lente. La desdiferenciación parece ser un requisito crucial en el proceso de transdiferenciación (Tsonis *et al.*, 2004).

Aparentemente, el blastema en vertebrados es un mosaico de células de plasticidad restringida con origen en diferentes tejidos, cada uno de los cuales aporta un conjunto distinto de progenitores celulares. Muchos vertebrados cuentan con una colección de progenitores restringidos a un linaje en diferentes tejidos. De este modo, no es necesario que las células reviertan a un estado pluripotente sino que retienen una memoria de su origen tisular (Li *et al.*, 2015).

Los mamíferos conservamos solo una cierta capacidad para renovar tejidos (como la sangre y muchos epitelios) y órganos que presentan en mayor o menor medida un recambio celular continuo, posible gracias a células multipotentes localizadas en el

propio tejido. En el caso del hígado, se da una regeneración compensatoria después de una amputación que parece involucrar una desdiferenciación y una proliferación limitadas de las células remanentes (Miyaoka & Miyajima, 2013). En el páncreas ocurre una expansión de las células  $\beta$  en respuesta a una mayor demanda de insulina, durante la gestación, por ejemplo. Su capacidad regenerativa tras una lesión parece darse fundamentalmente por replicación de las propias células  $\beta$  aunque experimentos recientes no permiten descartar una diferenciación a partir de progenitores pancreáticos (Pan *et al.*, 2013). En humanos, se calcula que el 45% de los cardiomiocitos se renuevan a un ritmo lento, mientras que el 55% restante perduran toda la vida (Bergmann *et al.*, 2009). Un estudio reciente reveló que la fuente de los nuevos cardiomiocitos en mamíferos son los ya preexistentes (Ali *et al.*, 2014; Senyo *et al.*, 2012). Durante la recuperación tras un infarto de miocardio su tasa de renovación de aproximadamente el 0,76% anual se incrementa cinco veces. Los cardiomiocitos adultos retienen una cierta capacidad para reentrar en el ciclo celular mientras que las células progenitoras cardíacas juegan un papel muy limitado tanto en homeostasis como tras una lesión.

Por tanto, los platelmintos no son únicos en poseer células madre somáticas pluripotentes, ni son los únicos que llevan a cabo un recambio continuo. Los sistemas de células madre a lo largo del reino animal forman un continuo en el que algunas especies de planarias representan el extremo dinámico. *Hydra* se sitúa cerca, mostrando un renovación continua del cuerpo, llevada a cabo por linajes celulares uni y multipotentes independientes. Los vertebrados, con un recambio celular extremadamente específico de tejido se sitúan en el medio, mientras que *C. elegans* carece completamente de células madre somáticas adultas y recambio somático, ocupando el extremo estático del segmento.

Basándose en estas observaciones se ha propuesto que las células madre somáticas pluripotentes son el estado ancestral de los sistemas de células madre animales (Agata *et al.*, 2006; Blackstone & Jasker, 2003; Extavour, 2007). Las células madre somáticas específicas de tejido podrían representar una adaptación secundaria al incremento de tamaño y la especialización funcional de las partes del organismo. Esta visión predeciría que los componentes moleculares de la pluripotencia fueron compartidos inicialmente entre las células somáticas pluripotentes y la línea germinal, antes de restringirse más y más a la línea germinal. La expresión de *piwi* y otros genes de la línea germinal en células madre pluri y multipotentes de invertebrados pero generalmente no en las células madre de vertebrados restringidas a un linaje, encajaría con esta concepción.

### 1.1.6. Presente y futuro en el estudio de la regeneración

Las aproximaciones “ómicas” (genómica, transcriptómica, proteómica) son de una importancia vital en el estudio de cualquier fenómeno biológico, como la regeneración, debido a que permite una búsqueda global de potenciales biomoléculas y las vías en las que participan. A pesar del extenso conocimiento ya disponible sobre los procesos fisiológicos y celulares relacionados con la regeneración, a nivel mo-

lecular persisten grandes lagunas, que no son subsanables mediante los estudios experimentales clásicos dirigidos a una sola o unas pocas moléculas. Habitualmente, estos dependen del uso de anticuerpos específicos, concretamente para ensayos de *Western blot*, inmunofluorescencia o ELISA. Es más, teniendo en cuenta el conocimiento presente, es predecible que un número significativo de eventos moleculares sean postranscripcionales (como eventos proteolíticos y un amplio rango de modificaciones postranscripcionales) y, consecuentemente, solo puedan ser detectados y potencialmente caracterizados utilizando herramientas proteómicas de elevada resolución, con una identificación precisa de aminoácidos. Desafortunadamente, en la actualidad, la cantidad de datos genómicos en la mayoría de las especies con más capacidades regenerativas son muy limitados, conduciendo a una falta de herramientas para la identificación de biomoléculas, concretamente proteínas, detectadas experimentalmente. Por tanto, es necesario un esfuerzo urgente para superar esta limitación (Franco *et al.*, 2013).

## 1.2. Cáncer

Como se ha visto, los metazoos basales mantienen típicamente un gran número de células madre pluripotentes que son capaces de diferenciarse en todos los tipos de células del organismo. Esto probablemente no es posible en animales más complejos debido al peligro de desarrollar cáncer (Pearson & Sánchez Alvarado, 2008).

Una solución consiste en restringir la proliferación usando un tipo de desarrollo más regulado, como por ejemplo, un sistema de telómeros que desactive la célula tras un número limitado de divisiones. El mismo mecanismo que conduce al envejecimiento celular (López-Otín *et al.*, 2013). Sin embargo, las estrategias para regular el desarrollo tisular y celular con el fin de prevenir la formación de tumores no se limitan a los metazoos superiores; incluso las planarias pueden desarrollar tumores y, por tanto, han desarrollado mecanismos para evitarlo (Oviedo *et al.*, 2008; Pearson & Sánchez Alvarado, 2008).

La proliferación celular necesita ser regulada para asegurar la supervivencia de los organismos multicelulares y, por tanto, los mecanismos para regular la supresión tumoral es probable que tengan un origen evolutivo ancestral. *S. mediterranea* tiene un único homólogo de p53 denominado *Smed-p53*, cuya expresión está restringida a la progenie de las células madre (Pearson & Alvarado, 2009). RNAi de p53 resulta en un incremento en el número de células madre y la proliferación a expensas de la diferenciación celular, de acuerdo con una función de supresión tumoral de *Smed-p53* en planarias. Además, a medida que el fenotipo de pérdida de función de *Smed-p53* progresa se observa una reducción en la población de células madre, lo que sugiere que esta molécula puede también actuar de forma similar a p63 en vertebrados. *Smed-p53* es el primer miembro de la familia p53 que ha mostrado tener un papel en el control de la proliferación de células madre y especificación de linaje. Conjuntamente, estos estudios respaldan la conclusión de que un miembro ancestral de la familia p53 ya funcionaba en el control de la proliferación y en la

biología de las células madre y que, por tanto, estas funciones no aparecieron por primera vez en los vertebrados.

*PTEN* es uno de los genes comúnmente mutados en cáncer humano. El uso de RNAi para inhibir este supresor tumoral evolutivamente conservado en *S. mediterranea*, *Smed-PTEN* (Oviedo *et al.*, 2008), conduce a una proliferación anormal de neoblastos, una desorganización del tejido, alteración de las interacciones epitelio-mesénquima y la presencia de células anormales que invaden tejidos distantes para formar sobrecrecimientos (*outgrowths*) agresivos y eventualmente letales a través de la sobreactivación de la vía de señalización de TOR (González-Estévez *et al.*, 2012b; Peiris *et al.*, 2012). Este fenotipo es consistente con los neoblastomas inducidos químicamente en planarias (Best & Morita, 1982). La regeneración puede verse inhibida en animales *Smed-PTEN(RNAi)*, indicando que bajo ciertas condiciones *PTEN* también se requiere para la regeneración. Un hallazgo importante es que la función de supresor tumoral de *PTEN* es similar en mamíferos pero está ausente en otros invertebrados (Pearson & Sánchez Alvarado, 2008).

Es interesante hacer notar que P53 en planarias y otras proteínas supresoras de tumores en vertebrados como PTEN (Oviedo *et al.*, 2008) están involucradas en la habilidad de estos organismos para permitir hiperproliferar de manera temporal a sus células madre después de una lesión. Esto sugiere la posibilidad de que este fuera el papel ancestral de estos genes y que se hayan seleccionado evolutivamente en vertebrados como genes supresores de tumores por su control sobre la proliferación celular.

Dos estudios investigando la función del homólogo en planaria del supresor tumoral p53 y el componente remodelador de la cromatina CHD4 ya han mostrado la utilidad de los marcadores de la progenie de neoblastos (Pearson & Alvarado, 2009; Scimone *et al.*, 2010). El RNAi de estos genes mostró que eran necesarios en los neoblastos para facilitar la regeneración y la homeostasis de los tejidos. Además, observando la dinámica de los neoblastos en proliferación y su descendencia por separado mostró que se producía una menor progenie seguida de una desaparición posterior de neoblastos. Así, estos genes podrían ser asignados con funciones específicas en la producción correcta de la progenie de los neoblastos (diferenciación temprana) con un efecto tardío en el mantenimiento de los neoblastos.

Entender cómo se controlan la interacción y el balance entre mantenimiento de las células madre y diferenciación y cómo puede ser manipulado es un área de investigación clave en biología de la regeneración y las planarias proporcionan un sistema ideal en el cual estudiarlo. Marcadores que identifiquen los precursores de linajes específicos son necesarios para entender cómo los neoblastos reemplazan los diferentes tipos celulares especializados presentes en planarias. Por ejemplo, *Smed-nanos* marca células madre de la línea germinal en planarias y se requiere para la formación de la línea germinal (Wang *et al.*, 2007). Establecer relaciones concluyentes entre la progenie de neoblastos y linajes diferenciados específicos requerirá de técnicas de marcado *in vivo* y probablemente dependerá del desarrollo de una técnica de transgénesis robusta.

A medida que se acumulan más datos sobre la genómica del cáncer, las funciones

de los genes implicados en otras especies menos estudiadas pueden proporcionar información no solo acerca de cómo estos genes han evolucionado sino también arrojar luz sobre los fundamentos de la enfermedad misma.

## 1.3. Envejecimiento

En la evolución de los bilaterados superiores, los animales más longevos se cuentan también entre aquellos más simples, los llamados metazoos basales, un grupo que incluye a esponjas, corales, medusas, hydras y anémonas marinas. La clave de la longevidad de estos animales es su gran número de células madre, que no solo les confieren sus remarcables capacidades regenerativas y una gran versatilidad para crecer, decrecer según sea necesario (Pearson & Sánchez Alvarado, 2008; Rando, 2006; Rink, 2012; Solana, 2013; Tanaka & Reddien, 2011) sino que, además, les permite rejuvenecer su propio organismo e incluso ser potencialmente inmortales en algunos casos (Rando, 2006).

Durante la evolución hacia formas de vida más complejas, estas habilidades se han ido reduciendo o perdiendo, aparentemente en un compromiso por producir estructuras más sofisticadas y de estrategias que protejan su ciclo celular del desarrollo de tumores (Pearson & Sánchez Alvarado, 2008; Rando, 2006). Sin embargo, no hay una relación directa entre un incremento en la complejidad corporal y una reducción en la longevidad. Por ejemplo, entre los bilaterados, algunos vertebrados como las tortugas pueden vivir más de 200 años, mientras que el nemátodo *Caenorhabditis elegans*, con un diseño poco más complejo que el de una planaria, carece de células madre somáticas y vive solo dos semanas.

Una característica asociada al envejecimiento en la evolución es que este va acompañado de una reducción gradual en la abundancia de células madre pluripotentes. Puesto que estas células mantienen el potencial de dar lugar a cualquier tipo celular del organismo, su reducción o ausencia en animales más complejos significa que la mayoría de sus células diferenciadas envejecerán y morirán llevando, a la postre, a la muerte del animal, impidiendo así cualquier posibilidad de inmortalidad.

El envejecimiento en planarias es un fenómeno todavía no bien comprendido. En el laboratorio, su longevidad ha sido variable, probablemente debido a las diferencias entre especies y en las condiciones de cultivo. Individuos de *Dugesia tigrina* o *Schmidtea polycrhoa* siguen produciendo huevos años después de haber sido recogidos de su medio natural sin mostrar signos de envejecimiento (Mouton *et al.*, 2011). Cepas asexuadas de *Schmidtea mediterranea* o *Dugesia japonica*, incluso líneas clonales, han sido mantenidas en el laboratorio durante décadas, hasta el punto de ser consideradas inmortales (Baguña & Romero, 1981; Ishizuka *et al.*, 2007; Newmark & Sánchez Alvarado, 2002; Rink, 2012; Saló, 2006; Tan *et al.*, 2012).

Así, las planarias pueden emplear tres mecanismos para evitar el envejecimiento: reproducción asexual por fisión, decrecimiento controlado durante el ayuno y regeneración a partir de fragmentos de su cuerpo. Todos parecen conllevar un rejuvenecimiento del animal. Por ello, las cepas asexuadas parecen estar exentas de



## 1 Células madre: regeneración, cáncer y envejecimiento

envejecimiento fisiológico con poblaciones de neoblastos con los telómeros siempre largos (Mouton *et al.*, 2011; Pearson & Sánchez Alvarado, 2008; Tan *et al.*, 2012).

## 2 La planaria como organismo modelo para el estudio de las células madre

*Man fears time,  
but time fears the pyramids.*

Arab proverb

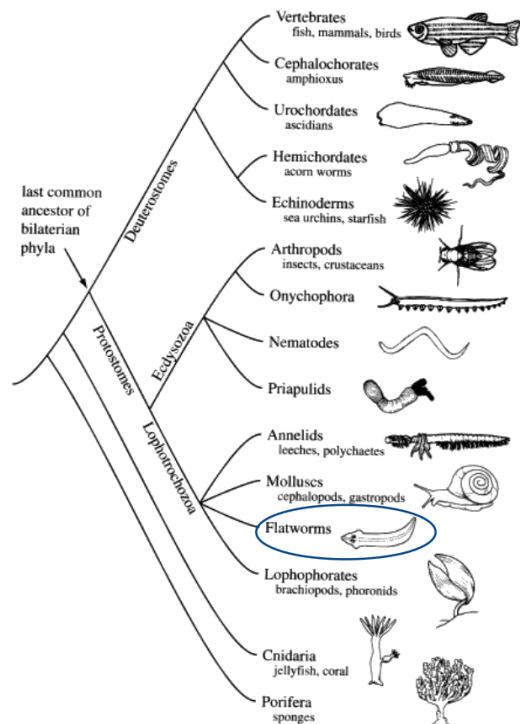
Haciendo una analogía similar a la del proverbio árabe podríamos decir que, además de a las pirámides, el tiempo debería temer también a las planarias, pues estas son inmortales o, al menos, guardan el secreto de la eterna juventud.

El fin evolutivo de cualquier especie es, presumiblemente, la perpetuación de la misma. Esto puede conseguirse alargando la vida de cada organismo individualmente, o bien, maximizando el número de generaciones. La primera estrategia consistió en alargar la vida del organismo manteniendo una población de células madre que pudieran reemplazar todos los tipos celulares para regenerar y rejuvenecer al individuo. A medida que avanzamos en la evolución hacia organismos más complejos, incluidos los humanos, se produce una reducción la longevidad en beneficio de la segunda estrategia y la reproducción sexual (Petralia *et al.*, 2014).

### 2.1. Filogenia y morfología de la planaria

“Planaria” es un término coloquial que generalmente se refiere a miembros de vida libre del orden Tricladida. Se conocen cientos de especies de planarias, marinas, de agua dulce e incluso terrestres. Las planarias son protóstomos miembros del filo de los platelmintos (Fig. 2.1). Como todos los Bilateria son animales triploblásticos, cuyo embrión cuenta con endodermo, mesodermo y ectodermo y simetría bilateral básica. De entre ellos, los platelmintos, junto con los acelomorfos, son los más simples y probablemente los Bilateria más primitivos; todos presentan simetría bilateral pero,

a diferencia de muchos de los bilaterados superiores, no cuentan con una cavidad corporal (acelomados). También son los primeros animales en haber evolucionado un auténtico cerebro. Cnidarios y ctenóforos representan un estadio intermedio de complejidad corporal entre poríferos y placozoos, y los Bilateria, que comprenden a la mayor parte del resto de filos del reino animal. Los platelmintos, además, incluyen a los clanes de parásitos Cestoda, Trematoda y Monogenea.



**Figura 2.1 Posición filogenética de las planarias en el reino animal.**

Los lofotrocozoos (Lophotrochozoa) son uno de los grupos de animales protostomados. Los platelmintos forman un clado denominado platizoos (Platyzoa), estrechamente relacionado con los lofotrocozoos y a veces se incluyen en ellos.

(Peter *et al.*, 2003; Reddien & Sánchez Alvarado, 2004; Rink, 2012; Shibata *et al.*, 2010; Tanaka & Reddien, 2011). En condiciones de homeostasis, los neoblastos se diferencian para reemplazar las células senescentes de todos los tejidos del animal y, si es necesario, contribuir a su crecimiento (Fig. 2.2), a la vez que se dividen para mantener su población constante. Ante una herida, responden aumentando su tasa de división para producir células que reemplacen el tejido perdido. Recientemente, gracias a las técnicas moleculares, se está pudiendo comprobar que la población de células colectivamente identificadas como neoblastos es más heterogénea de lo que originalmente se pensaba. Si bien muchas de las células que la componen son células madre pluripotentes, esta población también parece

Las planarias presentan un conjunto de órganos similar al de otros animales triploblásticos: un cerebro que comprende diferentes sistemas de neurotransmisores (Cebrià, 2007, 2008; Fraguas *et al.*, 2014, 2012), una cavidad gastrovascular muy ramificada que se encarga tanto de la digestión como de la distribución de los nutrientes (Rink *et al.*, 2011; Scimone *et al.*, 2011), un sistema excretor de protonefridios con similitudes evolutivas al riñón de vertebrados, diversos órganos sensoriales y un sistema reproductor hermafrodita (Newmark *et al.*, 2008).

## 2.2. La plasticidad de las planarias

comprender células ya comprometidas con un linaje.



**Figura 2.2** Crecimiento y decrecimiento en la planaria *S. mediterranea*.

Las planarias son capaces de aumentar o disminuir su tamaño manteniendo las proporciones corporales llegando a una variación de hasta 40 veces, entre 0,5mm y 2cm, en el caso de *S. mediterranea*, en función de la disponibilidad de alimento (Baguña & Romero, 1981; Newmark & Sánchez Alvarado, 2002; Saló, 2006). Barra de escala: 0,5cm.

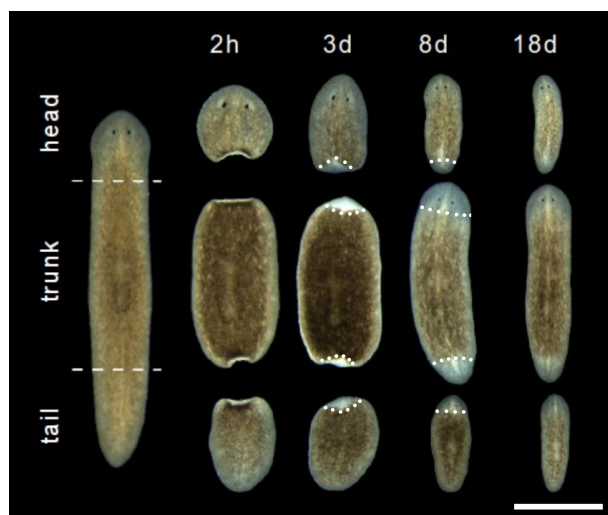
Se ha observado, además, que la regeneración depende del sistema nervioso (Kumar & Brockes, 2012) e incluye la expresión de numerosas metaloproteinasas de matriz (Isolani *et al.*, 2013), las cuales son responsables de degradar y reestructurar la matriz extracelular, especialmente relevante durante los procesos de cicatrización y reparación de la herida.

## 2.3. La planaria como organismo modelo

La principal característica que hace interesantes desde un punto de vista biológico a estos animales es su ilimitada capacidad de regeneración. Su potencial regenerativo varía entre las diferentes especies y, de entre ellas, destaca la especie *S. mediterranea*, capaz de regenerar un individuo completo a partir de un pequeño fragmento de casi cualquier parte de su cuerpo (Fig. 2.3).

Como organismo modelo, las planarias se han usado para el estudio de la regeneración durante más de cien años (Baguña, 2012; Morgan, 1898, 1901). Las dos especies más utilizadas actualmente en laboratorio son *S. mediterranea* y *Dugesia japonica*, de las cuales se utilizan líneas clonales creadas a partir de un único animal para disminuir la variabilidad en los experimentos. Ambas poseen excelentes capacidades regenerativas y el uso de una u otra se debe principalmente a razones históricas y geográficas. La diferencia entre las cepas sexuada y asexuada de *S. mediterranea* se atribuye genéticamente a la presencia en la segunda de una translocación en uno de sus cromosomas, si bien nunca se ha demostrado que esta sea la causa real. Los individuos asexuados se reproducen únicamente por fisión asimétrica mientras que los sexuales lo hacen como hermafroditas (Baguña *et al.*, 1999; Saló, 2006; Sánchez Alvarado *et al.*, 2002; Zayas *et al.*, 2005).

Entre las herramientas experimentales disponibles para trabajar con este modelo destacan la inhibición de la expresión génica mediante RNA interferente (*RNA*



**Figura 2.3** Diferentes estadios del proceso de regeneración de la planaria *S. mediterranea*.

En poco más de dos semanas, la planaria *S. mediterranea* es capaz de regenerar un individuo completo y con las proporciones correctas a partir de una parte amputada como cabeza, tronco o cola, o incluso fragmentos mucho menores. La línea discontinua indica las secciones de corte; la línea de puntos delimita el blastema de regeneración. Barra de escala: 0,5cm.

*interference*, RNAi) (Reddien *et al.*, 2005a; Sánchez Alvarado & Newmark, 1999), el marcaje con bromodesoxiuridina (BrdU) (Newmark & Sánchez Alvarado, 2000), hibridación *in situ* (*whole mount in situ hybridization*, WISH) (King & Newmark, 2013) y más recientemente el aislamiento de células mediante *fluorescence activated cell sorting* (FACS) (Hayashi & Agata, 2012; Hayashi *et al.*, 2006). Otro instrumento simple pero importante en el estudio de la biología de la planaria ha sido el uso de radiación ionizante (gamma o X) para eliminar los neoblastos puesto que, como se verá en el siguiente capítulo, son las únicas células que se dividen. Por desgracia, hay que mencionar dos técnicas largamente anheladas de las que todavía carece este modelo pese a los esfuerzos por desarrollarlas: la posibilidad de crear animales transgénicos y mantener cultivos celulares *in vitro*.

La definición práctica de los neoblastos como las únicas células que se dividen y la rápida eliminación de las mismas por medio de irradiación ha sido tomada para realizar comparaciones entre animales irradiados y no irradiados que han aportado ribosondas para la localización de los neoblastos (Reddien *et al.*, 2005b; Salvetti *et al.*, 2000; Shibata *et al.*, 1999), un criterio para aislar poblaciones de neoblastos mediante FACS (Hayashi *et al.*, 2006) y estudios a gran escala de expresión de genes sensibles a irradiación (Blythe *et al.*, 2010; Eisenhoffer *et al.*, 2008; Friedländer *et al.*, 2009; Rossi *et al.*, 2007; Solana *et al.*, 2012; Wagner *et al.*, 2012).

La pérdida de las células madre en planaria causa defectos severos en la regeneración. En animales intactos, la eliminación de células madre tras una irradiación letal, por ejemplo, conduce a la regresión de la cabeza, replegamiento ventral y finalmente lisis (Bardeen & Baetjer, 1904; Reddien *et al.*, 2005a). Este se ha convertido también en el fenotipo estereotípico de eliminación de neoblastos mediante RNAi (Reddien *et al.*, 2005a,b). Recientemente, la eliminación de neoblastos mediante RNAi contra una histona específica de neoblasto, la histona H2B, se ha propuesto como una alternativa más específica a la irradiación, que evitaría la respuesta transcripcional al daño en el DNA a nivel de todo el organismo (Solana *et al.*, 2012).

La investigación en células madre está huérfana de un organismo modelo para

el estudio de células madre *in vivo*. Por tanto, las planarias representan una oportunidad única de abordar un gran espectro de problemas en la investigación con células madre, incluyendo la conservación evolutiva de la pluripotencia, la organización dinámica de los linajes de diferenciación y los mecanismos subyacentes de la homeostasis de las células madre. Constituyen un sistema de células madre que es el paradigma de uno de los principios de diseño universal de los sistemas biológicos: el estado dinámico estable. El trabajo en organismos con grandes capacidades regenerativas, como las planarias, puede complementar los estudios *in vitro*, aportando información sobre la regeneración a nivel orgánico y tisular *in vivo*, además de la interacción entre diferentes tipos celulares, todo ello no factible mediante cultivos celulares. Además, la población de neoblastos debe estar sujeta a una regulación estricta que impida la proliferación descontrolada de estas células pluripotentes. Los mecanismos de control implicados pueden ser de un interés muy relevante en el estudio de la tumorigénesis y el cáncer.

## 2.4. Regeneración en planaria

La acumulación de neoblastos en la regeneración de heridas y su rápida pérdida después de una dosis de radiación que impedía la regeneración relacionó a los neoblastos con la regeneración (Wolff & Dubois, 1948). En planarias, fue descrito por Morgan ya en 1898 (Morgan, 1898) que los dos procesos de regeneración se dan conjuntamente teniendo lugar una remodelación de los tejidos preexistentes (morfalaxis) a la vez que se da la formación de un blastema (epimorfosis), que deriva de las ya mencionadas células madre somáticas, los neoblastos. Las planarias asexuadas utilizan esta habilidad regenerativa como su única forma de reproducción, fisionándose transversalmente a lo largo del eje antero-posterior.

### 2.4.1. Cicatrización de la herida

Durante el inicio del proceso de regeneración, el principal objetivo es reducir la superficie de la herida, así que la primera respuesta implica la contracción de la musculatura adyacente a la región de la amputación. Además, la herida es recubierta de una capa de mucopolisacárido protectora. A continuación, las células epidérmicas pierden su morfología característica y se extienden tanto dorsal como ventralmente para facilitar el cierre de la herida. Este proceso tarda solamente 30 minutos. La extensión de las células implica un contacto de las epidermis ventral y dorsal con el parénquima que podría ser uno de los factores desencadenantes de la formación en la región de la herida del tejido despigmentado donde se diferencian los nuevos tejidos y estructuras denominado blastema (Fig. 2.3) (Chandebois, 1980; Saló & Baguña, 1984).

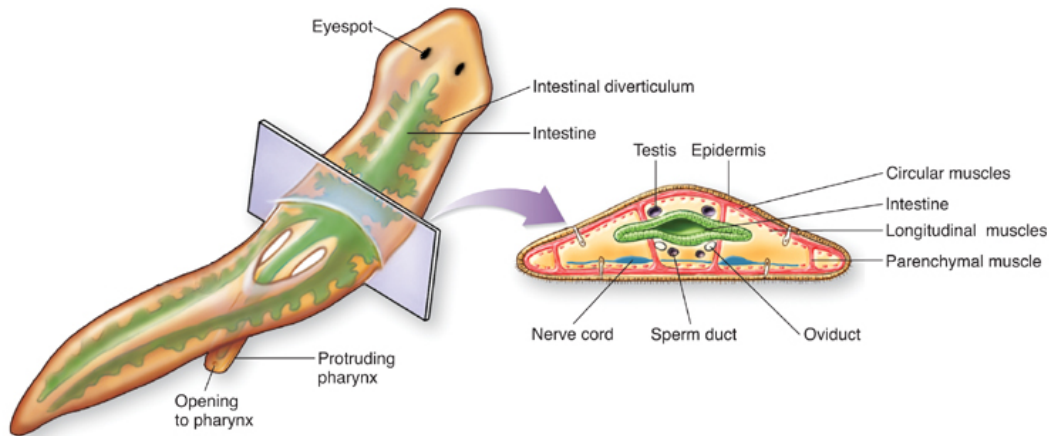
### 2.4.2. Formación del blastema

Durante la formación del blastema a consecuencia de una herida es indispensable el papel que juegan los neoblastos. Ante una pérdida de tejido, estas células madre se activan por mecanismos poco conocidos y se acumulan en la herida, donde su descendencia forma el blastema de regeneración (Eisenhoffer *et al.*, 2008; Wenemoser & Reddien, 2010). Los neoblastos proliferan de una manera espacial y temporal característica. Primero hay una respuesta generalizada a la herida; a continuación se da una proliferación espacialmente más localizada para reemplazar los tejidos perdidos (Saló & Baguñà, 1984; Wenemoser & Reddien, 2010). Después de la cicatrización se produce un pico de actividad mitótica detectado principalmente en una estrecha región del tejido antiguo que está justo por debajo del blastema y que se denomina por ello posblastema. Las células que van a formar los nuevos tejidos que sustituirán a los perdidos se generan en el posblastema por división de los neoblastos, migran al blastema y, una vez allí, dejan de dividirse (Saló & Baguñà, 1984), se diferencian e integran, reemplazando el tejido distal perdido mientras el tejido existente se remodela. En este proceso las células en apoptosis también juegan un papel importante, de manera que la tasa de muerte celular aumenta de forma importante (Pellettieri *et al.*, 2010; Wenemoser & Reddien, 2010).

### 2.4.3. Identidad posicional y especificación de los ejes corporales

Las planarias son organismos bilaterales que muestran una polaridad bien definida (Fig. 2.4). Este patrón corporal debe ser adoptado correctamente por los nuevos tejidos para formar las nuevas estructuras correctamente dependiendo de su posición, anterior o posterior y ventral o dorsal. ¿Cómo son reestablecidas durante la regeneración las señales que dirigen la identidad posicional del plan corporal de la planaria? Estas señales dirigen la progenie de los neoblastos para reemplazar tejidos perdidos distantes y remodelar los tejidos existentes para asegurar la restauración y la integración de todos los sistemas del organismo. Un entendimiento detallado de estos procesos proporciona un paradigma simple de cómo esto podría conseguirse en contextos donde tejidos enfermos, dañados o envejecidos podrían, potencialmente, ser reparados por terapias basadas en células madre.

Vías de desarrollo conservadas están implicadas en controlar la polaridad regenerativa y el control del correcto posicionamiento de la diferenciación de los neoblastos. Por ejemplo, la vía de BMP está claramente implicada en establecer correctamente el eje dorsoventral durante la regeneración, de manera análoga a su función en la embriogénesis en otros animales (Molina *et al.*, 2007; Orii & Watanabe, 2007; Reddien *et al.*, 2007). El patrón del eje mediolateral y la simetría bilateral del sistema nervioso central también requiere los homólogos de planaria de moléculas conocidas por estar implicadas en estos procesos durante el desarrollo embrionario de otros animales, como *slit* (Cebrià *et al.*, 2007), *robo* (Cebrià & Newmark, 2007), *netrin* (Cebrià, 2005) y *Wnt-5* (Adell *et al.*, 2009b; Gurley *et al.*, 2010).



**Figura 2.4 Morfología de una planaria.**

A lo largo del eje anteroposterior se encuentran distribuidos asimétricamente diversos órganos corporales, como el digestivo, con una única rama intestinal en la región anterior del animal y dos en la región posterior, con la faringe situada en la región central del cuerpo; los cordones nerviosos ventrales forman el sistema nervioso central y sensorial, con los ganglios cefálicos y los ojos en la región anterior del animal. De la misma manera, diversos órganos corporales se encuentran también distribuidos asimétricamente a lo largo del eje dorsoventral del organismo. En la región anterior, por ejemplo, desde una posición más dorsal hacia una más ventral encontramos los ojos, el cerebro y los cordones nerviosos. Asimismo, a lo largo de toda la longitud anteroposterior, el sistema digestivo se localiza en una posición dorsal en relación al sistema nervioso central. Adaptado de McGraw-Hill, *Trends in Evolution*.

Conocemos mucho sobre la formación del eje antero-posterior en planaria (Adell *et al.*, 2009a). La vía de señalización Wnt, requerida por todos los metazoos para formar el patrón A/P durante la embriogénesis (Petersen & Reddien, 2009), juega un papel central tanto durante la regeneración como la homeostasis y se requiere para la regeneración posterior y la identidad posterior (Adell *et al.*, 2009a; De Robertis, 2010; Forsthoefel & Newmark, 2009). La señalización a través de Hedgehog (Hh) también es necesaria para el correcto establecimiento de la identidad posterior promoviendo la actividad de Wnt en el blastema posterior (Oviedo *et al.*, 2010; Yazawa *et al.*, 2009). Un posible factor HOX de la clase TALE, *Smed-Prep*, ha sido descrito actuando en la vía de señalización Wnt (Felix & Aboobaker, 2010), proporcionando una conexión mecánica entre la progenie celular de los neoblastos y las señales de polaridad. La proteína Smed-Prep es normalmente suprimida postranscripcionalmente en el blastema posterior por Wnt. *Smed-prep(RNAi)* resulta en la pérdida de identidad anterior pero sin la adopción de identidad posterior.





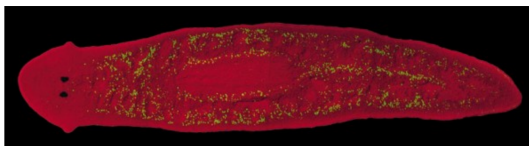
# 3 El neoblasto

*Omnis cellula e cellula.*

“Cellular Pathology” (1858)  
Rudolf Virchow

Como se ha visto en el capítulo anterior, las planarias son platelmintos de vida libre de una extraordinaria plasticidad. De entre ellas destaca la especie *S. mediterranea* que se ha convertido en el organismo modelo de elección. Su extraordinaria capacidad de regenerar depende de la presencia de un gran número de células madre adultas somáticas conocidas como neoblastos.

## 3.1. Citología del neoblasto

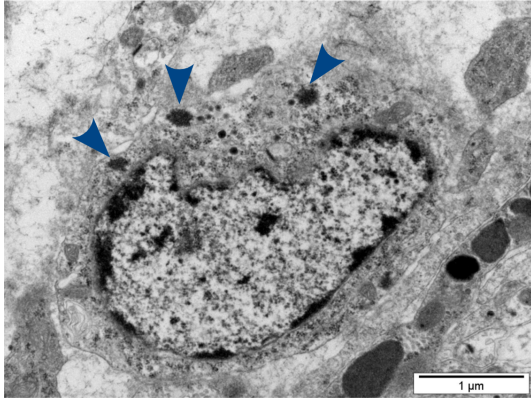


**Figura 3.1 Distribución de los neoblastos en la planaria *S. mediterranea*.**

Sección confocal en la que se observan las células en proliferación (neoblastos) marcadas con BrdU (verde). Adaptado de Newmark & Sánchez Alvarado (2000).

El término “neoblasto” empezó a usarse para describir unas pequeñas y abundantes células redondeadas distribuidas por todo el mesénquima de la planaria, excepto en el área anterior a los fotorreceptores y en la faringe (que son las únicas áreas incapaces de regenerar por si solas) (Fig. 3.1) (Baguñà, 2012; Reddien & Sánchez Alvarado, 2004). Constituyen alrededor una cuarta parte de todas las células de la planaria (Baguñà & Romero, 1981).

En imágenes de microscopía electrónica (Fig. 3.2), los neoblastos aparecen como células de 5 a 10µm de diámetro con poco citoplasma, muchos ribosomas libres, unos cuantos orgánulos distinguibles, cuerpos cromatoides prominentes y un gran núcleo con poca heterocromatina que ocupa la mayor parte del volumen celular.



**Figura 3.2 Imagen de un neoblasto en el microscopio electrónico de transmisión.** Las flechas señalan los cuerpos cromatoides en el citoplasma de la célula. Barra de escala: 1  $\mu\text{m}$ . Cortesía del Dr. Luca Gentile.

La división celular nunca ocurre en los confines del tejido diferenciado. Ni siquiera en tejidos con elevado recambio como el intestino o la epidermis contienen células en división (Baguña, 1976; Forsthoefel *et al.*, 2011; Newmark & Sánchez Alvarado, 2000). Las divisiones celulares en planaria se limitan estrictamente al poco organizado mesénquima que rodea todos los órganos y, más concretamente, a las células que cumplen la definición morfológica antes detalladas. Así, los neoblastos son las únicas células que se dividen en las planarias (Morita & Best, 1984) y, por tanto, las únicas capaces de proliferar. El flujo de progenitores

resultante reemplaza continuamente las células diferenciadas senescentes y esta elevada tasa de recambio celular conduce a una rápida renovación de todas las células del animal en cuestión de semanas (Rink, 2012), permitiendo a las planarias crecer y decrecer en respuesta a condiciones cambiantes de nutrientes, manteniendo las proporciones de tejidos y células constantes (Oviedo *et al.*, 2003; Saló, 2006; Takeda *et al.*, 2009).

### 3.2. El neoblasto como célula madre

Una célula madre se define como una célula indiferenciada que puede diferenciarse en tipos celulares especializados a la vez que puede dividirse por mitosis para producir más células madre, bien por un número limitado de divisiones o, potencialmente, como un clon celular inmortal. En función de la cantidad potencial de tipos celulares a los cuales pueden dar lugar, las células madre se han clasificado en varias categorías (Petralia *et al.*, 2014):

- **Células madre unipotentes/oligopotentes**  
Células madre que solo pueden diferenciarse en uno o unos pocos tipos celulares.
- **Células madre multipotentes**  
Células madre que pueden diferenciarse en muchos tipos celulares.
- **Células madre somáticas pluripotentes**  
Células madre que pueden diferenciarse en todos los tipos conocidos de células somáticas de un animal.

- **Células germinales primordiales**

Células iniciales del desarrollo de una línea germinal, que son solo capaces de formar células germinales (gametogénesis).

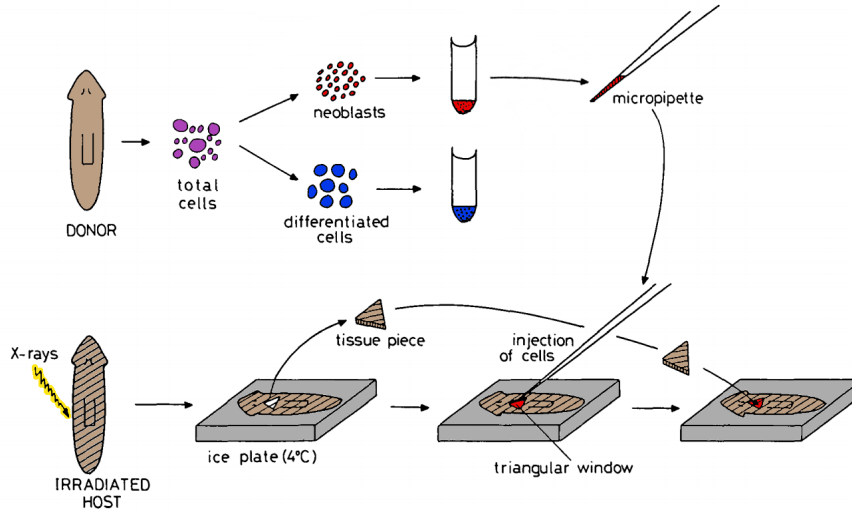
- **Células madre totipotentes**

Células madre pluripotentes que también son capaces realizar la gametogénesis.

Un ejemplo de células madres unipotentes son las células satélite musculares de vertebrados, que únicamente producen mioblastos que se transforman en células musculares. Las células intersticiales de *Hydra* comentadas en el primer capítulo son un caso de células madre multipotentes. El cigoto de mamífero y el embrión temprano están formados por células madre totipotentes. Los neoblastos de planaria (así como los coanocitos y los arqueocitos de esponjas) podrían ser totipotentes o pluripotentes, dependiendo de las definiciones y descripciones de varios autores.

Se han realizado varios estudios con el propósito de arrojar luz sobre esta cuestión. Revisando un experimento clásico de Baguñà *et al.* (1989) en el que una planaria sexuada irradiada letalmente se transformaba en asexuada después de ser trasplantada con neoblastos aislados de un individuo asexuado y viceversa (Fig. 3.3), Wagner *et al.* (2011) mostraron que el trasplante de un único neoblasto en una planaria irradiada letalmente (sin células madre) rescataba al receptor y daba lugar a un animal sano con el mismo genotipo que el donante. Por tanto, un solo neoblasto es capaz de diferenciarse en cualquier tipo de célula postmitótica (célula somática) conocida. Este experimento demostró de forma concluyente y elegante que al menos ciertos neoblastos son pluripotentes. Probablemente son incluso totipotentes, pero el uso neoblastos de un donante asexuado en el experimento anterior impide la reconstrucción de la reproducción sexual y la demostración formal de totipotencia. Por ello, nos referiremos a los neoblastos *sensu stricto* como células madre pluripotentes. Estos experimentos, sin embargo, no pueden esclarecer si algunos neoblastos no son pluripotentes puesto que el fallo de algunos neoblastos trasplantados para generar colonias podría deberse a razones técnicas.

Entender las bases mecanísticas de la pluripotencia de los neoblastos ha sido y continúa siendo uno de los mayores focos de la investigación en planaria.



**Figura 3.3 Experimento clásico de trasplante de neoblastos.**

Procedimiento de inyección de una fracción celular enriquecida en neoblastos en una planaria receptora cuyos neoblastos han sido eliminados mediante radiación X. Tras recuperarse, la planaria receptora se convierte en un clon genético del donante. Adaptado de Baguña *et al.* (1989).

### 3.3. Los neoblastos y la línea germinal

Además de los neoblastos somáticos pluripotentes, las planarias también poseen una línea germinal como parte de su sistema reproductor hermafrodita (Newmark *et al.*, 2008). Tanto la línea germinal como las gónadas somáticas pueden regenerar *de novo*, probablemente gracias a una población de células madre (neoblastos) germinales putativa. Estas células son muy similares a los neoblastos en términos de morfología y sensibilidad a la radiación, aunque podrían ciclar a un ritmo más lento (Handberg-Thorsager & Saló, 2007; Sato *et al.*, 2006; Wang *et al.*, 2007).

El factor que distingue a los neoblastos somáticos de los neoblastos de la línea germinal es la expresión de *nanos*, asociado al desarrollo de las gónadas y las células germinales y, hasta ahora, el único determinante de la línea germinal altamente conservado que no se expresa constitutivamente en todos los neoblastos (Handberg-Thorsager & Saló, 2007; Rink, 2012; Sato *et al.*, 2006; Wang *et al.*, 2007). Tanto las cepas sexuales como asexuales tienen células germinales que expresan *nanos*, de manera que la inhibición de los órganos reproductores maduros en las cepas asexuales debe ocurrir más allá de la función de *nanos*. De acuerdo con esto, *nanos(RNAi)* no afecta al mantenimiento de los neoblastos pero impide la regeneración de las gónadas (Wang *et al.*, 2007). Los neoblastos se posicionan, por tanto, entre la línea que separa a las células somáticas de la línea germinal y solo necesitan de la expresión de *nanos* para transicionar hacia la línea germinal. Las similitudes entre los neoblastos y la línea germinal pueden estar reflejando una peculiaridad de la biología

de la planaria. Aún así, la pluripotencia de los neoblastos proporciona un enlace conceptual entre los dos tipos celulares. Aunque las células germinales solo dan lugar a los gametos, la fusión de estos durante la fertilización inicia el desarrollo del organismo completo, necesitando, por tanto, de pluripotencia en la línea germinal.

En general, la viabilidad de las poblaciones de las células madre de la línea germinal en metazoos se mantiene a través de divisiones celulares que superan los problemas de la replicación terminal de los cromosomas gracias a la elongación de los telómeros, lo cual ocurre durante la embriogénesis pero ¿cómo mantienen las planarias asexuadas la viabilidad de sus neoblastos somáticos? Estas células pueden producir elevados niveles de telomerasa (que alarga los telómeros) a través de *splicing* alternativo de la telomerasa activa, de manera que la longitud de sus telómeros se mantiene durante la replicación celular que se da durante la fisión o la regeneración (Tan *et al.*, 2012). Esto podría explicar cómo las planarias asexuadas consiguen su potencial inmortalidad. Estos neoblastos también expresan numerosos homólogos del gen *Piwi* (Palakodeti *et al.*, 2008; Reddien *et al.*, 2005b), típicamente asociado a células de la línea germinal. Estos homólogos ayudan a mantener la integridad del genoma del neoblasto (Rink, 2012). Además, los neoblastos presentan cuerpos cromatoides, asociados también a la función de las células germinales. Muy probablemente, tanto las células germinales como los neoblastos pluripotentes somáticos derivan de neoblastos totipotentes capaces de autorenovarse (Bely & Sikes, 2010; Solana, 2013).

Estudios funcionales han revelado que muchos genes con un papel en células madre y células germinales en otros animales tienen funciones en neoblastos y su progenie temprana (Conte *et al.*, 2009; Ewen-Campen *et al.*, 2009; Fernández-Taboada *et al.*, 2010; Guo *et al.*, 2006; Palakodeti *et al.*, 2008; Reddien *et al.*, 2005b; Rossi *et al.*, 2006; Rouhana *et al.*, 2010; Salvetti, 2005; Shibata *et al.*, 2010; Solana *et al.*, 2009; Wang *et al.*, 2010; Yoshida-Kashikawa *et al.*, 2007). Muchos de estos genes están asociados con la unión al RNA y el metabolismo y están localizados en los cuerpos cromatoides de los neoblastos o del plasma germinal en otros animales.

Hace tres décadas que los trabajos publicados por Gremigni (1988), muy poco comentados en la bibliografía actual, demostraban de forma elegante la existencia de transdeterminación de células progenitoras de la línea germinal a neoblastos de la línea somática. Mucho más recientemente, un estudio ha profundizado en el papel que determinadas hormonas peptídicas juegan en el desarrollo de la línea germinal (Collins *et al.*, 2010) y un análisis funcional de expresión por *microarray* utilizando el fenotipo *Smed-nanos(RNAi)* ha identificado los genes necesarios para la diferenciación de las células madre de esta línea (Wang *et al.*, 2010). Aún así, se conoce poco sobre los procesos celulares autónomos requeridos para iniciar la diferenciación de los neoblastos para formar y mantener los tejidos somáticos de la planaria.

El desarrollo de herramientas de marcaje para el análisis de linajes será una importante dirección futura para caracterizar los linajes ya conocidos y otros nuevos.

### 3.3.1. Cuerpos cromatoides

Los cuerpos cromatoides son estructuras perinucleares características, densas al microscopio electrónico (Fig. 3.2), que se asemejan a los gránulos germinales asociados con los linajes de la línea germinal en embriones de animales, con los que comparten constituyentes moleculares conservados (Eddy, 1975; Ewen-Campen *et al.*, 2009; Parvinen, 2005; Saffman & Lasko, 1999; Strome & Lehmann, 2007). Estas estructuras son ricas en complejos proteicos de unión al RNA y son lugares prominentes del metabolismo citoplasmático del RNA. Los cuerpos cromatoides de la planaria desaparecen mayoritariamente durante su diferenciación, excepto en los linajes de la línea germinal en las cepas sexuadas, donde parece que se mantienen, y los linajes neuronales, donde presumiblemente reaparecen después de la diferenciación (Hay & Coward, 1975; Solana *et al.*, 2009; Yoshida-Kashikawa *et al.*, 2007). Las crecientes evidencias en planarias sugieren que los cuerpos cromatoides tienen un papel central en la regulación de la biología del neoblasto.

Miembros de la familia de piwi se expresan en los cuerpos cromatoides de la línea germinal en mamíferos (Kotaja, 2006; Wang *et al.*, 2009). Tres ortólogos en planaria, llamados *smedwi-1*, *-2* y *-3*, se expresan en los neoblastos (Palakodeti *et al.*, 2008; Reddien *et al.*, 2005b). La localización subcelular de los transcritos de *smedwi-2* muestra una posición perinuclear reminiscente de los cuerpos cromatoides (González-Estévez *et al.*, 2009a). Tanto *smedwi-2* como *-3* se requieren para la regeneración y la homeostasis, causando su inhibición mediante RNAi la pérdida de los neoblastos.

Aunque no se ha demostrado que los homólogos de piwi en planaria estén presentes en los cuerpos cromatoides, otras proteínas requeridas para el mantenimiento de los neoblastos sí lo están, incluyendo Smed-SmB, un miembro de la superfamilia LSm de proteínas de unión al RNA (Fernández-Taboada *et al.*, 2010). *Smed-SmB(RNAi)* conduce a una disminución en el número de cuerpos cromatoides antes de la pérdida de la proliferación de los neoblastos y la pérdida de sus marcadores de expresión. Esto sugiere que componentes en el cuerpo cromatoide actúan en el mantenimiento de los neoblastos y su renovación. En concordancia con esto, *Spoltud-1*, una proteína de unión al RNA homóloga de tudor en *Schmidtea polychroa*, también se expresa en los cuerpos cromatoides y es necesaria para el mantenimiento de los neoblastos (Solana *et al.*, 2009).

En resumen, los cuerpos cromatoides de los neoblastos contienen muchas proteínas de unión al RNA asociadas con la formación de la línea germinal en otros animales. En conjunto, estos datos sugieren que, aunque los neoblastos son células madre somáticas adultas y no células germinales, presentan un gran paralelismo con las células madre de la línea germinal.

Futuros trabajos sobre los cuerpos cromatoides de planaria podrían arrojar luz acerca de cómo el mantenimiento de las células madre y la pluripotencia pueden ser controlados en el contexto de las terapias regenerativas. Sin embargo, mantener la capacidad de autorenovación y la pluripotencia de las células madre es solo la mitad de la historia en la regeneración una planaria. Los neoblastos también deben producir

una progenie capaz de diferenciarse en los tipos celulares correctos y posicionarse adecuadamente a lo largo del eje corporal.

### 3.4. Dinámica de los neoblastos

El mantenimiento del número apropiado de células es un punto clave en todos los sistemas de células madre. Excesivas divisiones pueden derivar en cáncer; por contra, la pérdida de células, al detenimiento del recambio tisular y a un envejecimiento prematuro. La homeostasis de las células madre ha de conseguir un balance entre estos dos fundamentos: la autorenovación (por división celular) y la diferenciación en una progenie múltiple para conseguir que el número de células madre se mantenga constante (Weissman, 2000). Cada división de una célula madre desafía esta homeostasis: divisiones simétricas resultantes en dos células madre incrementan el número de células madre. Divisiones simétricas resultantes en dos células diferenciadas significan una pérdida neta de una célula madre. Solo las divisiones asimétricas (esto es, la producción de una célula madre y una diferenciada) mantienen el *statu quo*.

#### 3.4.1. Respuesta proliferativa de los neoblastos

En condiciones estables, los neoblastos se dividen a un ritmo basal. Durante periodos prolongados de ayuno las divisiones de los neoblastos continúan, aunque la planaria reduzca su tamaño continuamente debido a un decrecimiento en el número de sus células (Baguñà & Romero, 1981; González-Estévez *et al.*, 2012a; Oviedo *et al.*, 2003; Takeda *et al.*, 2009). El catabolismo de las células que mueren probablemente alimenta las divisiones durante el ayuno, manteniendo así el recambio tisular a costa de una progresiva pérdida de células (Baguñà & Romero, 1981; Baguñà *et al.*, 1990). La alimentación provoca un incremento en la fracción de neoblastos en fase M, con un máximo ocho horas después que tarda varios días en recuperar su estado basal (Baguñà, 1974; Kang & Alvarado, 2009). Este incremento transitorio en las divisiones de los neoblastos desplaza el estado de equilibrio hacia un incremento neto en el número de células, conduciendo a un crecimiento del animal.

Una herida tiene un efecto similar en la actividad mitótica de los neoblastos, causando un pico de neoblastos en fase M entre seis y ocho horas después de la lesión y un segundo pico más sostenido y localizado en la cercanía de la herida entre dos y tres días después (Baguñà, 1976; Wenemoser & Reddien, 2010). Se cree que los progenitores generados durante la lesión migran localmente hacia la herida y dan lugar al blastema de regeneración (Aboobaker, 2011; Baguñà, 2012; Forsthoefel & Newmark, 2009).

¿Qué constituye la capacidad de reserva de los neoblastos? ¿Una subpoblación de neoblastos que no cicla dedicada específicamente para su activación durante el crecimiento o la regeneración? ¿O una aceleración de la dinámica celular dentro de una población que cicla continuamente? El primer máximo mitótico que tiene



lugar después de haberse producido la lesión o la alimentación se ha interpretado a favor de una subpoblación detenida en G2 (Baguñà, 1974; Saló & Baguñà, 1984). La introducción de pulsos de marcaje con BrdU cuestionó esta idea (Newmark & Sánchez Alvarado, 2000), llevando a la conclusión de que todos los neoblastos se dividen de forma continua. Sin embargo la validez de estos ensayos quedó en suspenso después de comprobarse que la pequeña herida provocada durante la microinyección de BrdU ya activa a los neoblastos (Wagner *et al.*, 2011; Wenemoser & Reddien, 2010). Esto deja abierta la posibilidad de que en condiciones sin estimulación algunos neoblastos permanezcan días sin dividirse o incluso no lo hagan nunca.

El hecho de que el sistema de neoblastos pueda cambiar dramáticamente su progenie enfatiza la importancia del control de las señales de proliferación. La red de control de la división de los neoblastos puede comprender señales locales de auto-renovación, señales globales liberadas en respuesta a una herida o cambios en el estado metabólico, así como bucles de *feedback* negativo que medien el retorno al ritmo basal de división. Estos bucles de disminución de la respuesta de los neoblastos pueden incluir a SMED-EGFR-1, un receptor de EGF que se expresa en el intestino de la planaria (Fraguas *et al.*, 2011).

Por tanto, los mecanismos celulares por los cuales el sistema de células madre de la planaria ajusta la formación de la progenie a las necesidades fisiológicas cambiantes no está claro. La importancia de la cuestión también en términos de heterogeneidad de la población y la disponibilidad de más y más versátiles marcadores de neoblasto abalan el retorno experimental a estas cuestiones.

Además, las planarias no solo regulan la proliferación de los neoblastos, sino también el ritmo de eliminación de las células diferenciadas, probablemente de forma específica en cada tejido (González-Estévez *et al.*, 2012a).

## 3.5. La población de neoblastos es heterogénea

El término “neoblasto” a menudo se emplea implicando una homogeneidad funcional aún en ausencia de una evidencia sólida. Asignar una etiqueta (“neoblasto”) a esta población es útil para referirse a células con muchas características comunes pero es potencialmente problemático por la uniformidad implícita de la población celular.

Los neoblastos se definen como las únicas células que proliferan en el organismo. La división celular es comúnmente utilizada como una característica definitoria de neoblasto: todas las células adultas que se dividen son neoblastos. Consecuentemente, marcadores genéricos de división celular son comúnmente utilizados como marcadores de neoblasto. Las células en fase S o mitosis (neoblastos por definición) pueden ser fácilmente marcados experimentalmente y, por lo tanto, visualizados, mediante la incorporación de pulsos de BrdU, con un anticuerpo contra una modificación de histona específica de mitosis (fosfo-Histona H3-Ser10, H3P), o mediante sondas de RNA de genes que se transcriben durante la fase S, como *h2b*, *pcna* o *MCM2* (Newmark & Sánchez Alvarado, 2000; Shibata *et al.*, 2010). Sin embargo,

### 3.5 La población de neoblastos es heterogénea

se desconoce si todas los neoblastos identificados con estos marcadores son realmente células madre o si las planarias albergan también tipos celulares de amplificación transitorios que, consecuentemente, contribuirían a la población de células positivas.

Las células en G0/G1 que se dividirán de nuevo (y que pueden ser consideradas, por tanto, neoblastos) son más difíciles de marcar experimentalmente. La presencia de abundante RNA mensajero (mRNA) del gen *smewi-1* es, tal vez, el más comúnmente usado para marcar una célula como neoblasto en la actualidad (Reddien *et al.*, 2005b). Todas las células en fase S son *smewi-1*<sup>+</sup>, aunque no todas las células *smewi-1*<sup>+</sup> están en fase S (Wagner *et al.*, 2011). Aún así, todas las células *smewi-1*<sup>+</sup> son eliminadas 24 horas después de la irradiación, lo cual es consistente con la posibilidad de que el mRNA de *smewi-1* marque todas las células que se dividieron o que se dividirán, y no otras células (Eisenhoffer *et al.*, 2008; Reddien *et al.*, 2005b).

Debido a que las células que se dividen son sensibles a la radiación, la sensibilidad a la irradiación aguda puede utilizarse también como una característica, quizás más inclusiva, de los neoblastos (Hayashi *et al.*, 2006; Reddien *et al.*, 2005a).

Ni un criterio morfológico ni los actuales marcadores moleculares pueden dibujar un límite entre células que se dividen y una progenie temprana ya comprometida. Es importante, por tanto, enfatizar en que el término “neoblasto” debe ser entendido como una referencia general al sistema de células madre de la planaria, incluyendo células madre pluripotentes, progenie postmitótica temprana y cualquier tipo de estadio intermedio. La causa de esta ambigüedad no es el uso de un término histórico, sino la actual falta de conocimiento en relación al sistema de células madre de planaria.

¿Son los neoblastos una minoría entre una mayoría de progenitores en amplificación transitorios? Hay evidencia de la heterogeneidad en términos de ultraestructura y expresión génica de las poblaciones celulares sensibles a la radiación aisladas por FACS (Hayashi *et al.*, 2010; Higuchi *et al.*, 2007; Shibata *et al.*, 2012). Cambios dinámicos en el patrón de expresión durante la recuperación de una irradiación no letal se han interpretado a favor de una subpoblación de neoblastos más resistente asociada con los cordones nerviosos ventrales (Salveti *et al.*, 2009). De todas formas, una diferenciación concluyente entre los neoblastos como células madre y células de amplificación transitorias requiere discernir linajes celulares. La posible existencia de poblaciones de amplificación transitorias o restringidas a un linaje permanece desconocida. Responder estas cuestiones requerirá una nueva generación de experimentos que relacionen marcadores moleculares con ensayos del potencial de diferenciación. Una de las colaboraciones en el marco de esta tesis ha consistido en el subfraccionamiento de neoblastos vivos con marcadores de superficie, lo que representa un paso más en esta dirección (Moritz *et al.*, 2012, Artículo 1).

Se hace necesaria la identificación de marcadores moleculares específicos de cada tipo celular para poder caracterizarla y entender el proceso de diferenciación así como revelar los genes implicados en él. El descubrimiento de factores de transcripción que controlan las decisiones sobre el destino celular en los neoblastos en fase de diferenciación constituye un primer paso.

### 3 *El neoblasto*

Aunque los experimentos clásicos habían identificado hasta hace poco a estas células como pluripotentes, la expresión génica en esta población es heterogénea. Esto implica que el término “neoblasto” se refiere a una colección de diferentes tipos celulares en proliferación (Hayashi *et al.*, 2010). Así, se ha abierto el debate sobre si los neoblastos constan de un único tipo celular pluripotente, o bien, de progenitores distintos para cada linaje. La pregunta que subyace es simple: cómo los neoblastos son capaces de distinguir el destino celular correcto (Roberts-Galbraith & Newmark, 2015).

**Tabla 3.1 Algunos genes propios de neoblasto.**

Varios de estos genes son habitualmente usados como marcadores de neoblasto.

Gen	Patrón de expresión	Fenotipo RNAi (regeneración)	Función en células madre de planaria	Referencias
<b>Familia Argonaute/Piwi</b>				
<i>Djpiwi-1</i>	Línea media dorsal	Sin determinar	Desconocido	Rossi <i>et al.</i> (2006)
<i>Smedwi-1</i>	Mesénquima	Sin fenotipo	Desconocido	Reddien <i>et al.</i> (2005b)
<i>Smedwi-2</i>	Mesénquima	Formación de blastema, regresión, muerte en 2-3 semanas	expresión de piRNA, diferenciación de células madre	Reddien <i>et al.</i> (2005b) Palakodeti <i>et al.</i> (2008)
<i>Smedwi-3</i>	Mesénquima	Blastema reducido, regresión, muerte en 3 semanas	expresión de piRNA, diferenciación de células madre	Palakodeti <i>et al.</i> (2008)
<b>Otras proteínas de unión al RNA</b>				
<i>DjVlgA</i>	Células germinales, mesénquima, blastema	Sin determinar	Regulación postranscripcional	Shibata <i>et al.</i> (1999)
<i>DjPum</i>	Mesénquima, línea dorsal posterior, cerebro	Blastema reducido, muerte en 3-4 semanas	Regulación postranscripcional	Salveti (2005)
<i>Smed-Bruli</i>	Mesénquima, cerebro	Blastema reducido, regresión, muerte en 3 semanas	Autorenovación de células madre	Guo <i>et al.</i> (2006)
<i>Djnos</i> <i>Smednos</i>	Células germinales primordiales, espermatogonia, oogonia, células precursoras del ojo	No regeneración de gónadas ni aparato copulador	Proteína de unión al RNA, especificación de la línea germinal	Sato <i>et al.</i> (2006) Handberg-Thorsager & Saló (2007) Wang <i>et al.</i> (2007)
<i>Spolud-1</i>	Mesénquima, órganos reproductores, sistema nervioso central	Defectos en la regeneración, pérdida de células madre	Componente del cuerpo cromatoide, autorenovación de células madre	Solana <i>et al.</i> (2009)
<i>Djcbc-1</i>	Mesénquima, cerebro, células madre de la línea germinal	Sin fenotipo	Componente del cuerpo cromatoide	Yoshida-Kashikawa <i>et al.</i> (2007)
<i>Smed-SmB</i>	Mesénquima, cerebro	Sin blastema, pérdida de células madre, muerte en 2 semanas	Componente del cuerpo cromatoide, autorenovación de células madre	Fernández-Taboada <i>et al.</i> (2010)
<b>Replicación del DNA</b>				
<i>Djmcm2</i>	Mesénquima, línea media dorsal	Sin determinar	Replicación del DNA	Salveti <i>et al.</i> (2000)
<i>DjPCNA</i>	Mesénquima, línea media dorsal, líneas bilaterales (the bilateral lines)	Sin determinar	Replicación del DNA	Orii <i>et al.</i> (2005)
<b>Remodelación de la cromatina</b>				
<i>Smed-CHD4</i>	Mesénquima, sistema nervioso central	Respuesta proliferativa reducida, disminución de la progenie celular de los neoblastos, muerte en 3-4 semanas	Helicasa dependiente de ATP, diferenciación de células madre	Scimone <i>et al.</i> (2010)
<i>DjRbAp48</i>	Mesénquima, blastema	Blastema reducido, replegamiento ventral, motilidad reducida, muerte en 6 semanas	Proteína de unión a histona, remodelación del nucleosoma	Bonuccelli <i>et al.</i> (2010)
<b>Otras funciones</b>				
<i>Smedinx-11</i>	Mesénquima, faringe, región alrededor del cerebro, región anterior a los fotorreceptores	Sin blastema, inversión del gradiente de neoblasto, muerte en 5 semanas	Uniones gap (inexina)	Oviedo & Levin (2007)
Genes de las categorías 2 y 3	Blastema temprano, Mesénquima externo en animales intactos	Sin determinar	Diferenciación de la progenie temprana de las células madre	Eisenhoffer <i>et al.</i> (2008)
<i>Smed-PTEN-1</i> <i>Smed-PTEN-2</i>	Mesénquima	Sin blastema, regeneración inhibida, sobrecrecimientos anormales, hiperproliferación de neoblastos, falta de diferenciación celular, muerte en 2 semanas	Mantenimiento de la regulación de las células madre	Oviedo <i>et al.</i> (2008)
<i>Smed-p53</i>	Mesénquima, progenie de las células madre	Disminución de la progenie celular de los neoblastos	Diferenciación de la progenie temprana de las células madre	Pearson & Alvarado (2009)
<i>Djmot</i>	Mesénquima, línea media dorsal, blastema,	Blastema reducido, regeneración reducida, muerte en 5-7 semanas	Chaperona, secuestro de p53	Conte <i>et al.</i> (2009)

piRNA: Piwi-interacting RNA.  
Adaptado de Gentile *et al.* (2010).

## 3.6. Diferenciación de los neoblastos

Un factor crítico en el mantenimiento del *statu quo* entre las células madre y las células diferenciadas es la apropiada regulación de la diferenciación de la progenie de las células madre. La complejidad de esta tarea aumenta con el número de posibilidades disponibles. Las células madre adultas de vertebrados son específicas de tejido y sus destinos de diferenciación son, por tanto, limitados. Los neoblastos pluripotentes, que dan lugar a todos los tipos celulares de la planaria presentan un caso más complejo en términos de elección de destino. Esto arroja una serie de preguntas importantes: ¿está la elección del destino del neoblasto intrínsecamente programada o, por el contrario, está bajo el control de señales externas según las necesidades momentáneas del tejido de manera que tipos celulares particulares pueden diferenciarse a demanda?, ¿en qué punto de la diferenciación de los neoblastos se determina su destino?, o ¿cómo la elección de ese destino está coordinada con la organización global del plan corporal de la planaria? Se han propuesto dos posibles modelos que podrían responder a estas preguntas. Los veremos a continuación.

### 3.6.1. Los modelos de neoblasto nativo y especializado en la regeneración

Los neoblastos responden a una herida dividiéndose y migrando hacia las proximidades de la zona lesionada (Baguñà, 1976; Saló & Baguñà, 1984; Wenemoser & Reddien, 2010). La producción local de una progenie de neoblastos que no se dividen cerca de la herida resulta en la formación del blastema de regeneración. En condiciones normales, la proliferación del posblastema es suficiente para generar el blastema. Ahora bien, si la planaria es irradiada parcialmente puede haber una migración neoblastos desde una distancia mayor, produciéndose así una regeneración normal pero retardada en el tiempo.

Si los neoblastos comprenden una población uniforme de células durante la regeneración (esto es, todas son indeterminadas y pluripotentes), entonces la especialización de nuevos tipos celulares debe ocurrir en células multipotentes que no se dividen de la progenie de los neoblastos. Sin embargo, datos recientes indican que algunos neoblastos expresan factores de transcripción específicos de linaje durante la regeneración y en animales intactos. Estas observaciones apuntan la posibilidad de que uno de los primeros pasos claves en la regeneración de planarias sería la especialización de los neoblastos para producir células del blastema especializadas en lugar de indeterminadas.

#### El modelo de neoblastos nativos

Aunque la posibilidad de heterogeneidad entre los neoblastos parece cada vez más clara (Baguñà, 2012; Reddien & Sánchez Alvarado, 2004), la capacidad de los fragmentos de cualquier región de la planaria que contiene neoblastos para regenerar

sugiere que el atributo de pluripotencia está, si no en todos los neoblastos, al menos sí extendido en esta población.

Según este modelo, los neoblastos producen en el blastema células multipotentes que ya no se dividen. Las células en la población de neoblastos serían básicamente todas iguales, respondiendo todas a las lesiones simplemente migrando y dividiéndose, dando lugar a las células del blastema. Estas células posmitóticas nativas y multipotentes adoptarían las identidades apropiadas basándose en las señales externas que recibieran, por ejemplo, como consecuencia de su posición en el blastema.

#### **El modelo de neoblastos especializados**

A pesar de su apariencia similar, podrían existir diferentes subtipos de células en la población de neoblastos. La población de células adultas de planaria sensibles a irradiación incluye también a las células germinales, por tanto, esta población se sabe que ya contiene al menos dos tipos celulares (Handberg-Thorsager & Saló, 2007; Sato *et al.*, 2006; Wang *et al.*, 2007). En principio, podrían existir otros tipos celulares de neoblasto. Una pregunta recurrente durante mucho tiempo ha sido si existe un tipo celular pluripotente en la población de neoblastos, o bien si la regeneración se lleva a cabo mediante la acción colectiva de múltiples tipos celulares de linaje restringido. Dos aproximaciones han demostrado que algunos neoblastos son, efectivamente, pluripotentes (Wagner *et al.*, 2011). Primero, mediante irradiación se eliminaron la mayoría pero no todos los neoblastos. Los neoblastos supervivientes, esparcidos, produjeron clones o colonias de células descendientes. Estas colonias produjeron neuronas y células intestinales, así como muchos otros tipos de células descendientes. Segundo, el trasplante de neoblastos individuales reveló que algunos podían producir colonias de neoblastos y restaurar el recambio tisular y la capacidad regenerativa de animales irradiados con una dosis letal de radiación. Estos datos indican que al menos algunos neoblastos son pluripotentes.

En este modelo, los neoblastos implicados en la regeneración tendrían diferentes destinos. Estos neoblastos especializados producen en el blastema diversas células específicas de cada linaje que no se dividen. Qué células diferenciadas se producen finalmente por los diferentes neoblastos está, por tanto, predeterminado. La especialización de los neoblastos puede ocurrir antes o después de la lesión.

Los llamados marcadores de categorías (Eisenhoffer *et al.*, 2008) proporcionaron información sobre la orquestación espacio-temporal de la diferenciación de la progenie de los neoblastos. Basándose en la cinética de inhibición diferencial de los genes sensibles a la irradiación y en el marcaje con pulsos de BrdU, se definieron dos categorías de genes expresadas, probablemente, de forma secuencial en la progenie posmitótica en diferenciación de los neoblastos (progenie temprana y progenie tardía). Los distintos lugares de expresión de los marcadores de categoría sugieren que la progenie de los neoblastos nace dentro del mesénquima y experimenta una migración hacia la periferia durante el curso de su diferenciación. La expresión en neuronas se ha descrito para la mayoría de genes de células madre de planaria; sin embargo, su función en estas células es desconocida (Rouhana *et al.*, 2010).

### 3 El neoblasto

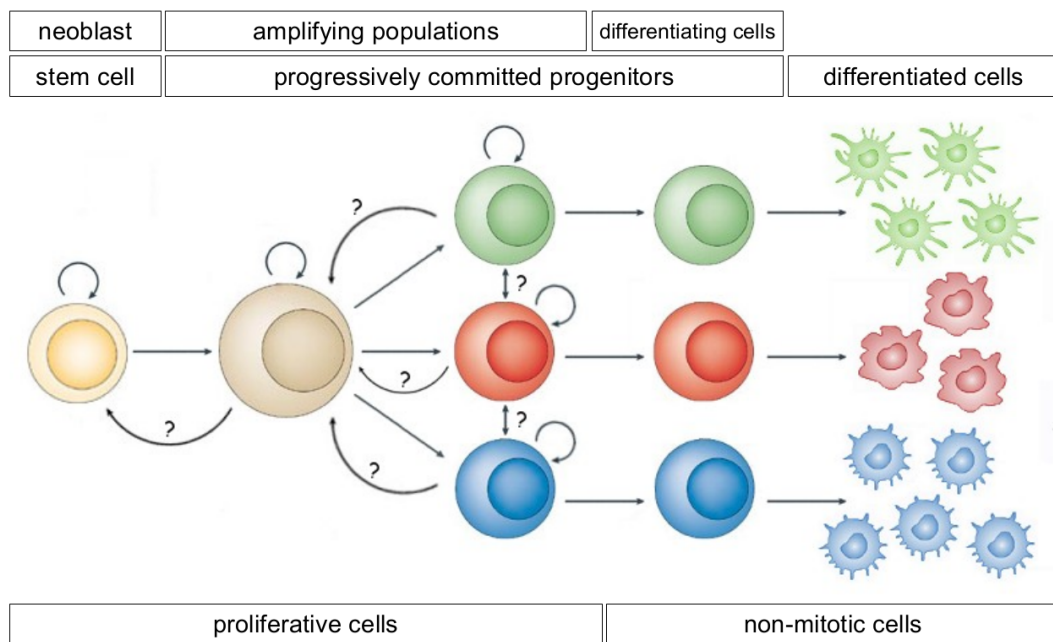
Un único punto de control parece improbable a la luz de la multitud de tipos celulares existentes en planaria. La elección del destino celular durante el desarrollo embrionario es el resultado de la restricción gradual del potencial de diferenciación. Una organización jerárquica de la diferenciación de los neoblastos parece más probable (Baguña *et al.*, 1990). Seguramente incorporando restricciones más generales de linaje al principio (como ectodermo, endodermo, mesodermo o músculo, neurona, intestino) y una especificación gradual del destino final en fases posteriores de la diferenciación.

Pueden aislarse las células de planaria en división con un contenido de DNA superior a 2N mediante marcaje y citometría de flujo. Las células aisladas de esta manera se denominan X1 (Hayashi *et al.*, 2010; Shibata *et al.*, 2012). Se ha observado heterogeneidad en la expresión de genes que codifican proteínas de unión al RNA (como *vlgA*) y la cadena pesada de la miosina. *p53* también se expresa en una minoría de las células que se dividen (Pearson & Alvarado, 2009) y estudios recientes han identificado neoblastos especializados en la regeneración del ojo *in vivo*, con factores de transcripción específicos (Lapan & Reddien, 2011) así como en el sistema excretor de protonefridios (Scimone *et al.*, 2011). Todas esas evidencias respaldan el modelo de que la población de neoblastos es heterogénea y que múltiples linajes siguen un modelo de neoblastos especializados durante la regeneración, probablemente siendo la lesión la que promueve su especialización y el contexto tisular el que determina la especialización que cada neoblasto debe adoptar. La expresión de varios factores de transcripción en subconjuntos de neoblastos durante la regeneración y el recambio tisular apuntan hacia la hipótesis de los neoblastos especializados. Esto implica que los destinos de las células del blastema que no se dividen están codificados en sus progenitores, al menos para algunos linajes. Esto implicaría que no existe una población blastémica multipotente sino que el blastema estaría formado por un mosaico de células comprometidas con un linaje. Aunque también podría tratarse de neoblastos especializados que pueden transdeterminarse, dependiendo de las circunstancias de regeneración. Hay que volver a recordar aquí los trabajos del grupo del profesor Gremigni en los años 80 (Gremigni, 1988), donde demostró la existencia de transdeterminación de células de la línea germinal a neoblastos de la línea somática. Todo ello abre una serie de nuevas cuestiones:

- ¿Los neoblastos especializados involucrados en la regeneración provienen de neoblastos especializados existentes antes de la lesión, o son neoblastos nativos inducidos a expresar factores de transcripción específicos en el lugar de la lesión?
- ¿Están los neoblastos especializados y/o las células del blastema de su progenie posmitótica determinadas de forma irreversible?
- ¿Qué grado de jerarquía existe en los linajes de la población de neoblastos?
- ¿Hay diversos grados de especialización y la determinación es progresiva o un neoblasto puede especializarse directamente?

- ¿Los neoblastos especializados se dividen y renuevan durante la recomposición de los tejidos perdidos en la regeneración y en el recambio tisular o, por contra, los neoblastos especializados representan sencillamente un estado transitorio entre las células pluripotentes que se dividen y las células diferenciadas?

En cualquier caso, ambos modelos no tienen porqué ser mutuamente excluyentes. Es posible que unos linajes sigan un modelo y otros otro. Así, las jerarquías de diferenciación pueden incorporar mecanismos de elección del destino tanto intrínsecos a la célula como mediados por señales externas en diferentes puntos de control, combinando estabilidad con flexibilidad, y es posible que una combinación de neoblastos especializados y neoblastos nativos produzca las células del blastema que conjuntamente reemplacen las células perdidas (Fig. 3.4).



**Figura 3.4 Modelo de diferenciación en distintos linajes celulares.**

Esquema en el que se combinan los dos posibles modelos de diferenciación (nativo y especializado) que podrían seguir los neoblastos. La posibilidad de dediferenciación y transdeterminación se indica con un signo de interrogación.





## 4 Proteómica

*The shift in thinking from genomics to proteomics comes with an appreciation of the difficulty of the task: Proteins are much more complicated than nucleic acids.*

Science, 291(5507):1221, 2001  
Stanley Fields

Las proteínas son los principales catalizadores, elementos estructurales, mensajeros y maquinaria molecular de la biología. El proteoma está formado por el conjunto total de proteínas expresadas en una célula o tejido (Kumar & Mann, 2009). El análisis proteómico, definido como el análisis del complemento proteico expresado por un genoma, constituye una aproximación a la descripción cuantitativa de un sistema biológico a través del análisis de los perfiles de expresión proteica. El análisis proteómico es conceptualmente atractivo por su potencial para determinar propiedades de sistemas biológicos que no son evidentes a través del análisis de las secuencias de DNA o mRNA a solas. Estas propiedades incluyen la cantidad de expresión proteica, la localización subcelular, el estado de modificación, la asociación con ligandos, así como el cambio de estas propiedades con el tiempo.

La proteómica persigue diversos objetivos, principalmente:

- Identificación de proteínas.
- Determinación de patrones específicos o diferenciales de expresión proteica.
- Determinación de modificaciones postraduccionales.
- Determinación de funciones proteicas y el patrón de las mismas.
- Determinación de interacciones proteicas.

La estrategia de los estudios proteómicos se basa en las técnicas para la separación de extractos proteicos a partir de lisados celulares bien por electroforesis bidimensional en gel de poliacrilamida, o bien por cromatografía líquida multidimensional;

ambas técnicas, unidas al desarrollo de las potentes herramientas de espectrometría de masas, permiten llevar a cabo con éxito la identificación de las proteínas y sus modificaciones postraduccionales.

La aplicación a gran escala de la proteómica no habría sido posible sin el avance en las técnicas de espectrometría de masas, que son el punto final de la mayoría de sus protocolos. Esta tecnología que permite la identificación de las proteínas que han sido antes seleccionadas por diferentes medios, basándose en la determinación de la proporción masa/carga de los fragmentos peptídicos ionizados de una proteína. Mediante espectrometría de masas pueden resolverse los péptidos que componen cada proteína y llevarse a cabo la identificación de esta. Para ello, determina la relación entre la masa y la carga de los fragmentos peptídicos ionizados de cada proteína separada, lo que genera una huella que, contrastada con un base de datos que contiene los patrones de la digestión *in silico* de las proteínas de la muestra, permite su identificación.

El objetivo último suele ser la comparación de los patrones de expresión proteica de diferentes muestras para obtener aquellas expresadas diferencialmente, candidatas a ser responsables de ese fenotipo particular. Así, se obtiene por un lado una visión global del proteoma que define un estado particular y puede pasarse, si se cree necesario, a analizar pormenorizadamente cada proteína particular con las técnicas que sean más convenientes.

El proceso que se realiza en un estudio proteómico típico consta de diferentes fases (Fig. 4.1):

### 1. Proceso de separación

Mediante electroforesis bidimensional se separan las proteínas en dos fases aprovechando las características biofísicas de las mismas. La primera fase separa las proteínas según su punto isoeléctrico (pI), punto en el que la carga total de la proteína es neutra. Para ello se hace pasar un campo eléctrico por una matriz que presenta diferente pH a lo largo del recorrido por el que transitarán las proteínas. Así, las proteínas migran movidas por el campo hasta alcanzar el pH que coincide con su pI. Es lo que se denomina proceso de isoelectroenfoque. Uno de los grandes avances de la técnica fue la descripción del protocolo para la electroforesis con gradientes de pH inmovilizados en tiras (Görg *et al.*, 2000), lo que facilitó de manera considerable el proceso.

La segunda parte, una vez definido el isoelectroenfoque, consiste en separar los grupos por peso molecular, haciendo una electroforesis clásica en gel de poliacrilamida (Görg *et al.*, 2004).

### 2. Detección y comparación

Una vez finalizada la electroforesis se marcan las proteínas con alguna tinción genérica de proteínas dependiendo del grado de sensibilidad deseado como, por ejemplo, azul de bromofenol o tinción con plata. A continuación, mediante un *software* específico, se comparan los geles para detectar los puntos o *spots* que contienen las proteínas diferencialmente expresadas entre las condiciones

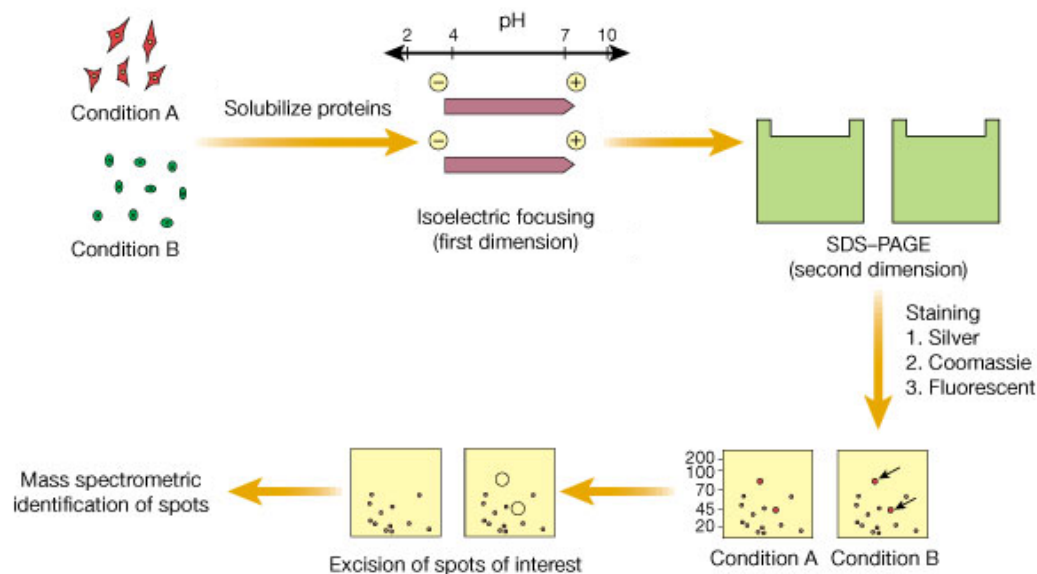
experimentales. En la práctica, a menudo es difícil correlacionar los puntos de los diferentes geles dada la ineludible variabilidad técnica que existe entre ellos, por mucho cuidado que se ponga en realizar el protocolo siempre de manera idéntica. Este problema se ha superado gracias a una técnica mediante la cual se pueden marcar dos muestras con fluorocromos diferentes, conocida como *Difference Gel Electrophoresis* (DIGE) (Unlu *et al.*, 1997). Con este método pueden analizarse ambas muestras de proteínas en el mismo gel, eliminando la variabilidad técnica entre geles y, por lo tanto, haciendo mucho más fácil y fiable la comparación.

### **3. Escisión, digestión y análisis experimental**

Los puntos seleccionados se recortan mecánicamente del gel de acrilamida. Mediante protocolos adecuados para eliminar la acrilamida y otros restos, se extraen las proteínas. Estas son sometidas a una digestión con endoproteasas específicas (como la tripsina) para generar una serie de péptidos que serán analizados con el espectrómetro de masas. Este instrumento determina la relación masa/carga de los péptidos convirtiendo los péptidos en iones en fase gaseosa y haciéndolos pasar por un campo electromagnético. El tiempo de vuelo de los iones y una serie de sensores permiten determinar la masa/carga, que es proporcional a la composición de los átomos constituyentes y su carga iónica.

### **4. Análisis computacional**

El espectrómetro de masas ofrece los valores referidos a la relación masa/carga del grupo de péptidos resultantes de la digestión la proteína presente en cada *spot*. La identificación de las proteínas se realiza gracias a programas diseñados específicamente para este fin, como MASCOT (Pappin *et al.*, 2013). A partir de una base de datos de proteínas, sus algoritmos realizan una digestión *in silico* de las mismas considerando el mismo enzima de restricción que el que se usó para fragmentar las proteínas experimentalmente y calculan los valores de la relación masa/carga teóricos de los péptidos resultantes. Comparando la información espectrométrica obtenida experimentalmente con los valores *in silico* pueden dilucidarse las proteínas presentes en cada *spot*.



**Figura 4.1 Esquema de la aproximación proteómica empleada en este proyecto.**

Tras obtener el tejido de cada condición experimental, A y B, se solubilizan las proteínas y estas se separan en una primera dimensión basándose en su pl. A continuación, se cargan en un gel SDS-PAGE donde las proteínas se desnaturalizan y se separan de acuerdo a su tamaño. Los gels se fijan y las proteínas se visualizan por tinción (azul de Coomassie, tinción de plata o fluorescencia). Después de la tinción, los *spots* resultantes se comparan con la ayuda de programas específicos y se escinden aquellos que se consideran de interés, generalmente los expresados diferencialmente entre condiciones. En el caso de utilizar fluorescencia con el sistema DIGE, la comparación es directa sobre un mismo gel. Los *spots* extraídos se someten finalmente a análisis mediante espectrometría de masas para identificar las proteínas que contienen. Así, este sistema de análisis de la expresión diferencial de proteínas permite la comparación de las proteínas presentes entre dos condiciones o muestras. Adaptado de Pandey & Mann (2000).

## 4.1. Separación de las proteínas de la muestra

### 4.1.1. 2D SDS-PAGE

La metodología 2D SDS-PAGE (*two-dimensional sodium dodecyl sulfate–polyacrylamide gel electrophoresis*) se basa en una separación mediante electroforesis bidimensional en gel de poliacrilamida por gradiente de pH y punto isoeléctrico. Obtener gels reproducibles mediante electroforesis en dos dimensiones es técnicamente complejo y a menudo complica la comparación entre gels de diferentes muestras, incluso preparados en las mismas condiciones. Proteínas muy básicas o muy ácidas, muy grandes o muy pequeñas, así como aquellas poco abundantes, son habitualmente difíciles o imposibles de analizar mediante 2D SDS-PAGE (Miller *et al.*, 2006; Patton & Beechem, 2002).

La implementación más común de análisis proteómico es la combinación de electroforesis en gel de dos dimensiones (2DE) (isoelectric focusing-sodium dodecyl sulfate [SDS]-polyacrylamide gel electrophoresis) para la separación y la cuantificación de proteínas con métodos analíticos para su identificación. 2DE permite la separación, visualización y cuantificación de miles de proteínas en un único gel de manera reproducible. Por sí mismo, 2DE es estrictamente una técnica descriptiva. La combinación de 2DE con técnicas de análisis proteico añade la posibilidad de establecer la identidad de cada proteína por separado y, así, en combinación con el análisis cuantitativo del mRNA, correlacionar también las medidas cuantitativas de proteína y expresión del mRNA.

Determinar los niveles de expresión relativos de una proteína mediante 2DE convencional requiere un enfoque isoelectrónico, electroforesis en gel de SDS-poliacrilamida, tinción, fijación, densitometría y un cuidadoso emparejamiento de los mismos *spots* en dos o más geles. Los *spots* diferencialmente expresados se escinden, se digieren enzimáticamente y los péptidos resultantes se identifican usando espectrometría de masas. Se requiere una gran pericia para obtener geles reproducibles y 2DE está limitada generalmente a proteínas que no son ni demasiado ácidas, ni demasiado básicas, ni demasiado hidrofóbicas, y con un tamaño de entre 10 y 200kDa, de manera que puedan ser separadas de manera fiable en el gel. Por otro lado, esta aproximación detecta solo aquellas proteínas que se expresan a un nivel relativamente alto y que tienen una vida media larga. Dadas todas estas limitaciones, la tecnología de 2DE convencional tiene un potencial limitado para análisis proteómicos a gran escala.

Existen diferentes técnicas de visualización para las proteínas separadas por SDS-PAGE. Los métodos de tinción clásicos no son lo bastante sensibles (azul de Coomassie) o no permiten cuantificar en un amplio rango de intensidades (tinción de plata) como para inferir directamente los niveles de abundancia de los *spots* (Maier *et al.*, 2009). La tinción de plata es menos cuantitativa que el azul de Coomassie pero es más sensitiva y compatible con el análisis por espectrometría de masas.

Los estudios clásicos sobre proteómica todavía utilizan esta aproximación de separación de las proteínas mediante electroforesis de dos dimensiones seguida de un análisis del gel para detectar los puntos diferencialmente expresados e identificación proteica por espectrometría de masas. Sin embargo, esta estrategia presenta ciertos inconvenientes, el más importante de los cuales se refiere a la reproducibilidad de la separación de proteínas, lo cual dificulta la detección de las proteínas expresadas diferencialmente.

#### 4.1.2. *Difference gel electrophoresis*

Los métodos de tinción fluorescentes resuelven estos y otros problemas. El método *Difference gel electrophoresis* (DIGE) utiliza tinciones fluorescentes de espectros diferentes (como Cy3 y Cy5) para marcar dos muestras de proteínas *in vitro* antes de la electroforesis bidimensional. Su principal ventaja respecto 2DE convencional es que tanto la muestra control como la problema corren en un único gel de poli-acrilamida. Las muestras son detectadas por separado pero pueden superponerse

perfectamente sin ninguna deformación de los geles. Gracias a que varias muestras pueden ser separadas en un mismo gel, la comparación directa entre diferentes muestras es mucho más sencilla. Esto aumenta considerablemente la fiabilidad con la que los cambios entre muestras pueden ser detectados y cuantificados. En consecuencia, es más sensible y permite el análisis cuantitativo de las proteínas en geles de dos dimensiones en un rango más amplio (Larbi & Jefferies, 2009; Timms & Cramer, 2008). En *spots* grandes pueden detectarse cambios en el nivel de expresión relativo de una proteína de hasta 1,2 veces. Debido a que la detección se basa en la fluorescencia, DIGE tiene un rango dinámico de 10.000, lo que permite el análisis de la expresión diferencial de proteínas que están presentes a una baja concentración. El límite de detección de DIGE para la cuantificación de la expresión de proteínas está entre 0,25 y 0,95ng de proteína, que es similar al de la tinción de plata.

Al igual que ocurre con la aproximación original, la cuantificación de los *spots* requiere de la espectrometría de masas para asignar información cuantitativa a cada proteína individualmente.

### 4.2. Espectrometría de masas

Las técnicas de espectrometría de masas para la identificación y la cuantificación de proteínas han experimentado grandes avances (Aebersold & Mann, 2003; Coombs, 2011; Ong & Mann, 2005). La complejidad de la muestra se reduce al separar la mezcla de las proteínas digeridas por cromatografía líquida antes del análisis mediante espectrometría de masas de los péptidos aislados. Existen varios diseños de espectrómetros de masas, cada uno con ventajas específicas en cuanto a precisión de masas, rango dinámico, poder de resolución y sensibilidad (Domon & Aebersold, 2006).

La introducción de las técnicas de análisis proteico de espectrometría de masas han mejorado la cantidad y la sensibilidad de las proteínas identificadas a un nivel que permite el análisis a gran escala de proteínas separadas por electroforesis bidimensional. Las técnicas han alcanzado un nivel de sensibilidad que permite la identificación de prácticamente cualquier proteína que sea detectable en un gel mediante una tinción convencional. La tecnología actual de análisis de proteínas se basa en la generación por espectrometría de masas de patrones de fragmentos peptídicos que son idiotípicos para la secuencia de una proteína. La identidad de la proteína se establece correlacionando dichos patrones con las bases de datos de secuencias con la ayuda de programas que se encargan de identificar las proteínas sin intervención humana.

La espectrometría de masas en muestras de proteínas no es intrínsecamente cuantitativa, aunque se han desarrollado varios métodos para solventar este problema. Básicamente se clasifican en dos grupos, dependiendo de si utilizan marcaje con isótopos estables o no. El marcaje con isótopos estables se puede conseguir bien vía marcaje metabólico de las proteínas *in vivo* o en cultivo celular (*stable isotope labeling by amino acids in cell culture*, SILAC; Ong *et al.*, 2002), o bien mediante marcaje

químico de los péptidos después de la digestión con tripsina (ICAT o iTRAQ; Gygi *et al.*, 1999b; Wiese *et al.*, 2007). Se ha utilizado SILAC para cuantificar muestras muy complejas, como el proteoma de levadura (de Godoy *et al.*, 2008) o células madre de ratón (Graumann *et al.*, 2007).

La cuantificación de proteínas sin marcaje se basa en la información obtenida directamente de las lecturas del espectrómetro de masas. Un método simple implica contar y comparar el número de fragmentos del espectro iónico de un péptido concreto. Este método se fundamenta en que el ratio MS/MS para un péptido en particular está directamente relacionado con la abundancia de su precursor iónico en la muestra. Una aproximación ligerametne diferente es el índice de abundancia proteica (PAI), en el que el número de péptidos identificados en un experimento MS/MS está en relación con el número de péptidos observables teóricos para cada proteína. Las proteínas también pueden cuantificarse midiendo y comparando la intensidad de la señal iónica de los precursores peptídicos integrada a lo largo de su tiempo de retención. Este último método se basa en el alineamiento de los cromatogramas y a menudo requiere de un software muy específico. En contraste con los isótopos estables, la cuantificación sin marcaje depende en mayor medida de réplicas técnicas. La información aportada por la identificación de los péptidos puede servir como una indicación aproximada de la cantidad de proteína en la muestra. La idea básica es que la abundancia de cada proteína escala con el número de péptidos identificados.

## 4.3. Análisis computacional

La obtención de las proteínas de una muestra suele incluir un paso de digestión enzimática, generándose péptidos como resultado de la proteólisis que, además, pueden incluir modificaciones postraduccionales (fosforilaciones, etc. . .). Los péptidos no son entidades biológicas funcionales. Pueden ser mapeados en las posiciones del genoma que los codifican. De esta manera los péptidos proporcionan evidencia de que el gen realmente se expresa y no se trata, por ejemplo, de un pseudogén. Esto es de vital importancia ya que una gran parte de los genes predichos en muchos genomas de eucarióticos todavía no cuentan con ninguna información experimental proteica asociada.

El análisis bioinformático típicamente incluye la integración de los datos proteómicos con bases de datos de anotaciones como Gene Ontology (Ashburner *et al.*, 2000), de dominios proteicos como InterPro (Mitchell *et al.*, 2014) y PFAM (Finn *et al.*, 2013), así como de rutas metabólicas (KEGG) (Kanehisa *et al.*, 2013). Una vez que se ha asignado una función a una proteína, pueden inferirse funciones relacionadas para otras proteínas cuya secuencia de aminoácidos es similar a la proteína original. Este método de búsqueda por homología (mediante el uso de BLAST (Camacho *et al.*, 2009), por ejemplo) se utiliza para extender las anotaciones funcionales de una proteína a otras parecidas que, presumiblemente, descienden de la misma proteína ancestral común. Usando esta metodología pueden asignarse funciones a entre



el 40 y el 70 % de de las nuevas secuencias genómicas. Aún así, es difícil predecir la función de una proteína partiendo únicamente de su estructura tridimensional y el abuso de la predicción basada en la homología puede provocar una onda expansiva de errores en la anotación (Jones *et al.*, 2007). Por esto, se hace necesario en último término un análisis funcional más fino de las proteínas halladas, usando técnicas más complejas que permiten modelar dominios como los modelos de Markov en los que se basa PFAM, por ejemplo.

### 4.4. Aproximación proteómica

La mayoría de lo que se conoce de la biología de los neoblastos proviene de los datos moleculares a partir de cDNA y análisis genómicos. De hecho, existe un borrador de genoma ensamblado (NCBI Genome Database: Schmidtea mediterranea, 2014; Robb *et al.*, 2015; Sánchez Alvarado *et al.*, 2002; The Schmidtea mediterranea Genome Sequencing Project, 2014) y varios estudios transcriptómicos en la planaria *S. mediterranea* (Abril *et al.*, 2010; Adamidi *et al.*, 2011; Blythe *et al.*, 2010; Kao *et al.*, 2013; Labbé *et al.*, 2012; Resch *et al.*, 2012; Rouhana *et al.*, 2012; Sandmann *et al.*, 2011). Sin embargo, el genoma está muy fragmentado todavía y las anotaciones existentes están basadas mayormente en predicciones computacionales generadas por MAKER (Cantarel *et al.*, 2007), dificultando tanto los análisis genómicos y transcriptómicos como los proteómicos.

Las aproximaciones proteómicas son reconocidas como extremadamente útiles para el estudio de la regeneración ya que hay una contribución relevante de procesos que frecuentemente implican un amplio rango de modificaciones postranscripcionales. La complejidad aumenta al considerar que las proteínas también pueden verse sometidas a cambios postraduccionales y que diferentes proteínas pueden interactuar entre ellas para dar lugar a complejos proteicos, formando una estructura cuaternaria.

La elección de esta estrategia se debió a diferentes razones. La primera es la falta de un estudio en profundidad de las proteínas de nuestro organismo modelo. En el momento de plantear este proyecto existía un vacío de casi dos décadas en lo que a estudios proteómicos en planaria se refiere. Desde los trabajos de Collet y Baguñà (Collet & Baguñà, 1987, 1993) no se había vuelto a abordar ningún estudio desde una vertiente proteómica. Las mejoras tecnológicas, especialmente en espectrometría de masas, así como la posibilidad de disponer de secuencias genómicas en nuestra especie de referencia, *S. mediterranea*, permitían, o casi exigían, retomar con una visión actualizada esos trabajos, en este caso, centrada en el proteoma del neoblasto.

Otra razón es la de complementar las técnicas genómicas que no son capaces de capturar la complejidad del conjunto de proteínas que están actuando en una situación concreta, dado que durante el trayecto de gen a proteína tienen lugar pasos de control y regulación muy importantes que escapan a las técnicas genómicas, como las muchas modificaciones postranscripcionales y postraduccionales, la movilización entre compartimentos celulares, la dinámica de proteólisis y una gran lista de proce-

Los hechos que hacen que las características de un sistema biológico no estén completamente relacionadas con el estado de la expresión génica. Todo ello no implica ni mucho menos que los datos aportados por los estudios genéticos no sean válidos, sino al contrario, que pueden ser mejor interpretados a la luz de los datos aportados por las técnicas proteómicas. Además, es interesante realizar este tipo de estudios ahora que el genoma de planaria está siendo ensamblado, dado que las nuevas secuencias pueden utilizarse para comprobar la validez del ensamblado de los fragmentos, así como para evaluar las anotaciones de los genes predichos computacionalmente para este genoma.

Como ya se ha mencionado, una de las aplicaciones de la proteómica es la determinación de proteínas específicas a través del análisis de la expresión diferencial entre muestras. Este proceso permite comparar la composición proteica de dos muestras diferentes, en nuestro caso con presencia y ausencia de neoblastos.

Por último, hasta hace poco la investigación biomédica ha estado centrada en los genes, fallando a menudo en capturar características y funciones específicas de las proteínas, las cuales surgen de la mayor diversidad de proteínas respecto a los genes. La relevancia de las técnicas de proteómica no se limita únicamente pues a modelos de interés para la investigación básica. Dos extensas revisiones remarcan la importancia de la proteómica en los estudios con células madre embrionarias humanas (Hughes *et al.*, 2011) y en con células madre cardíacas (Stastna *et al.*, 2009).



# 5 Secuenciación de RNA y análisis de la expresión génica

*A big computer, a complex algorithm and a long time does not equal science.*

Robert Gentleman

Como se comentó en la introducción, el dogma central de la biología sugiere que hay una relación directa entre los niveles de mRNA y de proteína. Esta presunción es la base de los estudios transcriptómicos de expresión, a menudo con el objetivo de identificar genes inhibidos o sobreexpresados entre varias condiciones. La asunción subyacente es que las diferencias en los niveles de mRNA se manifiestan en diferentes fenotipos como resultado de diferencias en los niveles de proteína (Gry *et al.*, 2009). Más adelante se discutirá que esto no siempre es necesariamente así.

## 5.1. *Microarrays*

Hasta hace pocos años, únicamente podía analizarse la expresión de un número significativo de genes simultáneamente mediante la tecnología de *microarrays* (Hoheisel, 2006). Esta se utilizó en dos especies diferentes de planarias para comparar la expresión de diferentes transcritos entre animales normales e irradiados, obteniendo así un catálogo de genes potencialmente involucrados en la biología del neoblasto (Eisenhoffer *et al.*, 2008; Rossi *et al.*, 2007). No es sorprendente que el conjunto de genes definido por estos estudios esté enriquecido con aquellos que intervienen en el ciclo celular (los neoblastos son las únicas células que proliferan) pero también con genes conocidos por regular mecanismos del desarrollo en otros animales. Estos últimos podrían estar implicados en el mantenimiento de la pluripotencia y la diferenciación de los neoblastos. Sin embargo, sin marcadores que permitan distinguir a los neoblastos de su descendencia es difícil asignar a estos genes una función u otra. El segundo de estos estudios (Eisenhoffer *et al.*, 2008), combinó el

análisis de la expresión en *microarrays* durante un *time course* de irradiación con un marcaje doble de las células con BrdU (Newmark & Sánchez Alvarado, 2000) y potenciales marcadores de la progenie de los neoplastos. Esta estrategia permitió la identificación de los primeros marcadores diferenciales de los neoplastos y su descendencia.

Sin embargo, los *microarrays* (Leung & Cavalieri, 2003) constituyen una tecnología limitada. En primer lugar, por la necesidad de un conocimiento previo de aquellas secuencias cuya expresión desea analizarse para la síntesis de sondas con las cuales debe hibridar el RNA de cada transcrito; lo que, además de los problemas de hibridación cruzada, solo permite explorar un número de genes relativamente bajo. Por otro lado, la cuantificación de la expresión se realiza midiendo la intensidad de la fluorescencia emitida, que será mayor cuanto mayor sea la expresión del gen y, por tanto, más moléculas de RNA hayan hibridado con su sonda correspondiente. Esta forma de medición es analógica y conlleva una serie de problemas intrínsecos asociados, como la dificultad para calcular con precisión la intensidad fluorescente, especialmente para genes cuya expresión es baja o, en el caso contrario, la saturación de la señal para genes con elevada expresión.

La secuenciación directa de los transcritos para determinar la expresión de los genes ha sobrepasado y reemplazado en la mayoría de análisis a los *microarrays*. Aunque estos siguen utilizándose rutinariamente en diagnóstico clínico, las nuevas tecnologías de secuenciación son más fiables y precisas en la cuantificación (Yu *et al.*, 2015). Además, con esta técnica, los transcritos se caracterizan por su secuencia, de manera que se pueden identificar prácticamente todos los transcritos presentes en una muestra, conocidos o no y, dado el enorme número de muestras que pueden ser analizadas simultáneamente, las técnicas basadas en secuenciación permiten detectar RNAs poco abundantes, *small RNAs*, etc. Otra ventaja es que, gracias a que la cuantificación es proporcional a las secuencias de la muestra, no hay posibilidad de saturación a niveles de expresión elevados. Aún así, esto puede representar un problema para la detección de transcritos muy poco expresados.

## 5.2. *Next-generation sequencing*

Una nueva generación de secuenciadores de alto rendimiento ofrecen la posibilidad de analizar y cuantificar transcriptomas de miles de transcritos en un solo experimento, proporcionando cada vez más cantidad de información y con mayor rapidez (Mortazavi *et al.*, 2008). La secuenciación se lleva a cabo de forma masiva gracias a la paralelización de la reacción de la replicación del DNA utilizando diferentes técnicas (Kukurba & Montgomery, 2015). Las secuencias así obtenidas se alinean contra el genoma o el transcriptoma de referencia de la especie y se cuantifican los transcritos. Dada la enorme cantidad de secuencias que pueden obtenerse, es posible detectar incluso RNAs muy poco abundantes. Con esta técnica, los transcritos se caracterizan por su secuencia, sin necesidad de conocerla previamente como ocurre con los *microarrays*. Los avances en esta tecnología están reduciendo el coste por

base secuenciada, mucho más barato que el de los *microarrays* (Maier *et al.*, 2009). Esta nueva generación de tecnologías de secuenciación de alto rendimiento se conoce en inglés como *Next-generation high-throughput sequencing* o sencillamente *Next-generation sequencing* (NGS) y está considerada la segunda generación por suceder a la secuenciación Sanger.

La técnica original de *Serial analysis of gene expression* (SAGE) (Velculescu *et al.*, 1995) demostró el gran potencial que esta tecnología presentaba pero requería la difícil ligación de un gran número de fragmentos cortos de DNA, su clonación y su posterior secuenciación por el método tradicional Sanger. Aplicando la tecnología de NGS al concepto de la metodología SAGE, se hizo posible el análisis masivo de transcritos de RNA en muestras biológicas, obteniendo directamente de cada transcrito secuencias cortas (*tags*) de 20-35 pares de bases de longitud (Mortazavi *et al.*, 2008).

La tecnología de NGS proporciona mayor velocidad y rendimiento, de manera que proyectos de secuenciación genómica que tardaban años con la técnica Sanger se pueden completar ahora en cuestión de semanas. El coste por base es, además, muy inferior comparado con la secuenciación Sanger y desde su inicio se ha ido reduciendo a la par que aumentaba la cantidad de *reads* capaz de obtenerse en un solo *run* y se reducía el tiempo de secuenciación. Sin embargo, la fidelidad es un factor a tener en cuenta en el que todavía no se ha igualado a la secuenciación Sanger. Otra limitación en algunas aplicaciones son los *reads* excesivamente cortos, una asignación de bases (*base calling*) no uniforme en la secuencia (*coverage* parcial), de peor calidad (particularmente en el extremo 3' en tecnologías de *reads* de longitud corta) y de baja precisión en general en segmentos de homopolímeros. El enorme volumen de datos generados por estos sistemas (por encima de una gigabase por *run*) en forma de *reads* cortos representa otro desafío a los desarrolladores de los algoritmos necesarios para filtrarlos, procesarlos y mapearlos o ensamblarlos.

Este proyecto se planteó con el nacimiento de estas nuevas tecnologías de secuenciación, a fin de complementar el estudio proteómico previo y encontrar respuesta a preguntas que la proteómica parece no poder responder en su actual estado de desarrollo. En el momento de iniciar el mismo eran tres las tecnologías disponibles, representada cada una por un sistema diferente de secuenciación. La principal ventaja de la tecnología de RNA-Seq, que estas primeras plataformas ya presentaban en común, es la capacidad para determinar la secuencia a partir de la amplificación de fragmentos únicos de DNA; evitando la necesidad de clonarlos como ocurría en el protocolo de SAGE original (Velculescu *et al.*, 1995). El RNA es amplificado después de convertirlo a cDNA.

A continuación repasaremos sus fundamentos y particularidades, deteniéndonos con más detalle en las dos primeras, que fueron las elegidas para llevar a cabo este proyecto por los motivos que se expondrán.

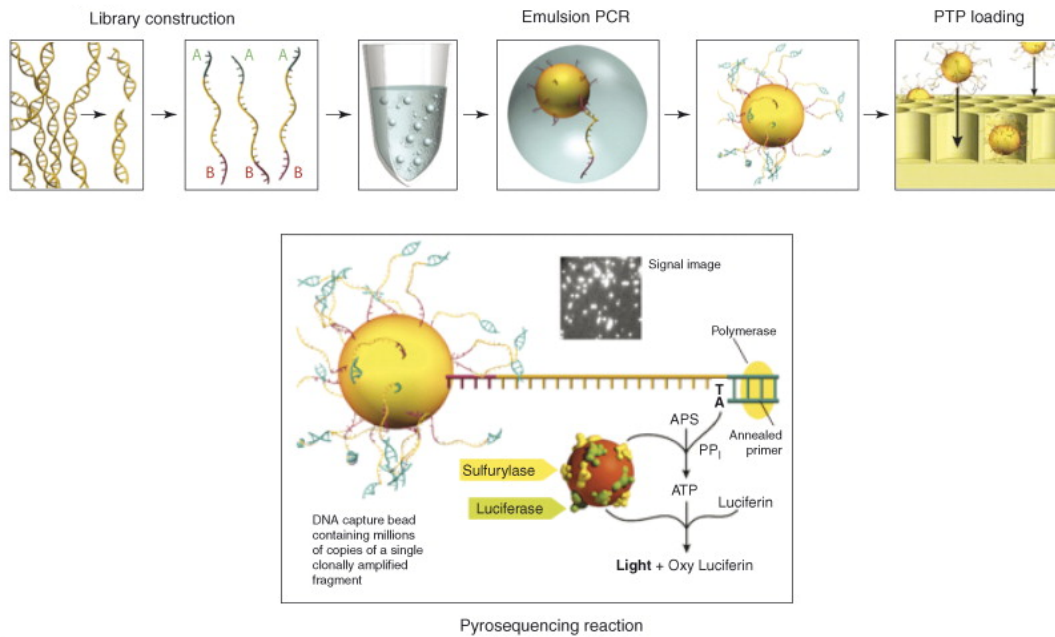
### 5.2.1. 454 GenomeSequencer FLX instrument (Roche Applied Science)

Esta tecnología de secuenciación se basa en el principio de detección mediante pirofosfato. El GenomeSequencer (GS) se introdujo en 2005, desarrollado por 454 Life Sciences, como el primer sistema de NGS en el mercado. En este sistema, los fragmentos de DNA se ligan con adaptadores específicos que provocan la unión del fragmento a una microesfera o *bead* (Fig. 5.1). Se carga una emulsión de PCR para amplificar el fragmento, con una microgota de agua que contiene un *bead* y los reactivos de PCR embebidos en aceite. La amplificación es necesaria para obtener una intensidad de señal luminosa suficiente que permita una detección fiable en los pasos de la reacción de secuenciación por síntesis. Cuando los ciclos de PCR se completan y después de la desnaturalización, cada *bead* con su propio fragmento amplificado se coloca al final del extremo de la fibra óptica de un chip compuesto por fibras de vidrio. Cada fibra de vidrio conduce la luz hasta el extremo que termina en una cámara CCD (*charge-coupled device*), que posibilita la detección de la luz emitida por medio de una reacción enzimática acoplada a la luciferasa. Así, cada *bead* se sitúa en una posición trazable en el chip, que contiene varios cientos de miles de fibras unidas a un *bead*. En el siguiente paso se añaden a los *beads* la enzima polimerasa y un *primer* (cebador), así como un nucleótido sin marcar a la mezcla de reacción a todos los *beads* del chip, de manera que la síntesis de la cadena complementaria pueda comenzar. La incorporación de la siguiente base por parte de la polimerasa en la cadena creciente libera un grupo pirofosfato, que se puede detectar como luz emitida. Conociendo la identidad del nucleótido aportado en cada paso, una señal luminosa indica la siguiente base incorporada en la cadena de DNA que se está alargando.

En su primera versión esta plataforma era capaz de generar secuencias de hasta 400-500 bases, en forma de *single-end reads*. Posteriormente se llegó a 800-1.000bp. Sus principales inconvenientes son el coste relativamente elevado y una precisión baja en las regiones de más de 4 o 5 bases consecutivas idénticas (homopolímeros).

### 5.2.2. Illumina (Solexa) Genome Analyzer

La plataforma de secuenciación Solexa fue comercializada en 2006, adquiriéndola Illumina a principios de 2007. Su principio se fundamenta en la química de secuenciación por síntesis, con nucleótidos de terminadores reversibles para las cuatro bases, cada uno con un marcador fluorescente distinto, y una polimerasa de DNA capaz de incorporarlos. Los fragmentos de ADN se ligan a ambos extremos a adaptadores y, después de su desnaturalización, son inmovilizados por un extremo a un soporte sólido (*flowcell*) (Fig. 5.2). La superficie del soporte está densamente recubierta con los adaptadores y sus adaptadores complementarios. Cada fragmento de cadena única, inmovilizado por un extremo en la superficie, crea una estructura de puente hibridando su extremo libre con el adaptador complementario en la superficie del soporte. En la mezcla que contiene los reactivos de amplificación de PCR, los adap-

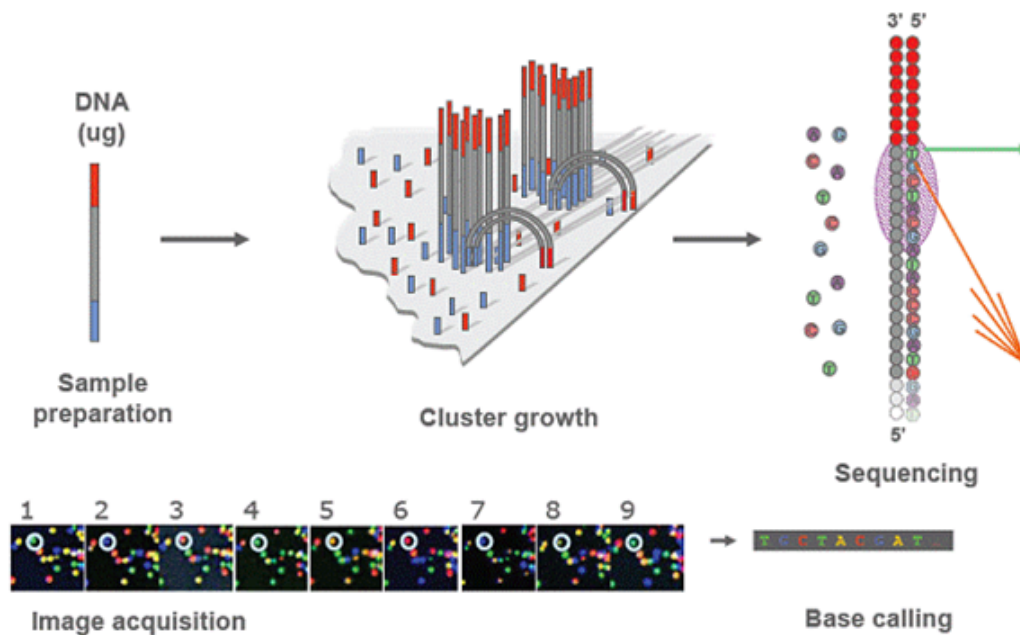


**Figura 5.1 Pirosecuenciación en 454 GS FLX.**

Durante la construcción de la librería se ligan adaptadores específicos a los fragmentos de DNA y se unen los *beads* de amplificación con el DNA en una emulsión de PCR para amplificar los fragmentos antes de secuenciarlos. Los *beads* se cargan en una placa (*picotiter plate*, PTP). El panel inferior ilustra la reacción de pirosecuenciación que tiene lugar durante la incorporación de los nucleótidos, que reporta la secuencia por síntesis. Adaptado de [www.454.com](http://www.454.com).

tadores en la superficie actúan como *primers* para la siguiente amplificación de PCR. Nuevamente, una amplificación es necesaria para obtener suficiente intensidad de señal luminosa que permita una detección fiable de las bases añadidas. Después de muchos ciclos de PCR se forman *clusters* aleatorios de alrededor de 1.000 copias de fragmentos de cadena única de DNA en la superficie. Los reactivos para la secuenciación y la síntesis del DNA se proporcionan en la superficie y comprenden *primers*, cuatro nucleótidos de terminador reversible, cada uno con un marcador fluorescente distinto, y la polimerasa de DNA. Después de su incorporación en la cadena de DNA, el nucleótido terminal, así como su posición en la superficie del soporte, se detectan e identifican a través de su color fluorescente mediante una cámara CCD. El grupo terminal del extremo 3' de la base y el marcador fluorescente se eliminan de la base y se repite el ciclo de síntesis. La longitud de *read* conseguida mediante reacciones repetidas era de unos 35 nucleótidos. En la actualidad se puede llegar hasta unos 150bp y variaciones de la técnica permiten obtener *reads* de hasta 300bp (MiSeq). Se pueden detectar al menos 40 millones de *clusters* en paralelo, lo que resulta en un rendimiento de varias gigabases por soporte.





**Figura 5.2 Secuenciación en Illumina Genome Analyzer.**

Durante la preparación de la librería la muestra se fragmenta y a cada fragmento se ligan adaptadores en sus extremos. La librería se aplica entonces a la superficie de la *flow cell* y, una vez unidos, los fragmentos de DNA forman puentes moleculares que son amplificados repetidamente a través de un proceso de amplificación isotérmica, lo que conduce a la formación de *clusters* de fragmentos idénticos. Finalmente, las moléculas de DNA amplificadas se desnaturalizan para permitir la unión del *primer* de secuenciación y se ven sometidas a una secuenciación por síntesis, utilizando nucleótidos marcados en 3' con un fluorocromo distinto cada uno. La señal luminosa de cada nucleótido añadido es recogida por una cámara, a partir de la cual se produce la secuencia de la cadena amplificada. Adaptado de [www.illumina.com](http://www.illumina.com).

### 5.2.3. Applied Biosystems ABI SOLiD system

El sistema de secuenciación ABI SOLiD fue introducido en otoño de 2007. Es una plataforma que se sirve también de la química basada en la ligación.

Los fragmentos de DNA se ligan a adaptadores y estos se unen a *beads*. Una microgota de agua en emulsión de aceite contiene los reactivos de amplificación y un único fragmento unido por *bead*; los fragmentos de DNA en los *beads* son amplificados por la emulsión de PCR. Después de la desnaturalización del DNA, los *beads* se depositan en un soporte con superficie de vidrio. En un primer paso, un *primer* hibrida con el adaptador. A continuación, una mezcla de oligonucleótidos octaméricos hibrida también con los fragmentos de DNA y se añade la mezcla de ligación. En estos octámeros, el doblete de la cuarta y quinta base se caracteriza por una de cuatro etiquetas fluorescentes al final del octámero. Después de la detección

de la fluorescencia de la etiqueta, se determinan las bases 4 y 5 de la secuencia. El octámero ligado se corta después de la quinta base, eliminando el marcaje fluorescente. Los ciclos de hibridación y ligación se repiten, esta vez determinando las bases 9 y 10 de la secuencia; en el siguiente ciclo las bases 14 y 15, etc. El proceso de secuenciación puede continuarse de la misma manera con otro *primer*, con una base menos que el anterior, permitiendo determinar en los sucesivos ciclos, las bases 3 y 4, 8 y 9, 13 y 14. La longitud del *read* así obtenido es de 35 bases. Puesto que cada base se determina con una marca fluorescente diferente, el ratio de error se reduce. Las secuencias se pueden determinar en paralelo para más de 50 millones de *clusters* de *beads*, resultando en un gran rendimiento, del orden de gigabases por *run*.

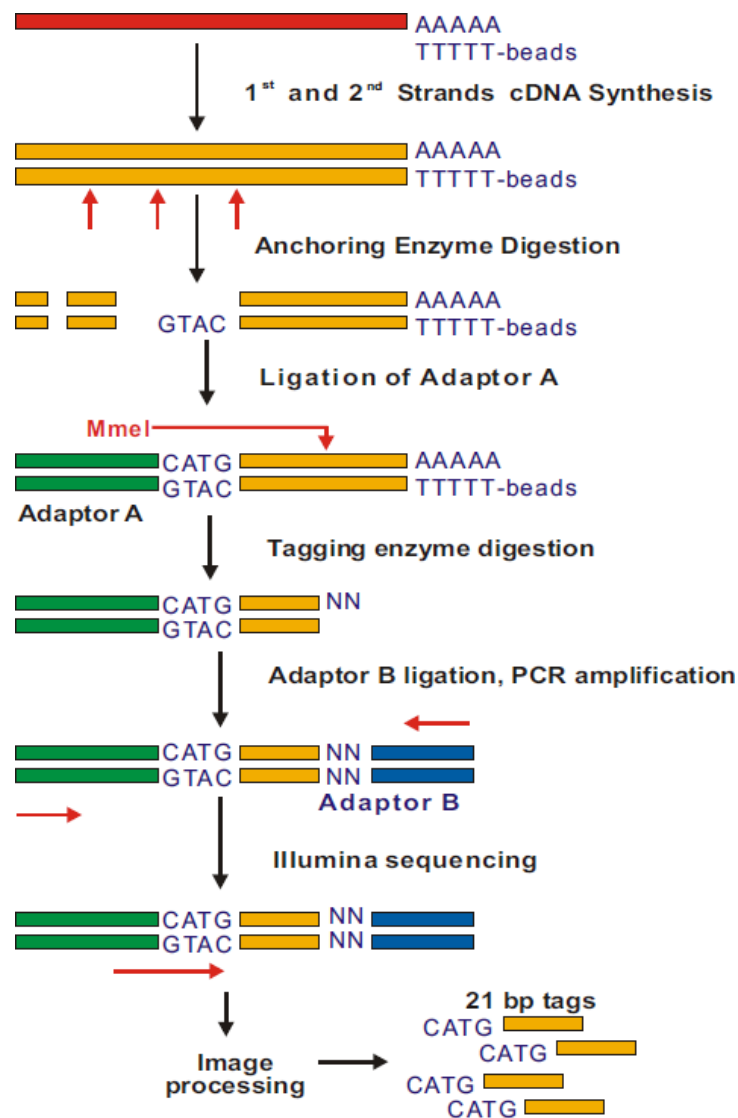
#### 5.2.4. Estado actual

Observando este escenario en retrospectiva, Illumina se ha convertido en la plataforma más utilizada, siendo actualmente los modelos de secuenciadores HiSeq y MiSeq, en sus diferentes series, casi un estándar de secuenciación mundial. ABI SOLiD perdió la carrera en gran parte debido al formato de salida de sus datos, ya que en lugar de generar ficheros de secuencia en el estándar FASTQ, utilizaba un formato propio codificado en lo que se denominó *color space*, para el cual ni existían ni se desarrollaron a tiempo los programas de análisis necesarios, mientras que sí existían numerosas herramientas bioinformáticas para procesar los archivos FASTQ. 454 de Roche, aún siendo la tecnología que proporcionaba una mayor longitud de secuencia, también resultaba la más cara y fue desplazada a medida que los fragmentos secuenciados por Illumina aumentaban de tamaño. La secuenciación Illumina permitía además, en comparación con la 454, cuantificar fácilmente las copias de los RNAs de las muestras.

Se ha abierto pues el camino al estudio de los perfiles de expresión génica a gran escala, más allá de la secuenciación genómica, objetivo original de su desarrollo y aplicación. La gran capacidad de secuenciación de las tecnologías NGS ha hecho posible el análisis de muestras complejas que contienen una mezcla de un gran número de transcritos. Pero su aplicabilidad no se reduce únicamente a la cuantificación de la expresión. Sus aplicaciones son múltiples: sirven como herramienta para la identificación y el análisis de regiones del DNA que interactúan con proteínas reguladoras de la expresión génica y regiones de unión a factores de transcripción, el estudio de la estructura de la cromatina, los patrones de metilación del DNA y son vitales en microbiología y metagenómica (Ansorge, 2009).

### 5.3. Digital Gene Expression en planaria

El objetivo de este proyecto consistía en la cuantificación y el análisis diferencial de expresión de los transcritos característicos de los neoblastos. Para ello era necesario utilizar una técnica cuantitativa lo bastante precisa que permitiera identificar incluso



**Figura 5.3 Protocolo experimental de DGE.**

Después de seleccionar los transcritos de mRNA por su cola poli-A, se producen *tags* específicas a partir del extremo 3' de cada uno gracias al uso de dos enzimas: la primera (denominada *anchoring enzyme*) reconoce un sitio de restricción concreto (en nuestro caso, NlaIII reconoce las bases CATG) y la segunda (el *tagging enzyme* MmeI) corta un fragmento a 21bp de distancia. Utilizando los adaptadores y la secuenciación de Illumina, se generan millones de *tags* que representan el estado transcripcional de la muestra en ese momento. Adaptado de [www.genomics.hk](http://www.genomics.hk).

aquellos genes menos expresados. Illumina (Solexa) era la tecnología más eficaz en este aspecto, mediante una técnica que producía un único *read* de 21 pares de bases a partir del extremo 3' de los transcritos de mRNA seleccionados por su cola de poli-A (Fig. 5.3). Esta estrategia, conocida como *Digital Gene Expression* (DGE), permitía una cuantificación precisa evitando secuenciar transcritos completos, lo cual proporciona el máximo rendimiento con una mínima cobertura. Además, el DGE evita el sesgo generado por la desventaja que sufren los transcritos más cortos frente a aquellos más largos al competir por ser secuenciados.

Sin embargo, 21 pares de bases no son suficientes para caracterizar un transcrito. Se hacía necesario, por tanto, contar con secuencias de referencia contra las que mapear esos *reads* cortos lo bastante largos que permitieran identificar transcritos completos. Ante la falta de un genoma de referencia debidamente anotado para *S. mediterranea* a partir del cual construir su transcriptoma, el primer reto consistió en

la generación de un transcriptoma de referencia para esta especie. En este caso se optó por la tecnología 454 de Roche, ya que era la que proporcionaba las mayores longitudes de secuencia. La secuenciación y posterior ensamblado del transcriptoma así como su análisis fueron publicados en el primer artículo de esta tesis (Abril *et al.*, 2010) y ha sido utilizado como transcriptoma de referencia tanto en este proyecto (Rodríguez-Esteban *et al.*, 2015), como en otros (Artículo 5; Fraguas *et al.*, 2014; Sandmann *et al.*, 2011). Este transcriptoma de referencia fue incluido posteriormente en un híbrido Illumina + 454 (Sandmann *et al.*, 2011).

Dada la facilidad de mapear los reads cortos de DGE sobre las secuencias de un transcriptoma, se optó por mapearlos sobre todos los conjuntos públicamente disponibles. El análisis comparativo está disponible en la web [planarian.bio.ub.edu/DGE](http://planarian.bio.ub.edu/DGE).



# Objetivos



# Objetivos

El objetivo general perseguido con esta tesis ha sido el de profundizar en el estudio de la biología del neoblasto y, en un sentido más amplio, aportar nuevos datos que puedan ser útiles a la comunidad científica en la investigación con células madre.

Este objetivo se concreta en la caracterización proteómica y transcriptómica de los neoblastos, que incluye:

- Generar anticuerpos monoclonales contra proteínas de membrana de neoblastos para discriminar distintas subpoblaciones.
- Identificación de proteínas características de neoblasto a través de una aproximación proteómica y validación funcional de las mismas.
- Construcción del transcriptoma de *S. mediterranea* que sirva, además, como referencia para nuevos estudios.
- Identificación mediante *Digital Gene Expression* de genes específicos de neoblasto, a partir de neoblastos aislados mediante la implementación de un protocolo de *cell sorting*, y caracterización funcional de los mismos.
- Desarrollo de las *pipelines* computacionales necesarias para el procesamiento y el análisis de los datos generados por los experimentos proteómicos y transcriptómicos anteriores y creación de herramientas que permitan a la comunidad científica acceder de forma sencilla y útil a los datos generados.





# Resultados



# **Informe de los directores sobre la participación del doctorando y el índice de impacto de las publicaciones**

La memoria de la tesis doctoral presentada por Gustavo Rodríguez Esteban, titulada "Caracterización del neoblasto como modelo de célula madre", de la que somos codirectores, incluye siete artículos, cuatro de ellos en el apartado de resultados y los tres restantes, incluyendo una revisión, en el de anexos. Todos los artículos han sido publicados en revistas internacionales que constan en la base de datos PubMed, la más importante en lo que a ciencias biomédicas se refiere, y con un índice de impacto (ISI) catalogado en la "Thomson Reuters Web of Science". Todos los artículos han sido evaluados por revisores anónimos designados por los editores de cada publicación. A continuación, se detalla la participación del doctorando en cada uno de los artículos, así como su índice de impacto y su posición en el listado ordenado de la especialidad correspondiente (según la "Thomson Reuters Web of Science", edición 2014).

## Artículo 1

Sören Moritz, Franziska Stöckle, Claudia Ortmeier, Henning Schmitz, **Gustavo Rodríguez-Esteban**, Göran Key And Luca Gentile. *Heterogeneity of planarian stem cells in the S/G2/M phase*. **International Journal of Developmental Biology** (2012) 56:117-125

El trabajo presentado en este manuscrito se realizó en una colaboración con el laboratorio del Dr. Luca Gentile en el Instituto Max Planck de Biomedicina Molecular en Münster (Alemania). GRE ha contribuido al análisis de los anticuerpos contra neoblasto participando en los protocolos de disociación celular y *cell sorting*, qPCR, inmunohistoquímica y *western blotting*.

<b>Factor de impacto</b>	Año	
Año anterior a la publicación	2011	2,823
Año de publicación	2012	2,614
Actual	2014	1,903
Últimos 5 años	2014	2,536

<b>Posición en el área</b>		
Developmental Biology	2014	35/41 Q4

## Artículo 2

Enrique Fernández-Taboada\*, **Gustavo Rodríguez-Esteban**\*, Emili Saló and Josep F Abril. *A proteomics approach to decipher the molecular nature of planarian stem cells*. **BMC Genomics** (2011) 12:133

\*Ambos autores han contribuido por igual.

EFT, ES y JFA concibieron el proyecto inicial presentado en este manuscrito. El trabajo experimental ha sido llevado a cabo por EFT y GRE. El análisis de los datos ha sido realizado por GRE y JFA. Todos los autores han participado en la redacción del artículo. Esta publicación fue incluida también en la tesis doctoral de EFT (2008).

<b>Factor de impacto</b>	Año	
Año anterior a la publicación	2010	4,206
Año de publicación	2011	4,073
Actual	2014	3,986
Últimos 5 años	2014	4,360

<b>Posición en el área</b>		
Biotechnology & Applied Microbiology	2014	26/162 Q1
Genetics & Heredity	2014	40/167 Q1

### Artículo 3

Josep F. Abril\*, Francesc Cebrià\*, **Gustavo Rodríguez-Esteban**, Thomas Horn, Susanna Fraguas, Beatriz Calvo, Kerstin Bartscherer and Emili Saló. *Smed454 dataset: unravelling the transcriptome of Schmidtea mediterranea*. **BMC Genomics** (2010) 11:731

\* Ambos autores han contribuido por igual.

JFA, FC, KB y ES concibieron el proyecto presentado en este manuscrito. GRE junto con SF ha realizado la validación de transcritos mediante RT-PCR presentados en la tabla suplementaria 3. GRE ha realizado también el análisis de *coverage* sobre genes conocidos mostrado en la tabla 1. Esta publicación fue incluida también en la tesis doctoral de SF (2014).

<b>Factor de impacto</b>	Año	
Año anterior a la publicación	2009	3.759
Año de publicación	2010	4.206
Actual	2014	3,986
Últimos 5 años	2014	4,360
 <b>Posición en el área</b>		
Biotechnology & Applied Microbiology	2014	26/162 Q1
Genetics & Heredity	2014	40/167 Q1

## Artículo 4

**Gustavo Rodríguez-Esteban**, Alejandro González-Sastre\*, José Ignacio Rojo-Laguna\*, Emili Saló and Josep F. Abril. *Digital Gene Expression approach over multiple RNA-Seq data sets to detect neoblast transcriptional changes in Schmidtea mediterranea*. **BMC Genomics** (2015) 16:361

\*Ambos autores han contribuido por igual.

GRE, ES y JFA diseñaron el proyecto presentado en este manuscrito. En cuanto al trabajo experimental, GRE ha realizado la implementación de los métodos de disociación, tinción y aislamiento de neoblastos mediante FACS, la extracción y purificación del RNA para la generación de las librerías de secuenciación a partir de las fracciones celulares X1, X1 y Xin, y la validación mediante WISH y RNAi de 47 genes candidatos de neoblasto. AGS ha llevado a cabo la caracterización del gen *Smed-meis-like* y JIRL la de los genes *Smed-nf-YA*, *Smed-nf-YB-2* y *Smed-nf-YC*. GRE ha llevado a cabo el análisis computacional de los datos desarrollando los programas y *pipelines* necesarios, así como las herramientas web para facilitar el acceso a los resultados, bajo la supervisión de JFA. GRE ha redactado el manuscrito con contribuciones de todos los autores.

### Factor de impacto

	Año	
Año anterior a la publicación	2014	3,986
Año de publicación	2015	-
Actual	2014	3,986
Últimos 5 años	2014	4,360

### Posición en el área

Biotechnology & Applied Microbiology	2014	26/162 Q1
Genetics & Heredity	2014	40/167 Q1

## Artículo 5 (Anexo 1)

Susanna Fraguas, Sara Barberán, Marta Iglesias, **Gustavo Rodríguez-Esteban** and Francesc Cebrià. *egr-4, a target of EGFR signaling, is required for the formation of the brain primordia and head regeneration in planarians.* **Development** (2014) 141:1835-1847

SF y FC diseñaron el proyecto presentado en este manuscrito. El trabajo experimental presentado en este manuscrito ha sido realizado mayoritariamente por SF bajo la supervisión de FC. SB ha contribuido con los experimentos presentados en la figura 6. GRE ha analizado los datos de DGE para identificar genes regulados por *Smed-egfr-3* y ha generado las figuras suplementarias 1 y 2 y la tabla suplementaria 2. SF y FC han redactado el manuscrito con contribuciones del resto de autores. Esta publicación fue incluida también en la tesis doctoral de SF (2014).

<b>Factor de impacto</b>	Año	
Año anterior a la publicación	2013	6,273
Año de publicación	2014	6,462
Actual	2014	6,462
Últimos 5 años	2014	6,741
<b>Posición en el área</b>		
Developmental Biology	2014	4/41 Q1

Los directores,

**Dr. Emili Saló Boix**

**Dr. Josep F. Abril Ferrando**

Barcelona, 2 de noviembre de 2015





# Artículo 1

## *Heterogeneity of planarian stem cells in the S/G2/M phase*

Sören Moritz, Franziska Stöckle, Claudia Ortmeier, Henning Schmitz, **Gustavo Rodríguez-Esteban**, Göran Key And Luca Gentile.

**International Journal of Developmental Biology** (2012) 56:117-125



## Resumen

## Heterogeneidad de las células madre de planaria en las fases S/G2/M

Las células madre adultas de la planaria, los neoblastos, pueden aislarse fácilmente mediante citometría de flujo. Sin embargo, la falta de anticuerpos contra antígenos de superficie de estas células impide un análisis más profundo de los conjuntos de células que forman esta población.

En este artículo se presenta la primera librería de anticuerpos monoclonales contra el proteoma de la membrana plasmática de células de planaria y se realiza un *immunoscreening* de esas proteínas de membrana en neoblastos. Para ello, se presenta un nuevo método de disociación de células en planaria basado en la papaína en lugar de la tripsina que incrementa el número y la viabilidad de las células recuperadas.

Las proteínas de membrana purificadas a partir de células individuales se utilizaron para inmunizar a ratones, obteniendo alrededor de 1000 hibridomas clonales. El sobrenadante de cada clon se analizó en primer lugar mediante ELISA y luego por inmunotinción. La mitad marcaron todas las células de planaria, mientras que la otra mitad marcaba fracciones de ellas. Se generaron 76 anticuerpos monoclonales a partir de ellos y el análisis detallado de los sobrenadantes de dos de los hibridomas reveló que las subpoblaciones X1, X2 y Xin expresan diferentes marcadores de membrana. Los resultados por PCR cuantitativa mostraron una correlación entre los resultados de inmunotinción y la expresión de marcadores de progenie temprana y tardía, incluidos los neoblastos en las fases S/G2/M del ciclo celular (subpoblación X1). Así, unos dos tercios de los neoblastos que ciclan mostraron una identificación de membrana característica junto con la expresión de marcadores que hasta la fecha se consideraban restringidos a la progenie posmitótica en diferenciación. El hecho de que un conjunto de la subpoblación de neoblastos X1 exprese marcadores de progenie temprana y tardía podría indicar que estas células están ya comprometidas mientras proliferan, lo que, además, refuerza la idea de heterogeneidad incluso en esta subpoblación de neoblastos en proliferación.



## Heterogeneity of planarian stem cells in the S/G2/M phase

SÖREN MORITZ<sup>1</sup>, FRANZISKA STÖCKLE<sup>1</sup>, CLAUDIA ORTMEIER<sup>1</sup>, HENNING SCHMITZ<sup>1</sup>,  
GUSTAVO RODRIGUEZ-ESTEBAN<sup>2</sup>, GÖRAN KEY<sup>3</sup> and LUCA GENTILE<sup>\*.1</sup>

<sup>1</sup>Planarian Stem Cell Laboratory, Max Planck Institute for Molecular Biomedicine, Münster, Germany,

<sup>2</sup>University of Barcelona, Department of Genetics, IBUB, Barcelona, Spain and

<sup>3</sup>Max Planck Institute for Molecular Biomedicine, Münster, Germany

**ABSTRACT** The planarian adult stem cell (pASC) population has a specific molecular signature and can be easily visualized and isolated by flow cytometry. However, the lack of antibodies against specific surface markers for planarian cells prevents a deeper analysis of specific cell populations. Here, if we describe the results of the immunoscreening of pASC plasma membrane proteins (PMPs). A novel papain-based method for planarian cell dissociation enabling both high yield and improved cell viability was used to generate single cell preparations for PMP purification. PMPs were used for intraperitoneal immunization of mice and thus about 1000 hybridoma clones were generated and screened. Supernatants collected from the hybridoma clones were first screened by ELISA and then by live immuno-staining. About half of these supernatants stained all the planarian cells, whereas the other half specifically labeled a subfraction thereof. A detailed analysis of two hybridoma supernatants revealed that large subfractions of the X1, X2 and Xin populations differentially express specific membrane markers. Quantitative PCR data disclosed a correlation between the immunostaining results and the expression of markers of the early and late progeny, also for those pASCs in the S/G2/M phase of the cell cycle (X1 population). Thus, about two thirds of the cycling pASCs showed a specific membrane signature coupled with the expression of markers hitherto considered to be restricted to differentiating, post-mitotic progeny. In summary, a library of 66 monoclonal antibodies against planarian PMPs was generated. The analysis of two of the clones generated revealed that a subset of cells of the X1 population expresses early and late progeny markers, which might indicate that these cells are committed while still proliferating. The findings demonstrate the usefulness of our PMP antibody library for planarian research.

**KEY WORDS:** *stem cell heterogeneity, cell surface marker, mAb library, planarian*

### Introduction

Planarians are the true masters of regeneration. Their remarkable ability to regenerate an entire animal from virtually any small body fragment (Sanchez Alvarado *et al.*, 2002; Handberg-Thorsager *et al.*, 2008) is unparalleled among both vertebrates and invertebrates. Planarian regeneration is based on planarian adult stem cells (pASCs) also named neoblasts. Experiments in the late 80s suggested that neoblasts behave pluripotent as population (Baguña, 1989), but only recently has it been elegantly shown by single cell transplantation into irradiated hosts that at least some (clonogenic) neoblasts are indeed pluripotent (Wagner *et al.*, 2011). Purification

of pASCs by FACS (fluorescence-activated cell sorting) is possible by staining isolated planarian cells for nuclear content (Hoechst 33342) and cytoplasmic size (Calcein AM). Since pASCs are the only proliferating cells in the asexual strains of *Schmidtea mediterranea*, irradiation efficiently ablates them. This allowed identifying one irradiation-insensitive (Xin) and two irradiation-sensitive (X1 and X2) cell populations (Reddien *et al.*, 2005; Hayashi *et al.*, 2006). While the X2 population is partially irradiation-sensitive and

*Abbreviations used in this paper:* ELISA, enzyme-linked immunosorbent assay; FACS, fluorescence-activated cell sorting; pASC, planarian adult stem cell; PMP, pASC plasma membrane proteins.

\*Address correspondence to: Luca Gentile, Planarian Stem Cell Laboratory, Max Planck Institute for Molecular Biomedicine, von-Esmarch-Str. 54, 48149 Münster, Germany. Tel: +49-251-8346-870. Fax: +49-251-8346-900. e-mail: luca.gentile@mpi-muenster.mpg.de

Supplementary Material (5 figures + 1 table) for this paper is available at: <http://dx.doi.org/10.1387/ijdb.113440sm>

Final, author-corrected PDF published online: 16 March 2012

hence heterogeneous by definition, the X1 population appears to be rather homogeneous, as many genes required for proper pASCs function are expressed by the entire stem cell population (reviewed in Gentile *et al.*, 2011). However, there is some evidence for heterogeneity of the X1 population at both the ultrastructural (Higuchi *et al.*, 2007) and the transcript levels (Eisenhoffer *et al.*, 2008; Hayashi *et al.*, 2010; Pearson and Sanchez Alvarado, 2010; Scimone *et al.*, 2010).

Isolation of specific cell types – including stem cells – could be achieved using antibodies specific to cell surface epitopes. One example are the long-term self-renewing human hematopoietic stem cells, which are defined by a remarkably long set of cell surface markers (Lin-/CD34+/CD38-/CD90+/CD45RA-; Notta *et al.*, 2010). Compared to other model organisms, available antibodies against planarian cells are scarce and directed against intracellular epitopes. Moreover, none has been described to be specific for pASCs so far (Bueno *et al.*, 1997). It is conceivable that different planarian cell types also differ in their cell surface epitopes, which is why we opted for generating a library of mouse monoclonal antibodies against the planarian membrane proteome. In order to produce such a library, we needed to ensure efficient and reliable isolation of large amounts of planarian cells. Papain (Papaya peptidase I) is a cysteine protease with a broad specificity, which preferentially hydrolyzes hydrophobic and aromatic residues (Kimmel and Smith, 1954), which is often used for the dissociation of neural tissues (Moritz *et al.*, 2008).

With this work, we raised the first library of mouse monoclonal antibodies directed against the plasma membrane proteome of planarian cells. The majority of these antibodies recognized surface epitopes of all or portions of the planarian cells. Among them, two antibodies (6-9.2 and 8-22.2) identified subfractions of the X1 population that we further characterized at the transcript level. All the stem cell markers tested did not reveal any difference in their expression compared to the negative counterparts, while the markers specific for both early and late progeny (Eisenhoffer *et al.*, 2008) were unequivocally upregulated. This finding was not related to a specific phase of the cell cycle, thus we hypothesized a correlation with the early commitment of proliferating stem cells. Further experiments are needed to substantiate this hypothesis, in order to establish when and how the planarian stem cells become committed and to elucidate the role of the early and late progeny genes, whose function in planarians is not known yet. All the genes mentioned in this paper are planarian genes, thus the prefix 'Smed-' is omitted.

## Results and Discussion

### **Planarian cell dissociation with papain improves both yield and cell viability**

In order to attain a sufficient amount of plasma membrane proteins (PMPs) for the downstream immunization, an efficient method for the dissociation of intact animals into a single cell suspension is required. In recent years, various methods have been used to dissociate planarian tissues – mechanical (Fernandez-Taboada *et al.*, 2010), trypsin-based (Reddien *et al.*, 2005; Hayashi *et al.*, 2006) and collagenase-based dissociation (Wagner *et al.*, 2011). Papain is a gentle protease that has been successfully used to isolate cells with rather complex morphologies, such as visual neurons from adult rats (Huettnner and Baughman, 1986). Thus,

the effectiveness of the papain-based planarian cell dissociation was compared to the trypsin-based (according to Hayashi *et al.*, 2010) and the mechanical dissociation. To quantify the efficiency of the dissociation protocols in terms of cell yield and viability, the animals were weighed before dissociation. Dissociated cells were stained with Hoechst 33342 and Calcein AM (H/C) and propidium iodide (PI) and analyzed by flow cytometry. The gating logic of our flow cytometric analysis is depicted in Fig. 1A. Papain-based dissociation yielded  $4.2 \times 10^5 \pm 8.3 \times 10^4$  cells/mg, trypsin-based dissociation  $1.3 \times 10^5 \pm 6.7 \times 10^4$  cells/mg and mechanical dissociation  $3.3 \times 10^5 \pm 4.2 \times 10^4$  cells/mg (Fig. 1B). Both papain-based and mechanical dissociation yielded significantly more cells per milligram of animal weight than trypsin-based dissociation ( $p \leq 0.01$ ). Papain-based dissociation yielded  $62.2 \pm 11.1\%$  of live cells on total events, whereas mechanical dissociation yielded  $49.1 \pm 14.6\%$  and trypsin-based dissociation  $38.7 \pm 2.1\%$  of live cells (Fig. 1C). Interestingly, the percentage of live cells in the samples dissociated with papain is significantly higher compared to those dissociated with trypsin ( $p \leq 0.01$ ; Fig. 1C). Also the percentage of single cells in the cell suspension is significantly higher for papain-based dissociation compared to both trypsin-based ( $p \leq 0.01$ , Fig. 1C) and mechanical dissociation ( $p \leq 0.05$ , Fig. 1C). Concerning the planarian FACS populations defined by H/C staining, a similar proportion of X1 and X2 cells was found in the papain-based and mechanical, but not in the trypsin-based preparation. ( $p \leq 0.05$ ; Fig. 1C). Also, the percentage of the Xin cells obtained by using papain-based dissociation ( $13.5 \pm 5.5\%$ ) was much higher than the one obtained with trypsin-based dissociation ( $5.2 \pm 0.5\%$ ;  $p \leq 0.01$ ; Fig. 1C).

As our immunization strategy required a large amount of planarian single cells for the isolation of immunogens, we aimed to optimize the dissociation protocol. Previously, we utilized mechanical dissociation (Fernandez-Taboada *et al.*, 2010), which produces a good yield of viable cells, but is time consuming and laborious. We therefore tested whether papain could be a suitable protease for the dissociation of planarian cells. All the dissociation methods tested yielded a cell suspension suitable for flow cytometry. However, the number of cells isolated per milligram of animal weight varied greatly among the methods, with papain-based dissociation being the most efficient one. The low efficiency of trypsin-based dissociation was in part due to the stickiness of the tissue fragments after incubation with the protease. Another reason is that many fragments could not be completely dissociated. This is also the case using mechanical and, to a lesser degree, papain-based dissociation. Nevertheless, the poor results obtained with trypsin-based dissociation might also be attributed to the fact that the protocol was optimized on another planarian species, *Dugesia japonica*. Flow cytometric analysis revealed that the cell suspension obtained using papain contains less debris (events with sub-G1 DNA content) and more viable cells compared to the other methods tested. Papain-based cell dissociation also performs best when the relative proportion of the planarian cell populations is taken into account. Our data suggest that Xin cells are particularly sensitive to both mechanical and trypsin-based dissociation. This may be due to the incomplete dissociation and/or the lysis of the large differentiated cells and implies that papain-based dissociation reflects the total planarian cell population more accurately.

In summary, we could convincingly show that our papain protocol

is a dissociation method that combines high yield, improved cell viability and a better representation of the planarian cell populations with ease of use. With our papain protocol it is possible to generate hundreds of millions planarian cells for any downstream purpose within approximately 1.5 hours. On the other hand, the dissociation of very small samples such as dissected blastemata is also possible without significant loss of cells.

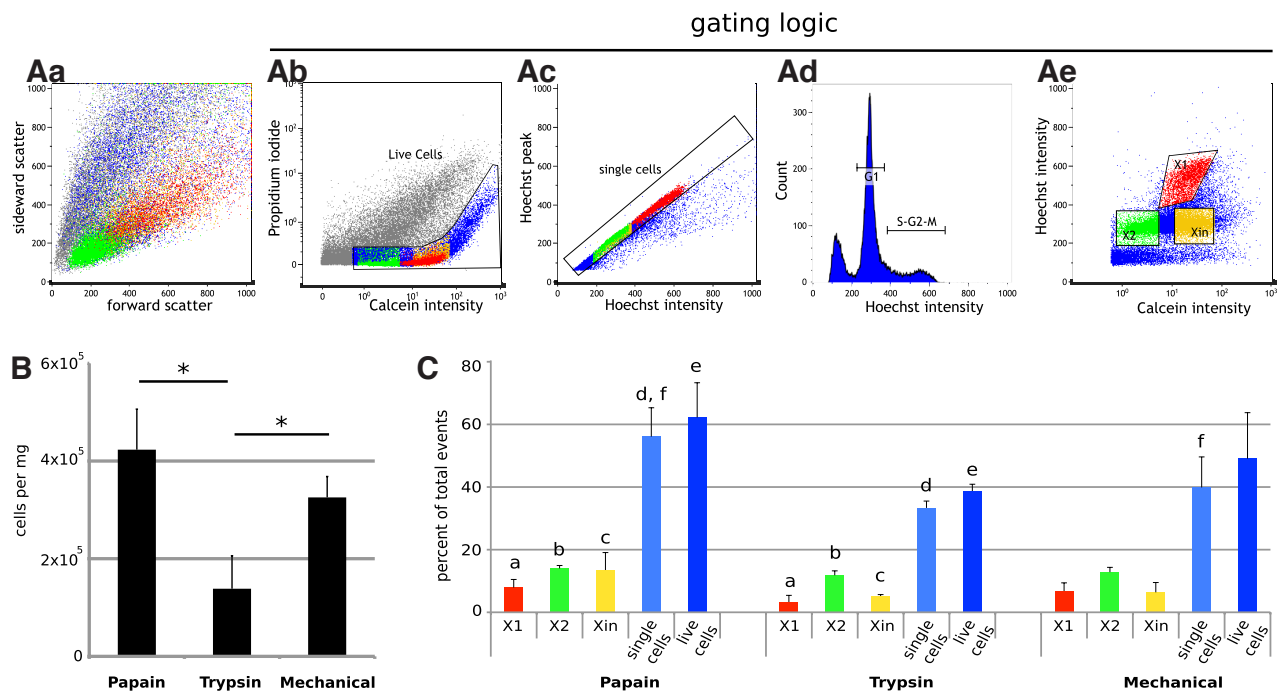
### Generation of antibodies against planarian plasma membrane proteins

We aimed to generate antibodies directed against cell surface epitopes of planarian cells, with a focus on antibodies specific to pASCs or subpopulations thereof. In a preliminary attempt, we immunized mice using sorted neoblasts of the X1 fraction. Although various antibodies – including cell type-specific ones – were generated, all of them ( $n=23$ ) were directed against intracellular epitopes (data not shown).

We therefore changed our strategy and immunized the mice using purified PMPs. The workflow for the generation of the antibody library is depicted schematically in Fig. 2A. Purified planarian PMPs were injected intraperitoneally into BALB/C mice and, after four immunizations, immune sera (IS) were taken and tested by Western blotting. Equal amounts of planarian PMPs and whole cell lysate were loaded and probed with two IS from

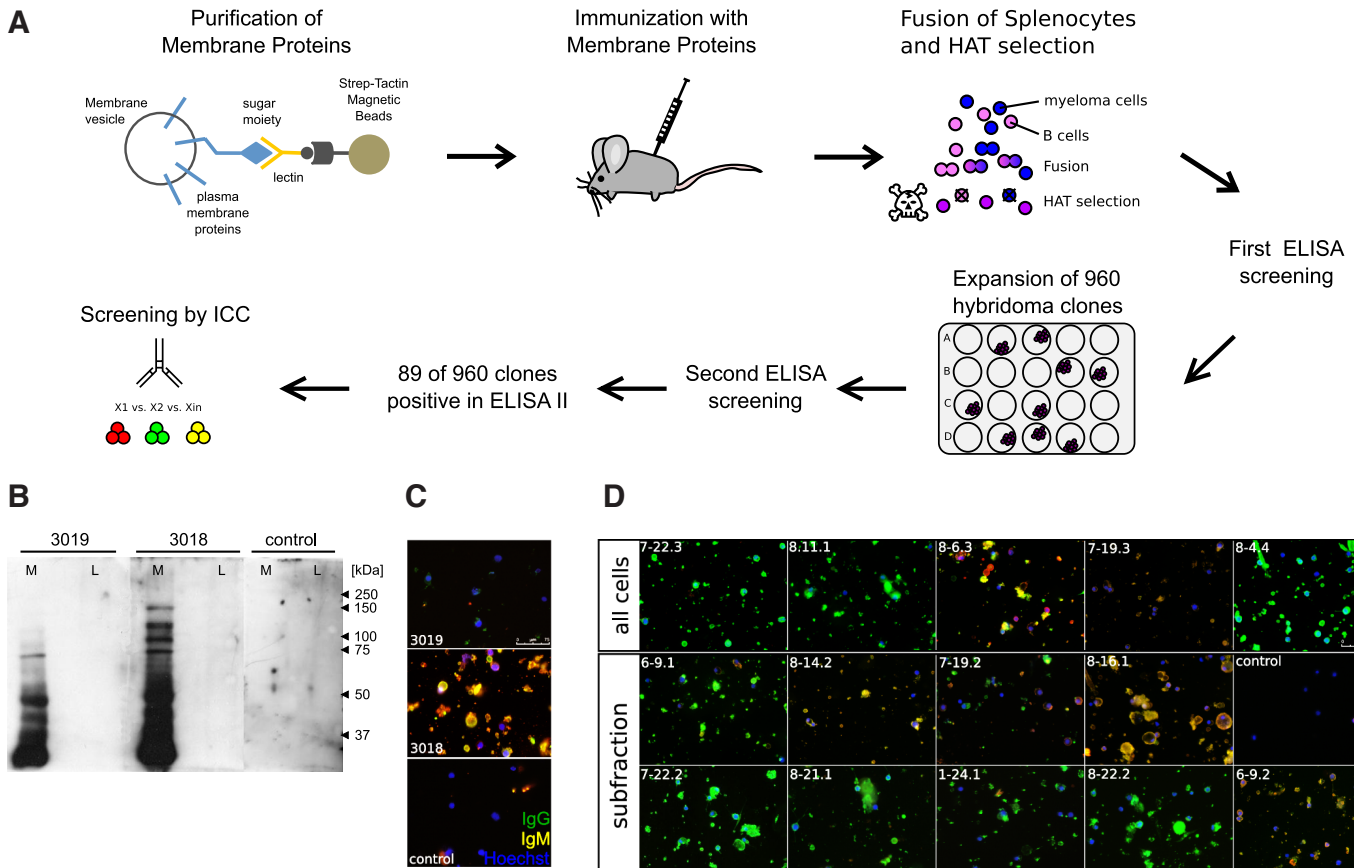
different mice (3018 and 3019). Several distinct bands ranging from 20 to 150 kDa were detected in the lanes where PMPs were loaded, whereas we found no signal in the whole cell lysate lanes (Fig. 2B). The absence of detectable signals in whole cell lysate is probably due to the low abundance of membrane proteins in the preparation. A control IS (from a non-immunized BALB/C mouse) did not reveal any signals. Although both mice showed comparable immune reactions, mouse 3018 displayed stronger signals, especially in the high molecular weight range (Fig. 2B). The IS were also tested by immunocytochemistry (ICC) on isolated live planarian cells. Immune sera from both the immunized mice showed membrane staining (Fig. 2C) but, similar to the Western blot, IS 3018 showed a stronger signal. Interestingly, IS 3018 produced signals for both the IgM and IgG antibody isotypes, while IS 3019 predominantly showed signals for the IgG isotype. Considering the strength and quality of the immune reaction observed, we therefore decided to use mouse 3018 for the generation of the hybridoma library.

In total, 960 hybridomas were picked from the cloning plates and expanded. The hybridoma supernatants were tested for immunoreactivity against PMPs by ELISA. Eighty-nine hybridoma supernatants reacted with PMPs and 76 hybridoma clones could be expanded. To investigate whether the antibodies could successfully label planarian cells, ICC was performed. Of 76 super-



**Fig. 1. Papain-based dissociation of planarian cells is highly efficient.** Planarians were dissociated using papain or trypsin as protease or by mechanical dissociation and stained with Hoechst 33342 / Calcein AM / Propidium iodide. Forward scatter versus side scatter plot of total events of our flow cytometric analysis is shown. **(Aa)** The events were triggered on Hoechst staining and gated for live cells (Calcein<sup>+</sup> / Propidium iodide; **(Ab)**), then for single cells (Hoechst<sup>intensity</sup> versus Hoechst<sup>peak</sup>; **(Ac)**). The Hoechst histogram for live single cells reveals different nuclear contents of the planarian cells, which correspond – for the majority of the cells – to different phases of the cell cycle (G1 = 2n; 2n < S-G2-M ≤ 4n; **(Ad)**). The dot plot in which Calcein intensity is plotted against Hoechst intensity allows the identification of the planarian FACS populations X1 (red), X2 (green) and Xin (yellow) **(Ae)**. The different fractions are mapped back with their respective colors to the plots shown in **(Aa-c)**. The animals were weighed before dissociation and the yield of each dissociation method was expressed as cells per milligram of animal **(B)**;  $n = 4$ . The cell suspensions obtained with either method were analyzed by flow cytometry **(C)**;  $n = 4$ . The number of the events that correspond to each gate (X1, X2, Xin, single cells and live cells) is expressed as percentage of total events. a, b, f:  $p \leq 0.05$ ; \*, c, d, e:  $p \leq 0.01$ .





**Fig. 2. Generation and screening of the mouse anti-planarian cell surface protein library of monoclonal antibodies.** Workflow followed for generating and screening the monoclonal antibody (mAb) library (A). Planarian plasma membrane proteins (PMPs) were isolated using the Qproteome plasma membrane kit. Two mice were immunized four times with 50  $\mu$ g of purified PMPs. Mouse 3018 was sacrificed and splenocytes were isolated. After PEG-assisted fusion with myeloma cells and HAT selection of the hybridomas, supernatants were screened by ELISA for reactivity against planarian PMPs. After expansion, clones were screened by a second round of ELISA. Eighty-nine hybridomas were found to produce mAbs with high affinity for PMPs. Planarian PMPs and whole cell lysates were loaded in equal amounts and separated by SDS-PAGE. The proteins were blotted onto PVDF membrane and probed with immune sera from mice 3019 and 3018 and control serum from a non-immunized mouse (B). Hybridoma clones were screened by immunocytochemistry on planarian cells (C). Signals for mouse-IgGs are shown in green, signals for mouse-IgMs are shown in yellow and nuclei were counterstained with Hoechst 33342 (shown in blue). A panel of antibodies that reacted with planarian live cells is shown (D). The upper row (all cells) depicts hybridomas producing antibodies that immunostained virtually every planarian cells. The lower rows (subfractions) depict hybridomas whose antibodies immunostained a subfraction of total planarian cells. Scale bars: 75  $\mu$ m; M, PMP; L, NP-40 whole cell lysate.

nantants tested, 51 showed reactivity against live planarian cells or subfractions thereof. In Fig. 2D a panel of hybridoma supernatants is presented that either recognize all planarian cells (upper panel) or subfractions of them (lower panel). Control mouse IgGs did not produce any staining (center-right). We also tested the hybridoma library on fixed and permeabilized cells. Sixty-six supernatants reacted to all planarian cells or subfractions thereof, indicating that 15 hybridomas solely recognized intracellular epitopes, whereas many of the extracellular epitopes seem to be resistant to para-formaldehyde fixation (data not shown).

In summary, PMPs were purified from planarian cells dissociated with papain and used for the immunization of two mice. The immunization succeeded, as shown by Western blotting and immunostaining and mouse 3018 was chosen to generate the antibody library according to the higher intensity and the overall better quality of the immunoreaction. The 3018 library consists of 66 hybridoma clones, 51 of which produce antibodies reacting

against all live planarian cells ( $n=26$ ) or subpopulations of them ( $n=25$ ), while the remaining 15 antibodies detect intracellular epitopes. The ratio of antibodies suitable for live cell immunostaining (77%) is relatively high, and underscores the high efficiency of our approach to generate cell surface-specific antibodies.

#### Some of the antibodies generated are specific for subsets of the planarian FACS populations

To further investigate our antibody library, we selected four clones (1-24.1, 6-9.2, 7-22.2 and 8-22.2) that reacted with subfractions of live planarian cells. In order to establish which cell population reacted with the respective antibody, planarian cells were sorted according to H/C/PI staining and immunostained with the hybridoma supernatants (Fig. 3 A-D). In each fraction, the percentage of cells positive for each of the 4 antibodies was determined. Hybridoma clone 1-24.1 (IgG1,  $\kappa$ ) did not stain any X1 cells, but  $11 \pm 2\%$  of the X2 cells and  $32 \pm 8\%$  of the Xin cells (Fig. 3D'), hybridoma

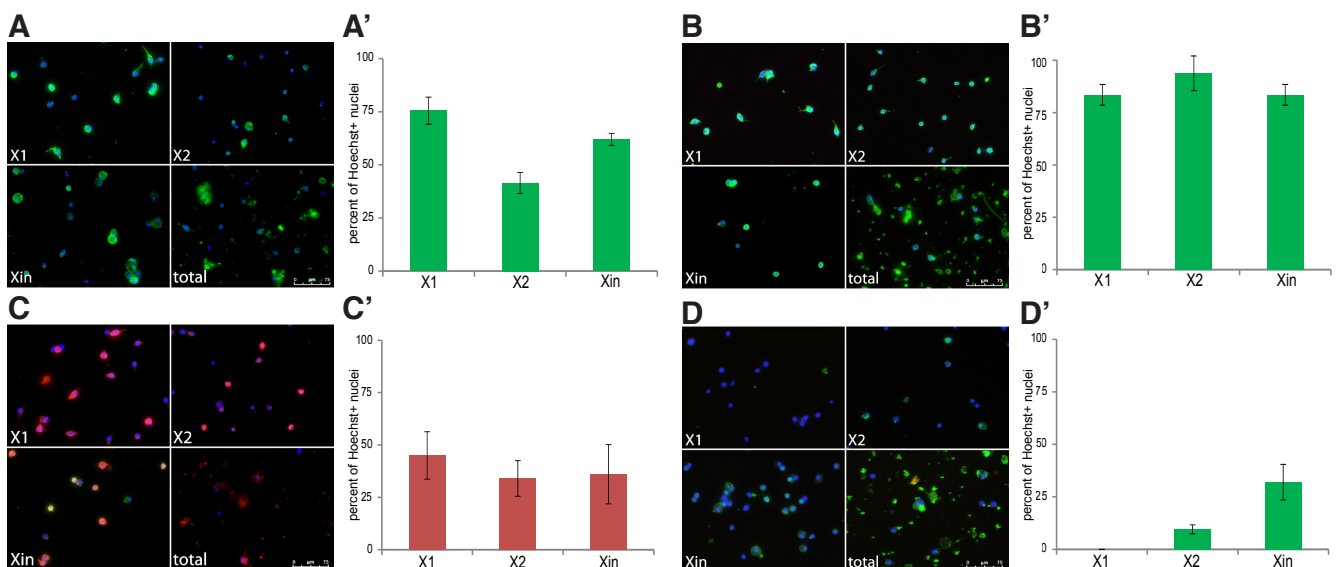
clone 6-9.2 (IgM,  $\kappa$ ) stained  $45 \pm 11\%$  of the X1 cells,  $35 \pm 8\%$  of the X2 cells and  $36 \pm 14\%$  of the Xin cells (Fig. 3C'), hybridoma clone 7-22.2 (IgG1,  $\kappa$ ) stained  $84 \pm 4\%$  of the X1 cells,  $94 \pm 8\%$  of the X2 cells and  $84 \pm 5\%$  of the Xin cells (Fig. 3B') and hybridoma clone 8-22.2 (IgG1,  $\kappa$ ) stained  $76 \pm 6\%$  of the X1 cells,  $42 \pm 4\%$  of the X2 cells and  $64 \pm 3\%$  of the Xin cells (Fig. 3A').

Determining the percentage of the sorted cells that were positive for each of the 4 antibodies, we found that hybridoma clone 1-24.1 did not stain any cells in the X1 fraction, a few cells in the X2 and approximately 30% of the cells in the Xin fraction. Although clone 1-24.1 does not recognize the stem cells, it fulfilled our expectations to find antibodies specific for subpopulations of differentiated cells. This could either be used to sort a specific differentiated cell type or to exclude it from a planarian cell suspension (negative selection). Eventually, none of the four clones analyzed displayed an immunostaining pattern compatible with a pure population of stem cells (high in X1, low in X2 and none in Xin); however, antibodies 6-9.2, 7-22.2 and 8-22.2 immunostained a different percentage of any given FACS populations (Fig. 3), and therefore we thought it unlikely that the three antibodies recognized the same cell surface antigen. It was equally unlikely that each of the antibodies labeled a specific cell type, because the added percentages of the cells positive for each antibody in each population largely exceeded 100%. Although we expected heterogeneity in the X2 fraction, which is a mixture of stem cells in the G1 phase, differentiating and differentiated cells (Higuchi *et al.*, 2007; Eisenhoffer *et al.*, 2008), and in the Xin fraction, which contains various types of large differentiating and differentiated cells, such a heterogeneity of the X1 fraction – as revealed by cell surface marker expression – was not anticipated. Thus, we further investigated the planarian cell fractions positive and negative for the respective antibodies using quantitative PCR (qPCR).

### Antibodies 6-9.2 and 8-22.2 revealed a subset of X1 cells expressing early and late progeny genes

In order to examine more closely the heterogeneity of the FACS populations as revealed by antibodies 6-9.2, 8-22.2 and, to a minor degree, 7-22.2, we assessed the expression signature of the cell subfractions by qPCR. We quantified the expression of several genes, such as markers of stem, differentiating and differentiated cells. After incubation with 6-9.2, 8-22.2 or 7-22.2 supernatants followed by appropriate secondary antibodies, live cells were sorted in accordance with the intensity of the immune signal. Normal mouse IgGs were used to set the gates in such a way as to have less than 4.0% of positive cells for each of the FACS fractions. Presumably, these few positive cells were the consequence of unspecific antibody binding (Fig. 4 A-C). Immunostaining with 6-9.2 produced a bimodal distribution of the signal for both the X1 and Xin populations, with negative and positive peaks. The signal of the X2<sup>6-9.2+</sup> cells is more broadly dispersed. Within the X1 population  $58.1 \pm 2.2\%$  of the cells were 6-9.2<sup>+</sup>, while  $30.1 \pm 1.3\%$  were 6-9.2<sup>-</sup> (Fig. 4A'). In the X2 population  $40.6 \pm 2.6\%$  of the cells were 6-9.2<sup>+</sup> and  $46.4 \pm 2.6\%$  were 6-9.2<sup>-</sup> (Fig. 4B'), while in the Xin population  $51.3 \pm 1.9\%$  of the cells were 6-9.2<sup>+</sup> and  $31.9 \pm 1.9\%$  were 6-9.2<sup>-</sup> (Fig. 4C'). Remarkably, both the 6-9.2 positive and negative subsets of cells plot back to the H/C gate without revealing any peculiar distribution related to the cell cycle (Fig. S1).

For each subfraction, the expression of *Piwi1* (*Piwil1*, stem cell marker; Reddien *et al.*, 2005), NB.32.1g, *Agat1* (*Gatm*), *Cyp1a1* (early and late progeny markers; Eisenhoffer *et al.*, 2008) and *Myhc* (T-mus, myocytes and differentiated muscle cell marker; Cebrià *et al.*, 1996) was assessed. In agreement with published data (Reddien *et al.*, 2005; Eisenhoffer *et al.*, 2008), *Piwi1* was expressed at high levels in the X1 population and at low and very low levels in the X2 and Xin populations, respectively. No differ-



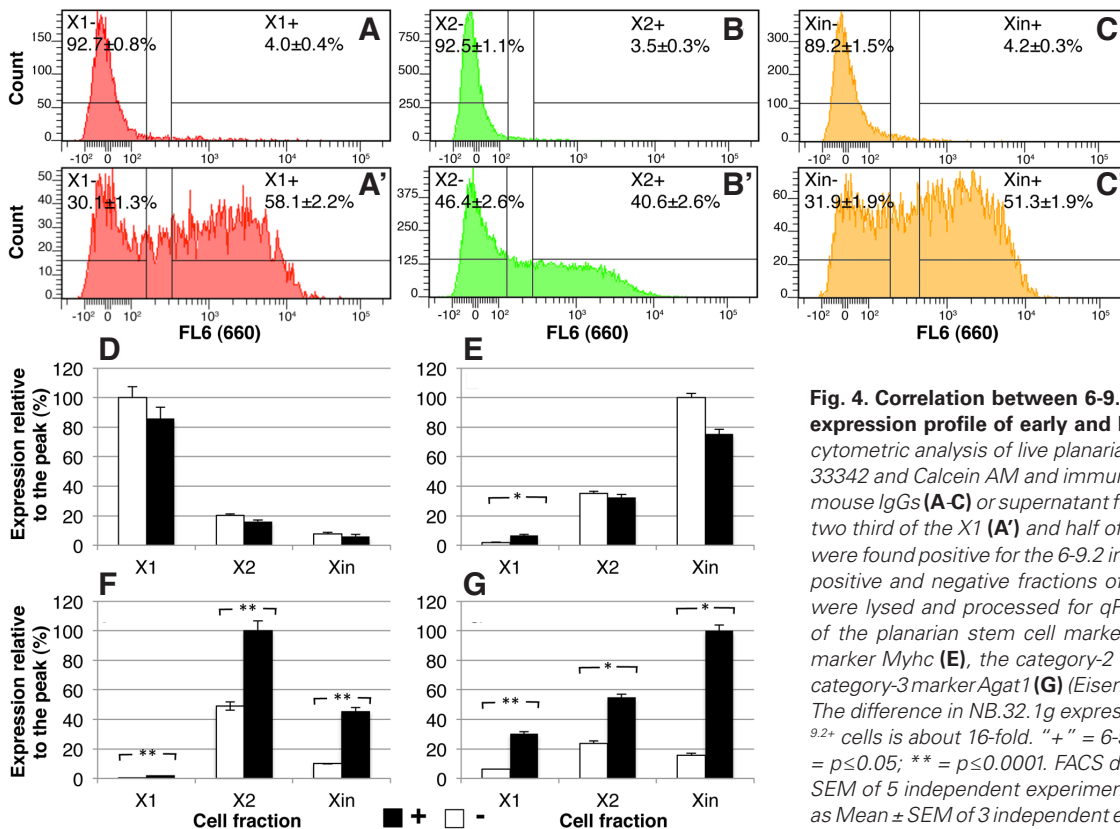
**Fig. 3. Immunocytochemistry of selected antibodies on X1, X2 and Xin FACS populations of planarian cells.** Total planarian cells were stained with Hoechst 33342 and CalceinAM and sorted according to the FACS gating depicted in Fig. 1A. Cells were plated on plastic dishes and live-immunostained with the supernatants of the hybridomas 8-22.2 (A), 7-22.2 (B), 6-9.2 (C) and 1-24.1 (D) and fixed. Photomicrographs of immunostained sorted planarian cells of the X1, X2 and Xin fractions as well as total planarian cells are shown. Signals for mouse-IgGs are shown in green, signals for mouse-IgMs are shown in yellow and nuclei were counterstained with Hoechst 33342 (shown in blue). The number of immunostained cells is expressed as percentage of positive nuclei per total nuclei (A'-D'). ICC data are presented as Mean  $\pm$  SD of 2 independent experiments; scale bar: 75  $\mu$ m.

ences were found between the 6-9.2<sup>+</sup> and 6-9.2<sup>-</sup> subfractions within each population (Fig. 4D). *Bruli* (*Celf3*; Guo et al., 2006), *Smb* (*Snrpb*; Fernandez-Taboada et al., 2010) and *cyclin B* (*Ccnb1*; Reddien et al., 2005) were also investigated, and none of them displayed a differential expression between the cells in the positive and negative subfractions (Fig. S2). Also the expression of *Myhc* did not show any significant differences, with the exception of the X1 fraction, where 6-9.2<sup>+</sup> cells showed a 3-fold upregulation compared to the 6-9.2<sup>-</sup> cells ( $p \leq 0.05$ , Fig. 4E). Remarkably, the expression of *Myhc* in some X1 cells was also recently reported in *D. japonica* (Hayashi et al., 2010). Interestingly, all markers of progeny cells tested (NB.32.1g, *Agat1*, *Cyp1a1*) were found upregulated in all the 6-9.2<sup>+</sup> subfractions. The early progeny marker NB.32.1g was expressed at moderate/high levels in X2<sup>+</sup> and Xin<sup>+</sup> cells, and at low level in X1<sup>+</sup> cells. Comparing positive to negative subfractions, NB.32.1g was upregulated about 16-fold in the X1<sup>+</sup> cells ( $p \leq 0.0001$ ), 2-fold in the X2<sup>+</sup> cells ( $p \leq 0.0001$ ) and 4-fold in the Xin<sup>+</sup> cells ( $p \leq 0.0001$ , Fig. 4F). *Agat1*, a late progeny marker, was also moderately upregulated in all the 6-9.2<sup>+</sup> subfractions, namely 7-fold in the X1<sup>+</sup> ( $p \leq 0.0001$ ), 2.5-fold in the X2<sup>+</sup> ( $p \leq 0.05$ ) and 5-fold in the Xin<sup>+</sup> subfraction ( $p \leq 0.05$ ; Fig. 4G). A similar albeit not statistically significant trend was observed for the other late progeny marker, *Cyp1a1* (Fig. S3A).

We repeated the qPCR on positive and negative subfractions for the antibodies 8-22.2 and 7-22.2. Although in the comparison with 6-9.2 the percentage of 8-22.2 positive cells in each fraction was different (Fig. S4 A'-C'), a striking similarity was found when qPCR data were compared. In fact, X1<sup>6-9.2+</sup> and X1<sup>8-22.2+</sup> cells are indistinguishable according to the expression of NB.32.1g and

*Agat1*, ( $p=0.3640$  and  $p=0.5179$ , respectively; Fig. 4 D-G and Fig. S4 D-G). Considering also that the immune signal intensity distributions for the two antibodies are similarly shaped, (cf. Fig. 4A'-C' with Fig. S4 A'-C'), we cannot exclude that the divergence found in the cell number might result from intrinsic differences between the two antibodies – such as antibody isotypes or affinity to the epitope – rather than from the recognition of a different antigen.

The scenario changes radically when the 7-22.2 antibody is considered. The signal intensity distribution was bimodal for all the FACS fractions (Fig. S5 A'-C'). However, for the X1 fraction we observed a large proportion of cells staining positive for the antibody, leaving few cells in the negative gate (Fig. S5A'). These data correlate very well with our ICC data (Fig. 3B'), while for the other two populations – X2 and Xin – we found that the positive cells were fewer than those counted in ICC. As for the 6-9.2 and 8-22.2 immunostainings, 7-22.2 positive and negative subfractions of both the X1 and Xin population also showed a similar expression of *Piwi1*, while X2<sup>+</sup> cells expressed the stem cell marker at a level doubling that of their negative counterpart ( $p \leq 0.05$ ; Fig. S5D). On the other hand, *Myhc* is expressed at very low levels in all the negative subfractions, whereas it is upregulated 4-, 6- and 14-fold in X1<sup>+</sup>, X2<sup>+</sup> and Xin<sup>+</sup> subfractions, respectively ( $p \leq 0.05$ ; Fig. S5E). Once more, significant differences were also found between positive and negative subfractions in the expression of both early and late progeny markers. However, contrary to what we observed for both 6-9.2 and 8-22.2 antibodies, NB.32.1g, *Agat1* and *Cyp1a1* were found upregulated in the negative subfractions of the 7-22.2 immunostained cells. Specifically, NB.32.1g was upregulated 5- and 7-fold in X2<sup>-</sup> and Xin<sup>-</sup> cells, respectively ( $p \leq 0.0001$ ; Fig S4F),



**Fig. 4. Correlation between 6-9.2 immunostaining and the expression profile of early and late progeny markers.** Flow cytometric analysis of live planarian cells stained with Hoechst 33342 and Calcein AM and immunostained with either generic mouse IgGs (A-C) or supernatant from clone 6-9.2 (A'-C'). About two third of the X1 (A') and half of the X2 (B') and Xin (C') cells were found positive for the 6-9.2 immunostaining. After sorting, positive and negative fractions of X1, X2 and Xin populations were lysed and processed for qPCR analysis. The expression of the planarian stem cell marker *Piwi1* (D), the muscle cell marker *Myhc* (E), the category-2 marker NB.32.1g (F) and the category-3 marker *Agat1* (G) (Eisenhoffer et al., 2008) are shown. The difference in NB.32.1g expression between X1<sup>6-9.2+</sup> and X1<sup>6-9.2-</sup> cells is about 16-fold. "+" = 6-9.2<sup>+</sup> cells; "-" = 6-9.2<sup>-</sup> cells; \* =  $p \leq 0.05$ ; \*\* =  $p \leq 0.0001$ . FACS data are presented as Mean  $\pm$  SEM of 5 independent experiments; qPCR data are presented as Mean  $\pm$  SEM of 3 independent experiments; peak expression is arbitrarily set to 100%.

while *Agat1* was upregulated 5-, 4- and 6-fold in the X1<sup>-</sup>, X2<sup>-</sup> and Xin<sup>-</sup> subfractions, respectively ( $p \leq 0.01$ ,  $p \leq 0.0001$  and  $p \leq 0.0001$ ; Fig. S5G). Expression of NB.32.1g did not differ between X1<sup>+</sup> and X1<sup>-</sup> subfractions, while *Cyp1a1* was significantly upregulated only in the Xin<sup>-</sup> fraction (Fig. S3C).

In an attempt to better understand the molecular features of the subfractions of cells labeled by some antibodies of the 3018 library, we assessed the expression of markers specific for stem, differentiating and differentiated cells in the subsets of cells defined by 6-9.2, 7-22.2 or 8-22.2 immunostaining. Interestingly, we never found any difference in the expression levels of the stem cell markers tested within the same FACS population, independently of the antibody used (Fig. 4D; Fig. S2, S4D, S5D). This suggests that none of the antibodies is specific for antigens solely expressed by the pASCs. However, all the cells positive for either the 6-9.2 or 8-22.2 antibody were highly enriched for the category-2 marker NB.32.1g, with a clear upregulation in the X1<sup>6-9.2+</sup> and the X1<sup>8-22.2+</sup> cells. A statistically significant upregulation was also observed for the category-3 marker *Agat1*. Regardless of whether 6-9.2 and 8-22.2 antibodies recognize the same antigen or not, they do define subfractions of the X1 population, as does the antibody 7-22.2. The percentage of 6-9.2<sup>+</sup> cells in the X1 population is quite high (about 60%), but the expression of both NB.32.1g and *Agat1* is relatively low, when compared to both X2 and Xin populations. According to published results (Eisenhoffer *et al.*, 2008; Table S2), only a small proportion of X1 cells expresses the progeny markers; specifically, the category-2 marker NB.21.11e and the category-3 marker *Agat1* are expressed by 0.7% and 2.4% of the X1 cells, respectively. It could be assumed that the category-2 marker that we considered (NB.32.1g) would behave in a similar way. Therefore, only a small portion of the X1<sup>6-9.2+</sup> cells expresses the progeny markers at high level, while the majority of the X1<sup>6-9.2+</sup> cells do not. These cells might express other progeny markers, either known or unknown, or represent a primed state of the stem cell, in which changes in the chromatin state are a prerequisite for changes in gene expression. This is consistent with the percentage of X1 cells expressing the chromatin remodeling protein CHD4, involved in, and essential for stem cell differentiation (Scimone *et al.*, 2010). This gene is required for the differentiation of *Agat1*-expressing cells; hence its upregulation in early committed X1 cells could act as a trigger for the expression of the progeny markers and the effective differentiation of the stem cells.

## Concluding remarks

The data presented in this paper indicate that: i) we successfully setup a papain-based dissociation protocol that improves both yield and cell viability compared to other published methods; ii) we established a library of 66 mouse monoclonal antibodies against planarian plasma membrane proteins, 51 of which show membrane-specific immunostaining of live cells; iii) approximately half of these antibodies recognize subpopulations of whole planarian cells and at least three of them (6-9.2, 7-22.2 and 8-22.2) recognize subfractions of the three FACS populations, demonstrating that the stem cells in S/G2/M phase (X1 cells) are heterogeneous at the level of surface markers; iv) the heterogeneity of the X1 population is not related to the progression throughout the cell cycle, since X1<sup>6-9.2+</sup> and X1<sup>8-22.2+</sup> do not cluster differently from X1<sup>6-9.2-</sup> and X1<sup>8-22.2-</sup> in relation to the Hoechst signal intensity; v) the integration of live-

immunostaining FACS data and qPCR data revealed the existence of at least two discrete subpopulations of X1 cells, both expressing all the stem cell markers tested at the same level: one (about two third of the X1 population) is made up of 6-9.2(8-22.2)<sup>+</sup>/NB.32.1g<sup>+</sup>/*Agat1*<sup>+</sup>/*Myhc*<sup>+</sup> cells and one (about one third of X1 population) is made up of 6-9.2(8-22.2)<sup>-</sup>/NB.32.1g<sup>-</sup>/*Agat1*<sup>-</sup>/*Myhc*<sup>-</sup> cells, which probably consists of uncommitted self-renewing stem cells. Whether the 6-9.2<sup>+</sup> (or 8-22.2<sup>+</sup>) cells in the X1 population are committed to differentiate (and therefore no longer self-renewing) can only be determined by putting their actual pluripotency to the test and by comparing the clonogenic efficiency of single 6-9.2<sup>+</sup> or 6-9.2<sup>-</sup> cells engrafted into a lethally irradiated host.

## Materials and Methods

### Species and maintenance

The animals (clonal line BCN10) used for this study were maintained as previously described (Fernandez-Taboada *et al.*, 2010). Animals used for the experiments were starved for at least one week.

### Cell dissociation

To determine the amount of cells isolated per milligram of live animal, the animals were weighted before dissociation. To this end, four 8 to 10 mm long animals were placed onto a glass slide and residual planarian water was completely removed using a KimWipe tissue. Unless differently stated, chemicals are from Sigma-Aldrich, Germany. Prior to dissociation the animals were incubated 1 minute in 2% L-cysteine hydrochloride pH 7.4 to remove the mucus. Mechanical and trypsin-based dissociation were performed as previously described (Fernandez-Taboada *et al.*, 2010; Hayashi *et al.*, 2010). For papain-based dissociation the animals were cut into small pieces on a glass slide. The pieces were then transferred using wide bore-hole 1000G tips (Art-tips, MβP, USA) into 1.5 ml reaction tubes using 250 μl CMFH (2.5 mM NaH<sub>2</sub>PO<sub>4</sub>·2H<sub>2</sub>O; 14.3 mM NaCl; 10.2 mM KCl; 9.4 mM NaHCO<sub>3</sub>; 15 mM Hepes; 0.1% BSA; 0.5% Glucose; pH 7.2). Then 250 μl of 2 x papain solution (30U/ml Papain (Worthington Biochemical Corp., USA); 2 mM L-Cysteine in CMFH) was added and the reaction was incubated for 1 hour at 25°C. After adding 500 μl of 2 x STOP solution (1mg/ml chicken ovomucoid; 40μg/ml DNaseI (Worthington Biochemical Corp., USA) in CMFH) the pieces were gently triturated by pipetting. The cells were then washed once, resuspended in CMFH and counted using a hemocytometer.

### Purification of planarian plasma membrane proteins

The purification of planarian plasma membrane proteins (PMPs) was performed using the Qproteome™ Plasma Membrane Protein Kit (Qiagen, Germany), according to the manufacturer's instructions. Because planarian cells are smaller than mammalian cells, 5x10<sup>7</sup> cells were used for each preparation. Briefly, planarians were dissociated into a single cell suspension using papain. Cells were incubated in hypotonic buffer and subsequently mechanically disrupted using a 27-gauge needle. Large organelles (nuclei, mitochondria) were removed from the lysate by centrifugation. The PMPs were isolated from the supernatant by adding lectin-based ligands coupled to avidin, which were then bound to streptactin-coupled magnetic beads. The beads were subsequently separated by applying a strong magnetic field (Magnetic Particle Concentrator MPC-S, Dynal, Germany).

### Immunization with plasma membrane proteins

Prior to immunization, isolated PMPs were mixed 1:1 with Adjuvant MM (GERBU, Germany). For each immunization approx. 20 μg of PMPs were used for intraperitoneal injection. After four rounds of immunization, immune sera were collected for testing.

### Production of the antibody library

To produce the antibody library, mouse 3018 was sacrificed and the spleen

isolated and dissociated. Splenocytes were mixed 5:1 with myeloma cells P3-X63-Ag8 (ATCC, USA). Cell fusion was performed in the presence of polyethylene glycol as described previously (PEG 1500 - Serva, Heidelberg, Germany; Key *et al.*, 1993). The hybridoma cells were seeded into 8 cloning plates (cloning plate 24 well, 16 compartments/well, Greiner Bio-one, Germany) and selected in HAT (hypoxanthine, aminopterin, thymidine) medium for about 10 days. Hybridoma supernatants were screened for immunoreactivity against planarian PMPs by ELISA.

#### ELISA

Planarian PMPs were diluted to a concentration of 1 µg/ml in PBS. Every well of a PVC microtiter plate was coated with 50 µl of the diluted membrane proteins and incubated for 1 hour at RT or overnight at 4°C. Coating solution was removed and unspecific binding sites were blocked by adding 1% BSA in PBS for 30 minutes at RT. In each well, 50 µl of the respective hybridoma supernatant was added and incubated for 1 hour at RT. After rinsing with PBS for three times, PMP-coated plates were incubated with 50 µl of anti-mouse IgG- (Fc specific, 1:20000) and anti-mouse IgM-HRP-conjugated (1:1000) in 1% BSA in PBS for 1 hour at RT. After incubation with the secondary antibodies, PMP-coated plates were washed three times with PBS. For detection, 50 µl of substrate solution containing 60 ng/ml 3,3',5,5'-tetramethylbenzidine (TMB) and 0.006% H<sub>2</sub>O<sub>2</sub> in 0.1 M citrate buffer, pH 6.0 was dispensed in each well. After sufficient color development, 50 µl 2 M H<sub>2</sub>SO<sub>4</sub> was added to the wells to stop the reaction. Absorbance was detected at 450nm with a plate reader (Synergy MX, Biotek Instruments, Germany).

#### Flow cytometry and cell sorting

Staining of planarian cells for flow cytometry was performed as previously described (Fernandez-Taboada *et al.*, 2010). Data were acquired using a Gallios flow cytometer (Beckmann Coulter, Germany). For live staining of planarian cells with the antibodies of the library, the cell suspension was incubated 90 minutes at RT with Hoechst 33342 and CalceinAM under agitation. Then, the respective hybridoma supernatant was added (1:4 in CMFH) and the cells were incubated for 20 additional minutes at RT under agitation. Subsequently, the cells were stained with rabbit anti-mouse Alexa Fluor 647 (1:1000 in CMFH) for 15 minutes. After filtering the cell suspension through a 30 µm nylon mesh, propidium iodide was added at a concentration of 1 µg/ml for dead cells exclusion. Fluorescence activated cell sorting was performed using a BD FACS Aria II (Becton Dickinson, Germany).

#### Immunocytochemistry of planarian cells

Planarian cells were plated onto cell culture dishes coated with polyornithine (10 µg/ml in water). The cells were allowed to adhere to the plates for 2 hours at RT or overnight at 4°C. Subsequently, the cells were live stained with the respective hybridoma supernatant diluted 1:1 in CMFH for 15 minutes at RT. After wash in CMFH and fixation with 4% paraformaldehyde for 15 minutes, the cells were incubated with goat anti-mouse IgG Alexa Fluor 488 (1:500, Invitrogen, Germany), goat anti-mouse IgM Alexa Fluor 568 (1:500, Invitrogen, Germany) and 1 µg/ml Hoechst 33342 (Invitrogen, Germany) in PBS containing 1% BSA and 5% normal goat serum (Sigma-Aldrich, USA) for 30 minutes at RT. Imaging was performed using a Leica AF6000 inverted fluorescence microscope (Leica, Germany). To stain permeabilized planarian cells, the cells were fixed with 4% paraformaldehyde and permeabilized with PBS containing 0.3% Triton X-100 and 5% normal goat serum. Subsequently, the cells were immunostained as described above.

#### Quantitative realtime PCR

Reverse transcription and qPCR were performed as previously described (Fernandez-Taboada *et al.*, 2010). Briefly, 2 x 10<sup>4</sup> cells from each FACS subpopulation were collected and RNA was extracted using RNA XS columns (Machery & Nagel, Germany) according to the manufacturer's instructions. After reverse transcription with random hexamer primers, qPCR was performed using Taqman chemistry. Gapdh (mk4.002051.00)

served as endogenous control. Relative quantification of gene expression was calculated using the  $\Delta\Delta C_t$  method, with calibration on the peak expression. Probes and primers are listed in supplemental Table 1. For each condition, three technical replicates for each of the three biological replicates were performed.

#### Western blotting

Planarians cells were lysed in 20 mM Tris-HCl pH 8.0; 150 mM NaCl; 1% Nonidet P-40; 10% Glycerol; 2 mM EDTA; 1x complete protease inhibitor (referred to as NP-40 lysis buffer). The lysate was homogenized through a 27 Gauge needle and cleared from cellular debris by centrifugation. Whole cell lysate and purified planarian membrane proteins were each mixed 1:1 with 2x Laemmli buffer (126 mM Tris-HCl pH 6.8; 20% Glycerol; 4% SDS; 10%  $\beta$ -Mercaptoethanol; 0.02% Bromphenol blue.) and denatured for 5 minutes at 99°C. The samples were separated by SDS-PAGE and transferred onto a PVDF membrane (Immobilon, Millipore, Germany). The membrane was blocked in PBS containing 0.1% Tween 20 and 5% non-fat dry milk powder (blocking solution; Bio-Rad, Germany) and probed with the immune sera 1:2000 in blocking solution. As secondary antibodies, goat anti-mouse IgM- and goat anti-mouse IgG-HRP conjugated 1:10.000 in blocking solution were used. Immunodetection was performed with ECL Plus Western Blotting Detection Reagents (Amersham Bioscience, Germany).

#### Statistical analysis

Prism 5.0 (Graphpad, USA) was used for the statistical analysis. Percentages of cells obtained by the different dissociation protocols and of the immunostained cells are expressed as Mean  $\pm$  SD; The percentages of live-immunostained (FACS) cells and gene expression data are expressed as Mean  $\pm$  SEM. Two-tailed t-test was used to compare side-by-side FACS and qPCR datasets. One-way ANOVA was used to weight the variance of the qPCR data from the 6-9.2 against 8-22.2 subfractions and to compare the cell dissociation methods (with Dunnet post-hoc test).

#### Acknowledgments

The authors would like to thank Karin Hübner for critically reading and Susanne Kölsch for editing the manuscript. This work was supported by the Max Planck Society.

#### References

- BAGUÑA, J., SALO, E., AULADELL, C. (1989). Regeneration and pattern formation in planarians. III. Evidence that neoblasts are totipotent stem cells and the cells. *Development* 107: 77-86.
- BUENO, D., BAGUNA, J. and ROMERO, R. (1997). Cell-, tissue-, and position-specific monoclonal antibodies against the planarian *Dugesia* (*Girardia*) *tigrina*. *Histochem Cell Biol* 107: 139-149.
- CEBRIÀ, F., VISPO, M., BUENO, D., CARRANZA, S., NEWMARK, P. and ROMERO, R. (1996). Myosin heavy chain gene in *Dugesia* (*G.*) *tigrina*: a tool for studying muscle regeneration in planarians. *Int J Dev Biol* 40: S177-S178.
- EISENHOFFER, G.T., KANG, H. and SANCHEZ ALVARADO, A. (2008). Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell Stem Cell* 3: 327-339.
- FERNANDEZ-TABOADA, E., MORITZ, S., ZEUSCHNER, D., STEHLING, M., SCHOLER, H.R., SALO, E. and GENTILE, L. (2010). Smed-SmB, a member of the LSM protein superfamily, is essential for chromatoid body organization and planarian stem cell proliferation. *Development* 137: 1055-1065.
- GENTILE, L., CEBRIA, F. and BARTSCHERER, K. (2011). The planarian flatworm: an *in vivo* model for stem cell biology and nervous system regeneration. *Dis Model Mech* 4: 12-19.
- GUO, T., PETERS, A.H. and NEWMARK, P.A. (2006). A Bruno-like gene is required for stem cell maintenance in planarians. *Dev Cell* 11: 159-169.
- HANDBERG-THORSAGER, M., FERNANDEZ, E. and SALO, E. (2008). Stem cells and regeneration in planarians. *Front Biosci* 13: 6374-6394.
- HAYASHI, T., ASAMI, M., HIGUCHI, S., SHIBATA, N. and AGATA, K. (2006). Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Dev*

- Growth Differ* 48: 371-380.
- HAYASHI, T., SHIBATA, N., OKUMURA, R., KUDOME, T., NISHIMURA, O., TARUI, H. and AGATA, K. (2010). Single-cell gene profiling of planarian stem cells using fluorescent activated cell sorting and its "index sorting" function for stem cell research. *Dev Growth Differ* 52: 131-144.
- HIGUCHI, S., HAYASHI, T., HORI, I., SHIBATA, N., SAKAMOTO, H. and AGATA, K. (2007). Characterization and categorization of fluorescence activated cell sorted planarian stem cells by ultrastructural analysis. *Dev Growth Differ* 49: 571-581.
- HUETTNER, J.E. and BAUGHMAN, R.W. (1986). Primary culture of identified neurons from the visual cortex of postnatal rats. *J Neurosci* 6: 3044-3060.
- KEY, G., BECKER, M.H., BARON, B., DUCHROW, M., SCHLUTER, C., FLAD, H.D. and GERDES, J. (1993). New Ki-67-equivalent murine monoclonal antibodies (MIB 1-3) generated against bacterially expressed parts of the Ki-67 cDNA containing three 62 base pair repetitive elements encoding for the Ki-67 epitope. *Lab Invest* 68: 629-636.
- KIMMEL, J.R. and SMITH, E.L. (1954). Crystalline papain. I. Preparation, specificity, and activation. *J Biol Chem* 207: 515-531.
- MORITZ, S., LEHMANN, S., FAISSNER, A. and VON HOLST, A. (2008). An induction gene trap screen in neural stem cells reveals an instructive function of the niche and identifies the splicing regulator sam68 as a tenascin-C-regulated target gene. *Stem Cells* 26: 2321-2331.
- NOTTA, F., DOULATOV, S. and DICK, J.E. (2010). Engraftment of human hematopoietic stem cells is more efficient in female NOD/SCID/IL-2R $\gamma$ -null recipients. *Blood* 115: 3704-3707.
- PEARSON, B.J. and SANCHEZ ALVARADO, A. (2010). A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. *Development* 137: 213-221.
- REDDIEN, P.W., OVIEDO, N.J., JENNINGS, J.R., JENKIN, J.C. and SANCHEZ ALVARADO, A. (2005). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science* 310: 1327-1330.
- SANCHEZ ALVARADO, A., NEWMARK, P.A., ROBB, S.M. and JUSTE, R. (2002). The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development* 129: 5659-5665.
- SCIMONE, M.L., MEISEL, J. and REDDIEN, P.W. (2010). The Mi-2-like Smed-CHD4 gene is required for stem cell differentiation in the planarian *Schmidtea mediterranea*. *Development* 137: 1231-1241.
- WAGNER, D.E., WANG, I.E. and REDDIEN, P.W. (2011). Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science* 332: 811-816.



## Artículo 2

### *A proteomics approach to decipher the molecular nature of planarian stem cells*

Enrique Fernández-Taboada\*, **Gustavo Rodríguez-Esteban\***, Emili Saló and Josep F Abril.

**BMC Genomics** (2011) 12:133

\*Ambos autores han contribuido por igual.





## Resumen

## Aproximación proteómica para descifrar la naturaleza molecular de las células madre de la planaria

En este artículo se presenta el proteoma del neoblasto, la célula madre de la planaria, y representa la primera lista de genes candidatos relacionados con neoblastos identificados utilizando una aproximación proteómica en planaria. El resultado muestra poca correspondencia con los estudios genómicos previos (Reddien *et al.*, 2005a; Rossi *et al.*, 2007).

Las células madre de planaria se pueden eliminar por medio de radiación, puesto que son las únicas células que se dividen en el organismo (Reddien & Sánchez Alvarado, 2004). A fin de extrapolar el proteoma del neoblasto, comparamos los patrones de proteínas extraídas de planarias normales con aquellos de planarias después de ser irradiadas en geles bidimensionales.

La herramienta de identificación de proteínas MASCOT asignó significativamente más péptidos usando la base de datos Smed\_URF que las bases de datos NCBI-nr o RefSeq, indicando que, a pesar del gran número de entradas en estas últimas, las bases de datos de secuencias como NCBI o UniProt están lejos de ser completas.

Aparte de la presencia de proteínas del metabolismo que corroboran la elevada actividad metabólica de los neoblastos, se identifican proteínas implicadas en la expresión de genes y la regulación de la transcripción, proteínas relacionadas con el desarrollo y la diferenciación, y proteínas involucradas en la respuesta al estrés y la apoptosis. Varios de los genes que se aportan en este artículo estaban presentes en estudios diseñados para identificar genes específicos de células madre en otros organismos modelo (Baharvand *et al.*, 2008; Hoffrogge *et al.*, 2006; Maurer *et al.*, 2004; Zenzmaier *et al.*, 2003). Además, cinco de los genes relacionados con neoblasto (*Hsp40*, *Hsp60*, *Hsp70*, *chaperonin containing TCP1 theta subunit* y *splicing factor 3b subunit 1*) habían sido analizados también en un microchip de transcripción (Rossi *et al.*, 2007) pero solo *Hsp60* había sido detectado, demostrando así la importancia de la proteómica como complemento a las aproximaciones genómicas.



RESEARCH ARTICLE

Open Access

# A proteomics approach to decipher the molecular nature of planarian stem cells

Enrique Fernández-Taboada<sup>†</sup>, Gustavo Rodríguez-Esteban<sup>†</sup>, Emili Saló<sup>\*</sup>, Josep F Abril<sup>\*</sup>

## Abstract

**Background:** In recent years, planaria have emerged as an important model system for research into stem cells and regeneration. Attention is focused on their unique stem cells, the neoblasts, which can differentiate into any cell type present in the adult organism. Sequencing of the *Schmidtea mediterranea* genome and some expressed sequence tag projects have generated extensive data on the genetic profile of these cells. However, little information is available on their protein dynamics.

**Results:** We developed a proteomic strategy to identify neoblast-specific proteins. Here we describe the method and discuss the results in comparison to the genomic high-throughput analyses carried out in planaria and to proteomic studies using other stem cell systems. We also show functional data for some of the candidate genes selected in our proteomic approach.

**Conclusions:** We have developed an accurate and reliable mass-spectra-based proteomics approach to complement previous genomic studies and to further achieve a more accurate understanding and description of the molecular and cellular processes related to the neoblasts.

## Background

As we move further into the post-genomic era it becomes increasingly clear that DNA sequence data alone is insufficient to explain complex cellular and molecular processes. Although the enormous volume of data generated by genome sequencing projects, expressed sequence tags (ESTs), and cDNA analyses has improved our understanding of many processes, they often fail to reflect the influence of post-transcriptional modifications and protein interactions or offer a true reflection of protein levels or activity. Consequently, the role of specific proteins is relatively difficult to determine with confidence on the basis of mRNA expression or genomic data alone [1,2].

Proteomic approaches offer a more realistic description of protein function and its influence on cell dynamics. Although comparative analysis of phenotypically different biological samples, such as in diseased versus healthy tissue [3], remains a challenge, those studies raise the possibility of identifying the protein

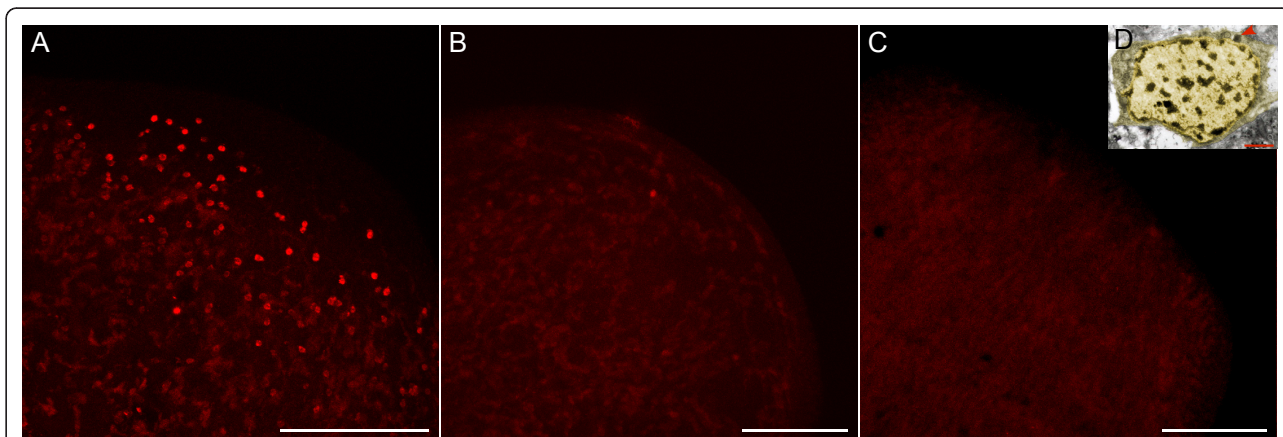
“signatures” that underlie key biological phenomena [4]. Furthermore, the use of bioinformatics to integrate data obtained using genomic and proteomic techniques could help to bypass the limitations of each approach and achieve a more comprehensive view of the information flow within cells.

Planarians, an emerging model system for the investigation of stem cell and regenerative biology, [5-7], have a unique population of stem cells called neoblasts (see Figure 1), which can give rise to all of the differentiated cell types present in the adult organism during regeneration or normal homeostasis [8,9]. Albeit a great deal is now known about the biology of these cells, most molecular data have come from cDNA and genomic analyses. The neoblasts are particularly suited to proteomic approaches, however, as they contain chromatoid bodies (CB) that are progressively lost during differentiation [10-12] and can be employed as a marker for undifferentiated cells. The CB complexes are mainly formed by proteins and latent mRNA molecules, which can distort the levels of gene expression in transcriptional analyses of neoblasts samples. Moreover, since the neoblasts are the only dividing cells in the planaria [5], they can be easily depleted by irradiation [13]. Thus, these unique

\* Correspondence: esalo@ub.edu; jabil@ub.edu

† Contributed equally

Departament de Genètica and Institute of Biomedicine (IBUB), Universitat de Barcelona, Av. Diagonal 645, 08028, Barcelona, Catalonia, Spain



**Figure 1 Neoblast depletion by irradiation and image of a neoblast shown by electron microscopy.** Immunostaining with anti-phosphorylated histone H3 ( $\alpha$ H3P), labelling mitotic neoblasts in 3-day head-regenerating organisms: A, control; B, 75 Gy irradiated 3 days after irradiation; and C, 75Gy irradiated 14 days after irradiation. Whereas a high number of proliferating cells appear in control animals next to the blastema and some mitotic cells still remain 3 days after irradiation, no divisions are detected after 14 days, showing that neoblasts are completely eliminated at that time. D, Electron microscopy image of a neoblast cell. Cytoplasm (dim yellow) and nucleus (yellow) are highlighted for clarity. The red arrow indicates a chromatoid body. Scale bars: A-C = 0.5 mm, D = 3  $\mu$ m.

characteristics make planarians an ideal system in which to explore the use of proteomics to analyze the biology of processes such cell differentiation, stem cell behavior, homeostasis and an array of other events. As a first step in the development of such an approach, here we describe the methodological establishment and validation of a proteomic analysis of the planarian neoblast.

## Results

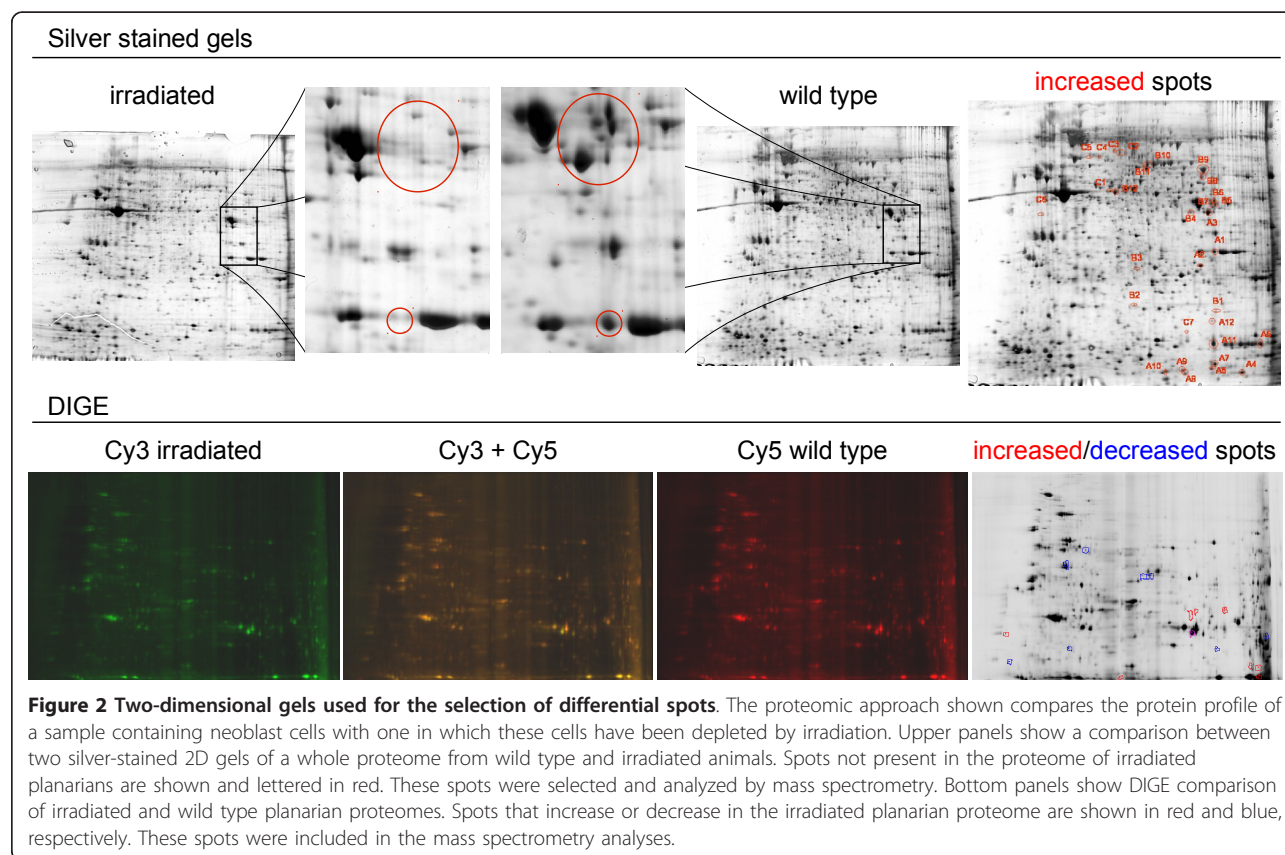
### Establishment of the planarian proteomic approach

Different methods were tested to achieve a consistent and reproducible pattern on two-dimensional (2D) gels. To optimize sample preparation, proteins were extracted from dissociated cells or from whole animals. The yield from dissociated cells was insufficient to establish an efficient 2D procedure. Furthermore, the reproducibility of the 2D gel pattern was poor (data not shown). Prior to extraction from whole animals, a short treatment with 2% cysteine chloride in planarian water was used to eliminate mucous production, which is known to interfere with molecular techniques [14]. Based on our tests and previous work by Collet and Bagnà [15], we established a consistent method for 2D analysis from planarian samples (Figure 2 and Additional File 1). The different lysis buffers and sample cleaning procedures tested are shown in Table 1. Between 50 and 1000  $\mu$ g of total planarian proteins were loaded on 2D gels to establish the best sample quantity in terms of spot definition. From 100 to 500  $\mu$ g the spot resolution was acceptable. We selected the 500  $\mu$ g as the optimal amount of protein to load onto 2D gels to achieve the maximum number of spots. A minimum of 100  $\mu$ g was necessary for spot detection. Different immobilized pH

gradient strips were used and the second-dimension protocol was modified to avoid streaking problems (Table 1). All these variables were tested on 12-cm 2D gels and scaled up to 24-cm gels for subsequent procedures.

### Proteomic data

In order to identify proteins specifically expressed in neoblasts, we compared 2D patterns of two samples: wild type (WT) versus irradiated animals (IA). This method has been extensively used to study the effects of neoblast depletion [8,13]. Extractions were done 14 days after irradiation, when animals remained viable but cell proliferation was absent (Figure 1). Once the protocol was set up and the spot patterns were reproducible (Figure 2 and Additional File 2), the spots were compared and selected. Although spot labelling by silver staining and DIGE was consistent in each case, we did not succeed in obtaining a uniform pattern with the two techniques. Follow-up analysis was therefore done separately. With the aim of establishing the real potential of the silver-staining technique, only clear and conserved qualitative comparison based on silver staining was considered (present in WT sample and not present in irradiated sample). Image master 2D<sup>TM</sup> software (from Amersham Biosciences) was used to analyze the scanned gels. However, the potential bottleneck of this proteomic approach is the image analysis. Many authors have highlighted the difficulties in obtaining good replicates [16], and this has now been partially overcome with the use of DIGE. Whereas our silver-staining results showed remarkable pattern conservation within replicates (Additional File 2), the numbers after spot image analysis showed some variability. In order to improve signal specificity we used two types of



gels, one loaded with 100 µg and another with 500 µg of sample protein. The differences between irradiated and non-irradiated samples that were conserved in both sample loads and also had three surrounding reference spots in both experimental conditions were selected after reviewing the correspondence in Ip and Mw (Figure 2). These

**Table 1 Variables taken into account for the establishment of the planarian proteomic protocol using 2D gels**

<b>Samples:</b>	Whole planarian extracts, dissociated cell extracts, dissociated cell and sub-fractionated extracts.
<b>Extraction Buffers:</b>	SDS, urea/thiourea.
<b>Processing Sample:</b>	(Precipitation procedure) Amersham 2D clean up kit, acetone, TCA-acetone.
<b>Isoelectric Focusing</b>	(Immobiline Dry strip gels 24 cm) Linear pH 4-7, <b>(1st Dimension):</b> Linear pH 7-11, Non-linear pH 3-11.
<b>Other Modifications:</b>	Trypsin inhibitors, general protease inhibitors, sonication.

All the different variables affecting protein sample production and 2D gel electrophoresis are listed on this table.

restrictions reduced the number of selected spots substantially, but ensured a high degree of confidence in the differences selected, providing a better platform for validation of the technique. For DIGE staining, the standard protocol was followed without modifications and the analysis software was used with the default parameters. Only clear and conserved quantitative changes (>2-fold changes) were selected, drastically reducing the number of final candidate spots (Figure 2). A total of 26 and 58 spots were selected for silver and DIGE staining, respectively (Table 2).

#### Computational analyses

MASCOT [17] was tested against different open reading frame (ORF) datasets derived from NCBI-nr/RefSeq

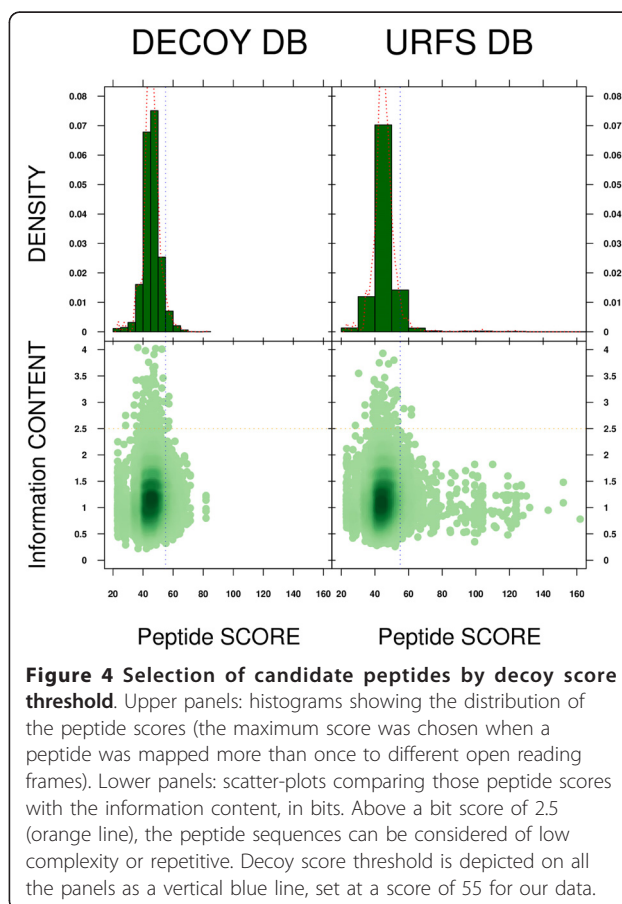
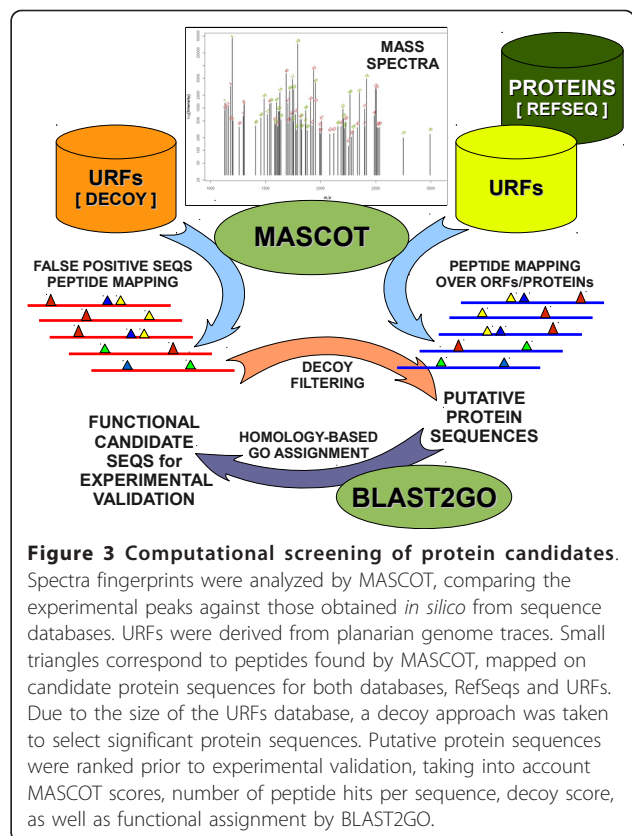
**Table 2 Spot counts for the 2D gels**

	Semi-Automatic Procedure		Final Selected Spots
	Irradiated	Wild type	
100-SIL	1182 ± 43.13	901 ± 77.07	26
500-SIL	1931 ± 92.63	1413 ± 81.31	
500-DIGE		2445	58

Summarized data are shown for the 2D gel analyses. Image master 2DTM software (from Amersham Biosciences) was used to analyze the scanned gels. SIL, Silver staining; 100-SIL, 100 µg of total protein extract loaded on the gel; 500-SIL, 500 µg of total protein extract loaded on the gel; DIGE, differential in gel electrophoresis.

[18,19], *Schmidtea mediterranea* ESTs [20], the contigs for the planarian genome WUSTL assembly version 3.1 [21], and *S. mediterranea* whole-genome shotgun reads (traces). Of those datasets only NCBI-nr and traces are discussed here; the former is routinely used on this kind of analyses, while the latter yielded the largest number of peptide assignments (unpublished results). MASCOT assigned 20,107 peptides to spectra for NCBI-nr, which mapped to 602 protein sequences. Sequences from traces contained in the “forward” database were reversed to produce a “decoy” database containing sequences of the same length and composition but a different distribution of trypsin targets to those from the “forward”; Figure 3 illustrates the whole process. MASCOT returned 50 hits per search on each trace database, both for “forward” and “decoy”. This resulted in 100 hits per search, for a total MS-fingerprint of 83 different spots.

MASCOT predicted a total of 44,712 and 36,956 peptides for the forward and decoy databases, respectively, and these were mapped to 8300 unique ORFs (URFs), corresponding to 23,376 and 26,741 unique peptide sequences. When the same peptide was mapped on two or more URFs, the highest score was retrieved. Figure 4 shows the score distribution of the two sets of unique peptides. Assuming that the decoy database comprised reversed sequences, it would be expected that none of



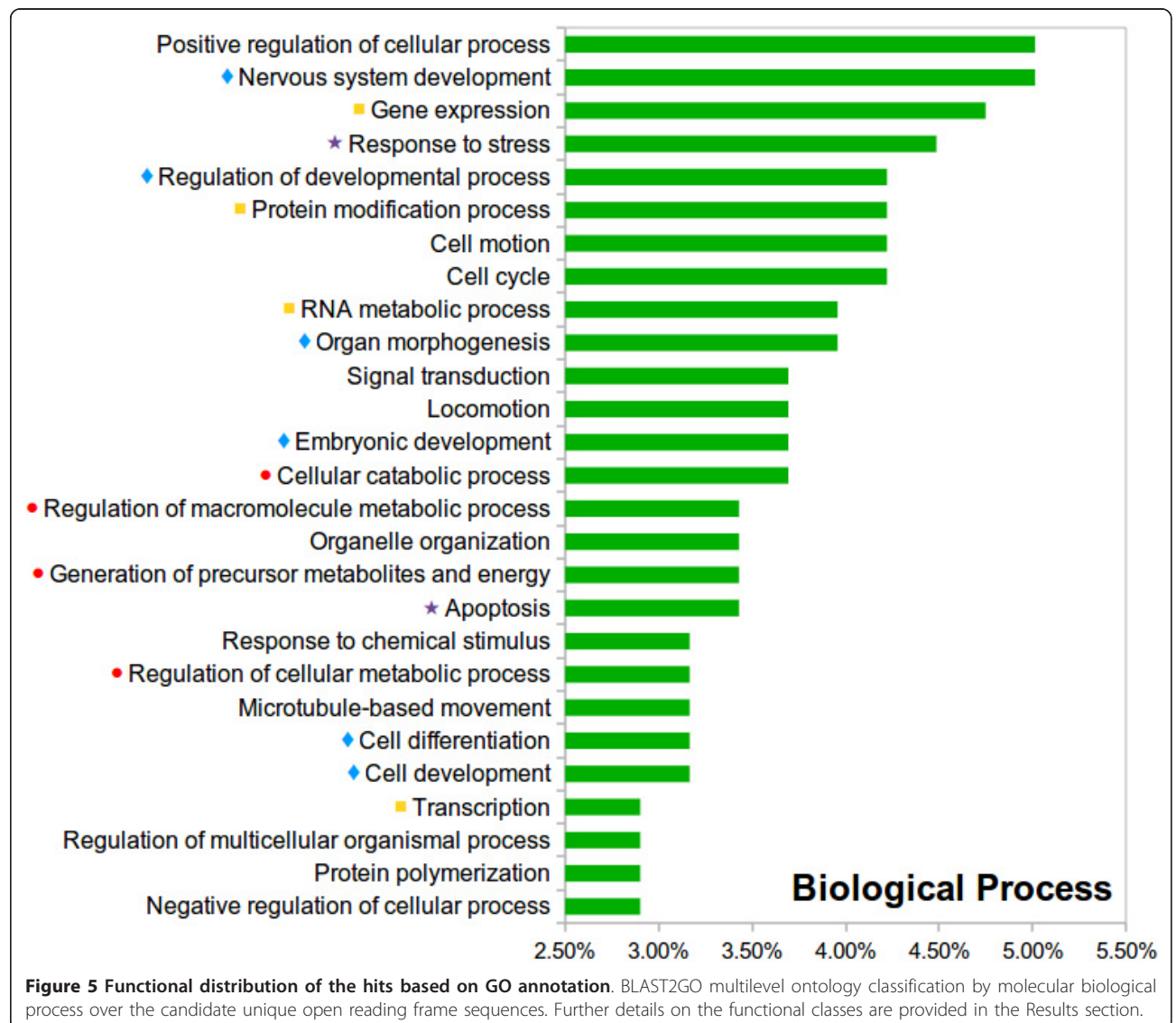
the peptide hits found there would be real. Assuming that by chance some of the peptide sequences predicted for this set could be similar to those from the forward database, we can thus consider a false-negative error rate in order to determine a score threshold for both datasets. On this basis, for a 5% false-negative error rate in the decoy database, 1337 peptides would be above the threshold. Ranking the list of peptides, sorting by score, and taking 5% of the highest scoring peptides, the score threshold was set at 55 (shown in all panels of Figure 4 as a vertical blue line). When applying that score cut-off to the peptides obtained from the forward database, 1249 of 23,376 unique peptides (5.34%) from that database were “decoy” filtered. Translating this to the 8300 URFs used to detect the peptides, 1728 of these had at least one significant “decoy” peptide mapped onto it or was aligned with one such URF sequence. Therefore, 20.82% of the URFs can be considered more reliable than the rest.

The sequences of all the URFs for the forward database were uploaded into the BLAST2GO software suite [22,23]. The first step was to compare those amino acid sequences to homologous proteins (using BLASTP against NCBI-nr, min e-value = 0.001, min hsp length = 25). Of the URFs

with scores above decoy threshold, 1416 (81.94%) had at least a significant BLAST hit. In contrast, only 636 out of 6572 URFs with scores below the decoy threshold (10.71%) also had one or more significant BLAST hits. It was possible then to obtain a functional Gene Ontology (GO) annotation for those URFs having a BLAST hit against a known functionally annotated protein. Results of the functional annotation are summarized in Figure 5.

After GO assignment and the corresponding functional annotation of the sequences derived from our approach, enzyme codes were mapped by BLAST2GO when possible. With such codes it was possible to retrieve the KEGG pathway where the protein may play its role on the planarian molecular biology. However, less than one third of the sequences had a homologous gene/protein BLAST hit—especially for URFs dataset—, and from those many had a GO functional assignment.

A fraction of the sequences with at least one GO hit was linked to an enzyme code, which would be related to a component of the KEGG pathways: 1,670 of 2,804 clusters, mapping to 118 pathways, and 131 of 5,528 clusters, mapping to 35 pathways, for MASCOT results on RefSeq and URFs respectively. All 35 pathways for URFs were also found using the RefSeq dataset. The lower ratio for the URFs set can be explained by species specific sequences, proteins or functions that are not yet annotated on the reference databases. 297 RefSeq clustered sequences had a match to 171 enzyme codes for proteins distributed on the 118 pathways. 16 URFs clustered sequences had a match to 9 enzyme codes for proteins distributed on the 35 pathways. The enzymes can appear on several pathways, due to the hierarchical structure of the KEGG a match can be found on both, a general route as “Metabolic pathway”, and a more





specific process, such as “Glycolysis/Gluconeogenesis”. Among the pathways found, metabolism routes of sugars and lipids were expected, as energy is required for cellular processes, regeneration among them. Nevertheless, there are few candidate sequences that will deserve further analyses, as they appear on pathways close to development and regeneration: “Selenoamino acid metabolism”, “Retinol metabolism in animals”, and “mTOR signaling pathway”. Additional data, including figures of all those pathways with color-highlighted boxes for proteins found, is available on the planarian proteomics web page [24].

### Gene profile

As depicted in Figure 5, the annotated proteins cover a wide range of biological processes, of which four main groups can be emphasized: proteins involved in energy production and metabolism (red dots in Figure 5); gene expression and transcription regulators (yellow squares); proteins related to development and differentiation (blue diamonds); and proteins involved in stress-response pathways and the apoptosis (purple stars). This functional distribution resembles the distributions described in previous studies of embryonic stem (ES) cells [25], proliferating cells [26], and differentiating neural stem cells [27], among others [28-30] (see corresponding table in Additional File 3). Additional protein sequence comparisons were performed using NCBI BLAST [31] (E-value < 10e-3) to extensively compare sets of candidate proteins from our RefSeq and URFs databases with the sequences described in those studies as stem-cell related. The same analysis was applied to the genes reported in two studies using high-throughput approaches to detect neoblast genes by RNAi-feeding [32] and by expression macrochip [33] (see corresponding table in Additional File 3). A total of 822 sequences out of 2801 (29.35%) from the RefSeq dataset and 50 out of 309 (16.18%) from the URFs dataset presented homology with at least one sequence in any of the studies. Yet only 52 (1.86%) from RefSeq and none from the URFs dataset had homology with sequences reported in the planarian studies.

### Functional studies

We performed functional analyses on some candidates from our lists to further assess the quality and accuracy of the approach used. Candidates were selected from the RefSeq and the URFs from the traces (see Table 3). In the case of RefSeq candidates, the sequence was mapped onto the draft genome and primers were designed to clone a longer fragment of the protein for subsequent characterization. Three main groups of genes were selected. The first two groups were proteins belonging to the Ras superfamily of small GTPases and the heat shock proteins

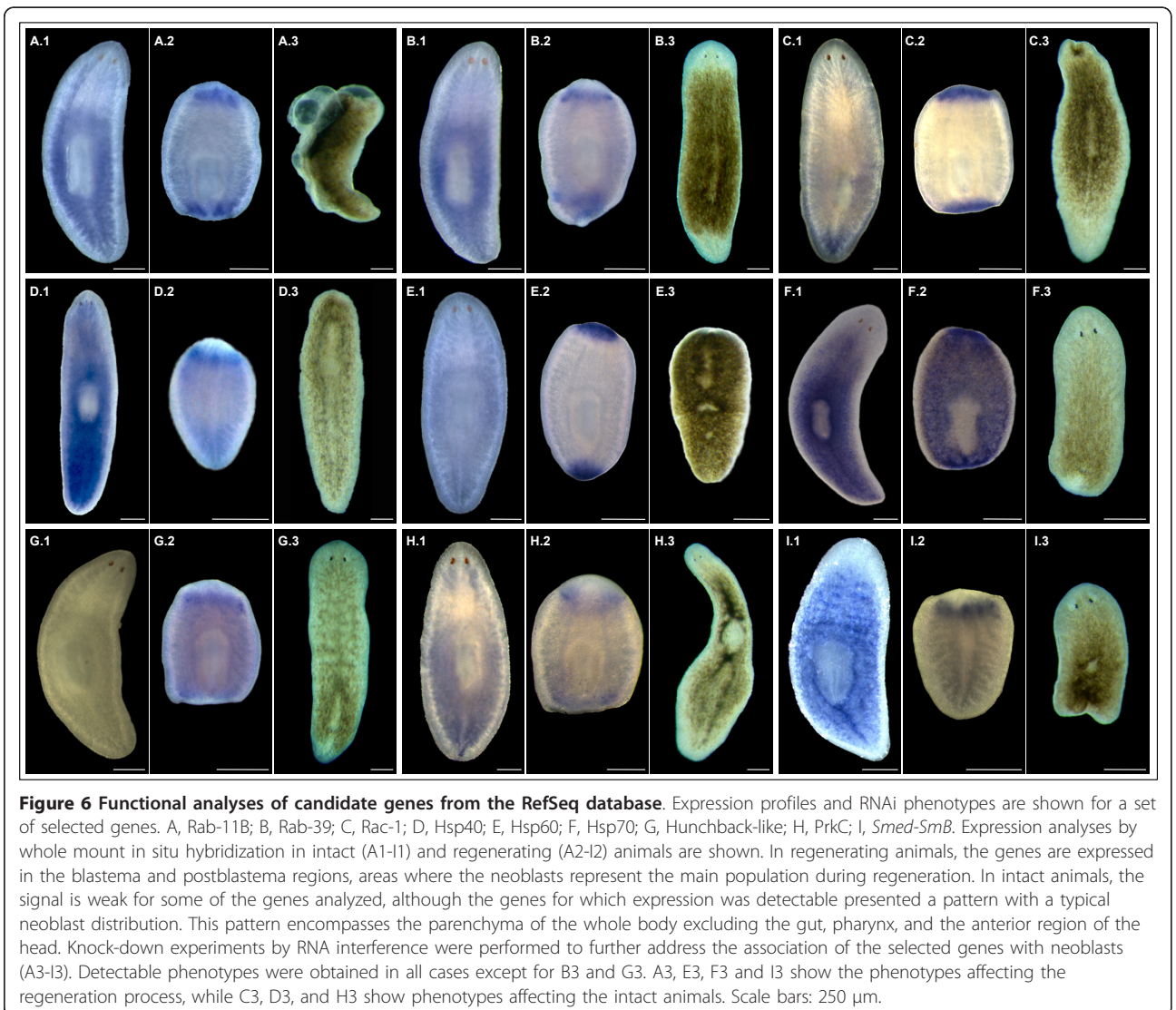
**Table 3 Summary of BLAST hits found for the analyzed candidate sequences**

RefSeq candidate sequences		
ACCESSION NUMBER	BLAST HOMOMOLOGY	E-VALUE
GU591870	Rab-11B, member RAS oncogene family	1e-79
GU591871	Rab-39, Ras-related protein	1e-23
GU591872	Rac-1, ras-related	3e-90
GU591873	Hsp40 (DnaJ)	7e-18
GU591874	Hsp60	3e-103
GU591875	Hsp70 (Mortalin-like protein)	0.0
GU591876	Hunchback-like	1e-50
GU591877	PrkC (cAMP-dependent protein kinase)	2e-57
GU562964	<i>Smed-Smb</i> [58]	4e-38
URFs candidate sequences		
ACCESSION NUMBER	BLAST HOMOMOLOGY	E-VALUE
GU591864	Chaperonin containing TCP1 theta subunit	1e-51
GU591865	Splicing factor 3b subunit 1	6e-109
GU591866	TNF receptor associated factor	3e-25
GU591867	Similar to pol polyprotein	2e-32
GU591868	<i>Unknown protein</i>	—
GU591869	Lectin-like	4e-28

BLAST homologies to both RefSeq and URFs candidate sequences are shown. Candidate sequences coming from MASCOT predictions over the RefSeq database were mapped onto the genome draft of *Schmidtea mediterranea* to retrieve the specific sequences for this species. BLASTP was performed with the species-specific protein sequences against NCBI-nr in order to annotate their function accurately.

(HSP) family. The third group encompassed unrelated genes from different spots. The first family includes the genes Rab-11B, Rab-39 (vesicle and membrane traffic) [34-36] and Rac-1 (cytoskeleton regulation and apoptosis) [37,38]. The second family contains HSPs (40, 60 and 70 kDa) involved in a wide variety of processes [39-41]. The last group contained the transcription factor Hunchback-like (related to *Drosophila* axial polarity and neuroblast lineage) [42], PrkC (a kinase linked to apoptosis and other processes) [43,44] and LSm proteins (RNA processing and regulation) [45-47]. This gene selection was done because no direct relation with neoblasts was established previously, with the exception of the HSPs.

To assess the relationship between these genes and the neoblasts, we analyzed their expression patterns and RNAi phenotypes (Figure 6). The observed expression patterns were variable. Some of the genes were expressed in the blastema (Figure 6C and 6E), where neoblasts migrate to after division in order to regenerate the missing body parts. Others were expressed in the post-blastema (Figure 6B, D, G, H and 6I), where the neoblast population is amplified by division to



generate the cells that will form the blastema. Finally, some genes were expressed in both blastema and postblastema (Figure 6A and 6F). These expression patterns disappeared in late stages of regeneration or developed over time to correspond to the typical expression pattern for neoblasts, distributed throughout the parenchyma with no expression in the pharynx or at the head tip anterior to the eyes [5]. In addition, for some of the genes, expression was only detectable under regeneration conditions, in which neoblasts are known to proliferate at higher rates. In that case, expression was barely detectable when only a basal number of neoblast cells was present in intact adult animals (Figure 6C, E and 6G). Therefore, the expression patterns for the candidate genes were consistent with neoblast expression.

Since neoblasts are known to be the only source of cells for homeostasis and regeneration, the relationship between the selected genes and the neoblasts was

validated by RNAi experiments [48,49]. All injected animals, both intact and regenerating, died within a few days or weeks, except in the case of Rab39 and Hunchback-like (Figure 6B and 6G), for which no phenotype was observed in RNAi experiments. Intact planarians showed a gradual head regression followed by lysis after several weeks, as shown in Figure 6C, D and 6H. This phenotype has been linked to a lack of neoblast cells available for cell renewal [50]. In addition, regeneration was completely absent in fragments from RNAi-treated animals, which produced small blastemas that never differentiated, or no blastema at all with indented wounds, as illustrated in Figure 6A, E, F and 6I.

In a second screen to validate candidate URFs from the traces, the expression of some of these genes was analyzed by comparing intact and irradiated organisms. Whole-mount in situ hybridization in intact adult organisms revealed parenchymal expression consistent

with a neoblast distribution, whereas this expression pattern was not present in irradiated animals (Figure 7A-B). This is consistent with neoblast-related genes, since high-dose irradiation destroys neoblasts. Some genes showed additional expression around the CNS that may have been associated with a non-dividing neural precursor cell type. While this expression pattern remained after irradiation, the signal in the parenchyma disappeared (Figure 7C-E). Finally, the planarian ortholog of C-type lectin-like was only expressed in the digestive system of irradiated organisms and never in intact animals (Figure 7F), suggesting a role in cell renewal under stress conditions, given that the gut has the fastest cell turnover of all tissues. These data provide further support for the involvement of these candidate genes in processes linked to neoblast biology, such as proliferation, cell migration or the regulation of differentiation.

## Discussion

The results of this study show that we have successfully developed a rapid and reliable method for 2D analysis of planarian protein samples (Figure 2 and Additional files 1 and 2). This approach will provide the basis for future proteomics studies that will increase our understanding of a number of biological processes, in planarians and beyond, building upon data obtained using genomics and cDNA-based approaches.

Proteomic studies can help to fill gaps on the annotation of the planarian genome. Despite the large number of entries already submitted, sequence databases such as NCBI [51] or UniProt [52] are far from complete. Recent metagenomic projects have identified novel putative protein sequences not present in current sequence databases, thus extending the range of biological functions that may be represented [53]. For instance, Yooseph et al [54] report up to 1 in 3 orphan ORFs from whole-genome shotgun sequencing of marine samples containing a mixture of prokaryotic organisms. Our findings indicate that MASCOT can assign substantially more peaks on those spots selected from 2D gels when using the Smed\_URF database than with NCBI-nr/RefSeq, as would be expected.

The use of ORF sequences in whole genomes without prior knowledge of where the genes, mainly the exons, are located presents a number of issues that can distort the measures used to discriminate between true and false peptide hits. These include the ratio of coding to non-coding sequences, which can be quite low (around 2% of coding regions for the human genome [55]), and the presence of more repetitive sequences in intergenic regions, despite the fact that some amino acid repeats are vital functional and structural regions in proteins [56]. Moreover, the experimental spectra are compared to simulated ones that

were computed from putative protein-coding regions directly translated from genomic sequences of the same species, not from related homologs from different organisms at different phylogenetic distances.

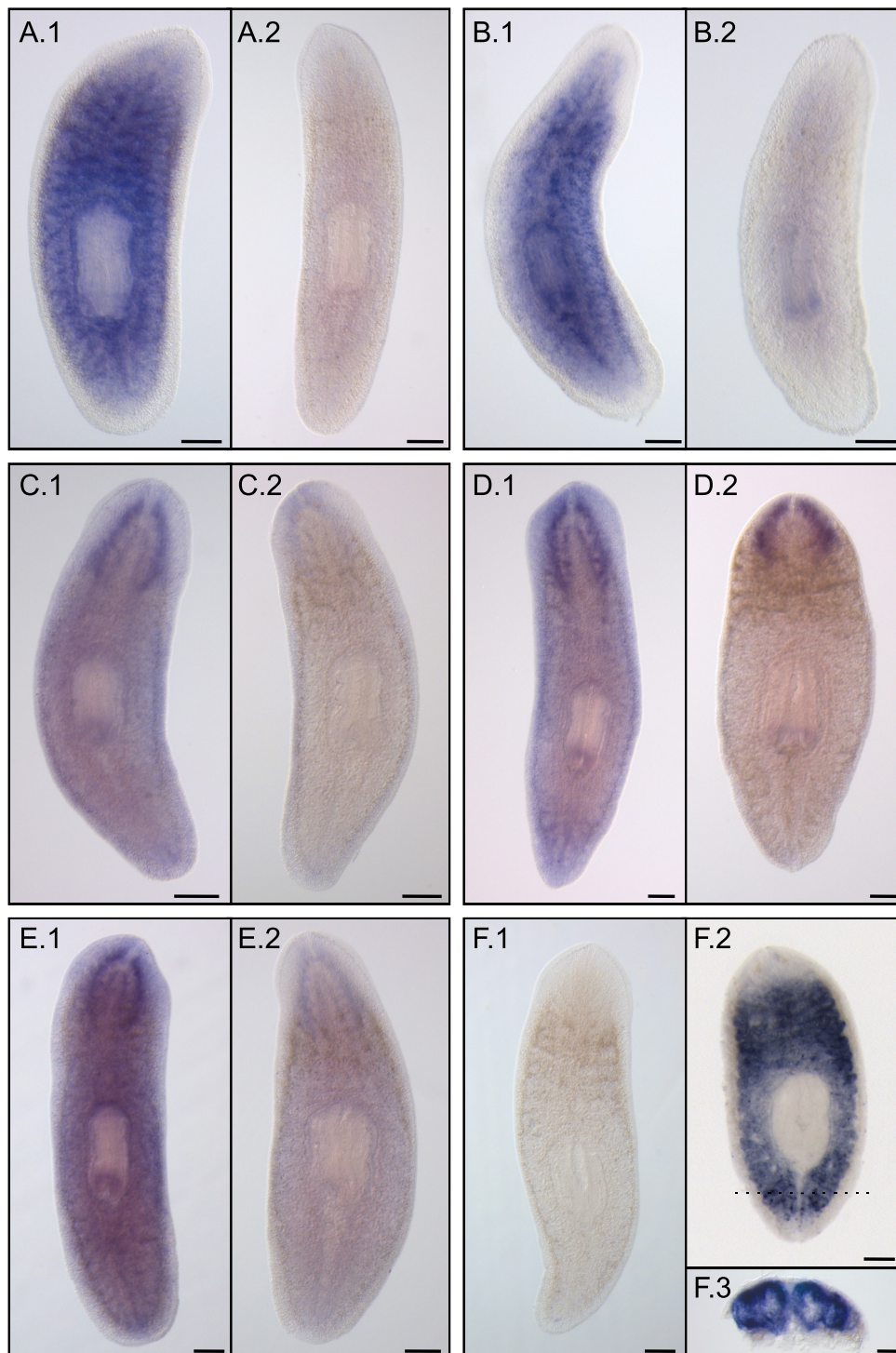
Galindo et al. [57] described a novel family of eukaryotic coding genes consisting of peptides shorter than 50 amino acids (small ORFs [smORFs]) with key biological functions during *Drosophila* development. Therefore, future searches will have to take this into account, for instance removing any length constraint when building up the ORF databases.

## Identification of proteins

Apart from the presence of metabolic proteins that indicate the high metabolic rate of neoblasts, several of the proteins detected in this analysis seem to be good candidates to be involved in neoblast-related functions, and thus in regeneration and tissue homeostasis. One of those, *Smed-SmB*, from the LSM family, has been analyzed in detail and shown to be essential for neoblast proliferation and maintenance [58]. Moreover, other candidates belonging to the HSP class of proteins have been linked to the biology of neoblasts in recent studies [59-61]. The experimental results described in this paper support the use of an ORF database built upon genomic sequences from the same species, which yields, as one might expect, more reliable results in subsequent proteomic searches, despite assuming nothing about the coding content of those ORFs. This will bridge the gap between proteomic and genomic approaches to extend our knowledge of the functional components of emerging model organisms.

## An initial proteomic picture of the neoblasts

The genes identified in this study represent the first list of neoblast-related candidate genes identified using a proteomic approach in planarias (Table 3 and Additional file 4). The results show little correspondence to those of previous genomic studies [32,33]. Interestingly, however, a number of the genes reported in this analysis were also present in studies designed to identify stem cell-specific genes in other model organisms [25-30]. In addition, five of the neoblast-related genes characterized through our proteomic approach (Hsp40, Hsp60, Hsp70, Chaperonin containing TCP1 theta subunit and Splicing factor 3b subunit 1) have also been analyzed in a planarian transcription macrochip, but only one of them was detected (Hsp60) [33]. These findings support our proteomic strategy as a complement to genomic approaches. Furthermore, the large number of putative neoblast-related proteins identified in this proteomic study will be of invaluable help in future research investigating the biology of the neoblast.



**Figure 7 Expression patterns of candidate genes from the *Schmidtea mediterranea* traces database.** Expression in whole mount in situ hybridization of different genes in (1) control and (2) 75 Gy irradiated planarians 6 days after irradiation. A, chaperonin containing TCP1 theta subunit homolog; B, splicing factor 3b subunit 1 homolog; C, TNF receptor-associated factor homolog; D, similar to pol polyprotein; E, unknown protein; F, lectin-like homolog. F.3 shows a higher magnification view of a transverse section from F.2 (dashed line), where the two posterior gut branches were labeled. Scale bars: 250  $\mu$ m in panels A.1 to F.2 and 100  $\mu$ m in panel F.3.

## Conclusions

We have developed a proteomic approach to characterize specific planarian stem-cell (neoblast) proteins. An accurate and reproducible method for protein purification, 2D gel electrophoresis and MS analysis was defined and an ORF database of species-specific genomic DNA was developed for peptide assignment of the retrieved MS spectra. Subsequent computational analyses yielded a list of annotated candidate proteins, some of which were functionally validated as neoblast-specific genes by RNAi and whole-mount *in situ* hybridization. Substantial overlap was observed between the candidate genes identified in our study and those reported from previous analyses of embryonic stem cells, thus validating the specificity of the approach. In addition, we detected novel sequence candidates and expression changes that merit further investigation in future studies to determine their role in stem-cell biology.

## Methods

### Sequences

The genome of *S. mediterranea* (strain S2F2) was sequenced and assembled at the Genome Sequencing Center (GSC) at Washington University in Saint Louis (WUSTL) [62,63]. It contains around 800 Mbp distributed on four chromosomes ( $2n = 8$ ). The latest assembly version, v3.1 [21], comprises up to 90,000 sequences, which were reduced to 45,000 by means of pair-ends sequencing. Lengths of those sequences range from thousands to hundreds of thousands of nucleotides. During the assembly process, sequencing errors can be fixed by aligning different traces, but the software can also reduce polymorphisms and misplace those trace sequences because of the repeats. In order to overcome those limitations, a database of ORFs was produced directly from the set of the whole-genome shotgun reads. About 16 million traces were downloaded from the NCBI Trace Archive [64] and translated, without prior masking, into the six possible reading frames, taking into account only those ORF sequences longer than at least 50 amino acids. The ORFs were stored in a MySQL relational database along with the original sequences, to make it possible to retrieve the original nucleotide sequences and design probes for experimental validation. To reduce the large amount of sequence data produced and thus speed up the peptide searches by MASCOT [65], a set of URFs was derived from the set of ORFs with a checksum function to generate hash keys as unique identifiers for every sequence. A total of 54,382,803 ORFs were retrieved from 16,580,722 shotgun reads. This resulted in 28,946,081 URFs with properly formatted sequences to populate a MASCOT database. As MASCOT was not able to work with databases larger than 24 million

entries, the original set was split into two databases. MASCOT results for both sets were then merged to get the final set of ORFs that had at least one peptide matching spectra. The probability of false matches increases when large databases, with millions of protein sequences, are used to detect a wide variety of possible candidate proteins in a sample [66,67]. To assess the significance of the peptide hits found by MASCOT, a decoy database was built by reversing all the URF sequences [68-70]. It was also split into two, as described above for the "forward" database. MASCOT was run separately on the decoy databases for all the mass fingerprints previously analysed with the original URF dataset.

### Irradiation

Intact asexual planarians were irradiated at 75 Gy (1,66 Gy/minute) with a Gammacell 1000 [Atomic Energy of Canada Limited] [71].

### Sample preparation

Protein samples were obtained from whole animals using a lysis buffer and heating. See Additional File 1 for further details.

### Running 2D gels

First-dimension isoelectric focusing was performed on immobilized pH gradient strips (24 cm, pH 3-10) using an Ettan IPGphor system. Second-dimension SDS-PAGE was performed by laying the strips on 12.5% isocratic Laemmli gels (24 × 20 cm) cast in low-fluorescence glass plates on an Ettan DALT system. Details of the procedure are available in the Additional File 1.

### Sample analysis

Gel spots were extracted and digested before analysis by MS. Then, MASCOT software (Matrix Science, London, UK) was used to search those spectra on different databases. All spectra were processed by PRIDE Converter software [72] and were submitted to the PRIDE database [73], project accession number is **15541**. For details see Additional File 1. After careful selection of score thresholds for the predicted peptides (see the Results section for the values chosen and the final numbers of the filtered datasets), the sequences that allowed detection of the URFs were uploaded into BLAST2GO [22,23]. This software tool facilitates high-throughput integration of sequence data, homology to related species via NCBI-BLAST [31] and functional annotations of DNA or protein sequences based on the Gene Ontology (GO) classification [74]. MASCOT output files, selected peptide and protein sequences, as well as BLAST2GO results and KEGG summary, are available at the planarian proteomics materials web page [24].

## Gene Cloning

Gene identifiers and corresponding forward/reverse primers (including nested primers). GU591870: F1.5' -TC TGGGATACTGCAGTCC-3', R1.5' -GATGGAATAATC GGTGCG-3'; GU591871: F1.5' -TTTTAATTGGTGATA GCATGG-3', R1.5' -CTTGACCTGCTGTATCCC-3'; GU 591872: F1.5' -TGTTGTTGGTGACGGAGC-3', R1.5' -GCACGAATTGCCTCATCG-3', R2.5' -TGTC GGACAG TGATGGG-3'; GU591873: F1.5' -GACTATTATTCAATA TTAGG-3', R1.5' -TACCTCATATGCTTCA GCAA-3'; GU591874: F1.5' -TTGCTGAAGATGTTGA CGG-3', R1.5' -AGAGCGGTACCTCCTCC-3', R2.5' -ACCTCA CTACTACCACCG-3'; GU591875: F1.5' -GAGACAAGCT ACCAAAGATGC-3', R1.5' -CATCCGTAACATCTCCAGC AAG-3'; GU591876: F1.5' -AACAAATATCTGGAATGC CC-3', R1.5' -GCTTAAAATTTCCGCGGAG-3'; GU591 877: F1.5' -CAATATGGCTGAGGCAGC-3', R1.5' -CTG GAGTTCCACACATCG-3', R2.5' -TGGATGGGAAATTT GCTCC-3'; GU562964: F1.5' -CAACACTTCAAGAT GGTGC-3', R1.5' -TTGCACCAGTACCTGGCA-3'; GU591864: F1.5' -CCCAGTTCTTTTCAAGGTTTGAAG- 3', F2.5' -CTGTCTTCCGAAATATCCAAGCATGC-3', R1.5' -CCAAAGATTTTGAATTTACTGCCGTTTCG-3', R2.5' -CTTACCAACAGATTCTTCGTCACG-3'; GU5918 65: F1.5' -GCTCATGCGCTTGGCATTTCGTATTTG-3', F2.5' -CGTTTCTGAAGGCTGTGTGCAAATC-3', R1.5' -CAATGGTGTCCGCGCCTTGAGCAAC-3', R2.5' -CAATTG CTCCTCCAACCGAATGTC-3'; GU591866: F1.5' -GCAA -CAGATGACCAACAATATAAAGG-3', F2.5' -CTAGAAAC CAACAATTTTATAGCCAG-3', R1.5' -CTTGTCCGGCCTC TCTACTTC-3', R2.5' -GATTATCTTCTCGCAAGAAT CCTTCTC-3'; GU591867: F1.5' -CCAGCTTCTCAACA AAGACGGGAC-3', F2.5' -GTTTCAACAGAATGCCGTTTG GAATTGC-3', R1.5' -CCGAAAAACATAAGATTGGCGC CGTC-3', R2.5' -GTTTCAAACCCTCAAACACGCTATT CG-3'; GU591868: F1.5' -GCACTAGATCAAAAAATAGAA GTGTTAGC-3', F2.5' -CTCAAGAAATGGAGGAACCAAGA TTGG-3', R1.5' -CGATCTACTTCTTCTACAATCTC-3', R2.5' -CTGTTTCTGTTCTTCTTCTGACACGTTTC-3'; GU 591869: F1.5' -GGCTAGGTAAGTATTGGATAGATGG-3', F2.5' -GGAAC TGACGATGGGTTGATAG-3', R1.5' -CC AATTTGTGTAGGTCATTTTGCATCC-3', R2.5' -CCATCA TTGAATGTCCATCTTCCAGTG-3'.

## In situ hybridization

Digoxigenin-labeled RNA probes were prepared using an *in vitro* labeling kit (Roche). Whole-mount *in situ* hybridization was performed as described by Agata et al [75], with some modifications: proteinase K (20 µg/ml) treatment for 10 min; triethanolamine treatment was performed as described by Nogi and Levin [76]; hybridization at 55°C for 18 or 30 h; and final probe concentration of 0.07 ng/µl.

## RNA interference

Double-stranded RNAs (dsRNA) were produced by *in vitro* transcription (Roche) and injected into the gut of the planarians as described in Sánchez-Alvarado and Newmark [49]. Three aliquots of 32 nl (400-800 ng/µl) were injected on three consecutive days with a Drummond Scientific Nanoject injector (Broomall, PA). On the fourth or fifth day, some of the planarians were amputated while the rest were left intact. Control organisms were injected with water.

## Additional material

**Additional file 1: Details on Material and Methods.** An extended description of the proteomics protocols applied to perform the analyses presented on this paper.

**Additional file 2: Image scans of all silver-stained 2D gel replicates.**

Image scans of different and independent silver-stained 2D gels used in the study. A to D and the respective zooms, for the regions delimited by red squares, I to L, come from 100 µg of loaded samples. E to H and the respective zooms M to P correspond to 500 µg loaded samples. A, C, E and G are control samples. B, D, F and H are irradiated samples. Although the staining and running conditions were not exactly equivalent, one can observe that the spot pattern shown by all the gels is repetitive, which is more evident on the zoomed regions.

**Additional file 3: Comparing the results presented in this manuscript with previously published studies relating to stem cells.**

Comparison of candidate neoblast protein sequences presented in this paper with genes reported in other proteomic studies to be related to stem cells [25-30] and with specific neoblast-related genes identified in two different high-throughput approaches [32,33]. From the URFs database, only those sequences with a positive decoy were selected. NCBI BLASTP [31] (min e-value = 0.001) was used on sequence comparison. Sequences were clustered according to their homology and they are listed in the table by their original GI identifier from the corresponding NCBI database.

**Additional file 4: Table of peptide candidates.** Listing of the sequence candidates obtained from the computational analysis of the raw proteomics data over the RefSeq and URF datasets (see the corresponding sheet on the spreadsheet file). Only those with a significant BLAST hit are shown (using BLASTP against NCBI-nr, min e-value = 0.001, min hsp length = 25). Genes described in detail in Table 3 are not included. The sequences in this table were built from sets of URFs derived from traces; we provide the corresponding trace identifiers from Genbank TraceDB [64].

## Abbreviations

EST: expressed sequence tags; MS: mass spectrometry; CB: chromatoid bodies; 2D gel: two-dimensional gel; DIGE: difference in gel electrophoresis; cm: centimeters; Ip: Isoelectric point; MW: Molecular weight; WT: wild type; IA: irradiated animals; H3P: phosphorylated histone H3; ORF: open reading frame; URF: unique ORF; NCBI-nr: NCBI non-redundant (database); WUSTL: Washington University in Saint Louis; hsp: high-scoring segment pair (BLAST); GO: Gene Ontology; EC: Enzyme Code (KEGG); ES: embryonic stem cells; HSP/Hsp: heat shock protein; kDa: kilodalton; RNAi: RNA interference; CNS: central nervous system; Gy: grays; dsRNA: double-stranded RNA.

## Acknowledgements

Genomic sequence data was produced by the Washington University Genome Sequencing Center in St. Louis, although trace sequences to generate the URFs database were downloaded from NCBI Trace server. We would like to thank Dr. Roger Florensa for his help in the protein sample preparation and setting up the 2D-gel running conditions, and Dr. Eliandra Oliveira and all members of the proteomic facility at the Parc Científic de

Barcelona for their help in the proteomic work and analyses. We thank all members of the Saló group for advice and critical reading of the manuscript and Dr. Iain Patten for editorial advice. We are also grateful to the reviewers of the earlier version of the manuscript for their helpful comments. This work was supported by grants BFU-2005-00422 and BFU2008-01544/BMC from the Ministerio de Educación y Ciencia, Spain, and grant 2009SGR1018 from AGAUR (Generalitat de Catalunya, Spain). JFA started this project as a Juan de la Cierva post-doctoral fellow. E.F.T. and G.R.E. received an FPI fellowship from the Ministerio de Ciencia y Cultura.

#### Authors' contributions

EFT, ES and JFA conceived of the study. EFT ran the 2D gels and counted the spots. JFA performed the computational analyses, compiled the sequence databases, processed the MASCOT results, and ran the GO functional and KEGG annotation. EFT ran the MASCOT searches and produced the initial BLAST annotation for RefSeq candidates. EFT and GRE performed the experimental validation of the selected protein candidates. All authors participated in its design and coordination, helped to draft the manuscript, and read and approved the final manuscript.

Received: 8 November 2010 Accepted: 28 February 2011  
Published: 28 February 2011

#### References

- Beyer A, Hollunder J, Nasheuer HP, Wilhelm T: **Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale.** *Mol Cell Proteomics* 2004, **3**(11):1083-1092.
- Pandey A, Mann M: **Proteomics to study genes and genomes.** *Nature* 2000, **405**(6788):837-846.
- Hanash S: **Disease proteomics.** *Nature* 2003, **422**(6928):226-232.
- Khan SM, Franke-Fayard B, Mair GR, Lasonder E, Janse CJ, Mann M, Waters AP: **Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology.** *Cell* 2005, **121**(5):675-687.
- Handberg-Thorsager M, Fernández-Taboada E, Saló E: **Stem cells and regeneration in planarians.** *Front Biosci* 2008, **13**:6374-6394.
- Saló E: **The power of regeneration and the stem-cell kingdom: freshwater planarians (Platyhelminthes).** *Bioessays* 2006, **28**(5):546-559.
- Sánchez-Alvarado A, Newmark PA, Robb SM, Juste R: **The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration.** *Development* 2002, **129**(24):5659-5665.
- Baguña J, Saló E, Auladell C: **Regeneration and pattern formation in planarians III. Evidence that neoblasts are totipotent stem cells and the source of blastema cells.** *Development* 1989, **107**:77-86.
- Newmark PA, Sánchez-Alvarado A: **Bromodeoxyuridine specifically labels the regenerative stem cells of planarians.** *Dev Biol* 2000, **220**(2):142-153.
- Coward SJ: **Chromatoid bodies in somatic cells of the planarian: observations on their behavior during mitosis.** *Anat Rec* 1974, **180**(3):533-545.
- Gremigni V: **Planarian regeneration: An overview of some cellular mechanisms.** *Zool Sci* 1988, **5**:1153-1163.
- Higuchi S, Hayashi T, Hori I, Shibata N, Sakamoto H, Agata K: **Characterization and categorization of fluorescence activated cell sorted planarian stem cells by ultrastructural analysis.** *Dev Growth Differ* 2007, **49**(7):571-581.
- Wolff E, Dubois F: **Sur la migration des cellules de régénération chez les planaires.** *Rev Suisse Zool* 1948, **55**:218-227.
- Bayascas JR, Castillo E, Muñoz-Mármol AM, Saló E: **Planarian Hox genes: novel patterns of expression during regeneration.** *Development* 1997, **124**(1):141-148.
- Collet J, Baguña J: **Optimizing a method of protein extraction for two-dimensional electrophoretic separation of proteins from planarians (Platyhelminthes, Turbellaria).** *Electrophoresis* 1993, **14**(10):1054-1059.
- Garbis S, Lubec G, Fountoulakis M: **Limitations of current proteomics technologies.** *Journal of Chromatography A* 2005, **1077**(1):1-18.
- MASCOT search engine to identify proteins from primary sequence databases using mass spectrometry data. [http://www.matrixscience.com/].
- Sayers EW, Barrett T, Benson DA, Bryant SH, Canese K, Chetvermin V, Church DM, DiCuccio M, Edgar R, Federhen S, *et al*: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2009, **37** Database: D5-15.
- Pruitt DKim, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2007, **35** Database: D61-D65.
- Zayas RM, Hernandez A, Habermann B, Wang Y, Stary JM, Newmark PA: **The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: analysis of ESTs from the hermaphroditic strain.** *Proc Natl Acad Sci USA* 2005, **102**(51):18491-18496.
- Robb SM, Ross E, Sánchez-Alvarado A: **SmedGD: the *Schmidtea mediterranea* genome database.** *Nucleic Acids Res* 2008, **36** Database: D599-606.
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674-3676.
- Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Res* 2008, **36**(10):3420-3435.
- Planarian neoblast proteomics online supplementary data. [http://compgen.bio.ub.es/tiki-index.php?page=Planarian+Proteomics].
- Baharvand H, Fathi A, Gourabi H, Mollamohammadi S, Salekdeh GH: **Identification of mouse embryonic stem cell-associated proteins.** *J Proteome Res* 2008, **7**(1):412-423.
- Hoffrogge R, Mikkat S, Scharf C, Beyer S, Christoph H, Pahnke J, Mix E, Berth M, Uhrmacher A, Zubrzycki IZ, *et al*: **2-DE proteome analysis of a proliferating and differentiating human neuronal stem cell line (ReNcell VM).** *Proteomics* 2006, **6**(6):1833-1847.
- Maurer MH, Feldmann RE Jr, Futterer CD, Butlin J, Kuschinsky W: **Comprehensive proteome expression profiling of undifferentiated versus differentiated neural stem cells from adult rat hippocampus.** *Neurochem Res* 2004, **29**(6):1129-1144.
- Kohler C, Wolff S, Albrecht D, Fuchs S, Becher D, Buttner K, Engelmann S, Hecker M: **Proteome analyses of *Staphylococcus aureus* in growing and non-growing cells: a physiological approach.** *Int J Med Microbiol* 2005, **295**(8):547-565.
- Nagano K, Taoka M, Yamauchi Y, Itagaki C, Shinkawa T, Nunomura K, Okamura N, Takahashi N, Izumi T, Isobe T: **Large-scale identification of proteins expressed in mouse embryonic stem cells.** *Proteomics* 2005, **5**(5):1346-1361.
- Zenzmaier C, Kollroser M, Gesslbauer B, Jandrositz A, Preisegger KH, Kungl AJ: **Preliminary 2-D chromatographic investigation of the human stem cell proteome.** *Biochem Biophys Res Commun* 2003, **310**(2):483-490.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
- Reddien PW, Bermange AL, Murfitt KJ, Jennings JR, Sánchez-Alvarado A: **Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria.** *Dev Cell* 2005, **8**(5):635-649.
- Rossi L, Salvetti A, Marincola FM, Lena A, Deri P, Mannini L, Batistoni R, Wang E, Gremigni V: **Deciphering the molecular machinery of stem cells: a look at the neoblast gene expression profile.** *Genome Biol* 2007, **8**(4):R62.
- Stenmark H, Olkkonen VM: **The Rab GTPase family.** *Genome Biol* 2001, **2**(5):REVIEWS3007.
- Segev N: **Ypt and Rab GTPases: insight into functions through novel interactions.** *Curr Opin Cell Biol* 2001, **13**(4):500-511.
- Cheng H, Ma Y, Ni X, Jiang M, Guo L, Ying K, Xie Y, Mao Y: **Isolation and characterization of a human novel RAB (RAB39B) gene.** *Cytogenet Genome Res* 2002, **97**(1-2):72-75.
- Aznar S, Lacal JC: **Rho signals to cell growth and apoptosis.** *Cancer Lett* 2001, **165**(1):1-10.
- Raftopoulos M, Hall A: **Cell migration: Rho GTPases lead the way.** *Dev Biol* 2004, **265**(1):23-32.
- Beere HM, Green DR: **Stress management-heat shock protein-70 and the regulation of apoptosis.** *Trends Cell Biol* 2001, **11**(1):6-10.
- Hartl FU, Hayer-Hartl M: **Molecular chaperones in the cytosol: from nascent chain to folded protein.** *Science* 2002, **295**(5561):1852-1858.
- Rutherford SL, Lindquist S: **Hsp90 as a capacitor for morphological evolution.** *Nature* 1998, **396**(6709):336-342.

42. Pearson BJ, Doe CQ: **Regulation of neuroblast competence in *Drosophila***. *Nature* 2003, **425**(6958):624-628.
43. Abdel-Raheem IT, Hide I, Yanase Y, Shigemoto-Mogami Y, Sakai N, Shirai Y, Saito N, Hamada FM, El-Mahdy NA, Elsisy Ael D, et al: **Protein kinase C- $\alpha$  mediates TNF release process in RBL-2H3 mast cells**. *Br J Pharmacol* 2005, **145**(4):415-423.
44. Nakajima T: **Signaling cascades in radiation-induced apoptosis: roles of protein kinase C in the apoptosis regulation**. *Med Sci Monit* 2006, **12**(10):RA220-224.
45. Beggs JD: **Lsm proteins and RNA processing**. *Biochem Soc Trans* 2005, **33**(Pt 3):433-438.
46. He W, Parker R: **Functions of Lsm proteins in mRNA degradation and splicing**. *Curr Opin Cell Biol* 2000, **12**(3):346-350.
47. Tharun S, He W, Mayes AE, Lennertz P, Beggs JD, Parker R: **Yeast Sm-like proteins function in mRNA decapping and decay**. *Nature* 2000, **404**(6777):515-518.
48. Pineda D, Gonzalez J, Callaerts P, Ikeo K, Gehring WJ, Saló E: **Searching for the prototypic eye genetic network sine oculis is essential for eye regeneration in planarians**. *Proc Natl Acad Sci USA* 2000, **97**(9):4525-4529.
49. Sánchez Alvarado A, Newmark PA: **Double-stranded RNA specifically disrupts gene expression during planarian regeneration**. *Proc Natl Acad Sci USA* 1999, **96**(9):5049-5054.
50. Salvetti A, Rossi L, Deri P, Batistoni R: **An MCM2-related gene is expressed in proliferating cells of intact and regenerating planarians**. *Developmental Dynamics* 2000, **218**(4):603-614.
51. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank**. *Nucleic Acids Res* 2008, **36** Database: D25-30.
52. The UniProt Consortium: **The universal protein resource (UniProt)**. *Nucleic Acids Res* 2008, **36** Database: D190-195.
53. Pignatelli M, Aparicio G, Blanquer I, Hernandez V, Moya A, Tamames J: **Metagenomics reveals our incomplete knowledge of global diversity**. *Bioinformatics* 2008, **1524**(18):2124-5.
54. Yoosuf S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W, et al: **The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families**. *PLoS Biol* 2007, **5**(3):e16.
55. International Human Genome Sequencing Consortium: **Finishing the euchromatic sequence of the human genome**. *Nature* 2004, **431**(7011):931-945.
56. Kalita MK, Ramasamy G, Duraisamy S, Chauhan VS, Gupta D: **ProtRepeatsDB: a database of amino acid repeats in genomes**. *BMC Bioinformatics* 2006, **7**:336.
57. Galindo MI, Pueyo JJ, Fouix S, Bishop SA, Couso JP: **Peptides encoded by short ORFs control development and define a new eukaryotic gene family**. *PLoS Biol* 2007, **5**(5):e106.
58. Fernández-Taboada E, Moritz S, Stehling M, Zeuschner D, HR S, Saló E, Gentile L: **Smed-5mB, a member of the (L)Sm protein superfamily, is essential for chromatoid body organization and planarian stem cell proliferation**. *Development* 2010, **137**(9):1583.
59. Conte M, Deri P, Isolani ME, Mannini L, Batistoni R: **A mortalin-like gene is crucial for planarian stem cell viability**. *Dev Biol* 2009, **334**(1):109-118.
60. Conte M, Isolani ME, Deri P, Mannini L, Batistoni R: **Expression of hsp90 mediates cytoprotective effects in the gastrodermis of planarians**. *Cell Stress Chaperones* 2011, **16**(1):33-39.
61. Sánchez Navarro B, Michiels N, Köhler H-R, D'Souza T: **Differential expression of heat shock protein 70 in relation to stress type in the flatworm *Schmidtea polychroa***. *Hydrobiologia* 2009, **636**:393-400.
62. ***Schmidtea mediterranea* genome sequencing project**. [http://genome.wustl.edu/genomes/view/schmidtea\_mediterranea/].
63. Belleville S, Beauchemin M, Tremblay M, Noiseux N, Savard P: **Homeobox-containing genes in the newt are organized in clusters similar to other vertebrates**. *Gene* 1992, **114**:179-186.
64. ***Schmidtea mediterranea* trace archive at NCBI**. [ftp://ftp.ncbi.nih.gov/pub/TraceDB/schmidtea\_mediterranea/].
65. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS: **Probability-based protein identification by searching sequence databases using mass spectrometry data**. *Electrophoresis* 1999, **20**(18):3551-3567.
66. Cargile BJ, Talley DL, Stephenson JL Jr: **Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pl predictability of peptides**. *Electrophoresis* 2004, **25**(6):936-945.
67. Resing KA, Meyer-Arendt K, Mendoza AM, Aveline-Wolf LD, Jonscher KR, Pierce KG, Old WM, Cheung HT, Russell S, Wattawa JL, et al: **Improving reproducibility and sensitivity in identifying human proteins by shotgun proteomics**. *Anal Chem* 2004, **76**(13):3556-3568.
68. Elias JE, Gygi SP: **Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry**. *Nat Methods* 2007, **4**(3):207-214.
69. Higdon R, Hogan JM, Kolker N, van Belle G, Kolker E: **Experiment-specific estimation of peptide identification probabilities using a randomized database**. *Omics* 2007, **11**(4):351-365.
70. Higdon R, Hogan JM, Van Belle G, Kolker E: **Randomized sequence databases for tandem mass spectrometry peptide and protein identification**. *Omics* 2005, **9**(4):364-379.
71. Saló E, Baguñà J: **Cell movement in intact and regenerating planarians. Quantitation using chromosomal, nuclear and cytoplasmic markers**. *J Embryol Exp Morphol* 1985, **89**:57-70.
72. Barsnes H, Vizcaino JA, Eidhammer I, Martens L: **PRIDE Converter: making proteomics data-sharing easy**. *Nat Biotechnol* 2009, **27**(7):598-599.
73. Vizcaino JA, Cote R, Reisinger F, Foster JM, Mueller M, Rameseder J, Hermjakob H, Martens L: **A guide to the Proteomics Identifications Database proteomics data repository**. *Proteomics* 2009, **9**(18):4276-4283.
74. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene ontology: tool for the unification of biology**. The Gene Ontology Consortium. *Nat Genet* 2000, **25**(1):25-29.
75. Agata K, Soejima Y, Kato K, Kobayashi C, Umehono Y, Watanabe K: **Structure of the planarian central nervous system (CNS) revealed by neuronal cell markers**. *Zool Sci* 1998, **15**:433-440.
76. Nogi T, Levin M: **Characterization of innexin gene expression and functional roles of gap-junctional communication in planarian regeneration**. *Dev Biol* 2005, **287**(2):314-335.

doi:10.1186/1471-2164-12-133

**Cite this article as:** Fernández-Taboada et al.: **A proteomics approach to decipher the molecular nature of planarian stem cells**. *BMC Genomics* 2011 **12**:133.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at  
www.biomedcentral.com/submit







## Artículo 3

### ***Smed454 dataset: unravelling the transcriptome of Schmidtea mediterranea***

Josep F. Abril\*, Francesc Cebrià\*, **Gustavo Rodríguez-Esteban**, Thomas Horn, Susanna Fraguas, Beatriz Calvo, Kerstin Bartscherer and Emili Saló.

**BMC Genomics** (2010) 11:731

\*Ambos autores han contribuido por igual.



## Resumen

**Conjunto de datos Smed454: revelando el transcriptoma de *Schmidtea mediterranea***

A pesar de la disponibilidad de la secuencia del genoma de *Schmidtea mediterranea* una descripción detallada de su transcriptoma es esencial para la investigación en este organismo modelo. En este artículo se presenta el primer transcriptoma de planaria, Smed454, como un recurso al que puede accederse libremente a través de su interficie web.

El método empleado para la secuenciación fue el de pirosecuenciación 454. Se obtuvieron 598.435 *reads*, con una longitud media de 327pb, que fueron ensambladas junto con las 10.000 secuencias de *S. mediterranea* de la base de datos UniGene utilizando diferentes límites de similitud. A continuación, el ensamblado fue proyectado sobre el genoma. Este nuevo transcriptoma contiene más de tres millones de nuevos nucleótidos secuenciados por primera vez. Asimismo, se llevó a cabo un análisis de los sitios de *splicing* en aquellos contigs que mapaban de manera unívoca sobre el genoma. El análisis de las secuencias nos permitió identificar genes que codifican proteínas putativas con propiedades estructurales definidas, como dominios transmembrana. Además, se anotó el transcriptoma mediante *Gene Ontology* y se identificaron homólogos putativos de muchas familias génicas que pueden tener un papel importante durante la regeneración, como neurotransmisores y receptores hormonales, genes homeobox y genes relacionados con la función de los fotorreceptores.



RESEARCH ARTICLE

Open Access

# Smed454 dataset: unravelling the transcriptome of *Schmidtea mediterranea*

Josep F Abril<sup>1,2†</sup>, Francesc Cebrià<sup>1,2†</sup>, Gustavo Rodríguez-Esteban<sup>1,2</sup>, Thomas Horn<sup>3</sup>, Susanna Fraguas<sup>1,2</sup>, Beatriz Calvo<sup>1,2</sup>, Kerstin Bartscherer<sup>4</sup>, Emili Saló<sup>1,2\*</sup>

## Abstract

**Background:** Freshwater planarians are an attractive model for regeneration and stem cell research and have become a promising tool in the field of regenerative medicine. With the availability of a sequenced planarian genome, the recent application of modern genetic and high-throughput tools has resulted in revitalized interest in these animals, long known for their amazing regenerative capabilities, which enable them to regrow even a new head after decapitation. However, a detailed description of the planarian transcriptome is essential for future investigation into regenerative processes using planarians as a model system.

**Results:** In order to complement and improve existing gene annotations, we used a 454 pyrosequencing approach to analyze the transcriptome of the planarian species *Schmidtea mediterranea*. Altogether, 598,435 454-sequencing reads, with an average length of 327 bp, were assembled together with the ~10,000 sequences of the *S. mediterranea* UniGene set using different similarity cutoffs. The assembly was then mapped onto the current genome data. Remarkably, our Smed454 dataset contains more than 3 million novel transcribed nucleotides sequenced for the first time. A descriptive analysis of planarian splice sites was conducted on those Smed454 contigs that mapped univocally to the current genome assembly. Sequence analysis allowed us to identify genes encoding putative proteins with defined structural properties, such as transmembrane domains. Moreover, we annotated the Smed454 dataset using Gene Ontology, and identified putative homologues of several gene families that may play a key role during regeneration, such as neurotransmitter and hormone receptors, homeobox-containing genes, and genes related to eye function.

**Conclusions:** We report the first planarian transcript dataset, Smed454, as an open resource tool that can be accessed via a web interface. Smed454 contains significant novel sequence information about most expressed genes of *S. mediterranea*. Analysis of the annotated data promises to contribute to identification of gene families poorly characterized at a functional level. The Smed454 transcriptome data will assist in the molecular characterization of *S. mediterranea* as a model organism, which will be useful to a broad scientific community.

## Background

One of the challenges that medical research must address in the near future is to understand why some animals are able to regenerate complex structures, including eyes and even whole bodies, from small body fragments, while others are not. With the recent emergence of the field of regenerative medicine, the future

biomedical ramifications of the study of animal regeneration are obvious.

Freshwater planarians are a classic model for studying the fascinating process of regeneration [1-4] because they are capable of re-building a complete organism from almost any small body fragment. This is made possible by a unique population of adult somatic stem cells called neoblasts. During regeneration and constant homeostatic cell turnover, neoblasts differentiate into all cell types, including germ cells in sexual species [5,6]. In recent years, several studies have begun to unravel the mechanisms by which regeneration is regulated at the molecular level. For example, different genes have been

\* Correspondence: esalo@ub.edu

† Contributed equally

<sup>1</sup>Departament de Genètica, Facultat de Biologia, Universitat de Barcelona (UB), Av. Diagonal 645, edifici annex, planta 1, 08028, Barcelona, Catalunya, Spain

Full list of author information is available at the end of the article

shown to play pivotal roles in axon guidance and neurogenesis [7], the regulation of neoblast proliferation and differentiation [8,9], and the re-establishment and maintenance of the anteroposterior (AP) and dorsoventral (DV) body axes [10]. *Schmidtea mediterranea* and *Dugesia japonica* are the two planarian species most often used in regeneration studies. There are about 78,000 ESTs (Expressed Sequence Tags) for *S. mediterranea* in NCBI generated in different projects [11,12]. Those sequences were clustered to produce a set of 10,000 putative mRNAs which are available from the NCBI Unigene database [13]. The *S. mediterranea* genome has also been sequenced and assembled [14] at the Genome Sequencing Center at Washington University in St. Louis (WUSL, USA) after approval of a white paper [15]. However, because of this genome's internal complexity (67% A+T, [16]) and the lack of a BAC library, its completeness and assembly still needs improvement. A step towards this end was taken when the *S. mediterranea* genome and EST information were integrated and approximately 30,000 genes were predicted using an annotation pipeline called MAKER [16]. Those gene models, together with ~9,000 mRNAs generated using next-generation sequencing technology, were mapped on the planarian genome and used to improve the assembly [17]. The current assembly contains 43,673 contigs. These are accessible, together with the MAKER annotation data, in the *S. mediterranea* genome database (SmedGD; [17]).

In order to expand our knowledge of the planarian transcriptome and to provide a new tool that can be used to improve the *S. mediterranea* genome annotation, we generated a new transcriptome dataset using 454 pyrosequencing technology [18]. The Smed454 dataset can be freely accessed via a website, and the complete sequence data can be downloaded by anyone from there. Mapping of the Smed454 ESTs onto the genome scaffolds shows that the Smed454 dataset contains more than 3 million nucleotides sequenced *de novo*. In addition, this mapping extends and connects currently fragmented genomic contigs. Finally, GO annotation of the Smed454 dataset assigns candidate functions to those sequences and facilitates their grouping into distinct gene families. In this way, whole gene families can be analyzed for putative roles in planarian regeneration. Thus we are confident that the Smed454 dataset will improve our understanding of how planarian regeneration works at the molecular level.

## Results and Discussion

### Construction and sequencing of the Smed454 dataset

In order to obtain the most representative set of planarian genes expressed under different physiological conditions, total RNA was isolated from a mixture of non-irradiated

and irradiated intact and regenerating planarians (see Methods). We used planarians regenerating both head and tail to identify the genes specifically expressed in a tissue-specific manner. Similarly, planarians at different stages of regeneration were used in order to isolate genes with different temporal expression profiles. Irradiation destroys planarian neoblasts within 1-2 days, and the animals die within a few weeks because they cannot sustain normal cell turnover. By including irradiated animals, potential transcripts specifically expressed under those conditions will be contained in the 454 dataset.

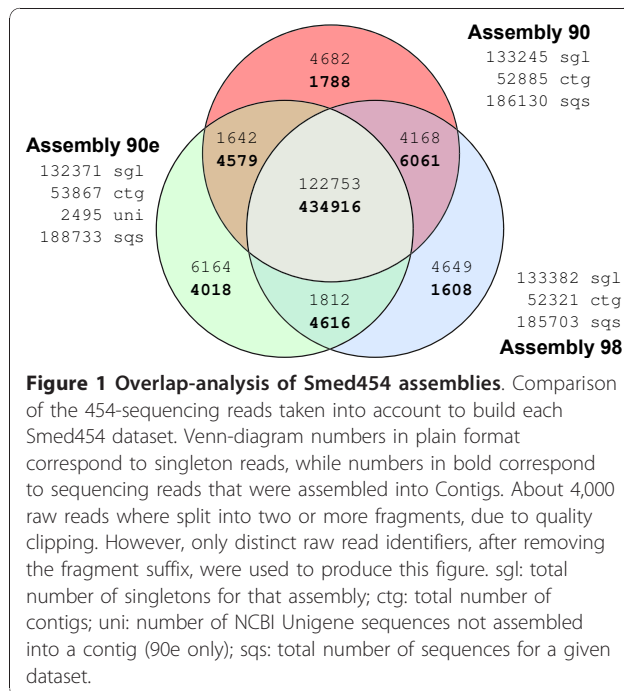
Using 454 pyrosequencing, 601,439 sequencing reads with an average length of 327 bp were obtained. After sequence cleaning to remove vector contamination, the remaining 598,435 sequences were assembled using different cut-off values for sequence similarity (90%, 95% and 98%). In addition, our 454 sequence reads were assembled together with the ~10,000 *S. mediterranea* UniGene set available at NCBI, using the 90% similarity criteria. This last set, which was used in most of the analyses reported, is referred to as the **90e** set. Table 1 summarizes the number of contigs and singletons obtained in each of those assemblies. The similarities between the three assemblies (**90**, **98** and **90e**) are illustrated in Figure 1 a Venn diagram which shows that 72.68% of the raw sequencing reads were integrated into contigs common to all three assemblies, and 20.51% of the sequencing reads make up a shared pool of single sequencing reads (singletons). Therefore, differences between the assemblies can be explained by differential inclusion corresponding to 6.81% of the sequencing reads.

Average GC content and sequence length and their respective distributions were similar for all three assemblies (Table 1 and Figure 2). GC content is distributed around 35%, the expected value for coding sequences in this species. The **90e** length distribution shape was slightly shifted towards larger sequences. This shift was mainly due to a set of long sequences (>800 bp) from

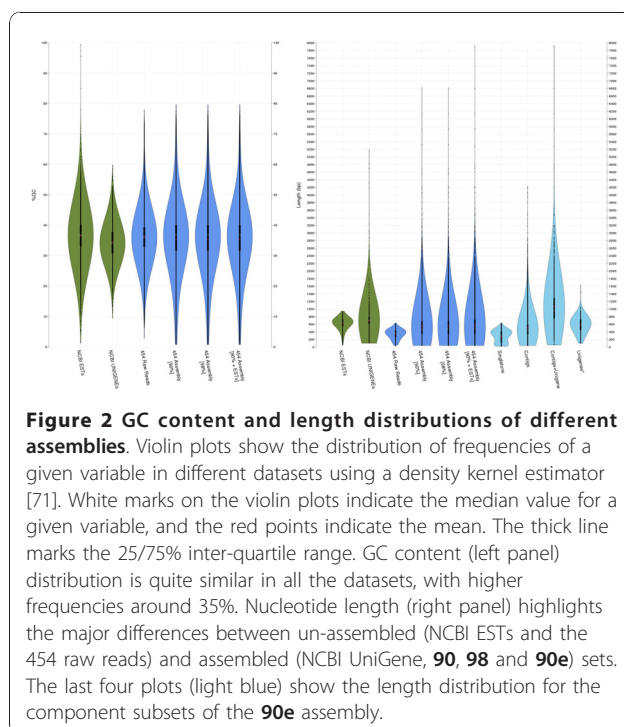
**Table 1 Summary of sequence statistics for each assembly.**

SET	Contigs	Singletons	TOTAL SEQs	GC%	LENGTHs [min/median/max/avg]			
<b>90</b>	52,885	137,213	190,098	35.130	20	354	6812	355.78
<b>95</b>	52,501	137,077	189,578	—	—	—	—	—
<b>98</b>	52,321	137,353	189,674	35.127	20	354	6812	355.82
<b>90e</b>	53,867	138,766	192,633	35.108	20	358	7918	364.81

Set names are related to the corresponding homology level cutoff value (90e stands for 90% similarity including the set of NCBI Unigene ESTs). Contigs are the result of at least two sequencing reads, and singletons of only one read. GC content is the average value for all sequences. Sequence lengths are shown as minimum, median, maximum and average values in nucleotides for each set.



the NCBI Unigene ESTs included in this assembly. This causal relationship was evident in the comparison of the following four subsets of sequences from the 90e set (lightblue violin plots on Figure 2 right panel): singletons (136,271), contigs that do not contain UniGene ESTs (46,958), contigs including Unigene ESTs (6,909),



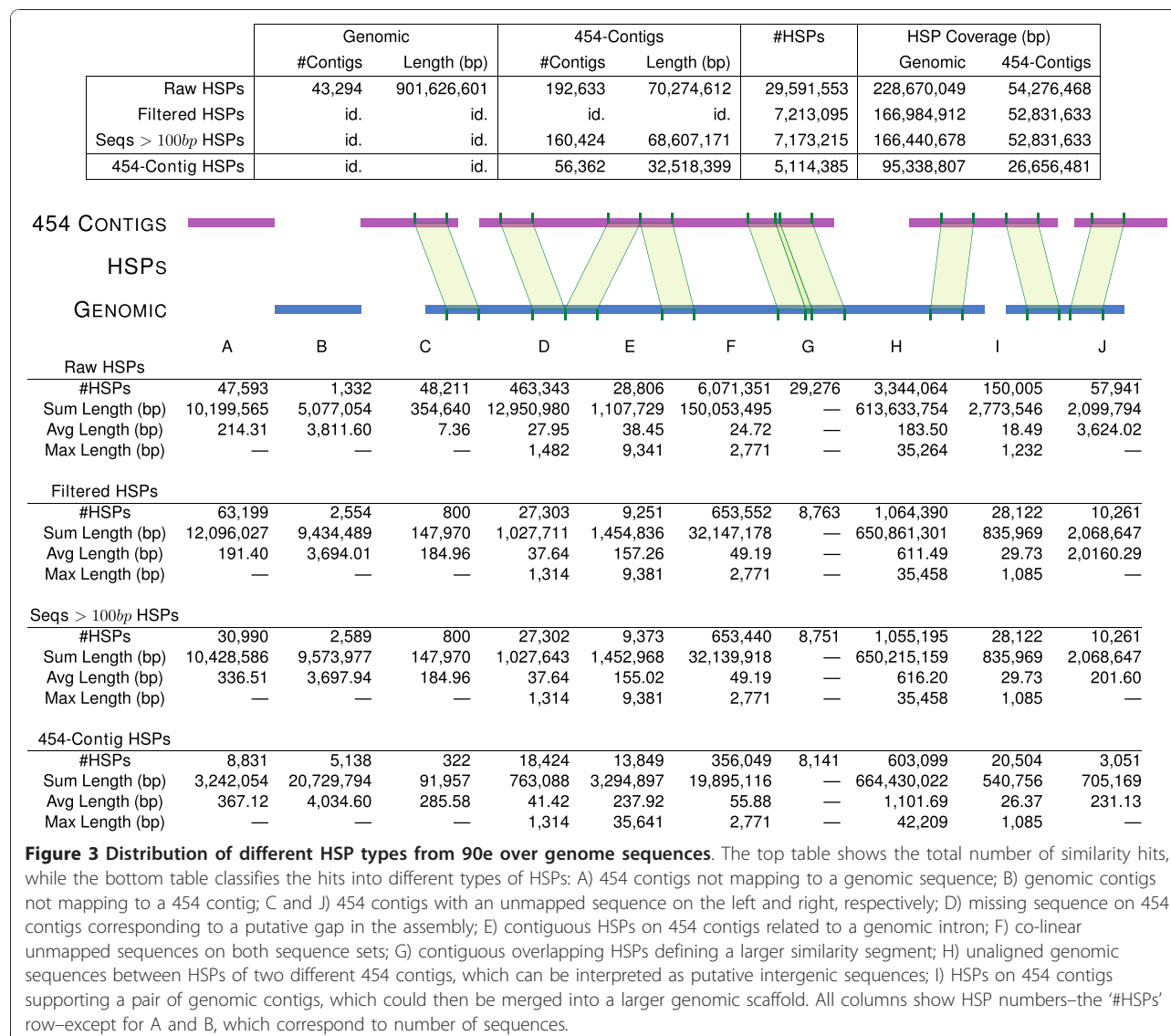
and finally, Unigene ESTs not assembled into a contig (2,495).

### Mapping the 90e assembly onto the genome

The 90e assembly (192,633 sequences, 70,274,612 bases in total, average length of 365 bp per sequence) was aligned to scaffolds from the *S. mediterranea* WUSL genome assembly, version 3.1 [14] (43,294 sequences, 901,626,601 bases in total, average length of 20,8 kilobases per scaffold). Figure 3 shows all possible high-scoring segment pair (HSP) relationships between those two sequence sets. From almost 30 million initial HSPs, around 7 million were selected using a combination of thresholds, as described in the Methods section. Discarding singleton sequences in a second round of filtering further reduced the number of HSPs to 5 million, and HSP coverage dropped from 25.36% and 77.24%, for scaffolds and 90e respectively, to 10.57% and 37.93%. However, when the total nucleotide length was considered only for the contigs (56,363 sequences, 32,518,399 bases in total, with an average of 577 bp per sequence), HSP coverage for 90e rose to 81.97%. This means that most of the significant HSP hits are retained after the second round of filtering. In total, 8,831 contigs from 90e did not map to the genomic contigs (3,242,054 bp that are completely novel and also transcribed, see column A in Figure 3). Conversely, 5,138 genomic contigs did not match a sequence from 90e (column B). Of the 90e contigs, 322 extended a genomic sequence from the left (column C) and 3,051 from the right (column J). The largest intergenic distance was 42,209 bp, with an average value of 1,102 bp (column H). The largest intron was estimated to be about 9,300 bp, the average length being 238 bp (column E). Finally, there were 20,504 HSPs connecting different genomic sequences via 8,604 different 90e contigs (column I). Of the 8,831 90e contigs not found on the genome, 3,480 had a BLAST hit to the NCBI NR protein database (39.41%), and, of those, 2,401 had a hit to a protein with GO annotation (27.19%). After discarding abundant actin-like sequences (1,503), ATP/ADP transporter proteins (722) and sequences matching bacterial, protozoan or fungal genes (1,234), 71 90e contigs remained as new sequences not mapping on the genome (see Additional File 1).

In order to validate exonic structures, 6,226 90e contigs mapping 1-to-1 over genome sequences were selected. After re-aligning the 90e/genomic sequence pairs, 4,739 contained at least one putative intron (see the corresponding splice sites boundaries in Additional File 2). In total 8,609 introns were retrieved from the genomic contigs. Figure 4 shows the number of introns per 90e contig, as well as the length distribution for those introns. Pictograms summarize the nucleotide frequencies for the donor and acceptor splice sites, both





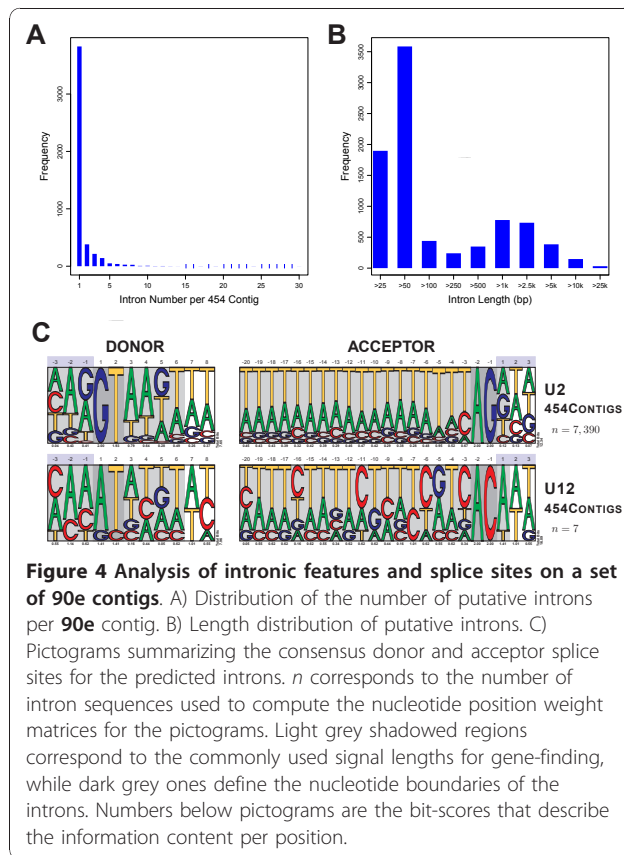
for the U2 (canonical) and U12 (non-cannonical) introns. The splice sites patterns resemble those from other metazoan [19], taking into account that the genome of *S.mediterranea* is A/T-rich [16].

Also, 50 randomly picked 90e contigs that either mapped or did not map to the genome were validated by RT-PCR (see Additional File 3 containing a list of the selected 90e contigs, as well as information on the primers used to amplify them). Additionally, 20 out of those 50 genes were further validated by sequencing. Finally, to further confirm the quality and coverage of the sequences from the 90e dataset, the *S. mediterranea* genes already annotated in NCBI GenBank [20] were compared with those sequences. After discarding 18 S and 28 S ribosomal RNA genes and alpha-tubulins, 124 known genes were aligned to the 90e sequences. In total, 108 of these genes

had at least one significant similarity hit with one 90e sequence, and two matched 5 sequences from 90e. On average, the known genes had co-linear similarity hits against 1.32 different Smed454 sequences. Minimum and average similarities were 8.35% and 85.34% respectively, and 71 sequences had more than 95% similarity. Mean coverage dropped to 77.63% when each hit was considered separately. A summary of these similarity analyses is shown in Additional File 4.

#### Browsing the Smed454 dataset

In order to make the Smed454 dataset useful and accessible to the planarian and non-planarian communities, a public database is available via web [21]. The web site allows users to view contig assemblies along with their read alignments, and to perform BLAST searches against assembled



sequences. The BLAST option in the home page menu (1 in Figure 5) allows the user to BLAST sequences of interest against the 90, 98, and 90e databases (1.2 in Figure 5). Both nucleotide (BLASTN) and protein (BLASTP) searches can be performed (1.1 in Figure 5). Clicking on the Search button (1.3 in Figure 5) brings up a new window displaying a list of hits. When a score value is selected (1.4 in Figure 5), the alignment between the query sequence and the Smed454 hit is shown. The site also offers the option of downloading Smed454 sequences of interest (1.5 in Figure 5). The contig or singleton accession number can be browsed directly from the main home page (2 in Figure 5). When the user searches for a specific contig, a new window appears showing the alignment of all the sequencing reads assembled in that contig. At the bottom of that window, the result of a pre-computed BLAST on the contig consensus sequence is displayed. When a contig, singleton or read name is selected (2.1 in Figure 5), a new window will display the requested sequence. All raw and assembled sequence data are available from that web site too.

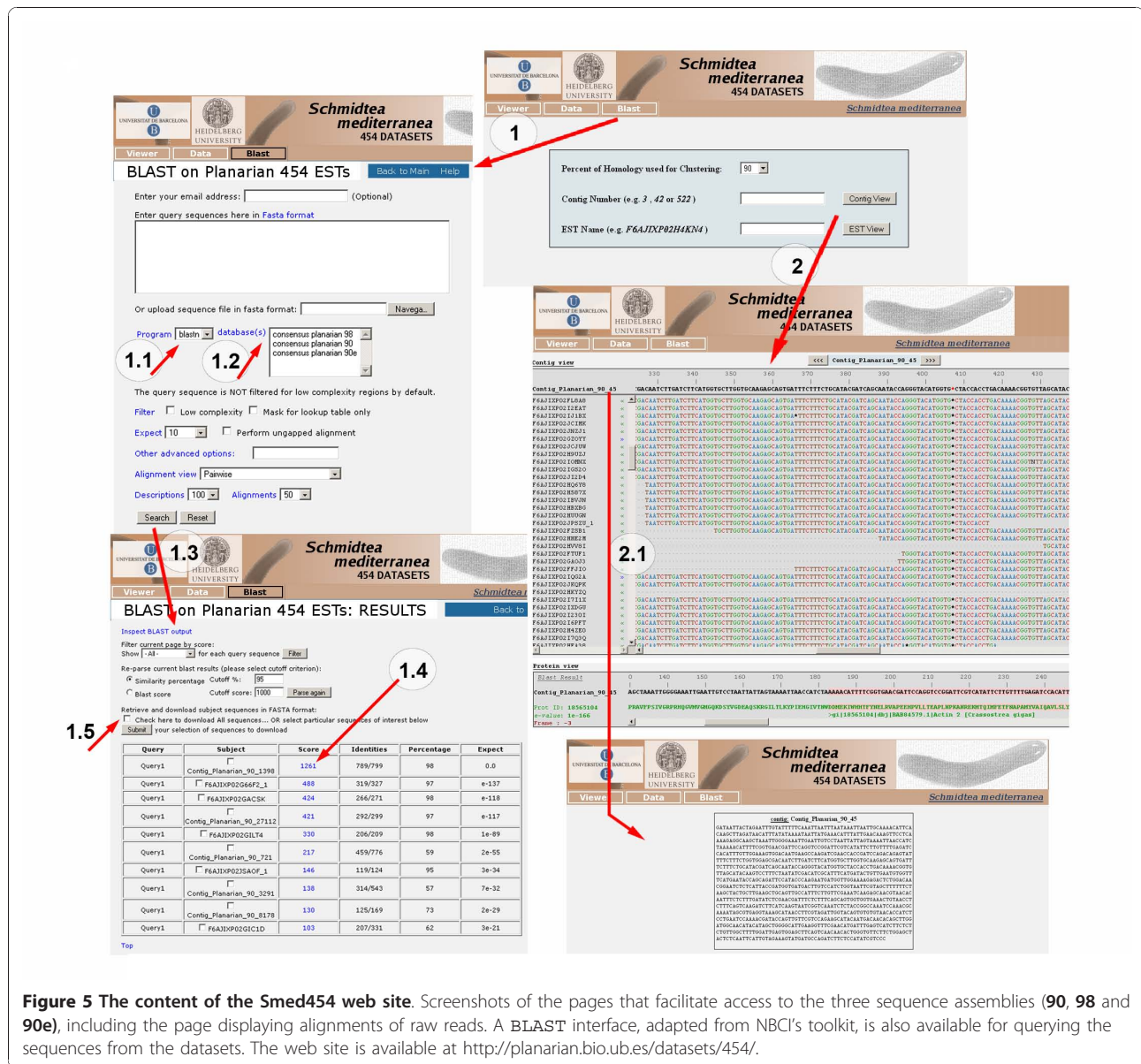
#### Functional annotation of 90e sequences

In order to characterize the gene families that can be found on Smed454, we annotated the three datasets; we will focus on 90e dataset here. In total, 42.42% of the

sequences had a similarity hit with at least one protein sequence in the NCBI NR protein database [20]. Of these, almost two-thirds had 250 or more hits (see Figure 6), but the BLASTX output was limited to a maximum of 250 hits per 90e sequence owing to the large number of HSPs reported by BLAST for some of them. The Gene Ontology (GO) [22] database was used to computationally annotate all the sequences (see Additional File 5 for 90, 98, and 90e datasets) by mapping onto them the functional codes already assigned to known proteins from NCBI NR. Many of these sequence hits matched to a short ATP-binding domain, in most cases corresponding to proteins of the actins family. Consequently, that functional class, which was also anomalously over-represented, was discarded from the total number of annotations for the 90e set, as shown in Table 2.

Among the most abundant GO annotations at the biological process level, leaving aside metabolism-related features, 'response to stress' was found for 1,070 sequences (6.58%). This finding was expected because the original biological sample was a mixture of intact and regenerating planarians, both normal and irradiated. 'Regulation of biological process' was in the same range, with 1,012 sequences (6.22%). At the GO molecular function level, 'binding' was the most common annotation, although where possible a more specific annotation was provided by drilling down to the 2<sup>nd</sup> level child annotations on the GO graph. It is interesting to find, among others, 3 'selenium binding' activities, since it has been reported that selenium may play an important role in cancer prevention, immune system function, male fertility, cardiovascular and muscle disorders, and prevention and control of the ageing process [23]. Finding selenium-binding proteins would be evidence of the presence of selenoproteins, which are thought to be responsible for most of the biomedical effects of selenium across eukaryota [24]. When looking at the cellular component level and discarding many of the 'intracellular organelles' due to their co-occurrence with 'nucleotide binding', there are a notably large number of 'protein complexes', 2,918 sequences (40.79%). With 819 sequences (11.45%), another important term on this level is 'vesicle', which correlates with secretory functions, apoptosis, and autophagy.

To prove the usefulness of the Smed454 dataset, we performed several searches on specific groups and gene families for which only scant data has been reported to date in planarians. Planarians are mainly known for their remarkable regenerative capabilities, which depend upon the presence of stem cells named neoblasts. Because of the unique properties of these cells, some studies have used a microarray-based strategy to detect neoblast-specific genes [25,26]. In our Smed454 dataset we were able to identify, in addition to known neoblast

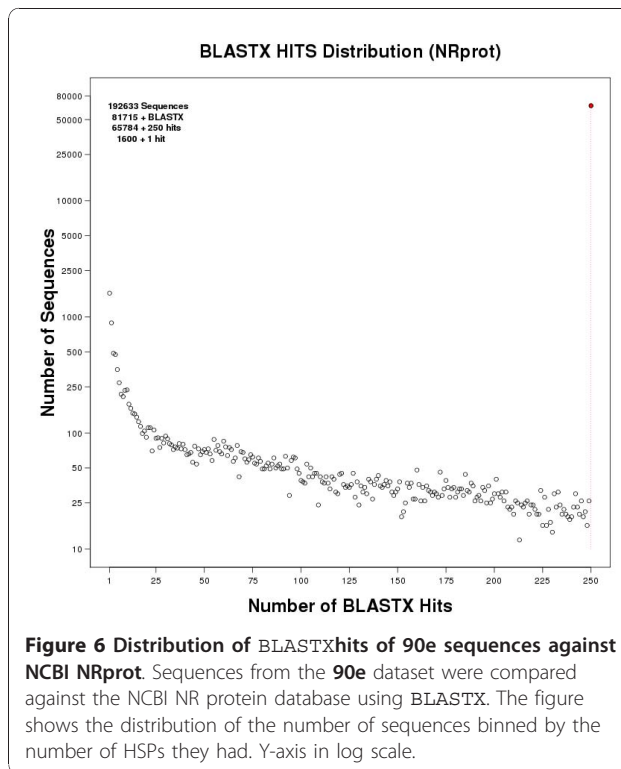


**Figure 5** The content of the Smed454 web site. Screenshots of the pages that facilitate access to the three sequence assemblies (90, 98 and 90e), including the page displaying alignments of raw reads. A BLAST interface, adapted from NBCI's toolkit, is also available for querying the sequences from the datasets. The web site is available at <http://planarian.bio.ub.es/datasets/454/>.

markers such as piwis, histones, bruli, vasa or tudor, several other genes annotated as involved in cell cycle or DNA damage and repair (Additional File 6). Within these gene set we find many cyclins and cell cycle division-related genes but also genes related to replication and chromosome maintenance. Finally, genes related to stress response and DNA damage were also identified, probably owing to the use of irradiated animals in the generation of the Smed454 dataset. In addition to these neoblast-related genes we were able to identify large collections of much less well-characterized families in planarians, such as neurotransmitter, peptide and hormone receptors, homeobox domain-containing genes, and genes related to eye function in other animals.

### Prediction of planarian transmembrane proteins

Transmembrane (TM) proteins regulate a number of biological processes ranging from catalytic processes in intracellular and extracellular transport to cell-to-cell communication. TM proteins have become particularly interesting as many of them are key initiators of signal transduction pathways, and they can be easily manipulated by small molecule- or antibody-based drugs. To identify putative TM proteins from the planarian transcriptome, we mined the 454 dataset for putative TM protein-encoding messages (see Methods). Considering only the proteins that at least two applications predicted would contain one or more transmembrane domains, resulted in a list of 8,597 predicted transmembrane



proteins (see Figure 7a), which represents 15,3% of the complete protein database. Protein-BLAST searches were then used to align sequences to each other, and redundant sequences were removed from the predicted transmembrane set. The resulting database contained 4,663 sequences. Functional categorization using the UFO web-server [27] allowed us to assign PFAM protein families to 1,474 of the sequences and gene ontology classifications to 2,464. The top ten PFAM domains (~33% of all assignments) included, for example, the classifications for 'major facilitator superfamily' (a ubiquitous transporter family), '7 transmembrane receptor (rhodopsin family)' and 'ion transport protein' (see Figure 7b). The top ten gene ontology profiles (~49% of all assignments) included 'membrane' (cellular component), 'transport', and 'G-protein coupled receptor protein signalling pathway' (both biological processes, see Figure 7c). The enrichment of our database with proteins that have a predicted function in transport and receptor signalling supports the reliability of our approach. A complete list of the 4,663 predicted transmembrane proteins, the number of predicted transmembrane domains, predicted topology, and functional categorizations (PFAM and GO) are shown in Additional File 7.

#### Neurotransmitter and hormone receptors in *Schmidtea mediterranea*

Despite our growing knowledge about how planarian neoblasts are regulated at the molecular level [9,25,26,28-31],

we are still far from characterizing the complete repertoire of factors that control neoblast biology. Receptors for neurotransmitters, peptides and hormones are among the candidates for a role in the regulation of neoblast proliferation, differentiation and migration. In planarians, some of the data suggest that molecules such as dopamine [32,33], serotonin [34], substance P [35], somatostatin [36] and FMRFamide [37] can accelerate or delay the regeneration rate, probably by regulating neoblast proliferation and/or differentiation. A model has been proposed in which neoblasts express receptors for some of these factors, which in turn regulate the fate of these cells [35]. We found 288 contigs and singletons in the annotated Smed454 dataset with significant homology to neurotransmitter and hormone receptors (Table 3 and Additional File 8), providing a list of potentially interesting candidates.

#### Homeobox-containing sequences in *Schmidtea mediterranea*

Since the first homeobox-containing genes were characterized in planarians [38], a large number of Hox and ParaHox genes that could be accommodated into the classical series of paralogous groups from Plhox1 to Plo-hox-9 and Xlox to cad/Cdx [39,40] have been described. Some of them show a differentially axial nested expression; while others are ubiquitously expressed [41-43]. Most of this work has been done in the planarians *Girardia tigrina* and *Dugesia japonica*. Recently, the first expression of an *S. mediterranea* Hox gene has been reported [44]. We identified 50 contigs and singletons with significant sequence similarity to homeobox gene sequences in the annotated Smed545 dataset (Table 4), including Hox genes and homeobox-containing genes, some already characterized in other planarian species.

#### Eye genes in *Schmidtea mediterranea*

The structural simplicity of the planarian eye in conjunction with the regenerative abilities of these organisms provides a unique system for dissecting the genetic mechanisms that allow a simple visual structure to be built [45,46]. Despite great morphological differences, there is evidence that the early morphogenesis of animal eyes requires the regulatory activity of *Pax6*, *Sine oculis* (*Six*), *Eyes absent* (*Eya*) and *Dachshund* (*Dach*), a gene network known as the retinal determination gene network (RDGN) [47-50]. Most of the genetic elements of the RDGN have been characterized in planarians [51-54]. In addition, the following planarian genes have been identified as being involved in eye regeneration: *Djeye53*, *Dj1020HH* [55]; *Smed-netR*, *Smed-netrin2* [56]; *Gt/Smed/Dj ops* [46,57]; *Djsnap-25* [58]; and *Smednos* [59]. In order to characterize new *S. mediterranea* eye network genes, we analyzed the Smed454 annotated

**Table 2 Gene Ontology annotation for 90e set sequences.**

GO	Molecular Function	Count	Freq %	GO	Biological Process	Count	Freq %	GO	Cellular Component	Count	Freq %
GO:0000166	nucleotide binding	54,823	—	—	—	—	—	GO:0043229	intracellular organelle	60,817	—
—	unannotated	9,709	—	—	unannotated	62,834	—	—	unannotated	11,131	—
GO:0016787	hydrolase activity	5,197	35.669	GO:0043170	macromolecule metabolic process	5,793	35.610	GO:0043234	protein complex	2,918	40.788
GO:0016740	transferase activity	2,030	13.933	GO:0022607	cellular component assembly	2,182	13.413	GO:0044424	intracellular part	2,314	32.346
GO:0043167	ion binding	1,323	9.080	GO:0006810	transport	1,213	7.456	GO:0031982	vesicle	819	11.448
GO:0003735	structural constituent of ribosome	874	5.999	GO:0006950	response to stress	1,070	6.577	GO:0044425	membrane part	469	6.556
GO:0005488	binding	761	5.223	GO:0050789	regulation of biological process	1,012	6.221	GO:0016020	membrane	210	2.935
GO:0016491	oxidoreductase activity	703	4.825	GO:0006807	nitrogen compound metabolic process	722	4.438	GO:0005622	intracellular	111	1.552
GO:0022857	transmembrane transporter activity	678	4.653	GO:0048869	cellular developmental process	655	4.026	GO:0044446	intracellular organelle part	91	1.272
GO:0030235	nitric-oxide synthase regulator activity	597	4.097	GO:0065009	regulation of molecular function	622	3.823	GO:0005576	extracellular region	63	0.881
GO:0043176	amine binding	580	3.981	GO:0009056	catabolic process	507	3.117	GO:0045211	postsynaptic membrane	21	0.294
GO:0005515	protein binding	532	3.651	GO:0044419	interspecies interaction between organisms	280	1.721	GO:0044420	extracellular matrix part	19	0.266
GO:0003676	nucleic acid binding	401	2.752	GO:0055114	oxidation reduction	236	1.451	GO:0043233	organelle lumen	18	0.252
GO:0005215	transporter activity	387	2.656	GO:0065008	regulation of biological quality	206	1.266	GO:0031012	extracellular matrix	16	0.224
GO:0016829	lyase activity	71	0.487	GO:0048856	anatomical structure development	193	1.186	GO:0042597	periplasmic space	15	0.210
GO:0016853	isomerase activity	55	0.377	GO:0051649	establishment of localization in cell	183	1.125	GO:0000267	cell fraction	15	0.210
GO:0048037	cofactor binding	52	0.357	GO:0044237	cellular metabolic process	182	1.119	GO:0044462	external encapsulating structure part	11	0.154
GO:0016874	ligase activity	49	0.336	GO:0023060	signal transmission	150	0.922	GO:0031975	envelope	8	0.112
GO:0004871	signal transducer activity	45	0.309	GO:0048870	cell motility	141	0.867	GO:0005615	extracellular space	7	0.098
GO:0003824	catalytic activity	32	0.220	GO:0008152	metabolic process	139	0.854	GO:0009986	cell surface	6	0.084
GO:0060589	nucleoside-triphosphatase regulator activity	32	0.220	GO:0023033	signaling pathway	107	0.658	GO:0043204	perikaryon	5	0.070
GO:0042277	peptide binding	28	0.192	GO:0044238	primary metabolic process	83	0.510	GO:0030427	site of polarized growth	4	0.056
GO:0022892	substrate-specific transporter activity	22	0.151	GO:0042221	response to chemical stimulus	69	0.424	GO:0042995	cell projection	3	0.042
GO:0019208	phosphatase regulator activity	12	0.082	GO:0006996	organelle organization	47	0.289	GO:0030312	external encapsulating structure	2	0.028
GO:0003712	transcription cofactor activity	12	0.082	GO:0007017	microtubule-based process	42	0.258	GO:0031594	neuromuscular junction	2	0.028
GO:0019207	kinase regulator activity	11	0.075	GO:0044281	small molecule metabolic process	39	0.240	GO:0045177	apical part of cell	2	0.028
GO:0008289	lipid binding	9	0.062	GO:0051301	cell division	37	0.227	GO:0019028	viral capsid	1	0.014

**Table 2 Gene Ontology annotation for 90e set sequences. (Continued)**

GO:0005201	extracellular matrix structural constituent	8	0.055	GO:0022613	ribonucleoprotein complex biogenesis	35	0.215	GO:0031252	cell leading edge	1	0.014
GO:0050840	extracellular matrix binding	6	0.041	GO:0019637	organophosphate metabolic process	34	0.209	GO:0044217	other organism part	1	0.014
GO:0061134	peptidase regulator activity	6	0.041	GO:0045184	establishment of protein localization	34	0.209	GO:0044297	cell body	1	0.014
GO:0030246	carbohydrate binding	6	0.041	GO:0009628	response to abiotic stimulus	23	0.141	GO:0044463	cell projection part	1	0.014
GO:0016248	channel inhibitor activity	5	0.034	GO:0019748	secondary metabolic process	21	0.129	<b>TOTAL</b>		7,154	
GO:0003702	RNA polymerase II transcription factor activity	5	0.034	GO:0009058	biosynthetic process	21	0.129				
GO:0005198	structural molecule activity	4	0.027	GO:0007155	cell adhesion	18	0.111				
GO:0016986	transcription initiation factor activity	4	0.027	GO:0061024	membrane organization	16	0.098				
GO:0042165	neurotransmitter binding	4	0.027	GO:0007275	multicellular organismal development	16	0.098				
GO:0003682	chromatin binding	3	0.021	GO:0016192	vesicle-mediated transport	12	0.074				
GO:0008430	selenium binding	3	0.021	GO:0043062	extracellular structure organization	10	0.061				
GO:0030234	enzyme regulator activity	2	0.014	GO:0034330	cell junction organization	9	0.055				
GO:0030528	transcription regulator activity	2	0.014	GO:0048609	reproductive process in a multicellular organism	9	0.055				
GO:0009055	electron carrier activity	2	0.014	GO:0007154	cell communication	8	0.049				
GO:0017056	structural constituent of nuclear pore	2	0.014	GO:0003008	system process	7	0.043				
GO:0017080	sodium channel regulator activity	2	0.014	GO:0016049	cell growth	7	0.043				
GO:0001871	pattern binding	2	0.014	GO:0016458	gene silencing	6	0.037				
GO:0019239	deaminase activity	2	0.014	GO:0008219	cell death	5	0.031				
GO:0043021	ribonucleoprotein binding	2	0.014	GO:0033036	macromolecule localization	4	0.025				
GO:0008538	proteasome activator activity	2	0.014	GO:0048610	reproductive cellular process	4	0.025				
GO:0030337	DNA polymerase processivity factor activity	1	0.007	GO:0051236	establishment of RNA localization	4	0.025				
GO:0031406	carboxylic acid binding	1	0.007	GO:0071684	organism emergence from protective structure	4	0.025				
GO:0042562	hormone binding	1	0.007	GO:0006955	immune response	4	0.025				
GO:0046906	tetrapyrrole binding	1	0.007	GO:0007049	cell cycle	4	0.025				
GO:0051540	metal cluster binding	1	0.007	GO:0009405	pathogenesis	4	0.025				
<b>TOTAL</b>		14,570		GO:0009607	response to biotic stimulus	4	0.025				
				GO:0009791	post-embryonic development	4	0.025				

**Table 2 Gene Ontology annotation for 90e set sequences. (Continued)**

GO:0048646	anatomical structure formation involved in morphogenesis	4	0.025
GO:0009790	embryonic development	3	0.018
GO:0015976	carbon utilization	2	0.012
GO:0032506	cytokinetic process	2	0.012
GO:0045103	intermediate filament-based process	2	0.012
GO:0070882	cellular cell wall organization or biogenesis	2	0.012
GO:0006413	translational initiation	2	0.012
GO:0009566	fertilization	2	0.012
GO:0001906	cell killing	1	0.006
GO:0043473	pigmentation	1	0.006
GO:0009987	cellular process	1	0.006
GO:0022402	cell cycle process	1	0.006
GO:0022411	cellular component disassembly	1	0.006
GO:0022415	viral reproductive process	1	0.006
GO:0023061	signal release	1	0.006
GO:0030030	cell projection organization	1	0.006
GO:0051606	detection of stimulus	1	0.006
GO:0000746	conjugation	1	0.006
GO:0007163	establishment or maintenance of cell polarity	1	0.006
GO:0009653	anatomical structure morphogenesis	1	0.006
<b>TOTAL</b>		16,268	

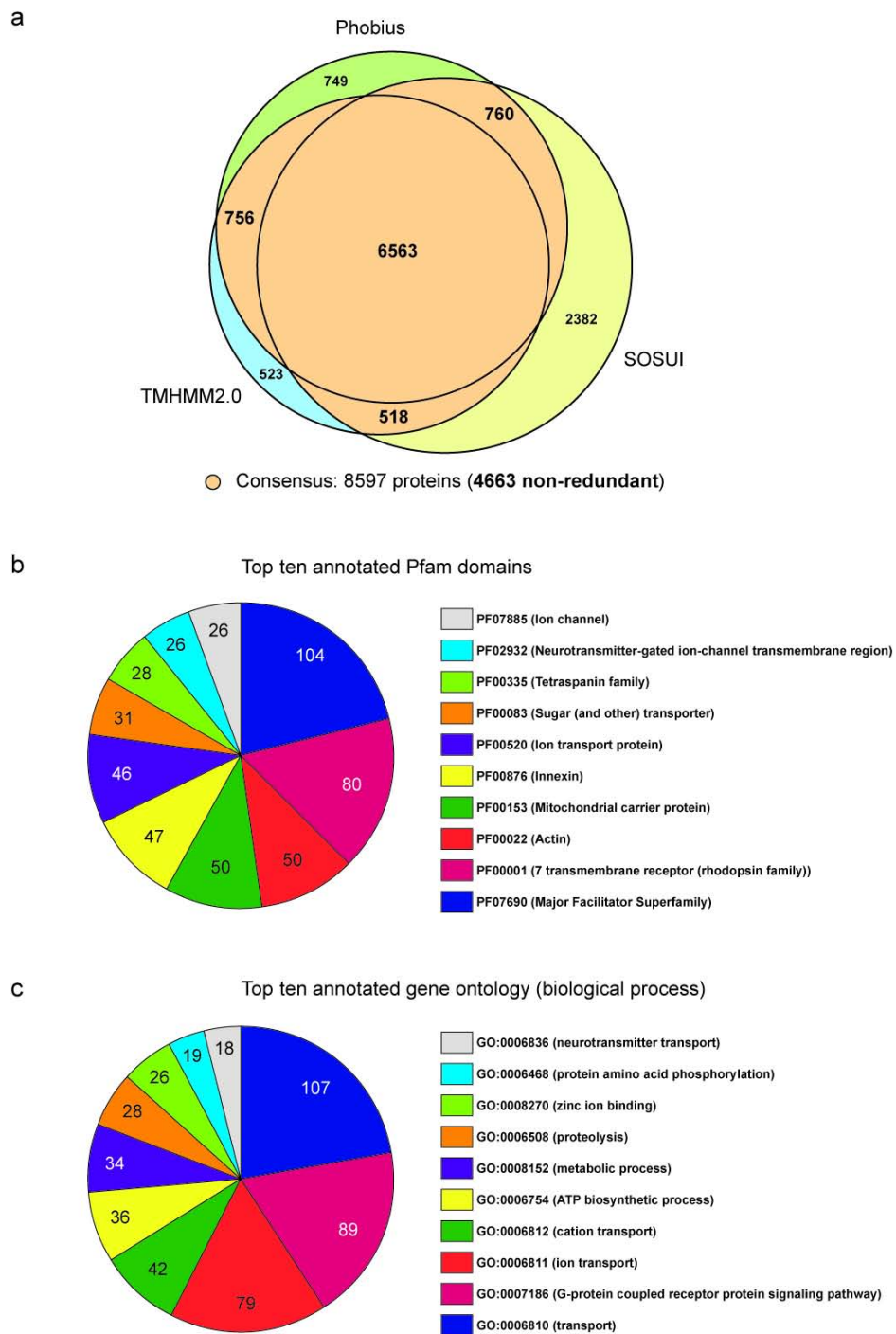
The most probable GO code in the three ontology categories—molecular function, biological process and cellular component—for each sequence in the 90e data set was selected. For simplicity, only level one and two codes are shown here in order.

dataset and found a collection of genes, ranging from transcription factors to eye-realizator genes, which have been implicated in eye development in other systems. These are good candidates for expanding our knowledge about the genetic network responsible for planarian eye regeneration (Table 5 and Additional File 9).

## Conclusions

The inherent complexity of the planarian genome and methodological difficulties initially prevented the complete genome assembly of *S. mediterranea*. High-throughput sequencing technologies are now well established and help molecular biologists to unravel the molecular components of organisms. We present a 454 sequencing dataset that can be used to decipher the transcriptome of the planarian *S. mediterranea*, an organism that has great potential for the study of regeneration processes.

We obtained more than half a million sequencing reads and assembled them into different datasets using a number of different similarity thresholds. The complete dataset has been made publicly available via web [21]. About 50,000 contigs in one of those sets (**90e**) were mapped against the most up-to-date genome scaffolds and to the set of known proteins from NCBI NR. Interestingly, we found a large number of transcribed sequences not covered by the genome sequence (more than 3 Mbp). The novel 454 contigs will allow us to extend current genomic sequences and connect up to 8,000 pairs of genome scaffolds. Furthermore, a preliminary analysis of the planarian splice sites was made on a collection of 454 contigs mapped univocally to the genome. Annotation of the sequences yielded a number of gene candidates in different functional categories that will be useful for further experimental studies. However,



**Figure 7 Prediction of planarian transmembrane proteins and functional annotations.** A) Venn-diagram showing the overlap between predictions of transmembrane proteins generated by the Phobius, TMHMM2.0 and SOSUI programs for a set of 56,362 protein sequences translated from planarian ESTs. Only proteins predicted to contain one or more transmembrane domains by at least two programs (colored orange, 8,597 proteins, of which 4,663 are non-redundant) were considered for further analysis. B) Top ten PFAM domains and C) gene ontologies (biological process) for the 4,663 non-redundant transmembrane-proteins predicted. The figures indicate the number of proteins contained in a given annotation group.



**Table 3 List of neurotransmitter, peptide and hormone receptor sequence candidates.**

ID	BLASTX HIT	ACCESSION NUMBER	E-VALUE
90_1623	adiponectin receptor (Schistosoma mansoni)	XP_002577010.1	2,00E-103
90_11706	allatostatin receptor, putative (Ixodes scapularis)	XP_002414997.1	2,00E-18
90_9653	amine GPCR (Schistosoma mansoni)	XP_002576533.1	8,00E-25
P02IKPED	atrial natriuretic peptide receptor (Aedes aegypti)	XP_001652228.1	7,00E-28
P02HWID8	beta adrenergic receptor (Aedes aegypti)	XP_001651714.1	2,00E-18
90_17484	similar to bombesin-like peptide receptor (Ornithorhynchus anatinus)	XP_001514235.1	2,00E-12
90_19322	C1A receptor, putative (Ixodes scapularis)	XP_002405845.1	4,00E-04
90_4815	calcitonin receptor, isoform CRA_d (Rattus norvegicus)	AAA65964.1	3,00E-50
90_20672	cardioexcitatory receptor (Lymnaea stagnalis)	AAB92258.1	6,00E-11
90_6224	class b secretin-like g-protein coupled receptor GPRmth5 (Pediculus humanus)	XP_002427184.1	2,00E-10
P02IZJB4	similar to putative diuretic hormone receptor II (Nasonia vitripennis)	XP_001606711.1	4,00E-22
90_7506	type I dopamine receptor (Panulirus interruptus)	ABB87183.1	8,00E-69
90_6802	dopamine receptor type D2 (Apis mellifera)	NP_001011567.1	2,00E-27
90_8536	dro/myosuppressin receptor (Schistosoma mansoni)	XP_002570000.1	7,00E-26
90_9052	FMRFamide receptor (Culex quinquefasciatus)	XP_001849293.1	2,00E-17
P02GUXTP	similar to galanin receptor type I (Danio rerio)	XP_690480.1	3,00E-06
90_6830	glutamate receptor kainate (Schistosoma mansoni)	XP_002576035.1	3,00E-70
P02GLFYW	glutamate receptor NMDA (Schistosoma mansoni)	XP_002572261.1	3,00E-21
90_15092	glutamate receptor, ionotropic, AMPA 1b (Danio rerio)	NP_991223.1	5,00E-78
90_13524	metabotropic glutamate receptor (Schistosoma mansoni)	XP_002572726.1	1,00E-12
90_18656	gonadotropin-releasing hormone receptor type I (Capra hircus)	ABL76162.1	7,00E-04
90_4098	growth hormone secretagogue receptor (Schistosoma mansoni)	XP_002569813.1	7,00E-36
90_976	growth hormone-inducible transmembrane protein (Osmerus mordax)	AC008873.1	2,00E-51
90_6465	putative insulin receptor (Echinococcus multilocularis)	CAD30260.1	6,00E-61
90_7253	lung seven transmembrane receptor (Culex quinquefasciatus)	XP_001868443.1	1,00E-68
90_17047	metabotropic GABA-B receptor subtype, putative (Ixodes scapularis)	XP_002406087.1	4,00E-21
90_6512	natriuretic peptide receptor (Xenopus laevis)	NP_001083703.1	4,00E-158
90_12800	muscarinic acetylcholine (GAR) receptor (Schistosoma mansoni)	XP_002575679.1	2,00E-39
90_1507	Nicotinic acetylcholine receptor alpha 1 subunit (Aplysia californica)	AF467898_1	3,00E-44
90_223	neuroendocrine protein 7b2 (Schistosoma mansoni)	XP_002578500.1	6,00E-25
90_6302	similar to neuromedin U receptor 2 (Strongylocentrotus purpuratus)	XP_001200425.1	4,00E-27
90_29452	neuropeptide FF receptor 2 isoform 3 (Homo sapiens)	NP_001138228.1	2,00E-09
90_6772	neuropeptide F-like receptor (Schistosoma mansoni)	XP_002573542.1	1,00E-28
90_5995	neuropeptide Y receptor Y7 (Oncorhynchus mykiss)	ABB54774.1	9,00E-18
90_25975	octopamine receptor (Aplysia californica)	AAF37686.1	1,00E-25
90_8498	odorant receptor (Tetraodon nigroviridis)	CAG08888.1	6,00E-05
90_5999	similar to olfactory receptor 355 (Bos taurus)	XP_610381.4	1,00E-05
90_2541	P2Y purinergic receptor (Meleagris gallopavo)	AAA18784.1	2,00E-04
90_8537	P2X receptor subunit (Schistosoma mansoni)	XP_002580774.1	1,00E-72
90_19040	pituitary adenylate cyclase activating polypeptide receptor (Oncorhynchus mykiss)	NP_001118113.1	1,00E-08
90_28219	parathyroid hormone 2 receptor (Danio rerio)	AAI62580.1	3,00E-11
90_6836	peptide (allatostatin)-like receptor (Schistosoma mansoni)	XP_002572656.1	2,00E-66
90_7984	peptide (allatostatin/somatostatin)-like receptor (Schistosoma mansoni)	XP_002575539.1	2,00E-32
90_10769	progesterone receptor membrane component 1 (Danio rerio)	NP_001007393.1	7,00E-04
90_5450	progesterone receptor membrane component 1 (Oryzias latipes)	BAE47967.1	2,00E-28
P02GZGVI	prolactin releasing hormone receptor (Homo sapiens)	BAG36078.1	2,00E-06
P02I1U9K	pyrokinin-like receptor (Dermacentor variabilis)	ACC99623.1	2,00E-11
90_10680	Rhodopsin-like GPCR superfamily, domain-containing protein (Schistosoma japonicum)	CAX73015.1	6,00E-37
90_2955	rhodopsin-like orphan GPCR (Schistosoma mansoni)	XP_002579928.1	2,00E-42
90_27829	ryanodine receptor 44F (Schistosoma japonicum)	CAX69439.1	8,00E-16
90_14326	serotonin receptor-like planarian receptor 1 (Dugesia japonica)	BAA22404.1	3,00E-54

**Table 3 List of neurotransmitter, peptide and hormone receptor sequence candidates. (Continued)**

90_15981	serotonin receptor 7 ( <i>Dugesia japonica</i> )	BAI44327.1	2,00E-14
90_11349	sex peptide receptor ( <i>Tribolium castaneum</i> )	NP_001106940.1	5,00E-25
P02HBR62	SIFamide receptor ( <i>Apis mellifera</i> )	NP_001106756.1	9,00E-10
90_19415	parathyroid hormone-related peptide receptor precursor ( <i>Tribolium castaneum</i> )	XP_969953.1	7,00E-20
P02FKOY5	parathyroid hormone receptor 2, isoform CRA_c ( <i>Mus musculus</i> )	EDL00229.1	3,00E-07
90_1140	somatostatin receptor ( <i>Culex quinquefasciatus</i> )	XP_001859671.1	7,00E-43
P02JZNR	tachykinin receptor 1 ( <i>Mus musculus</i> )	NP_033339.2	2,00E-06
P02FL51R	thyroid hormone receptor ( <i>Schistosoma mansoni</i> )	XP_002573733.1	2,00E-23
P02FHDMB	thyroid stimulating hormone receptor precursor ( <i>Canis lupus familiaris</i> )	NP_001003285.1	5,00E-04
90_3545	thyrotropin-releasing hormone receptor 1 ( <i>Catostomus commersonii</i> )	AAG31763.1	2,00E-51
90_26294	tyramine receptor ( <i>Bombyx mori</i> )	BAD11157.1	1,00E-11

many of the novel contigs have no similarity to known proteins and will require further validation if we want to understand the transcriptional inventory of the planarian at a functional level. We also provided a preliminary gene annotation for *S. mediterranea*, focusing our rankings on four different gene families; these serve as applied examples of the usefulness of this new sequence resource.

## Methods

### Animals and RNA isolation

*Schmidtea mediterranea* from the BCN-10 clonal line were used. Animals were starved one week prior to experiments and irradiated at a lethal dose of 100Gy. Total RNA was isolated from a mixed sample of planarians that contained non-irradiated intact and regenerating planarians (1, 3, 5 and 7 days of regeneration) as well as irradiated intact and regenerating animals (1, 3, 5 and 7 days of regeneration). RNA was extracted with TRIzol® (Invitrogen) following the manufacturer's instructions.

### cDNA library construction and 454 sequencing

First, 5 µg of total RNA was used to construct a cDNA library. RNA quality was assessed in a Bioanalyzer 2100 (Agilent-Bonsai Technologies). 5 µg of full-length double-stranded cDNA was then processed by the standard Genome Sequencer library-preparation method using the 454 DNA Library Preparation Kit (Titanium chemistry) to generate single-stranded DNA ready for emulsion PCR (emPCR™). The cDNA library was then nebulized according to the fragmentation process used in the standard Genome Sequencer shotgun library preparation procedure. The cDNA library was sequenced according to GS FLX technology (454/Roche). Reads were assembled by MIRA[60] version 3 using enhanced 454 parameters.

### Mapping to genomic and functional annotation

BLAT[61] was used with default parameters to map the Smed454 90e dataset on the *S. mediterranea* draft genome

assembly v3.1 [14] since the 454 sequences should be very similar to the corresponding genomic sequences, except for the lack of introns. Perl scripts were developed to classify all HSPs into the categories shown in Figure 3. 90e contigs having two or more collinear HSPs covering more than 100bp of the contig, and for which HSPs had more than 90% identity to the genomic contigs and length of the HSP larger than 50 bp, were chosen as 1-to-1 matches to genome. Once the sequences of the 90e/genomic contig pairs were retrieved, exonerate[62] was used to refine the alignments over the splice sites (using as parameters model = est2genome and bestn = 10). Perl scripts were used to retrieve the splice sites coordinates from exonerate output, as well as the sequences from genomic contigs. After clipping the donor and acceptor splice sites for each intron, nucleotide frequencies were computed and the corresponding position weight matrices for U2/U12 sites were drawn as pictograms using compi[19]. Known *S. mediterranea* genes were compared with contigs from 90e using BLASTN[63] with the following cut-offs: e-value = 0.001, identity score > 80%, HSP length > 50 bp.

GO functional annotation was computed on the BLASTX[63] results of the three assembly datasets (90, 98, and 90e) against all proteins from NCBI NR. BLASTX parameters were set to e-value = 10e-25 and maximum number of descriptions and alignments to report = 250, which produced around 26 million HSPs for each set. After that, only HSPs with a minimum length of 80 bp and a similarity score of at least 80% were considered. GO annotation was performed on those HSPs using the e-value selection criteria and supporting sequences described for Blast2GO[64]. Further Perl scripts were used to summarize the data shown in Table 2 and Additional File 3.

### RT-PCR

In order to validate the expression of a random subset of novel 454 transcripts, RT-PCRs were performed on planarian cDNA generated with Superscript III (Invitrogen) following the manufacturer's instructions. Additional File

**Table 4 Complete list of homeobox-containing gene sequence candidates.**

ID	BLASTX HIT	ACCESSION NUMBER	E-VALUE
F6AJXP02J3PG4	arrowhead [Schistosoma mansoni]	XP_002575389	6,00E-09
90_9219	barh homeobox protein [Schistosoma mansoni]	XP_002571667	9,00E-26
F6AJXP02J2YIH	brain-specific homeobox [Tribolium castaneum]	EFA05724	5,00E-05
90_23337	cut, isoform C [Drosophila melanogaster]	NP_001138174	3,00E-18
F6AJXP02HF7ZO	cut, isoform C [Drosophila melanogaster]	NP_001138174	2,00E-10
90_8368	Cut-like homeobox 1 [Mus musculus]	AAH14289	2,00E-23
90_3019	distalless, Dlx-1 [Platynereis dumerilii]	CAJ38799	8,00E-07
90_14605	DjotxB [Dugesia japonica]	BAF80446	4,00E-65
F6AJXP02FICZL	Eye absent protein [Dugesia japonica]	CAD89530	1,00E-74
F6AJXP02IV6Y0	gsx family homeobox protein [Schistosoma mansoni]	XP_002574396	3,00E-12
90_24312	H6-like-homeobox [Drosophila melanogaster]	NP_732244	2,00E-15
90_8293	homeobox protein distal-less dlx [Schistosoma mansoni]	XP_002574393	4,00E-07
F6AJXP02JJ1QK	Homeobox protein DTH-2 [Girardia tigrina]	Q00401	3,00E-40
90_8753	homeobox prox 1 [Danio rerio]	NP_956564	5,00E-19
90_12057	homeodomain protein Tlx [Capitella teleta]	ACH89436	1,00E-23
90_8083	Hox class homeodomain protein AbdBa Hox protein [Schmidtea mediterranea].	ABW79872	1,00E-26
90_7618	Hox class homeodomain protein DjAbd-Ba [Dugesia japonica]	BAB41079	2,00E-16
90_6369	Hox class homeodomain protein DjAbd-Bb [Dugesia japonica]	BAB41078	1,00E-108
F6AJXP02ILMDY	Hox class homeodomain protein DjAbd-Bb [Dugesia japonica]	BAB41078	3,00E-33
F6AJXP02HN15J_2	Hypothetical protein CBG18604 [Caenorhabditis briggsae]	XP_002638395	7,00E-05
90_28860	ladybird homeobox corepressor 1-like protein [Mus musculus]	NP_001103213 XP_001479028	8,00E-33
90_6629	lim domain binding protein [Schistosoma mansoni]	XP_002576324	6,00E-05
F6AJXP02GEYYP	lim domain homeobox 3/4 transcription factor [Saccoglossus kowalevskii]	NP_001158395	4,00E-23
90_10783	lim homeobox protein [Schistosoma mansoni]	XP_002579046	2,00E-13
90_11027	lim homeobox protein [Schistosoma mansoni]	XP_002579046	1,00E-26
90_10828	LIM homeobox transcription factor 1 alpha [Mus musculus]	EDL39177	2,00E-14
90_13775	LIM motif-containing protein kinase 1 [Schistosoma japonicum]	CAX72746	2,00E-11
90_9432	LIM-homeodomain protein AmphiLim1/5 [Branchiostoma floridae]	ABD59002	5,00E-05
90_8762	LIM-homeodomain transcription factor islet [Branchiostoma floridae]	AAF34717	2,00E-15
90_6339	Nk1 protein [Platynereis dumerilii]	CAJ38797	1,00E-11
F6AJXP02G077U	paired-like homeobox 2a [Danio rerio]	NP_996953	5,00E-16
90_6703	phtf [Drosophila melanogaster]	NP_610232	2,00E-55
90_25126	PLOX2-Dj [Dugesia japonica]	BAA77402	2,00E-42
90_21567	PLOX4-Dj [Dugesia japonica]	BAA77404	2,00E-21
90_23010	PLOX5-Dj [Dugesia japonica]	BAA77405	6,00E-22
F6AJXP02IVOTI	PLOX5-Dj [Dugesia japonica]	BAA77405	1,00E-17
90_21710	pre-B-cell leukemia transcription factor 1 2 3 4 (pbx) [Schistosoma mansoni]	XP_002572195	2,00E-27
90_3405	PREDICTED: similar to UBX domain protein 4 [Hydra magnipapillata]	XP_002162754	2,00E-06
F6AJXP02HI24E	PREP homeodomain-like protein [Schmidtea mediterranea]	ADB54565	2,00E-47
F6AJXP02JSRJD	PREP homeodomain-like protein [Schmidtea mediterranea]	ADB54565	1,00E-32
F6AJXP02GVFDM	prospero-like protein [Schistosoma mansoni]	XP_002578694	1,00E-21
F6AJXP02IUJ5Q	prospero-like protein [Schistosoma mansoni]	XP_002578694	4,00E-25
F6AJXP02HZIDG	short stature homeobox protein 2 isoform c [Homo sapiens]	NP_001157150	5,00E-08
90_7545	SIX homeobox 2 [Gallus gallus]	NP_001038160	7,00E-36
F6AJXP02HBGHT	SJCHGC06100 protein [Schistosoma japonicum]	AAW24487	6,00E-11
90_3395	UBX domain containing 8, isoform CRA_d [Mus musculus]	EDL41153	3,00E-14
90_1176	UBX domain-containing protein 4 [Mus musculus]	NP_080666	1,00E-06
90_2625	ubx6(yeast)-related [Schistosoma mansoni]	XP_002576054	2,00E-16
90_24438	visual system homeobox protein [Tribolium castaneum]	CAX64460	9,00E-23
F6AJXP02G5JJX_1	Zn finger homeodomain 2 [Tribolium castaneum]	EFA01350	1,00E-05

**Table 5 List of eye-related gene sequence candidates.**

ID	BLASTX HIT	ACCESSION Nr.	E-VALUE
90_7233	abl interactor 2 [Schistosoma japonicum]	CAX69750.1	6.00E-019
90_4001	adaptor-related protein complex [Schistosoma mansoni]	XP_002574891.1	3.00E-072
90_30923	arginine/serine-rich splicing factor [Schistosoma mansoni]	XP_002574990.1	2.00E-026
90_482	ATPase protein [Schistosoma japonicum]	AAW26203.1	3.00E-049
90_3152	beta-catenin-like protein 2 [Schmidtea mediterranea]	ABW79874.1	0
90_12909	BMP [Schmidtea mediterranea]	ABV04322.1	3.00E-090
90_120	cat eye syndrome protein [Schistosoma japonicum]	AAX27345.2	4.00E-035
P02FKNEB	CaTaLase family member (ctl-2) [Caenorhabditis elegans]	NP_001022473.1	1.00E-029
90_205	Chaperonin Containing TCP-1 family member (cct-3) [Caenorhabditis elegans]	NP_494218.2	1.00E-090
C90_6158	disks large homolog 1 isoform 1 [Homo sapiens]	NP_001091894.1	2.00E-027
P02GJNCV	extradenticle 1 protein [Schistosoma japonicum]	AAW24487.1	3.00E-013
P02JK4Z_2	eye53 [Dugesia japonica]	BAD20650.1	6.00E-016
90_8483	eyes absent protein [Dugesia japonica]	CAD89531.1	2.00E-064
90_651f	ascin protein [Schistosoma japonicum]	XP_002574990.1	5.00E-045
90_14368	heat shock protein 70 [Lumbricus terrestris]	ACB77918.1	4.00E-038
90_9533	Heparan sulfate 6-O-sulfotransferase 2 [Danio rerio]	AAH45453.1	1.00E-042
90_6564	histone-lysine n-methyltransferase suv9 [Schistosoma mansoni]	XP_002574171.1	3.00E-061
90_15456	homeodomain protein NK4 [Platynereis dumerilii]	ABQ10640.1	8.00E-023
90_12892	homeotic protein six3-alpha [Mus musculus]	S74256	1.00E-082
90_325	importin-7 [Culex quinquefasciatus]	XP_001843364.1	2.00E-147
90_4360	intraflagellar transport 57 homolog [Xenopus (Silurana) tropicalis]	NP_001016561.1	1.00E-044
90_11027	lim homeobox protein [Schistosoma mansoni]	XP_002579046.1	5.00E-027
90_8432	lozenge [Schistosoma mansoni]	XP_002580418.1	8.00E-032
90_8924	Male ABnormal family member (mab-21) [Caenorhabditis elegans]	NP_497940.2	1.00E-046
P02HSHWR	mothers against decapentaplegic homolog 4 [Mus musculus]	NP_032566.2	5.00E-018
90_5640	muscleblind-like protein [Schistosoma mansoni]	XP_002575346.1	3.00E-025
P02F0EF6	neurogenic differentiation [Platynereis dumerilii]	CAQ57533.1	2.00E-012
P02GMLJM	nuclear transcription factor X-box binding 1 (nfx1) [Schistosoma bovis]	XP_002577564.1	5.00E-014
90_828	phenylalanine hydroxylase [Caenorhabditis elegans]	AAD31643.1	2.00E-145
90_2925	protein [Schistosoma japonicum]	AAW24487.1	4.00E-126
90_7228	protein kinase [Schistosoma mansoni]	XP_002576342.1	2.00E-077
90_2256	Protein pob [Schistosoma japonicum]	CAX75988.1	4.00E-089
90_4436	Rab-protein 6 [Drosophila melanogaster]	NP_477172.1	8.00E-085
P02GENUT_1	retinaldehyde dehydrogenase 1 [Eleutherodactylus coqui]	ACE74542.1	8.00E-008
90_11988	runt protein [Branchiostoma lanceolatum]	AAN08565.1	4.00E-017
P02FN7BT	Septin-7 (CDC10 protein homolog) [Schistosoma japonicum]	CAX83064.1	3.00E-012
90_3747	serine/threonine protein kinase [Schistosoma mansoni]	XP_002580180.1	9.00E-094
P02FICZL	six1-2 protein [Dugesia japonica]	CAD89530.1	8.00E-86
P02IZDJZ_1	SRY-related HMG box B protein [Platynereis dumerilii]	CAY12631.1	3.00E-028
P02HE4J6	strabismus protein CBR-VANG-1 [Platynereis dumerilii]	CAJ26300.1	1.00E-006
90_9483	tetratricopeptide repeat protein 10 tpr10 [Schistosoma mansoni]	XP_002573898.1	4.00E-048
90_16088	tyrosine kinase [Schistosoma mansoni]	XP_002576978.1	2.00E-031
90_11388	ubiquitin conjugating enzyme E2 [Schistosoma mansoni]	XP_002578016.1	3.00E-053
90_1263	vacuolar ATP synthase proteolipid subunit 1 2 3 [Schistosoma japonicum]	XP_002571892.1	9.00E-049
90_12567	vermilion [Drosophila ananassae]	XP_001963597.1	2.00E-012
90_5500	white pigment protein [Drosophila melanogaster]	CAA26716.2	2.00E-020
90_13309	YY1 transcription factor [Schistosoma japonicum]	CAX73893.1	5.00E-049
90_10118	zinc finger protein 42 homolog [Homo sapiens]	NP_777560.2	6.00E-031
90_9460	14-3-3 zeta isoform [Schistosoma bovis]	AAT39382.1	2.00E-023
P02ILIK3	52-kD bracketing protein [Drosophila melanogaster]	CAA44483.1	1.00E-016

3 includes a list of the contigs validated and the primers used for each of them.

### Prediction of transmembrane proteins from ESTs

A total of 53,867 assembled ESTs (90e database) and 2,495 additional mRNAs were translated into all six reading frames using the 'transeq' program from the EMBOSS package [65]. The longest open reading frame for each EST/mRNA was then extracted and used as a protein database (containing 56,362 protein sequences overall) for the prediction of membrane-spanning proteins. We followed an approach described by Almen et al. [66] basing our analysis on consensus predictions of alpha-helices and using three applications: Phobius[67], TMHMM2.0[68], and SOSUI[69]. Phobius and TMHMM2.0 both use hidden Markov models based on different training sets to predict membrane topology. SOSUI evaluates proteins for their hydrophobic and amphiphilic properties to make its predictions. The use of all three programs should improve prediction accuracy. We first ran Phobius, which can predict both transmembrane helices and signal peptides. Signal peptide sequences are similar to transmembrane segments owing to their hydrophobic nature [70]. To avoid false positive predictions, we excluded signal peptides before running TMHMM2.0 and SOSUI.

### Additional material

**Additional file 1: GO annotation for 90e contigs not mapping onto the WUSL 3.1 genome assembly.** 8,831 90e contigs were not found in the genome. 3,480 had a BLASTX hit to a sequence of NCBI NRprot; yet only 2,401 had a hit to a protein functionally annotated in the GO database. This file contains the description of the best HSP for 71 of those annotated contigs, after filtering out as described above. (Header: CONTIG ID = Smed454 sequence identifier, E-VALUE = BLASTX HSP E-value, ALN\_SCORE = HSP alignment score, IDENTITIES = number of identical amino acids, POSITIVES = number of similar amino acids, SEQUENCE ID = Protein sequence identifier, ACCESSION NUMBER = Protein sequence full accession number, SEQUENCE DESCRIPTION = Full protein GenBank description).

**Additional file 2: Splice sites for a subset of Smed454 sequences mapped onto the Schmidtea mediterranea genome.** (Header: GID = Genomic contig Identifier from WUSLv3.1 genome assembly—including the start and end nucleotide coordinates for the complete match—, CIG=90e contig Identifier, INTNUM = Intron number within the 90e contig, EXO = splice signals found by exonerate, ORI = sequence orientation—here -1 means that the match was found on the reverse strand of the genomic contig—, CEXO = corrected splice site signals after reverse complementing the genomic sequence when required, ILEN = Intron length in bp, IORI = Intron start—relative to the match coordinates—, IEND = Intron end—relative to the match coordinates—, STRAND, SSSEQ = Splice sites sequences—where a point separates three nucleotides from the 5' and 3' exons, and the three dots in the middle denote intron sequence not shown for clarity—).

**Additional file 3: List of 90e transcripts validated by RT-PCR.** (Header: # = Number, CONTIG=90e contig ID, PRIMER\_FORWARD = 5' to 3' sequence of the forward primer used, REVERSE\_FORWARD = 5' to 3' sequence of the reverse primer used, AMPLICON SIZE = Size amplified in bp, SET = refers to the subset of origin of the 90e contig: no hit genome, hit genome, - blast (no BLASTX hit), +blast (BLASTX hit)).

**Additional file 4: Smed454 sequences matching known Schmidtea mediterranea genes.** (Header: ACCESSION NUMBER = Known gene sequence identifier as target, NAME = Description for that sequence, LENGTH = Nucleotide length for that sequence, A&T CONTENT = Sequence composition, 454 90e CONTIG/SINGLETON = Smed454 sequence identifier as query, LENGTH = Nucleotide sequence length for this sequence, ALIGNMENT LENGTH = HSP length, START = Start nucleotide of alignment on target, END = Final nucleotide of alignment on target, IDENTITY = Identity score, BITSCORE = Alignment bit score, E-VALUE = HSP BLAST e-value, HIT LENGTH = Un-gapped length of the alignment on the target, %COVERAGE = Sum of co-linear HSPs on target coordinates divided by the total length of the target, #SEQs = Number of co-linear HSPs considered, avg%COV = The coverage divided by the number of co-linear HSPs).

**Additional file 5: Gene Ontology for all three Smed454 sets: 90, 98 and 90e.** Level one and two GO codes are shown in order to simplify the listings. Although there are small changes in GO frequencies, annotation is consistent throughout all three sets. (Header: GO = Gene Ontology unique identifier, Count = Number of sequences with a given GO annotation, Freq% = Frequencies for every GO annotation. The total shown does not include the un-annotated and over-represented features, that is, the first two rows on each table).

**Additional file 6: List of cell cycle, cell division, DNA repair or DNA damage candidates.** Short list of candidates annotated as genes involved in cell cycle, cell division, DNA repair or DNA damage. (Header: ID = Smed454 sequence identifier, BLASTX HIT = Description of the best sequence hit, ACCESSION NUMBER = Sequence identifier of the best sequence hit, E-VALUE = BLASTX e-value for that sequence hit).

**Additional file 7: Summary report for the consensus set of 4,663 predicted transmembrane proteins including functional annotations.** (Header: Sequence\_ID = Protein sequence identifier, Sequence\_AA = Amino acid sequence, Length[aa] = Length of amino acid sequence, Phobius\_TM = Phobius prediction of number of transmembrane domains, Phobius\_SP = Phobius prediction of signal peptide, Phobius\_Top = Phobius prediction of membrane topology, TMHMM\_TM = TMHMM2.0 prediction of number of transmembrane domains, TMHMM\_Top = TMHMMv2.0 prediction of membrane topology, SOSUI\_TM = SOSUI prediction of number of transmembrane domains, SOSUI\_Top = SOSUI prediction of membrane topology, UFO\_PFAM = UFO annotation of Pfam protein families, UFO\_GO = UFO annotation of gene ontologies).

**Additional file 8: List of neurotransmitter, peptide and hormone receptor sequence candidates.** Complete complement of Smed454 dataset contigs and singletons showing homology to neurotransmitter and hormone receptors, totalling 287 sequences. (Header: ID = Smed454 sequence identifier, BLASTX HIT = Description of the best sequence hit, ACCESSION NUMBER = Sequence identifier of the best sequence hit, E-VALUE = BLASTX e-value for that sequence hit).

**Additional file 9: List of eye-related gene sequence candidates.** Complete complement of Smed454 dataset contigs and singletons showing homology to eye-related genes, totalling 95 sequences. (Header: ID = Smed454 sequence identifier, BLASTX HIT = Description of the best sequence hit, ACCESSION NUMBER = Sequence identifier of the best sequence hit, E-VALUE = BLASTX e-value for that sequence hit).

### Abbreviations

bp: base pairs (nucleotide length unit); EST: Expressed Sequence Tag; GC%: percent of guanine+cytosine sequence content; HSP: High-scoring Segment Pair; GO: Gene Ontology; WUSL: Washington University in St Louis; TM: transmembrane; RDGN: retinal determination gene network; Gy: gray.

### Acknowledgements

Genomic sequence data was produced by the Washington University Genome Sequencing Center in St. Louis. DNA library construction, 454 pyrosequencing and the contig browser were generated by Skuldtech (Montpellier, France). We thank Ann King for language revision and editorial advice, and Franziska Konert for technical assistance. SF and GRE are

supported by FPI fellowships, and BC is supported by an FPU fellowship (both MICINN, Spain). TH is supported by a fellowship of the Studienstiftung des deutschen Volkes. FC is a Ramón y Cajal researcher (MICINN, Spain). This work was funded by grants BFU2008-01544 (MICINN) to ES, BFU2008-00710 (MICINN) to FC, BFU2009-09102(MICINN) to JFA, and 2009SGR1018 (AGAUR) to ES and FC. Finally, K.B. was funded by a FRONTIER grant from the University of Heidelberg DFG Excellence Initiative.

#### Author details

<sup>1</sup>Departament de Genètica, Facultat de Biologia, Universitat de Barcelona (UB), Av. Diagonal 645, edifici annex, planta 1, 08028, Barcelona, Catalunya, Spain. <sup>2</sup>Institut de Biomedicina de la Universitat de Barcelona (IBUB), Av. Diagonal 645, edifici annex, planta 1, 08028, Barcelona, Catalunya, Spain. <sup>3</sup>Division of Signaling and Functional Genomics German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, 69120 Heidelberg, Germany. <sup>4</sup>Max Planck Research Group Stem Cells and Regeneration Max-Planck-Institute for Molecular Biomedicine, Von-Esmarch-Strasse 54, 48149 Muenster, Germany.

#### Authors' contributions

JFA performed the computational analyses on the assemblies, the GO characterization, the mapping into the genome and the analysis of spliced sites, and prepared all the corresponding figures and tables. GRE analyzed the coverage of known annotated genes and generated the corresponding table. TH performed the sequence analysis of planarian transmembrane proteins, generated the corresponding figure and table and designed primers for RT-PCRs. GRE, SF, FC and KB designed the primers and performed the RT-PCRs. FC and SF analyzed the annotated data to characterize the neurotransmitter, peptide and hormone receptors and prepared the corresponding tables. ES and BC analyzed the annotated data to characterize homeobox-containing and eye-related genes and prepared the corresponding tables. JFA, KB, FC and ES conceived of the study, participated in its design and coordination, and helped draft the manuscript. All authors read and approved the final manuscript.

Received: 31 August 2010 Accepted: 31 December 2010  
Published: 31 December 2010

#### References

1. Agata K, Tanaka T, Kobayashi C, Kato K, Saitoh Y: Intercalary regeneration in planarians. *Dev Dyn* 2003, **226**(2):308-316.
2. Newmark PA, Sánchez-Alvarado A: Not your father's planarian: a classic model enters the era of functional genomics. *Nat Rev Genet* 2002, **3**(3):210-219.
3. Reddien PW, Sánchez-Alvarado A: Fundamentals of planarian regeneration. *Annu Rev Cell Dev Biol* 2004, **20**:725-757.
4. Saló E: The power of regeneration and the stem-cell kingdom: freshwater planarians (Platyhelminthes). *Bioessays* 2006, **28**(5):546-559.
5. Bagnà J, Saló E, Auladell C: Regeneration and pattern formation in planarians. III. Evidence that neoblasts are totipotent stem cells and the source of blastema cells. *Development* 1989, **107**:77-86.
6. Newmark PA, Sánchez-Alvarado A: Bromodeoxyuridine specifically labels the regenerative stem cells of planarians. *Dev Biol* 2000, **220**(2):142-153.
7. Cebrià F: Regenerating the central nervous system: how easy for planarians! *Dev Genes Evol* 2007, **217**(11-12):733-748.
8. Scimone ML, Meisel J, Reddien PW: The Mi-2-like Smed-CHD4 gene is required for stem cell differentiation in the planarian *Schmidtea mediterranea*. *Development* 2010, **137**(8):1231-1241.
9. Fernández-Taboada E, Moritz S, Zeuschner D, Stehling M, Scholer HR, Saló E, Gentile L: Smed-SmB, a member of the LSM protein superfamily, is essential for chromatoid body organization and planarian stem cell proliferation. *Development* 2010, **137**(7):1055-1065.
10. Forsthoeft DJ, Newmark PA: Emerging patterns in planarian regeneration. *Curr Opin Genet Dev* 2009, **19**(4):412-420.
11. Sánchez-Alvarado A, Newmark PA, Robb SM, Juste R: The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development* 2002, **129**(24):5659-5665.
12. Zayas RM, Hernandez A, Habermann B, Wang Y, Stary JM, Newmark PA: The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: analysis of ESTs from the hermaphroditic strain. *Proc Natl Acad Sci USA* 2005, **102**(51):18491-18496.
13. Pontius JU, Wagner L, Schuler GD: UniGene: a unified view of the transcriptome. In *The NCBI Handbook* Edited by: Information NCfB. Bethesda (MD) 2003.
14. *Schmidtea mediterranea* genome sequencing project. [[http://genome.wustl.edu/genomes/view/schmidtea\\_mediterranea/](http://genome.wustl.edu/genomes/view/schmidtea_mediterranea/)].
15. Proposal for the Sequencing of a New Target Genome: White Paper for a Planarian Genome Project. [[http://genome.wustl.edu/images/uploads/Whitepaper/Schmidtea\\_mediterranea\\_WP.pdf](http://genome.wustl.edu/images/uploads/Whitepaper/Schmidtea_mediterranea_WP.pdf)].
16. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sánchez-Alvarado A, Yandell M: MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* 2008, **18**(1):188-196.
17. Robb SM, Ross E, Sánchez-Alvarado A: SmedGD: the *Schmidtea mediterranea* genome database. *Nucleic Acids Res* 2008, **36** Database: D599-606.
18. Emrich SJ, Barbazuk WB, Li L, Schnable PS: Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Res* 2007, **17**(1):69-73.
19. Abril JF, Castelo R, Guigo R: Comparison of splice sites in mammals and chicken. *Genome Res* 2005, **15**(1):111-119.
20. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: GenBank. *Nucleic Acids Res* 2009, **37** Database: D26-31.
21. Smed454 datasets web site. [<http://planarian.bio.ub.es/datasets/454/>].
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000, **25**(1):25-29.
23. Hatfield DL: Selenium: Its Molecular Biology and Role in Human Health. Dordrecht: Kluwer Academic Publishers; 2001.
24. Castellano S, Novoselov SV, Kryukov GV, Lescure A, Blanco E, Krol A, Gladyshev VN, Guigo R: Reconsidering the evolution of eukaryotic selenoproteins: a novel nonmammalian family with scattered phylogenetic distribution. *EMBO Rep* 2004, **5**(1):71-77.
25. Eisenhoffer GT, Kang H, Sánchez-Alvarado A: Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell Stem Cell* 2008, **3**(3):327-339.
26. Rossi L, Salvetti A, Marincola FM, Lena A, Deri P, Mannini L, Batistoni R, Wang E, Gremigni V: Deciphering the molecular machinery of stem cells: a look at the neoblast gene expression profile. *Genome Biol* 2007, **8**(4):R62.
27. Meinicke P: UFO: a web server for ultra-fast functional profiling of whole genome protein sequences. *BMC Genomics* 2009, **10**:409.
28. Bonuccelli L, Rossi L, Lena A, Scarcelli V, Rainaldi G, Evangelista M, Iacopetti P, Gremigni V, Salvetti A: An RbAp48-like gene regulates adult stem cells in planarians. *J Cell Sci* 2010, **123**(Pt 5):690-698.
29. Guo T, Peters AH, Newmark PA: A Bruno-like gene is required for stem cell maintenance in planarians. *Dev Cell* 2006, **11**(2):159-169.
30. Pearson BJ, Sánchez-Alvarado A: A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. *Development* 2010, **137**(2):213-221.
31. Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Sánchez-Alvarado A: SmedWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science* 2005, **310**(5752):1327-1330.
32. Franquinet R: [The role of serotonin and catecholamines in the regeneration of the Planaria *Polycelis tenuis*]. *J Embryol Exp Morphol* 1979, **51**:85-95.
33. Franquinet R, Martelly I: Effects of serotonin and catecholamines on RNA synthesis in planarians; in vitro and in vivo studies. *Cell Differ* 1981, **10**(4):201-209.
34. Franquinet R, Le Moigne A, Hanoune J: The adenylate cyclase system of planaria *Polycelis tenuis*: activation by serotonin and guanine nucleotides. *Biochim Biophys Acta* 1978, **539**(1):88-97.
35. Bagnà J, Saló E, Romero R: Effects of activators and antagonists of the neuropeptides substance P and substance K on cell proliferation in planarians. *Int J Dev Biol* 1989, **33**(2):261-266.
36. Bautz A, Schilt J: Somatostatin-like peptide and regeneration capacities in planarians. *Gen Comp Endocrinol* 1986, **64**(2):267-272.
37. Kreshchenko ND: Functions of flatworm neuropeptides NPF, GYIRF and FMRF in course of pharyngeal regeneration of anterior body fragments of planarian, *Girardia tigrina*. *Acta Biol Hung* 2008, **59**(Suppl):199-207.
38. García-Fernández J, Bagnà J, Saló E: Planarian homeobox genes: cloning, sequence analysis, and expression. *Proc Natl Acad Sci USA* 1991, **88**(16):7338-7342.

39. Saló E, Baguña J: **Regeneration in planarians and other worms: New findings, new tools, and new perspectives.** *J Exp Zool* 2002, **292**(6):528-539.
40. Saló E, Tauler J, Jiménez E, Bayasas JR, González J, García-Fernández J, Baguña J: **Hox and parahox genes in flatworms: characterization and expression.** *Amer Zool* 2001, **41**:652-663.
41. Bayasas JR, Castillo E, Saló E: **Platyhelminthes have a hox code differentially activated during regeneration, with genes closely related to those of spiralian protostomes.** *Dev Genes Evol* 1998, **208**(8):467-473.
42. Nogi T, Watanabe K: **Position-specific and non-colinear expression of the planarian posterior (Abdominal-B-like) gene.** *Dev Growth Differ* 2001, **43**(2):177-184.
43. Orii H, Kato K, Umeson Y, Sakurai T, Agata K, Watanabe K: **The planarian HOM/HOX homeobox genes (Plox) expressed along the anteroposterior axis.** *Dev Biol* 1999, **210**(2):456-468.
44. Iglesias M, Gómez-Skarmeta JL, Saló E, Adell T: **Silencing of *Smed-betacatenin1* generates radial-like hypercephalized planarians.** *Development* 2008, **135**(7):1215-1221.
45. Saló E, Batistoni R: **The planarian eye: a simple and plastic system with great regenerative capacity.** *Animal models in eye research* Tsonis PA: Academic Press; 2008, 15-26.
46. Saló E, Pineda D, Marsal M, González J, Gremigni V, Batistoni R: **Genetic network of the eye in Platyhelminthes: expression and functional analysis of some players during planarian regeneration.** *Gene* 2002, **287**:67-74.
47. Gehring WJ: **New perspectives on eye development and the evolution of eyes and photoreceptors.** *J Hered* 2005, **96**(3):171-184.
48. Gehring WJ, Ikeo K: **Pax 6: mastering eye morphogenesis and eye evolution.** *Trends Genet* 1999, **15**(9):371-377.
49. Silver SJ, Rebay I: **Signaling circuitries in development: insights from the retinal determination gene network.** *Development* 2005, **132**(1):3-13.
50. Treisman JE: **A conserved blueprint for the eye?** *Bioessays* 1999, **21**(10):843-850.
51. Mannini L, Rossi L, Deri P, Gremigni V, Salvetti A, Saló E, Batistoni R: **Djeyes absent (*Djeya*) controls prototypic planarian eye regeneration by cooperating with the transcription factor *Djsix-1*.** *Dev Biol* 2004, **269**(2):346-359.
52. Pineda D, González J, Callaerts P, Ikeo K, Gehring WJ, Saló E: **Searching for the prototypic eye genetic network: *Sine oculis* is essential for eye regeneration in planarians.** *Proc Natl Acad Sci USA* 2000, **97**(9):4525-4529.
53. Pineda D, González J, Marsal M, Saló E: **Evolutionary conservation of the initial eye genetic pathway in planarians.** *Belg J Zool* 2001, **131**:77-82.
54. Pineda D, Rossi L, Batistoni R, Salvetti A, Marsal M, Gremigni V, Falleni A, González-Linares J, Deri P, Saló E: **The genetic network of prototypic planarian eye regeneration is *Pax6* independent.** *Development* 2002, **129**(6):1423-1434.
55. Inoue T, Kumamoto H, Okamoto K, Umeson Y, Sakai M, Sánchez-Alvarado A, Agata K: **Morphological and functional recovery of the planarian photosensing system during head regeneration.** *Zoolog Sci* 2004, **21**(3):275-283.
56. Cebrià F, Newmark PA: **Planarian homologs of netrin and netrin receptor are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture.** *Development* 2005, **132**(16):3691-3703.
57. Sánchez Alvarado A, Newmark PA: **Double-stranded RNA specifically disrupts gene expression during planarian regeneration.** *Proc Natl Acad Sci USA* 1999, **96**(9):5049-5054.
58. Takano T, Pulvers JN, Inoue T, Tarui H, Sakamoto H, Agata K, Umeson Y: **Regeneration-dependent conditional gene knockdown (Readyknock) in planarian: Demonstration of requirement for *Djsnap-25* expression in the brain for negative phototactic behavior.** *Dev Growth Differ* 2007, **49**(5):383-394.
59. Handberg-Thorsager M, Saló E: **The planarian nanos-like gene *Smednos* is expressed in germline and eye precursor cells during development and regeneration.** *Dev Genes Evol* 2007, **217**(5):403-411.
60. Chevreaux B: **MIRA: An automated genome and EST assembler.** Duisburg: Heidelberg; 2005.
61. Kent WJ: **BLAT—the BLAST-like alignment tool.** *Genome Res* 2002, **12**(4):656-664.
62. **exonerate: A generic tool for sequence alignment.** [http://www.ebi.ac.uk/~guy/exonerate/].
63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**(3):403-410.
64. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite.** *Nucleic Acids Res* 2008, **36**(10):3420-3435.
65. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**(6):276-277.
66. Almen MS, Nordstrom KJ, Fredriksson R, Schioth HB: **Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin.** *BMC Biol* 2009, **7**:50.
67. Kall L, Krogh A, Sonnhammer EL: **Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server.** *Nucleic Acids Res* 2007, **35** Web Server: W429-432.
68. Krogh A, Larsson B, von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J Mol Biol* 2001, **305**(3):567-580.
69. Hirokawa T, Boon-Chieng S, Mitaku S: **SOSUI: classification and secondary structure prediction system for membrane proteins.** *Bioinformatics* 1998, **14**(4):378-379.
70. Ahram M, Litou ZI, Fang R, Al-Tawallbeh G: **Estimation of membrane proteins in the human proteome.** *In Silico Biol* 2006, **6**(5):379-386.
71. Hintze JL, Nelson RD: **Violin plots: a box plot-density trace synergism.** *The American Statistician* 1998, **52**(2):181-184.

doi:10.1186/1471-2164-11-731

Cite this article as: Abril et al.: *Smed454* dataset: unravelling the transcriptome of *Schmidtea mediterranea*. *BMC Genomics* 2010 **11**:731.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit



## Artículo 4

***Digital Gene Expression approach over multiple RNA-Seq data sets to detect neoblast transcriptional changes in Schmidtea mediterranea.***

**Gustavo Rodríguez-Esteban**, Alejandro González-Sastre<sup>\*</sup>, José Ignacio Rojo-Laguna<sup>\*</sup>, Emili Saló and Josep F. Abril.

**BMC Genomics** (2015) 16:361

<sup>\*</sup> Ambos autores han contribuido por igual.





## Resumen

**Aproximación mediante *Digital Gene Expression* sobre múltiples conjuntos de RNA-Seq para detectar cambios transcripcionales en *Schmidtea mediterranea***

En este artículo se realizó un análisis de expresión transcripcional de las células madre de la planaria *Schmidtea mediterranea*, los neoblastos, con la ayuda de las tecnologías de secuenciación *high-throughput* para RNA.

La tecnología de *Digital Gene Expression* (DGE) permitió comparar los perfiles de expresión de los neoblastos con el de células diferenciadas para extraer aquellos transcritos característicos de estas células madre adultas.

Se presentan 42 nuevos genes de neoblasto, incluyendo varios factores de transcripción y genes relacionados con cáncer, y se analizan en profundidad el gen *Smed-meis-like* y tres subunidades del Factor Nuclear Y: *Smed-nf-YA*, *Smed-nf-YB-2* y *Smed-nf-YC*.

Esta aproximación permite concluir que los neoblastos son una población compleja de células pluripotentes y multipotentes reguladas tanto por factores de transcripción como por genes relacionados con cáncer.



RESEARCH ARTICLE

Open Access

# Digital gene expression approach over multiple RNA-Seq data sets to detect neoblast transcriptional changes in *Schmidtea mediterranea*

Gustavo Rodríguez-Esteban, Alejandro González-Sastre<sup>†</sup>, José Ignacio Rojo-Laguna<sup>†</sup>, Emili Saló<sup>\*</sup> and Josep F Abril<sup>\*</sup>

## Abstract

**Background:** The freshwater planarian *Schmidtea mediterranea* is recognised as a valuable model for research into adult stem cells and regeneration.

With the advent of the high-throughput sequencing technologies, it has become feasible to undertake detailed transcriptional analysis of its unique stem cell population, the neoblasts. Nonetheless, a reliable reference for this type of studies is still lacking.

**Results:** Taking advantage of digital gene expression (DGE) sequencing technology we compare all the available transcriptomes for *S. mediterranea* and improve their annotation. These results are accessible via web for the community of researchers.

Using the quantitative nature of DGE, we describe the transcriptional profile of neoblasts and present 42 new neoblast genes, including several cancer-related genes and transcription factors. Furthermore, we describe in detail the *Smed-meis-like* gene and the three Nuclear Factor Y subunits *Smed-nf-YA*, *Smed-nf-YB-2* and *Smed-nf-YC*.

**Conclusions:** DGE is a valuable tool for gene discovery, quantification and annotation. The application of DGE in *S. mediterranea* confirms the planarian stem cells or neoblasts as a complex population of pluripotent and multipotent cells regulated by a mixture of transcription factors and cancer-related genes.

**Keywords:** Planaria, Neoblast, Stem cell, Transcriptome, Transcription factor

## Background

During the last decade, there has been increasing interest in the use of *Schmidtea mediterranea* as a model organism for the study of stem cells. These freshwater planarians contain a population of adult stem cells known as neoblasts, which are essential for normal cell renewal during homeostasis and which confers them with amazing regeneration capabilities [1-4]. Although a number of studies based on massive RNA interference (RNAi) [5], gene inhibition [6], microarray [7], and proteomics [8,9]

approaches have been carried out to identify the crucial neoblast genes responsible for their stemness, our understanding of their biology is far from complete. The use of next generation sequencing (NGS) technologies provides an opportunity to study these cells in depth at a transcriptional level. For that to be accomplished, however, a reliable transcriptome and genome references are required. Up to eight versions of the transcriptome for this organism have been published to date, making use of different RNA-Seq technologies [10-16], including one meta-assembly which slightly improves each one separately [17]. Despite all these efforts, a consistent reference transcriptome is still lacking.

Some studies have provided quantitative data on transcripts and their respective assemblies, focusing

\*Correspondence: esalo@ub.edu; jabil@ub.edu

<sup>†</sup>Equal contributors

Departament de Genètica, Facultat de Biologia, Universitat de Barcelona (UB), and Institut de Biomedicina de la Universitat de Barcelona (IBUB), Av. Diagonal 643, 08028 Barcelona, Catalonia, Spain

on regeneration [13,17,18] or directly on neoblasts [11,14,15,19]. However, RNA-Seq suffers from an intrinsic bias that affects the quantification of transcript expression in a length-dependent manner. This bias is independent of the sequencing platform and cannot be avoided nor removed by increasing the sequencing coverage or the length of the reads. Furthermore, it cannot be corrected a posteriori during the statistical analysis (by transcript length normalization, for instance). Consequently, the quantification of the transcripts and the detection of differentially expressed genes is compromised [20-22]. Digital gene expression (DGE) [23] is a sequence-based approach for gene expression analyses, that generates a digital output at an unparalleled level of sensitivity [22,24]. The output is highly correlated with qPCR [25-27] and does not suffer from sequence-length bias. The combination of DGE and RNA-Seq data has been shown to help overcome the specific limitations of RNA-Seq [28], and the usefulness of DGE has been thoroughly demonstrated in research ranging from humans [26,29] to non-model organisms [22,24]. However, to date, DGE has not been extensively applied to the study of the planarian transcriptome.

Here, we have compiled and analyzed all the transcriptomic and genomic data available for *S. mediterranea* using DGE. This has facilitated an improved annotation and provided tools to ease the comparison and browsing of all the information available for the planarian community.

We have taken advantage of the resolution of DGE to quantitatively characterize isolated populations of proliferating neoblasts, their progeny, and differentiated cells through fluorescence-activated cell sorting (FACS) [30,31]. The resulting changes in transcription levels were analyzed to obtain transcript candidates for which an extensive experimental validation was performed. This has yielded new neoblast-specific genes, including many transcription factors and cancer-related homologous genes, confirming the validity of our strategy and the utility of the tools that we have implemented. Moreover, we provide a deeper molecular description of four of those candidates, the *Smed-meis-like*, and the three subunits of the Nuclear Factor Y (NF-Y) complex *Smed-nf-YA*, *Smed-nf-YB-2*, and *Smed-nf-Y-C*. Both families of genes are attractive candidates to be studied in planaria. The Meis family of transcription factors specify anterior cell fate and axial patterning [32], whereas the NF-Y complex is a heterotrimeric transcription factor that promotes chromatin opening and is involved in the regulation of a wide number of early developmental genes [33].

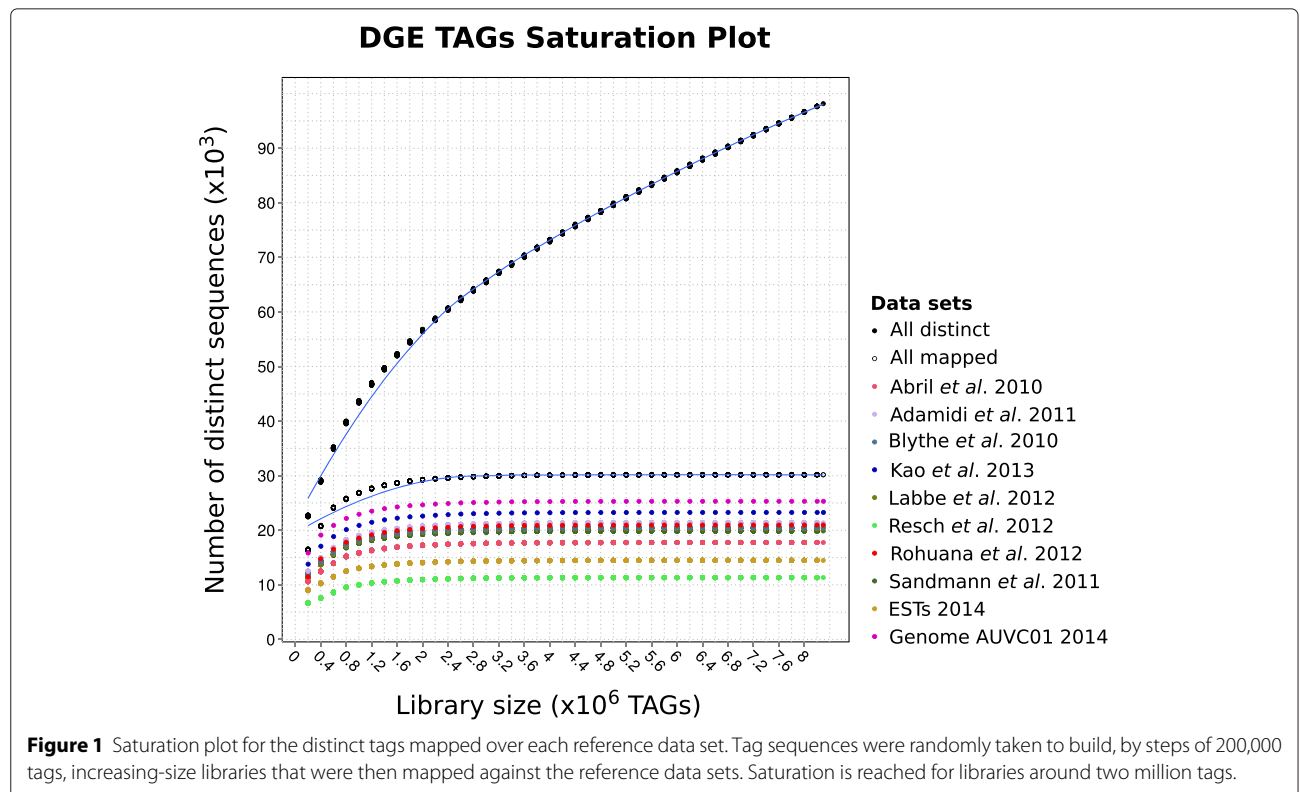
## Results and discussion

Three DGE libraries were obtained from FACS-isolated cell populations X1 (proliferating stem cells, S/G2/M), X2

(a mix of stem cell progeny and proliferating, G0/G1), and Xin (differentiated cells, G0/G1) [30] (Additional file 1). 8,298,210 total reads were sequenced (X1: 3,641,099; X2: 3,488,712; Xin: 1,168,399), representing 98,156 distinct tags (X1: 70,849; X2: 24,621; Xin: 25,221), with an average of 84.5 reads per tag (X1: 51.4; X2: 141.7; Xin: 46.3). The distribution of the tags in each cell population can be observed in Additional file 2A. DGE is reported to achieve near saturation in genes detected after 6-8 million tags [22]. Furthermore, for moderately to very highly expressed genes (>2 cpm) it occurs with three or even just two million tags [22,34]. Figure 1 shows that saturation was reached at around two million tags for most of the data sets which the distinct tags were mapped to, although the slope for the total number of distinct tags decreases without saturating. It is worth noting that all the reference transcriptome sets performed similarly, achieving a maximum near 20,000 mapped tags. However, when looking at how many distinct tags map to any of those transcriptomes, about 5,000 tags appear not to be shared among all of them (see the "All mapped" and the "All distinct" data series on Figure 1, and further details on mapping below).

A critical point in this kind of experiment has to do with the number of times a tag has to be seen so that it can be considered reliable. Discarding too many tags in an attempt to increase reliability will result in a loss of information whereas keeping all of them may generate background noise. To estimate the specificity of our tags and to establish an optimal cutoff for the minimum number of counts a tag should have in order not to be considered artefactual, we performed a series of simulations mapping iteratively randomized sets of our data. The results are summarized in Additional file 3 for the different cutoffs tested (1, 5, 10, 15 and 20 minimum occurrences of tags). For cutoffs higher than five there is no substantial gain in terms of specificity (the number of hits decreases less than one order of magnitude). Thus, we defined reliable tags as those sequenced five times or more and discarded the rest. Thereafter, for the subsequent computational and experimental analyses, only those tags occurring at least five times were considered. From the initial set of 98,156 distinct tags, 40,670 passed that cutoff (Additional file 2B).

The low technical variability of DGE and its high reproducibility, together with the digital quantification of transcripts, enables direct comparison of samples across different experiments, even from different laboratories [21,22,24-26,29,35]. That property allowed us to contrast our results with those from Galloni [36], who used DGE to identify neoblast genes by comparing irradiated versus control animals over the same strain of clonal *S. mediterranea*. A Venn diagram showing the similarity of the strategies can be seen in Additional file 4. From the total distinct tags, 31.38% (30,806 out of 98,156) were sequenced 10 times or more in our study, compared with



just 11,28% (42,159 out of 373,532) in the irradiation strategy, indicating a greater representation of each tag. This suggests, as expected, that the cell-sorting approach has higher specificity. In addition, the strand-specific nature of DGE allows the discrimination of sense and antisense transcripts. Almost 30% of the transcripts successfully identified also presented antisense transcription, even though at lower levels than canonical transcription. This confirms the findings of the aforementioned study in planarians [36] and others [37], and shows that a large proportion of the genome is transcribed from both strands of the DNA. Although the purpose of these transcripts is still open to debate, evidences point to a post-transcriptional gene regulatory function [38].

#### Tag mapping to reference sequence data sets

An essential step in DGE is the recovery of the transcript represented by each tag. The nature of the DGE methodology, which generates reads of only 21 nucleotides, implies mapping short reads against a reference genome or a collection of ESTs to retrieve full-length sequences for the original transcripts. On the other hand, the short length facilitates the fast mapping of the tags against the reference sequence data set. To obtain the maximum number of transcripts, tags were mapped against the 94,876 *S. mediterranea* ESTs from the NCBI dbEST [39-42] and all the available transcriptomes (formally

those can also be considered as ESTs libraries). 26,822 tags (65.95%) mapped over at least one set of ESTs/transcripts, leaving a huge number (34.05%) unmapped.

In an attempt to recover tags that did not map over the transcripts, tags were also mapped over the *S. mediterranea* genome assembly draft AUVCO1 masked with the *S. mediterranea* repeats [23,43-45] (Table 1 and Figure 2). The overlap between transcriptomes was high. Although in most cases sets of reads mapping over a single transcriptome has a very low incidence, there were two cases where one could find a relatively small number of tags mapping to only one transcriptome: 327 tags (1.1%) for Labbé et al. 2012; 208 tags (0.7%) for Rohuana et al. 2012; 3,231 tags (10.7%) remarkably mapping only over the genome; and 26.1% of tags (10,617 out of 40,670) not mapping at all. For tags sequenced 10 times or more, the proportion of unmapped tags is similar: 20.5% (6,327 out of 30,806) (Additional file 2B). Even allowing up to two mismatches, 9.36% of the reads remain not mappable to the genome. This is still an important amount, considering that two mismatches is very permissive (it represents almost a 10% of nucleotide substitution in the read with respect to the reference sequence).

These results indicate that there will be a significant number of transcripts that are not represented yet neither in the current transcriptomic sets nor in the reference

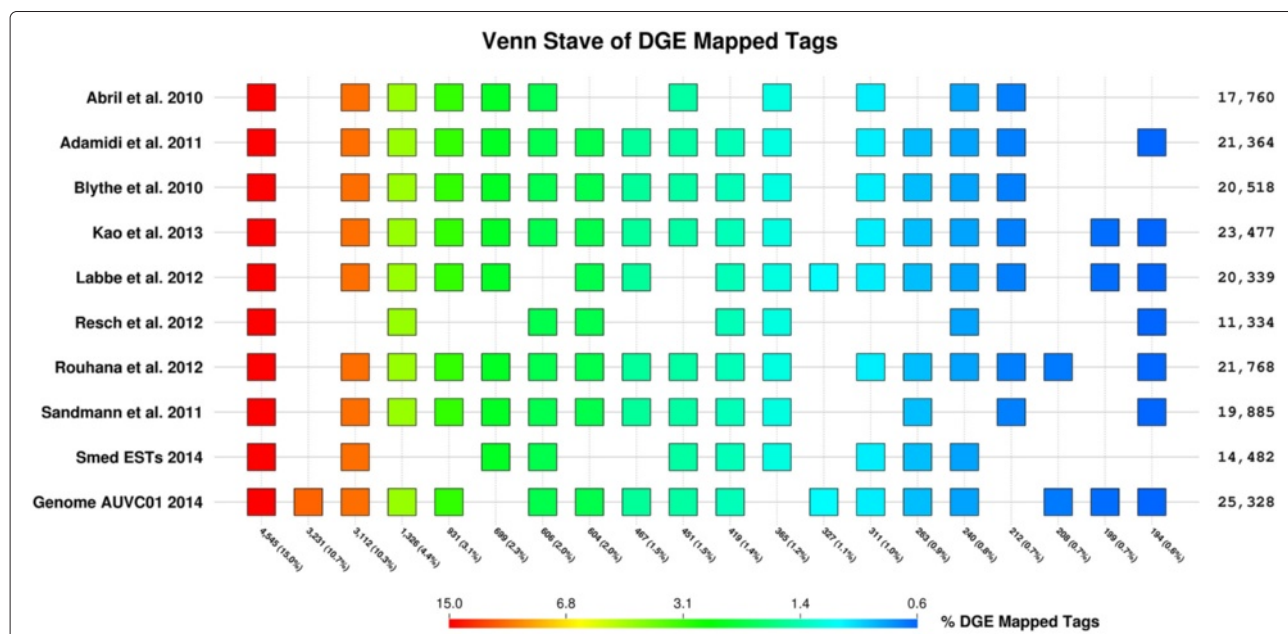
**Table 1 Summary of mapped tags**

Reference	Mapped	One match	More than one match	Orphan	Contigs per tag
Abril et al. 2010	17,760	12,848	4,912	22,910	1.616
Adamidi et al. 2011	21,364	18,024	3,340	19,306	1.282
Blythe et al. 2010	20,518	17,649	2,869	20,152	1.204
Kao et al. 2013	23,477	15,791	7,686	17,193	1.444
Labbé et al. 2012	20,339	19,513	826	20,331	1.040
Resch et al. 2012	11,334	9,789	1,545	29,336	1.158
Rouhana et al. 2012	21,768	14,891	6,877	18,902	1.579
Sandmann et al. 2011	19,885	14,774	5,111	20,785	1.407
ESTs 2014	14,482	3,650	10,832	26,188	5.442
Genome AUVC01 2014	25,328	19,019	6,309	15,342	1.272

Counts for the tags mapping over the reference data sets depicted in Figure 2. Total (distinct) tags: 40,670; mapped tags: 30,053; orphan tags (tags not mapped): 10,617.

genome, despite their coverage depth [46-49], and may correspond, for instance, to weakly expressed genes [50]. Mapping tags are expressed on average at 50.78 cpm, while non-mapping tags only at 19.85 cpm. Nonetheless, since transcriptomes currently available lack the complete annotation of 3'-UTR regions and the DGE libraries were made from the 3'-ends, reads that map to genomic sequences but not to current transcripts may potentially come from the 3'-UTR ends not yet sequenced. To evaluate this possibility, we have projected the transcriptome from Kao et al. 2013 [17] over the genome and looked

for the proximity of the tags mapping next to the 3'-end of the transcripts (Additional file 5). Downstream sequenced DGE tags account for 4.12% of all the possible CATG targets. This small amount of sequenced tags only mapping to the genome may correspond to potential novel unsequenced transcripts, alternative 3'-UTR exons of splicing isoforms, misannotated or alternative poly-adenylation sites, or even to non-coding RNAs not represented yet in the present transcriptome sets. Future RNA-Seq experiments may provide further sequence evidences supporting transcripts for those tags.



**Figure 2** Venn stave showing the proportions of the distinct tags mapped over the different reference data sets. Integrating data for Venn diagrams for sets larger than four or five can be a challenging task, so that, a linear projection of such a diagram is provided in the stave—showing the 20 topmost scoring comparisons from 752 different subsets, accounting for 62.26% (18,710 out of 30,053) of total mappings—for ten reference sequence sets: eight transcriptomes, the *S. mediterranea* ESTs from NCBI dbESTs [39-42], and the latest genome draft AUVC01 [43,44]. Color gradient scale is provided on the bottom bar and it is proportional to the number of unique tags mapped over each sequence subset. X-axis ticks present the number of tags and their relative percent; the numbers on the right Y-axis correspond to the total number of tags mapped into a given sequence sets comparison. It is easy to spot that 15% of the unique reads are mapping onto all the sequence sets.

### Functional annotation

8,903 contigs from Smed454\_90e—Smed454 from now on—[10] showing significant expression changes ( $p < 0.001$ ) were selected and, from those, 7,735 contigs presented a hit to a Pfam domain model (Figure 3). For those sequences having a significant hit to a known domain/protein, gene ontology (GO) analysis was performed in order to summarize changes on the biological processes and molecular functions due to the observed expression patterns of the enriched sets of transcripts. Those transcripts were classified according to the cell type in which they were mostly expressed, then their significant GO annotations were clustered (also taking into account their parent nodes in the ontology), to calculate the terms abundance log-odds ratio. Comparison of GO categories between transcripts predominantly expressed in X1, X2 or Xin cell fractions revealed significant patterns of enrichment as indicated in Additional file 6 (see also the “Transcriptomes” tables available from the web site—[planarian.bio.ub.edu/SmedDGE](http://planarian.bio.ub.edu/SmedDGE)—for specific GO terms assigned to each transcript).

The GO comparison between the neoblast population (X1) and the differentiated cells (Xin) reflects distinct functional signatures: X1 is enriched in ubiquitin-dependent protein catabolic process, nucleic acid binding, RNA-binding, helicase activity, ATP binding, translation, and nucleosome assembly; Xin most represented categories include actin binding, actin cytoskeleton organization, small GTPase mediated signal transduction, proteolysis, and calcium ion binding; whereas in X2, markers of secretory activity such as vacuolar transport are more abundant.

### Browsing data

All tag mappings over the different transcriptome versions are available in the form of dynamic tables from our web site ([planarian.bio.ub.edu/SmedDGE](http://planarian.bio.ub.edu/SmedDGE), Figure 4A). The relationship between Smed454, along with their domains and functional annotation, with the other reference transcriptomes described in this manuscript can be browsed on a subset of those tables. In order to establish the correspondence between the transcriptomes, a megablast—NCBI BLAST+ 2.2.29 [51]—was performed, filtering the resulting hits afterwards by three levels of coverage (90%, 95% and 98%). Although the focus is set on Smed454, the user can reorder those tables by columns containing identifiers for other transcriptome versions or she can choose to jump to the transcriptome version specific summary table.

Moreover, the Smed454 contig browser [10,52] has been revamped into a more flexible interface based on GBrowse2 ([planarian.bio.ub.edu/gbrowse/smed454\\_transcriptome](http://planarian.bio.ub.edu/gbrowse/smed454_transcriptome)). One can find there different types of annotation tracks: reads coverage, homology to known

genes/proteins, hits to Pfam domains, and also the information of the tags mapped over the sequence. One track-specific GBrowse2 Perl module was modified to display DGE tags data, such as the sequence, counts and rank position. Further customization of the GBrowse2 configuration facilitates the access to most of that information in the form of pop-up summary boxes, but also by means of additional “Details” page (see yellow panel on the right side of Figure 4B).

This browser has been developed under the principle of easy accessibility, in the hope that it will become a useful and informative user friendly tool for experimental researchers in their daily work.

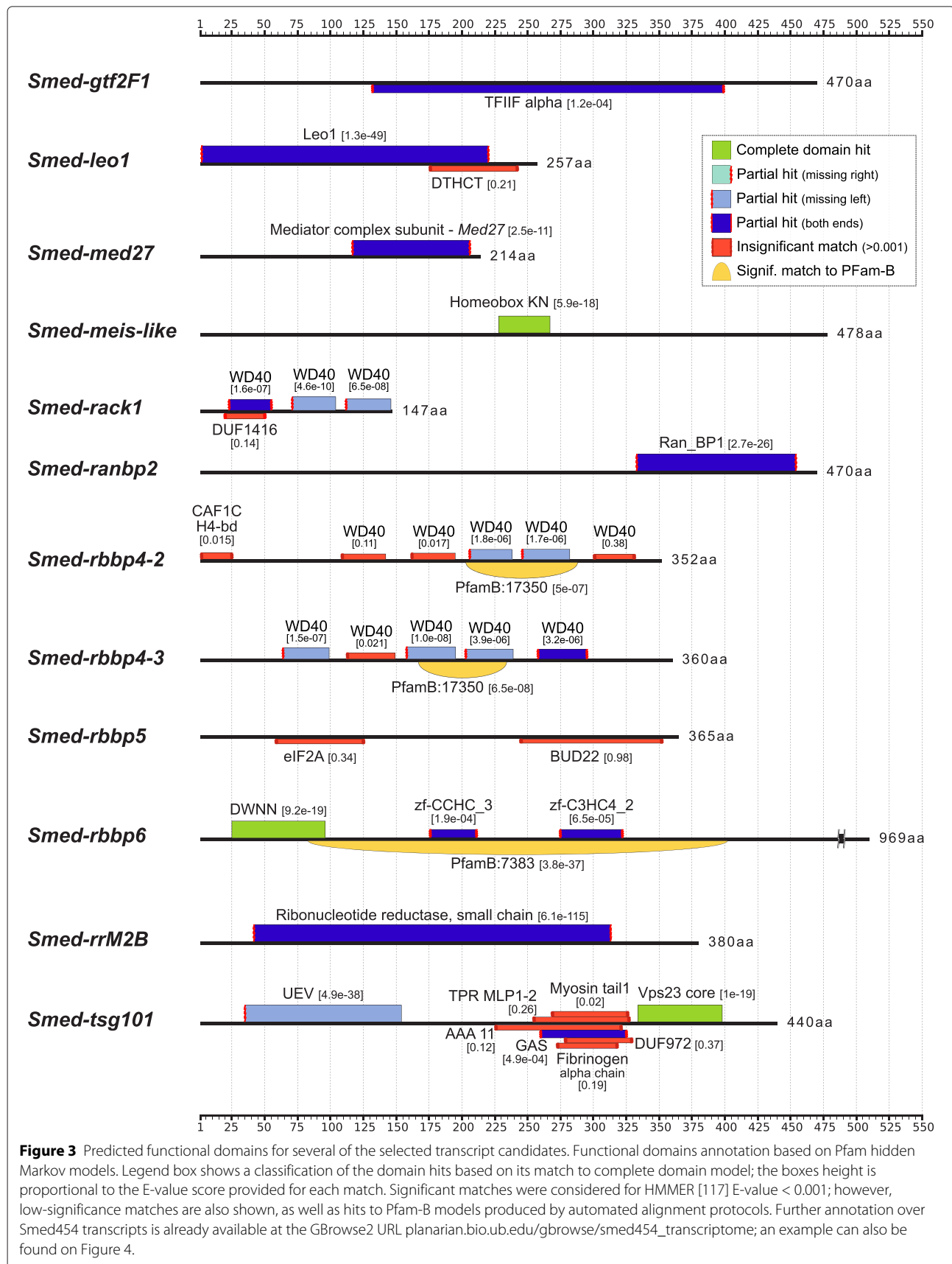
### Experimental validation

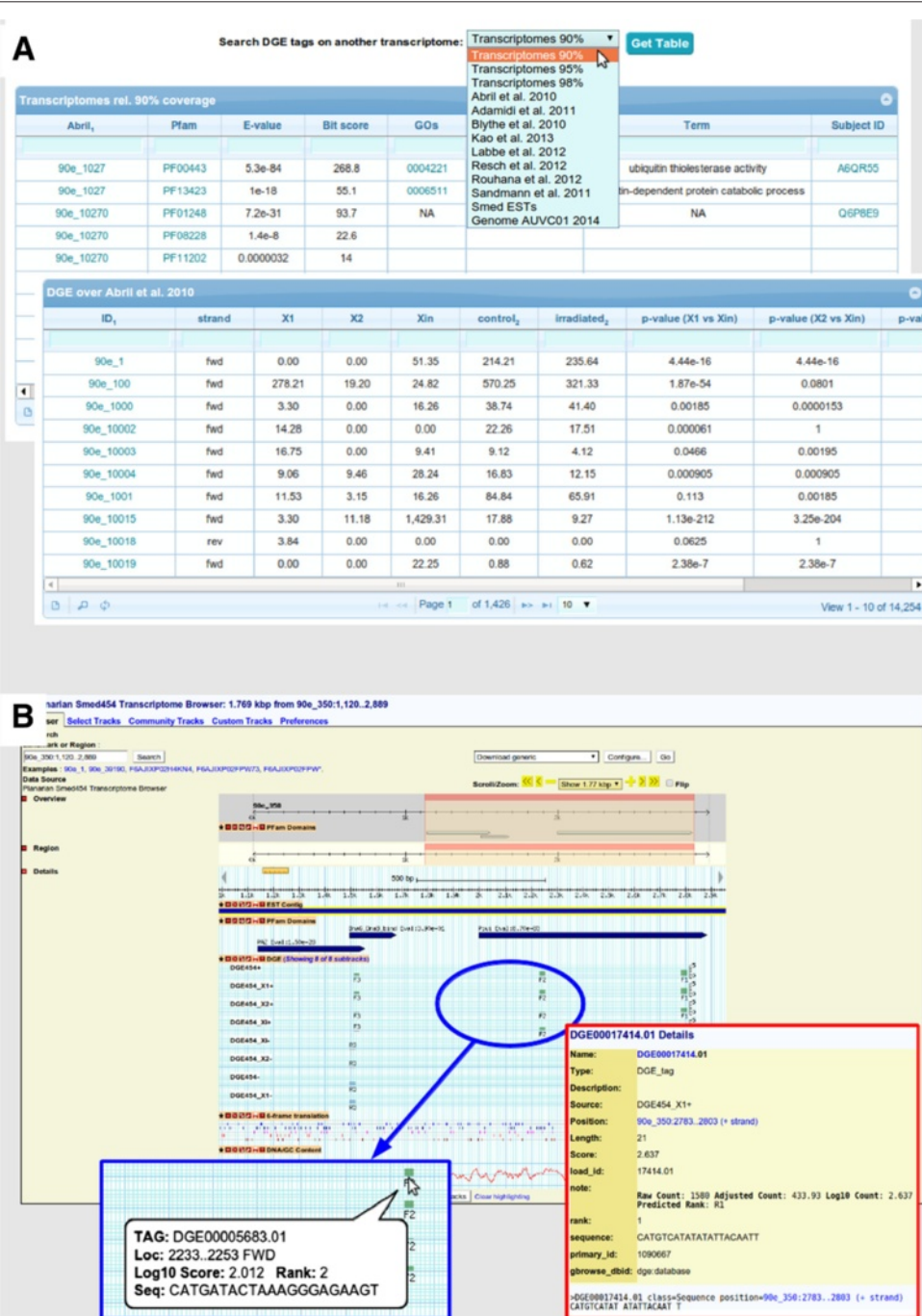
The validity of our approach is corroborated by the expression levels detected in 40 already known and well-characterized neoblast genes (Table 2), plus another 29 genes described in the literature with evidence of also being neoblast related (Table 3). As can be observed in Figure 5, both sets of genes show the expected expression pattern along the vertical right hyperbola, indicating a clear X1 specificity, with two exceptions overrepresented in X2: *Smed-nlk-1* and *Smed-prog-1*, which is described to be found in postmitotic cells [53]. *Smed-dlx* and *Smed-sp6-9* are key genes in eye formation [54]; despite their localized activation, DGE was sensitive enough to identify both of them predominantly in the X1 subfraction. Moreover, we could detect expression of genes such as *Smed-smg-1*—which is described as broadly expressed through all tissues, including neoblasts [55]—in both neoblasts and differentiated cells. Finally, 133 clones from two different studies [6,56] focussing on regeneration, stemness and tissue homeostasis are, indeed, significantly overexpressed in neoblasts (Additional file 7).

Based on their X1/Xin expression ratio, we selected a collection of potential new neoblast genes among the most represented in the X1 population. With the chosen candidates we performed expression pattern analysis by whole mount in situ hybridization (WISH) in irradiated animals. At different times after irradiation, as the neoblasts and its progeny decline, the hybridization signal disappears [57]. The expression of 42 out of 47 genes tested was diminished or completely lost in irradiated animals (Table 4 and Additional file 8).

Although neoblasts are essential also during homeostasis for normal cell renewal, the phenotype becomes more evident during regeneration. Functional analyses were therefore carried out by RNAi followed by head and tail amputation in order to visualize defects in the regenerating process. From the 42 genes whose expression was affected by irradiation, 24 showed a phenotype after RNAi (Additional file 9), most of them preventing a successful







**Figure 4** Online data sets and DGE data on Smed454 GBrowse2. **A** - To facilitate browsing of mapped tags over the transcripts we have worked with, we provide a dynamic table interface that paginates through the huge lists of records. This jQuery [112] interface allows the user to easily sort the output table by a given column—just by clicking on the column label—or to search for specific values on the cells—using either the form box just below the column labels or the advanced search available from the magnifying glass icon at the bottom of the table. Three tables, like the one in the background, contain the equivalences between contigs from different transcriptomes, as well as functional annotations, always focusing on the Smed454 data set. The other tables, like the one in the foreground, contain the tag mappings for each single transcriptomes considered to date. **B** - Previously published Smed454 database [10] has been ported to GBrowse2 in order to facilitate navigating through the transcripts annotations, such as predicted domains from Pfam, assembly reads mapping, etc. This panel shows the annotations on *Smed-wi-3* homologous contig as an example. A customized track allows the integration of information about mapped DGE tags into single or combined tracks; tags are represented as boxes with height proportional to log of the normalized tag counts, the rank and the strand for the tag hit are shown in the label just below that box. Bottom left blue box zooms into one of those combined tracks to visualize the pop-up box that the user can recover when moving the mouse over a given tag feature. In addition, bottom right red box displays the details page one can get when clicking on a tag feature.

**Table 2 Neoblast genes**

Gene	X1	X2	Xin	p-val X1-Xin	p-val X2-Xin	Accession	PubMed
<i>Smed-bruli</i>	212.57	122.68	0	2.20e-062	1.58e-035	DQ344977	16890156
<i>Smed-chd4</i>	159.84	18.34	13.69	5.91e-032	1.10e-001	GU980571	20223763
<i>Smed-coe</i>	10.16	0	0	9.77e-004	1	KF487109	25356635
<i>Smed-cycD</i>	18.95	0	0	1.91e-006	1	JX967267	23123964
<i>Smed-dlx</i>	5.22	0	0	3.12e-002	1	JN983829	21852957
<i>Smed-e2f4-1</i>	141.72	23.50	29.96	6.24e-018	7.79e-002	JX967265	23123964
<i>Smed-egr-3</i>	19.50	0.57	0	1.91e-006	5.00e-001	HM777016	21458439
<i>Smed-egr-1</i>	510.01	37.26	153.20	9.16e-045	3.06e-017	JF914965	21846378
<i>Smed-foxA</i>	15.65	0	0	1.53e-005	1	JX010556	24737865
<i>Smed-hdac-1</i>	1086.49	0	60.77	4.19e-122	4.34e-019	JX967266	23123964
<i>Smed-hnf4</i>	30.21	8.31	8.56	3.85e-004	1.85e-001	JF802199	21566185
<i>Smed-hsp60</i>	113.43	10.32	33.38	8.64e-011	2.18e-004	GU591874	21356107
<i>Smed-hsp70</i>	326.28	0	11.13	3.98e-081	4.88e-004	GU591875	21356107
<i>Smed-jnk</i>	87.61	13.47	11.98	8.29e-016	1.55e-001	KC879720	24922054
<i>Smed-lst8</i>	43.12	1.43	0	1.14e-013	5.00e-001	JN815261	22479207
<i>Smed-msh2</i>	57.13	2.58	0	6.94e-018	1.25e-001	JF511467	21747960
<i>Smed-nanos</i>	39.27	1.15	0	1.82e-012	5.00e-001	EF153633	17390146
<i>Smed-ncoa5</i>	48.34	30.38	0	1.46e-011	5.96e-008	KF668097	24268775
<i>Smed-nf-YB</i>	11.26	2.58	0	4.88e-004	1.25e-001	HM100653	20844018
<i>Smed-p53</i>	5.22	5.73	0	3.12e-002	1.56e-002	AY068713	12421706
<i>Smed-papbc</i>	46.96	0	0	7.11e-015	1	HM100651	20844018
<i>Smed-pbx</i>	226.03	38.12	19.69	1.17e-044	6.41e-003	KC353351	23318635
<i>Smed-pcna</i>	728.63	24.08	0	3.51e-217	5.96e-008	EU856391	18786419
<i>Smed-prmt5</i>	43.67	0.57	0	5.68e-014	5.00e-001	JQ035529	22318224
<i>Smed-prog-1</i>	1.92	389.54	37.66	7.09e-010	7.42e-074	JX122762	18786419
<i>Smed-runt-1</i>	16.48	0	0	1.53e-005	1	JF720854	21846378
<i>Smed-sd-1</i>	14.28	0.57	0	6.10e-005	5.00e-001	KF990481	24523458
<i>Smed-sd-2</i>	4.67	0	0	3.12e-002	1	KF990482	24523458
<i>Smed-smB</i>	461.12	0	29.96	1.72e-099	9.31e-010	GU562964	20215344
<i>Smed-smg-1</i>	72.78	11.47	26.53	1.51e-006	4.38e-003	JF894292	22479207
<i>Smed-soxP-1</i>	15.11	3.15	0	3.05e-005	1.25e-001	JQ425151	22385657
<i>Smed-sp6-9</i>	38.72	0.57	0	1.82e-012	5.00e-001	JN983830	21852957
<i>Smed-srf</i>	40.37	0.29	16.26	5.78e-004	1.53e-005	JX010474	22549959
<i>Smed-tert</i>	19.22	0	0	1.91e-006	1	JF693290	22371573
<i>Smed-tor</i>	31.86	0	10.27	3.35e-004	9.77e-004	JF894291	22479207
<i>Smed-vasa-1</i>	1209.52	22.93	22.25	3.39e-162	1.17e-001	JQ425140	22385657
<i>Smed-wi-1</i>	644.59	13.47	0	6.01e-192	1.22e-004	DQ186985	16311336
<i>Smed-wi-2</i>	724.78	50.45	26.53	1.41e-176	2.90e-003	DQ186986	16311336
<i>Smed-wi-3</i>	433.93	76.82	21.40	9.76e-101	4.01e-009	EU586258	18456843
<i>Smed-xin-11</i>	26.64	0	0	7.45e-009	1	DQ851133	17670787

X1, X2 and Xin DGE expression levels of already known and deeply characterized neoblast genes.

**Table 3 Likely neoblast genes**

Gene	X1	X2	Xin	p-val X1-Xin	p-val X2-Xin	Accession	PubMed
<i>Smed-armc1</i>	20.60	2.01	0	4.77e-007	2.50e-001	JQ425158	22385657
<i>Smed-ash2</i>	17.58	0.86	0	3.81e-006	5.00e-001	KC262336	23235145
<i>Smed-cpsf3</i>	19.77	0	0	9.54e-007	1	KJ573358	24737865
<i>Smed-da</i>	13.46	0	0	1.22e-004	1	KF487093	24173799
<i>Smed-eed-1</i>	42.02	0	0	2.27e-013	1	JQ425136	22385657
<i>Smed-ezh</i>	31.03	2.01	0	4.66e-010	2.50e-001	JQ425137	22385657
<i>Smed-fer3l-1</i>	12.36	1.15	0	2.44e-004	5.00e-001	KF487094	24173799
<i>Smed-fhl-1</i>	158.19	8.31	23.11	2.79e-025	3.67e-003	JQ425148	22385657
<i>Smed-hcf1</i>	20.60	0	0	4.77e-007	1	KC262343	23235145
<i>Smed-hesl-3</i>	26.09	0.57	0	1.49e-008	5.00e-001	KF487112	24173799
<i>Smed-junl-1</i>	173.30	4.59	0	1.97e-050	3.12e-002	JQ425155	22385657
<i>Smed-khd-1</i>	29.94	4.87	8.56	3.85e-004	1.22e-001	JQ425142	22385657
<i>Smed-mcm7</i>	351.82	24.08	0	5.22e-104	5.96e-008	KJ573361	24737865
<i>Smed-mll5-2</i>	97.50	15.48	32.52	7.09e-008	3.88e-003	KC262344	23235145
<i>Smed-mrg-1</i>	53.28	3.44	0	1.11e-016	1.25e-001	JQ425133	22385657
<i>Smed-nlk-1</i>	0	30.10	0	1	9.31e-010	JQ425157	22385657
<i>Smed-nsd-1</i>	135.40	8.03	0	4.23e-039	3.91e-003	JQ425134	22385657
<i>Smed-pabp2</i>	191.98	2.58	8.56	2.84e-045	5.37e-002	KJ573359	24737865
<i>Smed-rbbp4-1</i>	121.67	0	0	3.13e-035	1	JQ425135	22385657
<i>Smed-sae2</i>	19.77	6.02	0	9.54e-007	1.56e-002	KJ573350	24737865
<i>Smed-setd8-1</i>	10.99	0.57	0	4.88e-004	5.00e-001	JQ425139	22385657
<i>Smed-soxP-2</i>	37.63	4.01	0	3.64e-012	6.25e-002	JQ425152	22385657
<i>Smed-soxP-3</i>	14.01	14.62	0	6.10e-005	3.05e-005	JQ425153	22385657
<i>Smed-sz12-1</i>	76.90	0	0	6.62e-024	1	JQ425138	22385657
<i>Smed-tcf15</i>	47.51	16.05	0	3.55e-015	1.53e-005	JQ425150	22385657
<i>Smed-vasa-2</i>	491.06	184.02	55.63	6.25e-087	2.39e-016	JQ425141	22385657
<i>Smed-wdr82-2</i>	195.55	16.63	0	2.66e-057	7.63e-006	KC262342	23235145
<i>Smed-zmym-1</i>	180.99	6.31	0	8.05e-053	1.56e-002	JQ425146	22385657
<i>Smed-znf207-1</i>	44.77	3.73	0	2.84e-014	6.25e-002	JQ425147	22385657

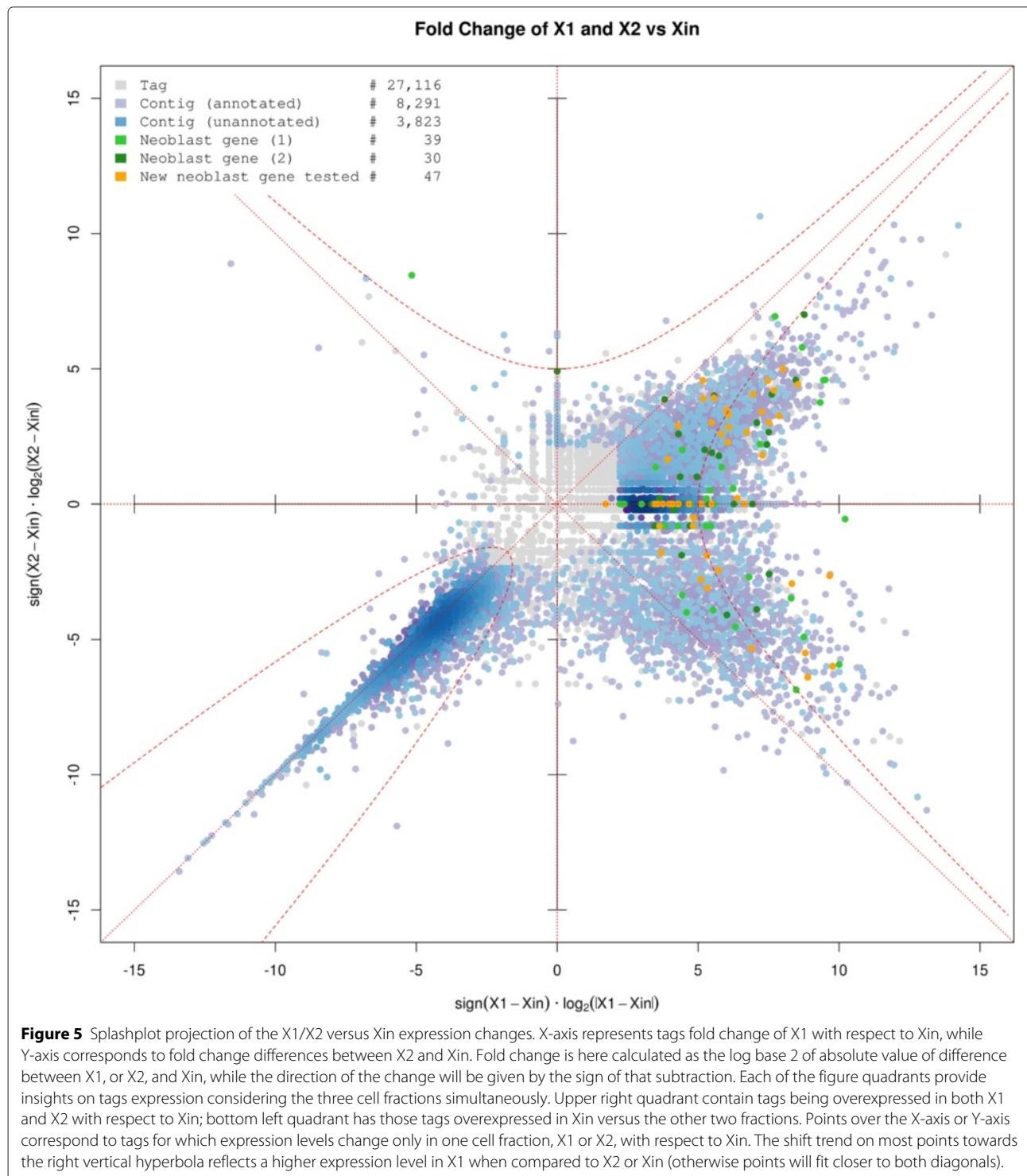
X1, X2 and Xin DGE expression levels of genes described in the literature with some evidences of being neoblast genes.

regeneration and leading to the death of the animals, the usual phenotype for neoblast genes [58,59].

### New neoblast genes

Interestingly, several of the new genes identified as neoblast genes correspond to transcription factors, which are key elements implicated in cell fate decisions. Furthermore, many are also homologous to cancer related genes. We briefly describe those that produce planarian regeneration impairment after RNAi (Additional file 9). The inhibition of six of them produce a reduced blastema with defective head and eyes. *Smed-atf6A*, is a cyclic AMP-dependent transcription factor, which interacts with the Nuclear Transcription Factor Y (NF-Y) complex (further analyzed

later). *Smed-ccar1*, is a perinuclear phospho-protein that functions as a p53 coactivator modulating apoptosis and cell cycle arrest [60]. *Smed-hnrnpA1/A2B1*, a component of the ribonucleosome, is involved in the packaging of pre-mRNA into hnRNP particles in embryonic invertebrate development [61] and in stem cells [62]. *Smed-srrt*, modulates arsenic sensitivity, a carcinogenic compound that inhibits DNA repair [63]. *Smed-med7* and *Smed-med27* belong to a mediator complex essential for the assembly of general transcription factors. *Smed-ranbp2* is a member of the nuclear pore complex and is implicated in nuclear protein import. Within the same family, *Smed-nup50* shows also a stronger phenotype. The knock-down of the other 14 genes prevents the formation of the blastema completely. *Smed-gtf2E1* and *Smed-gtf2F1*, are



components of the general transcription factors IIE and IIF. *Smed-ncapD2* is necessary for the chromosome condensation during mitosis [64]. *Smed-pes1*, is required in zebrafish for embryonic stem cell proliferation [65]. *Smed-rack1*, is an intracellular adaptor of the protein kinase C in a variety of signaling processes. *Smed-lin9*, is related to the

retinoblastoma pathway interacting with Retinoblastoma 1, which is required for cell cycle progression [66]. All six different retinoblastoma binding proteins produce a non-blastema phenotype. The retinoblastoma pathway has been described to regulate stem cell proliferation in planarians [67] and some of its genes are already identified.

**Table 4 New neoplast genes experimentally validated**

Gene	X1	X2	Xin	p-val X1-Xin	p-val X2-Xin	TR	TF	ED	CC	OG	Accession	PubMed
<i>Smed-atf6A (Smed-atf11)</i>	12.36	0.57	0	2.44e-004	5.00e-001	•	•				JX010554	22549959
<i>Smed-ccar1</i>	184.01	79.11	0	1.02e-053	1.65e-024	•		•	•	•	KM981922	
<i>Smed-dnaJA3</i>	133.75	26.94	10.27	2.65e-028	2.53e-003	•				•	KM981923	
<i>Smed-ergic3</i>	66.74	4.87	0	6.78e-021	3.12e-002					•	KM981924	
<i>Smed-got2    Smed-maspat</i>	106.29	7.17	0.86	1.78e-030	3.12e-002					•	KM981925	
<i>Smed-gtf2E1</i>	36.25	14.91	0	1.46e-011	3.05e-005	•	•				KM981926	
<i>Smed-gtf2F1</i>	25.54	0	0	1.49e-008	1	•	•				KM981927	
<i>Smed-hadhB</i>	153.80	12.33	1.71	4.86e-043	5.55e-003					•	KM981928	
<i>Smed-hnmpA1/A2B1</i>	341.93	13.76	21.40	3.70e-075	6.75e-002			•			KM981929	
<i>Smed-leo1 (NBE.6.06A)</i>	377.36	26.66	5.14	3.79e-104	4.69e-005	•		•		•	AY967650	15866156
<i>Smed-lin9</i>	47.51	13.76	2.57	9.25e-012	5.19e-003	•			•	•	KM981930	
<i>Smed-maf</i>	19.50	7.45	0	1.91e-006	7.81e-003	•	•	•			KM981931	
<i>Smed-med7</i>	162.86	11.18	7.70	3.59e-038	1.44e-001	•					KM981932	
<i>Smed-med27</i>	72.51	14.91	5.14	6.99e-017	1.48e-002	•					KM981933	
<i>Smed-meis-like</i>	10.99	0	0	4.88e-004	1	•	•	•		•	KM981934	
<i>Smed-mlx</i>	160.12	0	40.23	1.81e-017	9.09e-013	•	•		•		KM981935	
<i>Smed-ncapD2</i>	84.86	0	0.86	1.11e-024	5.00e-001				•		KM981936	
<i>Smed-nfx1    Smed-stc</i>	28.29	0.57	0	3.73e-009	5.00e-001	•	•	•			KM981937	
<i>Smed-nf-YA</i>	31.31	1.15	2.57	3.48e-007	2.50e-001	•	•	•	•	•	KM981938	
<i>Smed-nf-YB-2</i>	17.03	0	0	7.63e-006	1	•	•	•	•	•	KM981939	
<i>Smed-nf-YC</i>	589.38	97.74	142.93	4.54e-064	1.46e-002	•	•	•	•	•	KM981940	
<i>Smed-nme1    Smed-nm23H1</i>	603.39	45.00	129.24	4.69e-073	6.54e-010			•		•	KM981941	
<i>Smed-nup50</i>	45.32	8.89	0.86	6.54e-013	9.77e-003			•			KM981942	
<i>Smed-pes1 (Smed-pescadillo-1)</i>	228.23	46.15	18.83	5.45e-046	3.34e-004			•	•	•	JX010566	22549959
<i>Smed-rack1</i>	115.90	30.10	0	1.91e-033	9.31e-010			•	•	•	KM981943	
<i>Smed-ranbp2    Smed-nup358</i>	45.32	0.86	0.86	6.54e-013	5.00e-001					•	KM981944	
<i>Smed-rbbp4-2 (Smed-rbbp-1)</i>	100.24	0	0	1.08e-028	1	•				•	JX010613	22549959
<i>Smed-rbbp4-3 (NBE.6.02C)</i>	254.04	27.52	17.97	7.94e-054	4.01e-002	•				•	AY967644	15866156
<i>Smed-rbbp4-4</i>	56.30	6.02	0	1.39e-017	1.56e-002	•			•	•	KM981945	
<i>Smed-rbbp5</i>	43.94	0.57	4.28	6.91e-010	1.56e-001	•		•		•	KM981946	
<i>Smed-rbbp6</i>	64.27	11.18	0	5.42e-020	4.88e-004					•	KM981947	
<i>Smed-rM2B (Smed-rM2-2)</i>	826.40	7.45	13.69	1.07e-111	5.54e-002				•	•	JX010501	22549959



Despite that, most of them are yet to be analyzed. Finally, *Smed-rrM2B*, is a subunit of the ribonucleotide reductase (RNR) complex required for DNA repair [68]. Details on these genes as well as the rest of the genes tested from the X1 population can be examined in the Additional file 10.

The four remaining genes presenting an aberrant phenotype during regeneration when inhibited by RNAi are described in detail in the following two sections: the *Smed-meis-like*, a new member of the Meis family, and the three components of the Nuclear Factor Y complex, all of them found to be overexpressed in neoblasts.

### ***Smed-meis-like***

*Smed-meis-like* is a member of the TALE-class homeobox family, similar to Meis genes, which was found to be overexpressed in the X1 subpopulation. This gene family is characterized by the presence of a homeobox domain with three extra amino acids between helices 1 and 2 [69]. Some of its members can act as cofactors for *Hox* genes [32]. In *S. mediterranea*, other members of the family have been described: *Smed-prep* [70], *Smed-meis* [54] and *Smed-pbx* [71,72].

WISH on intact animals shows that it is expressed in the cephalic ganglia, the pharynx, the tip of the head, and the parenchyma (Figure 6A). The downregulation observed three days after irradiation suggests that the parenchyma-associated expression is related to neoblasts and early postmitotic cells. To corroborate this, a double fluorescence in situ hybridization (FISH) together with the neoblast marker *Smed-h2b* [59] has been carried out (Figure 6B and Additional file 11A). Confocal microscopy shows colocalization of both genes in some cells, which confirms the expression of *Smed-meis-like* in neoblasts and, thus, the DGE results. Nevertheless, not all *Smed-meis-like* positive cells are expressing *Smed-h2b*, reinforcing the idea that *Smed-meis-like* is not exclusive of neoblasts.

Knockdown of *Smed-meis-like* through RNAi produced a diverse range of anterior regeneration phenotypes (Figure 6C), which can be explained by a different penetrance. The mildest phenotype produced a squared head with elongated and disorganized eyes. This phenotype was also clearly visible with fluorescence in situ hybridization (FISH) against *Smed-opsin* [5] and *Smed-tph* [73], which label the photoreceptor and the pigment cells of the eye (Figure 6D). In an intermediate phenotype, cyclopic animals are obtained, whereas in the strongest one there is no anterior blastema formation. This range of phenotypes can also be observed with the marker of brain branches *Smed-gpas* [74], which shows a gradual reduction of brain regeneration after *Smed-meis-like* inhibition. These results are also confirmed by the reduction of the brain signal of the pan-neural marker  $\alpha$ -SYNAPSIN (Additional file 11B). Posterior regeneration was normal.

In the strongest phenotype, there is also no expression of the anterior markers *Smed-notum* [75] and *Smed-sfrp-1* [76,77], and the marker of sensory-related cells *Smed-cintillo* (Figure 6E) [78]. This indicates that *Smed-meis-like* is necessary for anterior identity. In contrast, expression of the posterior marker *Smed-wnt-1* [77] remains after *Smed-meis-like* inhibition. Thus, we can conclude that *Smed-meis-like* is necessary for anterior, but not for posterior regeneration.

Finally, immunohistochemistry against H3P (Figure 6F) shows a slight—but significant—decrease in proliferation in the whole animal ( $133.8 \pm 5.22$  mitosis/mm<sup>2</sup> in  $n=9$  controls versus  $94.6 \pm 4.06$  cells/mm<sup>2</sup> in  $n=9$  *Smed-meis-like*(RNAi), mean  $\pm$  s.e.m.). This decline in mitosis is matched by the lack of progenitors of some anterior structures, indicating also defects in differentiation. Thus, eye progenitor cells, which are labeled with *Smed-ovo* [54], are not present in *Smed-meis-like*(RNAi) animals (Figure 6E).

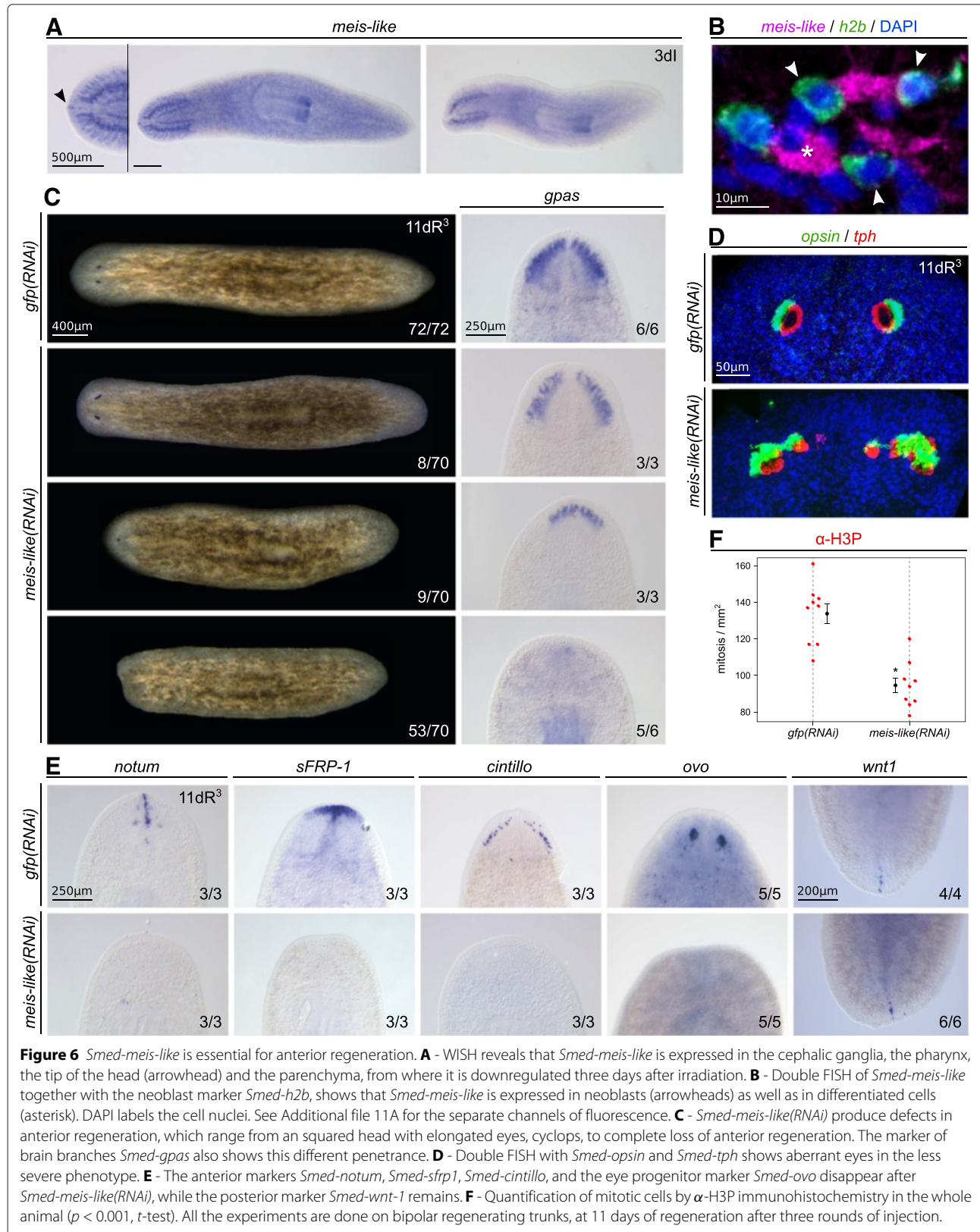
The requirement for *Smed-meis-like* in anterior regeneration is similar to another member of the family, *Smed-prep* [70]. This differential phenotype is also observed after the inhibition of other genes, such as *Smed-egr4* [79], *Smed-zicA* [80,81] and *Smed-FoxD* [82]. The milder phenotype, showing elongated eyes, is similar to the effect of *Smed-meis*(RNAi) [54], and also to the mild inhibition of *Smed-bmp4* [83]. Altogether, these results suggest that *Smed-meis-like* is important for eye and anterior regeneration, similarly to other members of the TALE-class homeobox family. However, given the lack of expression of *Smed-meis-like* in the eyes, the abnormal eye formation could be a consequence of the anomalous brain regeneration.

### **Nuclear Factor Y complex**

The Nuclear Factor Y complex (NF-Y) is an important transcription factor composed by three subunits (NF-YA, NF-YB and NF-YC), each one encoded by a different gene. This heterotrimeric complex acts as both an activator and a repressor, and it regulates other transcription factors, including several growth-related genes, through the recognition of the consensus sequence CCAAT localized in the promoter region [84–88]. In addition, it has been reported that the NF-Y complex regulates the transcription of many important genes like *Hoxb4*, *y-globin*, *TGF-beta receptor II*, or the *Major Histocompatibility Complex class II* and *Sox* gene families [89]. This large number of interactions makes the NF-Y complex an important mediator in a wide range of processes, from cell-cycle regulation and apoptosis-induced proliferation to development and several kinds of cancer [90].

In the sexual strain of *S. mediterranea*, an NF-YB is necessary to maintain spermatogonial stem cells [91]. We have isolated a different NF-YB subunit (NF-YB-2), and also a member of the other two subunits (NF-YA and





NF-YC). WISH shows that the three genes are expressed ubiquitously and in the cephalic ganglia (Figure 7A). Moreover, the expression decrease one day after irradiation indicating a linkage with stem cells, as described in other organisms [92]. Double FISH of each NF-Y subunit together with *Smed-h2b* confirms the expression of this complex in neoblasts and also in some determined cells (Figure 7B and Additional file 12A).

It has been suggested that each NF-Y component could have a specific role [93]. Therefore, to better understand the function of this complex, we knocked down each subunit separately. Although the penetrance varies depending on the subunit inhibited, the phenotype observed after RNAi treatment is the same. In intact non-regenerating animals, RNAi resulted in head regression, ventral curling and, finally, death by lysis (data not shown), as described for other neoblast-related genes [58,59]. After 11 days, head and tail amputated animals failed to regenerate properly, with a smaller brain and fewer brain ramifications as revealed by *Smed-gpas* (Figure 7C) and by  $\alpha$ -SYNAPSIN (Additional file 12B). Furthermore, we observe an increase in the number of *Smed-h2b*<sup>+</sup> cells (Figure 7C,E), also in the area in front of the eyes, where there should not be undifferentiated neoblasts, even though mitosis are reduced (Figure 7D). There is also a decrease in the number of early postmitotic cells (*Smed-nb.21.11e*<sup>+</sup>) (Figure 7C,E), whereas late postmitotic cells (*Smed-agat-I*<sup>+</sup>) do not present significant differences (Figure 7E) [53]. These early progeny markers have recently been associated with epidermal renewal [94]. Hence, the accumulation of neoblasts and the decrease of the subepidermal postmitotic population suggest a defect in the early stages of the differentiation process affecting the epidermal lineage. The neural lineage may also be compromised according to the atrophied cephalic ganglia.

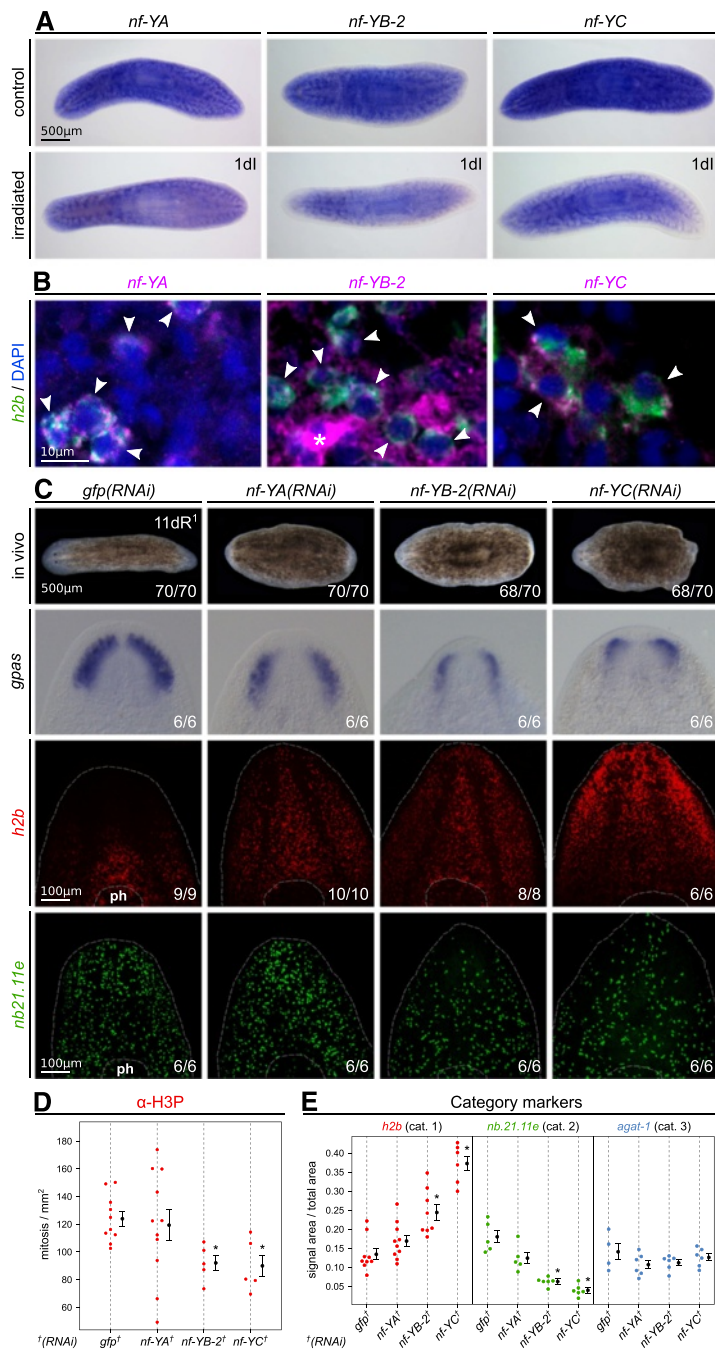
## Conclusions

This work presents experimental validation of a collection of putative neoblast genes obtained from a DGE assay on cell fractions. As clearly depicted in the splashplot for the comparison of expression levels between X1, X2 and Xin fractions (Figure 5 and Additional file 13A), there are only a few transcripts specific to X2. The plot produced with the data provided by Labbé [14] from their RNA-Seq analysis on X1, X2 and Xin cell fractions for *S. mediterranea* shows a similar pattern (Additional file 13B). Moreover, comparison among the three sets using Pearson and Spearman correlations indicates that X1 and X2 are the most correlated populations (Additional file 14). Following these results, most of the transcripts expressed in X2 are also expressed in X1. Hence, X2 is a heterogeneous population that cannot be transcriptionally differentiated from X1 without a deeper discrimination method. In this regard, the strategy recently applied by van Wolfswinkel

and collaborators using the last sequencing technology to obtain the transcriptome of individual cells [94], represents the most promising approach to deciphering the heterogeneity of the neoblast progeny.

Randomization simulations also illustrate the specificity of the 21bp tags to detect real transcripts, corroborating previous estimations [29,46,48,49,95,96]. Furthermore, those results reinforce the assumption that most of the non-mapping tags will correspond to real transcripts [46-49], still lacking from reference data sets for this species. Antisense transcription was also detected, confirming previous reports [25,36,49]. Although further analysis will be required to determine whether this could explain a fraction of the “novel” tags, our primary focus was to characterize the canonical protein-coding transcripts. Due to the heterogeneity of this species genome, we would expect some variability—both at sequence and expression arising from individuals (the pool of animals taken for the samples), and cells (as they do not come from a cell culture). This could explain another fraction of tags not mapping onto the reference transcriptomes. Consequently, we were quite strict in the current manuscript to look for exact tag matches, taking into account that one or more mismatches represents a mappability issue even for finished transcriptomes of the quality of human [97] or *Drosophila melanogaster* [98].

DGE has proven to be reliable for transcript quantification and new gene identification in planaria. In this work, we have described a new member of the TALE-class homeobox family, *Smed-meis-like*. Similar to other members of this family, this gene seems to be involved exclusively on anterior polarity determination during regeneration. Given that the expression of this gene is not restricted to neoblasts, its role can also be important in committed cells. Our results with the NF-Y complex suggest that the knockdown of this complex blocks early differentiation of the epidermal and, probably, neural lineages, both belonging to the ectodermal line, generating a neoblast accumulation and deregulation. This effect has been shown in other organisms such as *Drosophila*, in which NF-Y knockout blocks differentiation of R7 neurons through *senseless* [89,99]. The majority of the new neoblast genes reported and validated in this study were found to participate in cell proliferation, cell cycle regulation, embryogenesis or development in other models, and many of them are involved in processes related to cancer. The pathways participating in tumorigenic processes and stem cell regulation are often the same, as has been proposed previously for planarians [100]. These genes are probably fundamental for stem cell maintenance and the control of proliferation in organisms with the capacity to regenerate [101], thus reinforcing the potential value of *S. mediterranea* as an in vivo model for stem cell research [102].



**Figure 7** *Smed-nf-Y* gene complex is required for the proper neoblast differentiation and localization. **A** - WISH shows that the three *Smed-nf-Y* genes are expressed ubiquitously and in the cephalic ganglia, and one day after irradiation their expressions decrease. **B** - Double FISH of *Smed-nf-YA*, *Smed-nf-YB-2*, and *Smed-nf-YC* together with the neoblast marker *Smed-h2b* shows colocalization with the NF-Y subunits (arrowheads), demonstrating the expression of this complex in neoblasts as well as in differentiated cells (asterisk). DAPI labels the cell nuclei. See Additional file 12A to check each channel of fluorescence separately. **C** - *Smed-nf-Y(RNAi)* animals regenerate thinner blastemas with non well formed eyes and shape defects, and fail to differentiate a proper brain, with reduced cephalic ganglia as revealed with *Smed-gpas*. FISH with the neoblast marker *Smed-h2b* shows an accumulation of neoblasts in the region in front of the eyes while the early progeny marker *Smed-nb.21.11e* reveals a decrease of early postmitotic cells in *Smed-nf-Y(RNAi)* animals. **D** - Immunohistochemistry with the mitotic marker  $\alpha$ -H3P shows a reduction in the number of mitosis. **E** - Quantification with category markers indicate a significant increase of *Smed-h2b*<sup>+</sup> cells in *Smed-nf-YB-2(RNAi)* and *Smed-nf-YC(RNAi)* animals and a significant decrease of *nb.21.11e*<sup>+</sup> cells in all of the RNAi animals, whereas *Smed-agal-1*<sup>+</sup> cells do not show significant changes ( $p < 0.001$ , *t*-test). Counts are referred to the whole body. ph: pharynx. All the experiments are done on bipolar regenerating trunks, at 11 days of regeneration after one round of injection.

Our DGE analysis pointed out a high resemblance among all the transcriptomes available for *S. mediterranea*. We have also shown the redundancy of the transcriptomes currently available for *S. mediterranea* in agreement with Kao [17], together with their incompleteness under the light of the DGE data. Although our results provide a comprehensive comparison among them, it would be desirable to agree on a unique transcriptome to be used by the whole community. To this end, the PlanMine initiative [103] is attempting to obtain consensus among the researchers on an appropriate reference. Nonetheless, the need for a completely sequenced and well-annotated genome remains. The DGE strategy can help in this endeavour, since short sequences can be rapidly projected over the reference genome or the transcriptome, even from different laboratories, in order to improve their annotation [46]. Similarly, DGE allows the data generated to be reassessed as many times as required, as a more complete genome and transcriptome references for this species become available. Hence, the quantitative data provided here by DGE will prove useful in order to recover and annotate more undescribed genes in the future.

## Methods

### Animal samples

Planarians used in this study were from the asexual clonal line of *S. mediterranea* BCN10. Animals were maintained in artificial water and were starved at least seven days prior to experimentation.

### Cell dissociation, cell sorting and RNA extraction

To trigger neoblast proliferation and differentiation, two days head and tail regenerating animals were used for the preparation of the libraries. Three animals per library were used in order to obtain the required amount of RNA. Cell dissociation and FACS were carried out as described by Möritz [31] and Hayashi [30]. Briefly, after cell staining with Calcein AM and Hoechst 33342 (Molecular Probes, Life Technologies), one million cells were separated for each population in a FACSAria sorter (Becton Dickinson) at the Scientific and Technological Centers of the University of Barcelona (CCiTUB) cytometry facilities. A representative plot of the cell populations after the sorting can be seen in Additional file 1A. Cells were directly collected in TRIzol LS (Life Technologies) at 4°C and maintained in ice to preserve RNA integrity. RNA extraction followed to obtain 1 µg of total RNA for each library. Quantification of RNA was assessed with a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) and quality check was performed by capillary electrophoresis in an Agilent 2100 Bioanalyzer (Agilent Technologies) prior to library preparation.

### DGE sequencing

Unlike RNA-Seq, this method only sequences a short read of a fixed length, named tag, derived from a single site proximal to the 3'-end of polyadenylated transcripts. This short read is later used to identify the full transcript. The number of times that the very same tag has been sequenced—its number of occurrences—is proportional to the abundance of the transcript which it belongs to. Since it only counts one sequence per transcript, its ability to quantify is not affected by the transcript length. For that reason, DGE is better suited for the detection of short transcripts and low expressed genes when compared with RNA-Seq [20-22].

Sequence tag preparation was done with Illumina's DGE Tag Profiling Kit according to the manufacturer's protocol as described [104]. In short, the most relevant steps included the incubation of 1 µg of total RNA with oligo-dT beads to capture the polyadenylated RNA fraction followed by cDNA synthesis. Then, samples were digested with NlaIII to retain a cDNA fragment from the most 3' CATG proximal site to the poly(A)-tail. Subsequently, a second digestion with MmeI was performed, which cuts 17 bp downstream of the CATG site, generating, thus, the 21 bp tags.

Cluster generation was performed after applying 4pM of each sample to the individual lanes of the Illumina 1G flowcell. After hybridization of the sequencing primer to the single-stranded products, 18 cycles of base incorporation were carried out on the 1G analyzer according to the manufacturer's instructions. Image analysis and base calling were performed using the Illumina pipeline, where tag sequences were obtained after purity filtering. Generation of expression matrices, data annotation, filtering and processing were performed by using the Biotag software (SkuldTech, France) [104].

Raw sequencing data in FASTQ format as well as processed tag sequences and their associated expressions have been deposited at NCBI Gene Expression Omnibus (GEO) [105] and are accessible through GEO Series accession number GSE51681 [106].

### Comparison of expression data

Tag raw expression was normalized to counts per million (cpm). The statistical value of DGE data comparisons, as a function of tag counts, was calculated by assuming that each tag has an equal chance to be detected, in fair agreement with a binomial law. An internal algorithm allows the comparison between different libraries and measures the significance threshold for the observed variations and p-value calculation (see Mathematical Appendix of Piquemal et al. 2002 [104]).

Different Perl [107] scripts were designed for the subsequent analyses. All of them are available from the web site [planarian.bio.ub.edu/SmedDGE](http://planarian.bio.ub.edu/SmedDGE).

### Tag mapping

A database with all the possible CATG + 17bp theoretical tag sequences was constructed for each one of the reference data sets. Tags were compared to these databases to identify all perfect matches and, when more than one tag mapped over the same transcript, only the tag closer to the 3'-end was considered. For the genome reference, 2 mismatches were also considered for unmappable tags with the SeqMap mapper [108,109].

In addition, tags were also mapped against a database of 8,662,308 CDS and 5,189 genomic sequences from bacteria directly downloaded from GenBank [110] repositories to check sample contaminations. Only two tags mapped on bacterial transcripts, confirming the purity of our libraries.

For the 3'-UTR prediction, all 23,020 contigs of the transcriptome from Kao et al. 2013 [17], were mapped over the genome using Exonerate 2.2.0 [111] to characterize the putative 3'-UTR ends (poly-A sites were not predicted though). Apart from aligning the transcripts to the genomic contigs, the strand for the longest ORF contained was also considered to ensure proper transcript orientation. For each transcript, 1,000bp upstream and downstream regions around the genomic coordinate for the putative 3'-UTR ends found were considered to retrieve DGE tags (noted as transcripts 3'-end relative position in Additional file 5).

### Libraries and reference sequence data sets randomization

Libraries and reference data sets were randomized using Perl [107] scripts and the Inline::C library to generate analogous sets of random sequences. This method resembles the original data sets in terms of size and nucleotide abundance in comparison with other approximations which generate virtual sequences based on mathematical distributions [49]. 500 and 100 randomizations for each library and data sets respectively were generated. Mapping was performed using cutoffs of 1, 5, 10, 15 and 20 occurrences (Additional file 3).

### Browsing data sets

Mapped tags are also available from the web site through a set of dynamic tables (Figure 4A). They were implemented using the jQuery jqGrid-4.5.2 [112] library, an Ajax-enabled JavaScript control to represent and manipulate tabular data on the web. Those tables summarize the tags along with their mappings on the different transcriptomes publicly available (which were downloaded from the locations cited at the respective papers [10-17]), their correspondence with the Smed454 transcriptome, and their annotation.

The transcriptome browser shown in Figure 4B was initialized with the Smed454 [10] contigs using the GBrowse2 engine [113]. The browser also includes

high-scoring segment pairs (HSPs) from whole-transcriptome BLAST searches performed over the UniProt database [114] (NCBI BLAST+ 2.2.29 [51] with default parameters), as well as the Pfam [115,116] domains mapped by HMMER—with E-val=1 and domain E-val=1—[117] on the six-frame translations for the contigs sequences. DGE tag sequences—together with the corresponding counts, normalized scores, their ranks, etc.—were uploaded to the GBrowse2 MySQL database, and they are shown in the browser using a customized version of the Bio::Graphics::Glyph::xyplot module.

Functional annotation was projected from the UniProt GO annotations over the homologous Smed454 contig sequences. Two-tailed hypergeometric test, which accounts for significant overrepresented (positive-tail) or under-represented (negative-tail), was performed by comparing the set of GO assigned to transcriptome contigs over-represented on each of the cell fractions against the set of GO annotations for the whole set of contigs. Significance threshold was set to  $p < 10^{-5}$  and the results are summarized in Additional file 6 for the different cell fraction sets.

### Gene nomenclature

New genes were named following the nomenclature proposed for *S. mediterranea* [118] based on their BLASTx homology—NCBI BLAST+ 2.2.29 [51] with default parameters against the UniProt database [114]—to its human homologous gene according to the official gene name approved by the HUGO Gene Nomenclature Committee (HGNC) [119] whenever possible, and trying to honor the names of other members of the family if they were already stated for *S. mediterranea*. When no significant homology for the corresponding gene was available, its characteristic domain found at the Pfam site [115,116] was used to identify it.

Gene sequences and primers used for cloning are deposited at the GenBank [110] site—see Table 4 for the accession numbers of the sequences.

### Irradiation

For experimental protocols requiring irradiated animals, irradiation was carried out at 75 Gy (1,66 Gy/minute) in a X-ray cabinet MaxiShot 200 (Yxlon Int.) at the facilities of the Scientific and Technological Centers of the University of Barcelona (CCiTUB).

### In situ hybridization

WISH was conducted for gene expression analysis, as previously described [120,121]. Images from representative organisms of each experiment were captured with a ProgRes C3 camera (Jenoptik) through a Leica MZ16F stereomicroscope. Animals were fixed and hybridized at the indicated time points.

### Fluorescence in situ hybridization

For double FISH animals were treated as described elsewhere [122]. Confocal laser scanning microscopy was performed with a Leica SP2.

### Immunohistochemistry

Immunostaining was carried out as described previously [123]. The following antibodies were used:  $\alpha$ -SYNORF-1, a monoclonal antibody specific for SYNAPSIN, which was used as a pan-neural marker [124] (1:50; Developmental Studies Hybridoma Bank); and  $\alpha$ -phosphohistone H3 (H3P), which was used to detect mitotic cells (1:500; Cell Signaling Technology). Alexa 488-conjugated goat  $\alpha$ -mouse (1:400) and Alexa 568-conjugated goat  $\alpha$ -rabbit (1:1000; Molecular Probes) were used as secondary antibodies.

### RNAi experiments

Double-stranded RNAs (dsRNA) were produced by in vitro transcription (Roche) and injected into the gut of the planarians as previously described [5]. Three aliquots of 32 nl (400–800ng/ $\mu$ l) were injected on three consecutive days with a Drummond Scientific Nanoject II injector. Head and tail ablation pre- and post-pharyngeally followed the fourth day. If no phenotype was observed after two weeks, a second round of injection and amputation was carried out in the same manner, unless otherwise stated. Control organisms were injected with *gfp* dsRNA.

### Availability of supporting data

All data sets are fully available without restriction. Yet relevant data sets were already included within this article and its additional files, further supporting material, as well as updates, will be publicly available through the project web site [https://planarian.bio.ub.edu/SmedDGE].

Raw sequencing data in FASTQ format, along with processed tag sequences and their associated expressions, have been deposited at NCBI Gene Expression Omnibus (GEO) [105]; they are accessible through GEO Series accession number GSE51681 [106]. Gene sequences and primers used for cloning are deposited at the GenBank [110] repository, the corresponding accession numbers for the gene sequences are listed on Table 4.

### Additional files

#### Additional file 1: Fluorescence-activated cell sorting plot.

**A** - Representative FACS plot of the cell sorting experiment carried out in this study showing the selection criteria applied for the isolation of the X1, X2, and Xin cell populations as described by Möritz [31] and Hayashi [30]. The cytoplasm of the cells is stained with Calcein AM while their nuclei are labeled with Hoechst 33342. Then, cells are separated by their nucleus/cytoplasm ratio. **B** - Same FACS plot from a cell dissociation experiment with lethally irradiated planarians four days after irradiation. A complete ablation of X1 and an important reduction of cells in X2 can be

observed. The sensitivity of the cells in these populations to irradiation responds to their composition of neoblasts in different stages of the cell cycle and distinct levels of determination: X1, proliferating stem cells in S/G2/M, and X2, stem cell progeny and proliferating neoblasts in G0/G1. Neoblasts are the only proliferating cells in this organism.

**Additional file 2: Distribution of mapped and orphan tags by number of occurrences.** **A** - Venn diagrams showing the tags overlap between the three cell populations, by occurrence (top), and by significant p-value, ( $p < 0.05$ , bottom). The number of mapping tags is detailed in italics.

**B** - Frequency distribution of tags grouped by its number of occurrences, i.e., sequencing events, in all libraries. Tags detected in a low copy number are prone to be produced by sequencing errors—likely from more abundant tags. As can be appreciated, most of the tags with less than five occurrences do not map over any of the reference data sets, suggesting that those tags are less reliable [49], which is in agreement with the results of the randomization simulations (see the text and Additional file 3). Due to that, tags detected less than five times were discarded in further analysis.

**Additional file 3: Randomization simulations.** Number of tags mapped over the randomized reference data sets, and vice versa, at different occurrences cutoffs. When compared with the theoretical number of matches expected by chance, this facilitates the assessment of the minimum number of counts for a tag to be considered reliable.

**Additional file 4: X1 and X2 in irradiated animals.** Venn diagram showing the overlap between the results presented here and the DGE study conducted over irradiated planarians of the same clonal line by Galloni [36]. The number of mapping tags out of the total is detailed in italics. It can easily be appreciated that most of the tags present in X1 and X2 are not detected by the irradiation approach.

**Additional file 5: Tags potentially mapping in the 3'-UTR regions.** Y-axis represents the number of tags (tag counts) per nucleotide genomic position. The sequenced DGE tags were then classified in two groups: those mapping within the genomic region delimited by the transcript exons (green area), and those mapping outside (blue area). As position 0 depicts the last nucleotide for all the transcripts, we can only observe green marks upstream; blue marks can distribute across all the downstream region too. Background is defined by all those genomic CATG target sequences that do not match to any of the sequenced DGE tags (red areas). Dashed line depicts the average value for the downstream background tag counts.

**Additional file 6: Bar plots of the GO significant terms for different comparisons among X1, X2 and Xin annotation sets.** Each panel presents a list of the significant functional annotations ( $p < 10^{-5}$ , hypergeometric test), along with the corresponding GO code, that are over- or under-represented (computed as log-odds of the term abundance by sequence set) on each of the three ontology domains (Biological Processes, BP; Molecular Functions, MF; and Cellular Components, CC). Bar plots compare results obtained when considering the following four non-overlapping sets: X1-only (red bars), X2-only (green bars), the intersection between X1 and X2 not in Xin (orange), and Xin-only (blue bars). Bars color-filling is proportional to the p-value for the given GO code, thus darker colors corresponds to smaller p-values (all below the significant threshold anyway). A Venn diagram on top of each page represents the comparison made among the fraction sets.

**Additional file 7: Genes involved in stemness, regeneration or tissue homeostasis overexpressed in neoblasts and their progeny.** DGE expression of clones reported in two experimental high-throughput screenings by Reddien [6] and Wenemoser [56] related to regeneration, stemness or tissue homeostasis identified as being overexpressed in neoblasts ( $p < 0.001$ ).

**Additional file 8: Whole mount in situ hybridization of new neoblast genes.** Expression by WISH of new neoblast genes in control (left panel) and irradiated planarians (right panel). 38 out of the 42 genes tested are presented here. The remaining four are characterized in Figures 6A and 7A. Time after irradiation in days is shown in the top right corner for each gene. As expected for neoblast genes, expression is reduced or disappears after irradiation.

**Additional file 9: RNA interference of new neoblast genes showing defects in regeneration.** The stronger and most representative phenotype obtained after RNAi for those new neoblast genes producing

aberrant regeneration after head and tail ablation. Days of regeneration and round of injection in superscript, and number of individuals affected with respect to the total are shown in the top right and bottom right corners of each panel. All pictures are dorsal except *Smed-rbbp4-4*, which illustrates the typical ventral curling of dying animals. The inhibition of most of the genes completely prevented the formation of the blastema. For those cases in which a small blastema was allowed to develop, a detail of the anterior part is shown to appreciate the defective head and eyes. For a regenerating control animal see Figures 6C and 7C.

**Additional file 10: Literature review of the new neoblast genes presented in this study.** A description is provided for each one of the new neoblast genes proposed in this study (summarized in Table 4) based on the literature about their homologs in other species.

**Additional file 11: Double fluorescence in situ hybridization of *Smed-h2b* with *Smed-meis-like*.** **A** - Double FISH of *Smed-meis-like* together with the neoblast marker *Smed-h2b* shows colocalization of both genes, demonstrating the expression of *Smed-meis-like* in neoblasts. Expression is also detected in differentiated cells. **B** - The pan-neural marker  $\alpha$ -SYNAPSIN shows the different penetrance of phenotypes of *Smed-meis-like*(RNAi).

**Additional file 12: Double fluorescence in situ hybridization of *Smed-h2b* with *Smed-nf-YA*, *Smed-nf-YB-2*, and *Smed-nf-YC*.** **A** - Double FISH of *Smed-nf-YA*, *Smed-nf-YB-2*, and *Smed-nf-YC* shows colocalization of the NF-Y subunits with the neoblast marker *Smed-h2b*, corroborating the expression of this complex in neoblasts. Expression is also detected in differentiated cells. **B** - The pan-neural marker  $\alpha$ -SYNAPSIN shows reduced cephalic ganglia of RNAi animals compared with *gfp* controls.

**Additional file 13: Splashplot projection of the X1/X2 versus Xin expression changes of upregulated contigs by cell population.** **A** - Splashplot for overrepresented contigs in the three cell fractions X1, X2 and Xin according to our DGE data over the *Smed454* transcriptome [10]. **B** - Same representation using the data published by Labbé [14]. Both plots show a similar composition, revealing a low number of transcripts overexpressed specifically in X2/progeny cells.

**Additional file 14: Pearson and Spearman correlations of the normalized expression levels among X1, X2 and Xin.** Diagonal panels show violin plots with the distribution of the normalized expression levels for each of the three cell populations data sets. Panels on the upper diagonal summarize both Pearson (parametric) and Spearman (non-parametric) correlations, along with the p-values and the linear regression model estimates for the pairwise comparison between data sets. On the bottom diagonal panels, for each pair of cell fractions the scatterplots show differences in expression for each DGE tag. Blue dotted line is defined by the intercept and slope values for the linear regression model presented on the corresponding upper panel, confidence interval is drawn as a grey shadow along that regression line. Those tags having a normalized expression value of zero in one or both of the cell types, when considering each pair-wise comparisons, were removed before computing correlations and for the plots. One can notice that X1 and X2 are the more correlated pair of cell fractions, then X2 and Xin, and finally X1 and Xin. Those results match to what would be expected.

#### Abbreviations

3'-UTR: Three prime untranslated region; cpm: Counts per million; DGE: Digital gene expression; EST: Expressed sequence tag; FACS: Fluorescence-activated cell sorting; FISH: Fluorescence in situ hybridization; GO: Gene ontology; ORF: Open reading frame; RNAi: RNA interference; TALE: Three amino acid loop extension; WISH: Whole mount in situ hybridization.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

ES, GRE and JFA conceived the project. GRE prepared the cell fractions for the DGE sequencing, and did the screening of the selected genes. AGS and JIR carried out the experimental characterization of the *Smed-meis-like* and *Smed-nf-Y* genes respectively. GRE and JFA performed the computational analyses and set up the web material. GRE drafted the manuscript with contributions from all authors. All authors read and approved the final manuscript.

#### Acknowledgements

DGE libraries were produced by Skuldtech under supervision of David Piquemal. We thank Mireille Galloni (University of Montpellier 2 and CNRS) for sharing the DGE libraries of irradiated organisms before they were publicly available. Further thanks to Jaume Comas and Ricard Álvarez (Scientific and Technological Centers of the University of Barcelona) for their excellent technical assistance at the cytometry facilities. We are especially grateful to Luca Gentile and Sören Moritz (Max-Planck-Institut für molekulare Biomedizin) for the cell dissociation protocol and their inestimable advice. We also thank all the members of the Emili Saló and Francesc Cebrià labs for encouragement and discussion, and Enrique Blanco and Iain K. Patten for critical comments on the manuscript. We want to acknowledge the helpful comments made by the referees in order to improve the final manuscript version. This work was funded by grants BFU2008-01544 and BFU2011-22786 (MICINN), and 2009SGR-1018 and SGR2014-687 (AGAUR) to ES and BFU2009-09102 (MICINN) to JFA. GRE and JRL were supported by FPI (MICINN and MECC respectively), and AGS by FPU (MECD) fellowships. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this manuscript.

Received: 14 November 2014 Accepted: 13 April 2015

Published online: 08 May 2015

#### References

- Reddien PW. Specialized progenitors and regeneration. *Development*. 2013;140(5):951–7.
- Rink J. Stem cell systems and regeneration in planaria. *Dev Genes Evol*. 2013;223(1–2):67–84.
- Elliott S, Sánchez Alvarado A. The history and enduring contributions of planarians to the study of animal regeneration. *Rev Dev Biol*. 2013;2(3):301–26.
- Baguña J. The planarian neoblast: the rambling history of its origin and some current black boxes. *Int J Dev Biol*. 2012;56(1–3):19–37.
- Sánchez Alvarado A, Newmark PA. Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc Natl Acad Sci USA*. 1999;96(9):5049–54.
- Reddien PW, Bermange AL, Murfitt KJ, Jennings JR, Sánchez Alvarado A. Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Dev Cell*. 2005;8(5):635–49.
- Rossi L, Salvetti A, Marincola FM, Lena A, Deri P, Mannini L, et al. Deciphering the molecular machinery of stem cells: a look at the neoblast gene expression profile. *Genome Biol*. 2007;8(4):62.
- Fernández-Taboada E, Rodríguez-Esteban G, Saló E, Abril JF. A proteomics approach to decipher the molecular nature of planarian stem cells. *BMC Genomics*. 2011;12:133.
- Böser A, Drexler HC, Reuter H, Schmitz H, Wu G, Schöler HR, et al. SILAC proteomics of planarians identifies *ncoa5* as a conserved component of pluripotent stem cells. *Cell Reports*. 2013;5(4):1142–55.
- Abril JF, Cebrià F, Rodríguez-Esteban G, Horn T, Fraguas S, Calvo B, et al. *Smed454* dataset: unravelling the transcriptome of *Schmidtea mediterranea*. *BMC Genomics*. 2010;11(1):731.
- Blythe MJ, Kao D, Malla S, Rowsell J, Wilson R, Evans D, et al. A dual platform approach to transcript discovery for the planarian *Schmidtea mediterranea* to establish RNAseq for stem cell and regeneration biology. *PLoS One*. 2010;5(12):15617.
- Adamidi C, Wang Y, Gruen D, Mastrobuoni G, You X, Tolle D, et al. De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics. *Genome Res*. 2011;21(7):1193–200.
- Sandmann T, Vogg MC, Owlarn S, Boutros M, Bartscherer K. The head-regeneration transcriptome of the planarian *Schmidtea mediterranea*. *Genome Biol*. 2011;12(8):76.
- Labbé RM, Irimia M, Currie KW, Lin A, Zhu SJ, Brown DD, et al. A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells*. 2012;30(8):1734–45.
- Resch AM, Palakodeti D, Lu YC, Horowitz M, Graveley BR. Transcriptome analysis reveals strain-specific and conserved stemness genes in *Schmidtea mediterranea*. *PLoS One*. 2012;7(4):34447.
- Rouhana L, Vieira AP, Roberts-Galbraith RH, Newmark PA. PRMT5 and the role of symmetrical dimethylarginine in chromatoid bodies of planarian stem cells. *Development*. 2012;139(6):1083–94.

17. Kao D, Felix D, Aboobaker A. The planarian regeneration transcriptome reveals a shared but temporally shifted regulatory program between opposing head and tail scenarios. *BMC Genomics*. 2013;16(14):797.
18. Solana J, Kao D, Mihaylova Y, Jaber-Hijazi F, Malla S, Wilson R, et al. Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNAseq, RNA interference and irradiation approach. *Genome Biol*. 2012;13(3):19.
19. Scimone M, Kravarik K, Lapan S, Reddien P. Neoblast specialization in regeneration of the planarian *Schmidtea mediterranea*. *Stem Cell R*. 2014;3(2):339–52.
20. Oshlack A, Wakefield MJ. Transcript length bias in RNA-seq data confounds systems biology. *Biol Direct*. 2009;4:14.
21. Raz T, Kapranov P, Lipson D, Letovsky S, Milos PM, Thompson JF. Protocol dependence of sequencing-based gene expression measurements. *PLoS One*. 2011;6(5):19287.
22. Hong LZ, Li J, Schmidt-Kuntzel A, Warren WC, Barsh GS. Digital gene expression for non-model organisms. *Genome Res*. 2012;21(11):1905–15.
23. Hanriot L, Keime C, Gay N, Faure C, Dossat C, Wincker P, et al. A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. *BMC Genomics*. 2008;16(9):418.
24. Veitch NJ, Johnson PC, Trivedi U, Terry S, Wildridge D, MacLeod A. Digital gene expression analysis of two life cycle stages of the human-infective parasite, *Trypanosoma brucei* gambiense reveals differentially expressed clusters of co-regulated genes. *BMC Genomics*. 2010;11(1):124.
25. 't Hoen PAC, Ariyurek Y, Thygesen HH, Vreugdenhil E, Vossen RHAM, de Menezes RX, et al. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. *Nucleic Acids Res*. 2008;36(21):141.
26. Asmann YW, Klee EW, Thompson EA, Perez EA, Middha S, Oberg AL, et al. 3' tag digital gene expression profiling of human brain and universal reference RNA using Illumina Genome Analyzer. *BMC Genomics*. 2009;10(1):531.
27. de Lorgeril J, Zenagui R, Rosa RD, Piquemal D, Bachelère E. Whole transcriptome profiling of successful immune response to vibrio infections in the oyster *Crassostrea gigas* by digital gene expression analysis. *PLoS ONE*. 2011;6(8):23142.
28. Philippe N, Bou Samra E, Boureux A, Mancheron A, Ruffe F, Bai Q, et al. Combining DGE and RNA-sequencing data to identify new polyA+ non-coding transcripts in the human genome. *Nucleic Acids Res*. 2014;42(5):2820–32.
29. Dong H, Ge X, Shen Y, Chen L, Kong Y, Zhang H, et al. Gene expression profile analysis of human hepatocellular carcinoma using SAGE and LongSAGE. *BMC Med Genomics*. 2009;2(1):5.
30. Hayashi T, Asami M, Higuchi S, Shibata N, Agata K. Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Dev Growth Differ*. 2006;48(6):371–80.
31. Moritz S, Stöckle F, Ortmeier C, Schmitz H, Rodríguez-Esteban G, Key G, et al. Heterogeneity of planarian stem cells in the S/G2/M phase. *Int J Dev Biol*. 2012;56(1-3):117–25.
32. Mukherjee K, Bürglin TR. Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution. *J Mol Evol*. 2007;65(2):137–53.
33. Bhattacharya A, Deng J, Zhang Z, Behringer R, de Crombrughe B, Maity S. The B subunit of the CCAAT box binding transcription factor complex (CBF/NF-Y) is essential for early mouse development and cell proliferation. *Cancer Res*. 2003;63(23):8167–72.
34. Zhang X, Hao L, Meng L, Liu M, Zhao L, Hu F, et al. Digital Gene Expression tag profiling analysis of the gene expression patterns regulating the early stage of mouse spermatogenesis. *PLoS ONE*. 2013;8(3):58680.
35. Gowda M, Jantasuriyarat C, Dean RA, Wang GL. Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. *Plant Physiol*. 2004;134(3):890–7.
36. Galloni M. Global irradiation effects, stem cell genes and rare transcripts in the planarian transcriptome. *Int J Dev Biol*. 2012;56(1-3):103–16.
37. Taft A, Vermeire J, Bernier J, Birkeland S, Cipriano M, Papa A, et al. Transcriptome analysis of *Schistosoma mansoni* larval development using serial analysis of gene expression (SAGE). *Parasitology*. 2009;136(05):469.
38. Pelechano V, Steinmetz L. Gene regulation by antisense transcription. *Nat Rev Genet*. 2013;14(12):880–93.
39. Boguski MS, Lowe TMJ, Tolstoshev CM. dbEST - database for "expressed sequence tags". *Nat Genet*. 1993;4(4):332–3.
40. Sánchez Alvarado A, Newmark PA, Robb SM, Juste R. The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development*. 2002;129(24):5659–65.
41. Zayas RM, Bold TD, Newmark PA. Spliced-leader trans-splicing in freshwater planarians. *Mol Biol Evol*. 2005;22(10):2048–54.
42. Zayas RM, Hernández A, Habermann B, Wang Y, Stary JM, Newmark PA. The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: analysis of ESTs from the hermaphroditic strain. *Proc Natl Acad Sci U S A*. 2005;102(51):18491–6.
43. The *Schmidtea mediterranea* Genome Sequencing Project. <http://genome.wustl.edu/genomes/detail/schmidtea-mediterranea>.
44. NCBI Genome Database: *Schmidtea mediterranea*. <http://www.ncbi.nlm.nih.gov/genome/232>.
45. Reverter A, McWilliam S, Barris W, Dalrymple B. A rapid method for computationally inferring transcriptome coverage and microarray sensitivity. *Bioinformatics*. 2004;21(1):80–9.
46. Saha S, Sparks A, Rago C, Akmaev V, Wang C, Vogelstein B, et al. Using the transcriptome to annotate the genome. *Nat Biotechnol*. 2002;20(5):508–12.
47. Chen J, Sun M, Lee S, Zhou G, Rowley J, Wang S. Identifying novel transcripts and novel genes in the human genome by using novel SAGE tags. *Proc Natl Acad Sci U S A*. 2002;99(19):12257–62.
48. Pleasance E, Marra M, Jones S. Assessment of SAGE in transcript identification. *Genome Res*. 2003;13(6):1203–15.
49. Keime C, Sémon M, Mouchiroud D, Duret L, Gandrillon O. Unexpected observations after mapping LongSAGE tags to the human genome. *BMC Bioinf*. 2007;8(1):154.
50. Hene L, Sreenu V, Vuong M, Abidi S, Sutton J, Rowland-Jones S, et al. Deep analysis of cellular transcriptomes-LongSAGE versus classic MPSS. *BMC Genomics*. 2007;8(1):333.
51. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinf*. 2009;10(1):421.
52. Smed454 Contig Browser. [https://planarian.bio.ub.edu/datasets/454/planarian\\_menu.php](https://planarian.bio.ub.edu/datasets/454/planarian_menu.php).
53. Eisenhoffer G, Kang H, Sánchez Alvarado A. Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell Stem Cell*. 2008;3(3):327–39.
54. Lapan S, Reddien P. Transcriptome analysis of the planarian eye identifies ovo as a specific regulator of eye regeneration. *Cell Rep*. 2012;2(2):294–307.
55. González-Estévez C, Felix DA, Smith MD, Paps J, Morley SJ, James V, et al. SMG-1 and mTORC1 act antagonistically to regulate response to injury and growth in planarians. *PLoS Genet*. 2012;8(3):1002619.
56. Wenemoser D, Lapan SW, Wilkinson AW, Bell GW, Reddien PW. A molecular wound response program associated with regeneration initiation in planarians. *Genes Dev*. 2012;26(9):988–1002.
57. Wolff E, Dubois F. Sur la migration des cellules de regeneration chez les planaires. *Rev Suisse Zool*. 1948;55:218–27.
58. Reddien P, Oviedo N, Jennings J, Jenkin J, Sánchez Alvarado A. SMEDWI-2 is a PIWI-Like protein that regulates planarian stem cells. *Science*. 2005;310(5752):1327–30.
59. Guo T, Peters A, Newmark P. A Bruno-like gene is required for stem cell maintenance in planarians. *Dev Cell*. 2006;11(2):159–69.
60. Kim JH, Yang CK, Heo K, Roeder RG, An W, Stallcup MR. CCAR1, a key regulator of mediator complex recruitment to nuclear receptor transcription complexes. *Mol Cell*. 2008;31(4):510–9.
61. Ball EE, Rehm EJ, Goodman CS. Cloning of a grasshopper cDNA coding for a protein homologous to the A1, A2/B1 proteins of mammalian hnRNP. *Nucleic Acids Res*. 1991;19(2):397.
62. Rigbolt KT, Prokhorova TA, Akimov V, Henningsen J, Johansen PT, Kratchmarova I, et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci Signal*. 2011;4(164):3.
63. Rossman T, Wang Z. Expression cloning for arsenite-resistance resulted in isolation of tumor-suppressor *fau* cDNA: possible involvement of the ubiquitin system in arsenic carcinogenesis. *Carcinogenesis*. 1999;20(2):311–6.



64. Kimura K, Cuvier O, Hirano T. Chromosome condensation by a human condensin complex in *Xenopus* egg extracts. *J Biol Chem*. 2001;276(8):5417–20.
65. Allende ML, Amsterdam A, Becker T, Kawakami K, Gaiano N, Hopkins N. Insertional mutagenesis in zebrafish identifies two novel genes, pescadillo and dead eye, essential for embryonic development. *Genes Dev*. 1996;10(24):3141–55.
66. Knight AS, Notaridou M, Watson RJ. A Lin-9 complex is recruited by B-Myb to activate transcription of G2/M genes in undifferentiated embryonal carcinoma cells. *Oncogene*. 2009;28(15):1737–47.
67. Zhu SJ, Pearson BJ. The Retinoblastoma pathway regulates stem cell proliferation in freshwater planarians. *Dev Biol*. 2013;373(2):442–52.
68. Nakamura Y, Tanaka H, Arakawa H, Yamaguchi T, Shiraiishi K, Fukuda S, et al. A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature*. 2000;404(6773):42–9.
69. Bertolino E, Reimund B, Wildt-Perinic D, Clerc RG. A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. *J Biol Chem*. 1995;270(52):31178–88.
70. Felix DA, Aboobaker AA. The TALE class homeobox gene *Smed-prep* defines the anterior compartment for head regeneration. *PLoS Genet*. 2010;6(4):1000915.
71. Chen CC, Wang IE, Reddien PW. *pbx* is required for pole and eye regeneration in planarians. *Development*. 2013;140(4):719–29.
72. Blassberg RA, Felix DA, Tejada-Romero B, Aboobaker AA. PBX/extradenticle is required to re-establish axial structures and polarity during planarian regeneration. *Development*. 2013;140(4):730–9.
73. Fraguas S, Barberán S, Cebrià F. EGFR signaling regulates cell proliferation, differentiation and morphogenesis during planarian regeneration and homeostasis. *Dev Biol*. 2011;354(1):87–101.
74. Iglesias M, Almuedo-Castillo M, Aboobaker AA, Saló E. Early planarian brain regeneration is independent of blastema polarity mediated by the Wnt/ $\beta$ -catenin pathway. *Dev Biol*. 2011;358(1):68–78.
75. Petersen CP, Reddien PW. Polarized *notum* activation at wounds inhibits wnt function to promote planarian head regeneration. *Science*. 2011;332(6031):852–5.
76. Gurley KA, Rink JC, Sánchez Alvarado A. Beta-catenin defines head versus tail identity during planarian regeneration and homeostasis. *Science*. 2008;319(5861):323–7.
77. Petersen C, Reddien P. Smed-betacatenin-1 is required for anteroposterior blastema polarity in planarian regeneration. *Science*. 2008;319(5861):327–30.
78. Oviedo NJ, Newmark PA, Sánchez Alvarado A. Allometric scaling and proportion regulation in the freshwater planarian *Schmidtea mediterranea*. *Dev Dyn*. 2003;226(2):326–33.
79. Fraguas S, Barberán S, Iglesias M, Rodríguez-Esteban G, Cebrià F. *egr-4*, a target of EGFR signaling, is required for the formation of the brain primordia and head regeneration in planarians. *Development*. 2014;141(9):1835–47.
80. Vogg MC, Owlarn S, Pérez Rico YA, Xie J, Suzuki Y, Gentile L, et al. Stem cell-dependent formation of a functional anterior regeneration pole in planarians requires Zic and Forkhead transcription factors. *Dev Biol*. 2014;390(2):136–48.
81. Vázquez-Doorman C, Petersen CP. *zic-1* expression in planarian neoblasts after injury controls anterior pole regeneration. *PLoS Genet*. 2014;10(7):1004452.
82. Scimone ML, Lapan SW, Reddien PW. A forkhead transcription factor is wound-induced at the planarian midline and required for anterior pole regeneration. *PLoS Genet*. 2014;10(1):1003999.
83. González-Sastre A, Molina MD, Saló E. Inhibitory Smads and bone morphogenetic protein (BMP) modulate anterior photoreceptor cell number during planarian eye regeneration. *Int J Dev Biol*. 2012;56(1–2–3):155–63.
84. Dolfini D, Mantovani R. Targeting the Y/CCAAT box in cancer: YB-1 (YBX1) or NF-Y? *Cell Death Differ*. 2013;20(5):676–85.
85. Maity S, Golumbek P, Karsenty G, de Crombrughe B. Selective activation of transcription by a novel CCAAT binding factor. *Science*. 1988;241(4865):582–5.
86. Hatamochi A, Golumbek P, Van Schaftingen E, de Crombrughe B. A CCAAT DNA binding factor consisting of two different components that are both required for DNA binding. *J Biol Chem*. 1988;263(12):5940–7.
87. Hooft van Huijsduijnen R, Bollekens J, Dom A, Benoist C, Mathis D. Properties of a CCAAT box-binding protein. *Nucleic Acids Res*. 1987;15(18):7265–82.
88. Kim C, Sheffery M. Physical characterization of the purified CCAAT transcription factor, alpha-CP1. *J Biol Chem*. 1990;265(22):13362–9.
89. Yoshioka Y, Ly L, Yamaguchi M. Transcription factor NF-Y is involved in differentiation of R7 photoreceptor cell in *Drosophila*. *Biol Open*. 2012;1(1):19–29.
90. Ly L, Yoshida H, Yamaguchi M. Nuclear transcription factor y and its roles in cellular processes related to human disease. *Am J Cancer Res*. 2013;3(4):339–46.
91. Wang Y, Stary J, Wilhelm J, Newmark P. A functional genomic screen in planarians identifies novel regulators of germ cell development. *Genes Dev*. 2010;24(18):2081–92.
92. Zhu J, Zhang Y, Joe G, Pompetti R, Emerson S. NF-Ya activates multiple hematopoietic stem cell (HSC) regulatory genes and promotes HSC self-renewal. *Proc Natl Acad Sci*. 2005;102(33):11728–33.
93. Benatti P, Dolfini D, Viganò A, Ravo M, Weisz A, Imbriano C. Specific inhibition of NF-Y subunits triggers different cell proliferation defects. *Nucleic Acids Res*. 2011;39(13):5356–68.
94. van Wolfswinkel J, Wagner D, Reddien P. Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell*. 2014;15(3):326–39.
95. Unneberg P, Wennborg A, Larsson M. Transcript identification by analysis of short sequence tags—Influence of tag length, restriction site and transcript database. *Nucleic Acids Res*. 2003;31(8):2217–26.
96. Li B, Ruotti V, Stewart R, Thomson J, Dewey C. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*. 2009;26(4):493–500.
97. Koehler R, Issac H, Cloonan N, Grimmond SM. The uniqueome: a mappability resource for short-tag sequencing. *Bioinformatics*. 2010;27(2):272–4.
98. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet*. 2014;15(2):121–32.
99. Morey M, Yee S, Herman T, Nern A, Blanco E, Zipursky SL. Coordinate control of synaptic-layer specificity and rhodopsins in photoreceptor neurons. *Nature*. 2008;456(7223):795–9.
100. Pearson BJ, Sánchez Alvarado A. A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. *Development*. 2009;137(2):213–21.
101. Oviedo N, Beane W. Regeneration: the origin of cancer or a possible cure? *Semin Cell Dev Biol*. 2009;20(5):557–64.
102. Pearson B, Sánchez Alvarado A. Regeneration, stem cells, and the evolution of tumor suppression. *Cold Spring Harb Symp Quant Biol*. 2008;73(0):565–72.
103. PlanMine. <http://planmine.mpi-cbg.de>.
104. Piquemal D, Commes T, Manchon L, Lejeune M, Ferraz C, Pugnère D, et al. Transcriptome analysis of monocytic leukemia cell differentiation. *Genomics*. 2002;80(3):361–71.
105. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(D1):991–5.
106. NCBI Gene Expression Omnibus. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE51681>.
107. The Perl Programming Language. <https://www.perl.org>.
108. Jiang H, Wong W. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics*. 2008;24(20):2395–6.
109. SeqMap - A Tool for Mapping Millions of Short Sequences to the Genome. <http://www-personal.umich.edu/~jianghui/seqmap>.
110. Benson D, Clark K, Karsch-Mizrachi I, Lipman D, Ostell J, Sayers E. Genbank. *Nucleic Acids Res*. 2013;42(D1):32–7.
111. Slater G, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinf*. 2005;6(1):31.
112. jQuery jqGrid JavaScript Library. <http://plugins.jquery.com/jqGrid>.
113. GBrowse2. <http://gmod.org/wiki/GBrowse>.
114. Dimmer E, Huntley R, Alam-Faruque Y, Sawford T, O'Donovan C, Martin M, et al. The UniProt-GO Annotation database in 2011. *Nucleic Acids Res*. 2011;40(D1):565–70.
115. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42(D1):222–30.

116. Pfam Database of Protein Families. <http://pfam.xfam.org>.
117. HMMER: Biosequence Analysis Using Profile Hidden Markov Models. <http://hmmer.org>.
118. Reddien PW, Newmark PA, Sánchez Alvarado A. Gene nomenclature guidelines for the planarian *Schmidtea mediterranea*. *Dev Dyn*. 2008;237(11):3099–101.
119. Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res*. 2013;41(D1):545–52.
120. Molina M, Saló E, Cebrià F. The BMP pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. *Dev Biol*. 2007;311(1):79–94.
121. Cebrià F, Guo T, Jopek J, Newmark P. Regeneration and maintenance of the planarian midline is regulated by a slit orthologue. *Devel Biol*. 2007;307(2):394–406.
122. King R, Newmark P. In situ hybridization protocol for enhanced detection of gene expression in the planarian *Schmidtea mediterranea*. *BMC Dev Biol*. 2013;13(1):8.
123. Cebrià F, Newmark P. Planarian homologs of *netrin* and *netrin receptor* are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. *Development*. 2005;132(16):3691–703.
124. Cebrià F. Organization of the nervous system in the model planarian *Schmidtea mediterranea*: an immunocytochemical study. *Neurosci Res*. 2008;61(4):375–84.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)





## **Discusión general**



# Discusión

Cuanto más aprendemos sobre la regeneración en cualquier sistema biológico, más nos percatamos de cuán complejo es este fenómeno. Una célula depende de multitud de vías metabólicas y señales reguladoras para su supervivencia, así como para su interacción a corto y largo alcance con las células de su entorno en el contexto de un tejido, un órgano o el sistema complejo definido por un organismo. Estos aspectos son claves para entender los mecanismos moleculares en los que participan los neoblastos.

## La población de neoblastos es heterogénea

Estudios recientes han puesto más énfasis en la influencia del nicho celular, es decir, el microambiente donde físicamente se encuentra la célula madre dentro del organismo, algo muy estudiado en *Drosophila* y vertebrados (Scadden, 2014) pero no así en platelmintos (Mouton *et al.*, 2009).

¿Qué hace a los neoblastos pluripotentes? ¿Está la pluripotencia conservada evolutivamente? ¿Qué principios y mecanismos orquestan la generación ordenada de todos los tipos celulares del organismo a partir de una población pluripotente de células madre? El esfuerzo retomado para impulsar las planarias como sistema *in vivo* que sirva de modelo molecular en regeneración, está empezando a proporcionar conocimiento sobre estas cuestiones.

Retos clave incluyen la identificación de sistemas de señales que controlan la proliferación de los neoblastos y el recambio de las células diferenciadas, la orquestación de la diferenciación de las células progenitoras a partir de los neoblastos y un análisis sistemático de los movimientos celulares durante la homeostasis y la regeneración. El desarrollo de una técnica de transgénesis en planaria es especialmente importante para responder a estas preguntas. Los años venideros revelarán más acerca de la diversidad de tipos celulares en todos los tejidos de la planaria así como de los reguladores que definen la identidad celular, y se conocerán los mecanismos implicados en esta morfogénesis continua que tiene lugar en las planarias adultas.

## Proteómica y transcriptómica

Un punto de partida importante para investigar la biología del neoblasto es la identificación de genes activamente expresados. Estos sirven como genes candidatos

para regular aspectos importantes su biología, así como de marcadores de neoblasto y su progenie indiferenciada.

En esta era posgenómica, la proteómica junto con la transcriptómica han de proporcionar los fundamentos sobre los cuales se sustentará la Biología de Sistemas. La adquisición e interpretación de grandes cantidades de datos provenientes de múltiples fuentes es, probablemente, el mayor desafío actual para la Biología de Sistemas (Maier *et al.*, 2009). El futuro de la bioinformática se centrará en una biología de sistemas integrativa, pues muchos aspectos de las funciones celulares sólo pueden resolverse adoptando esta estrategia. Los retos futuros comprenden el desarrollo de bases de datos y la integración computacional de los conjuntos de datos “ómicos” de diferentes fuentes permitiendo, por ejemplo, mapear las proteínas identificadas sobre los correspondientes datos de mRNA (Maier *et al.*, 2009). Por otro lado, cabe destacar el reciente incremento de la complejidad transcripcional, fruto del descubrimiento de los transcritos no-codificantes como los lncRNA, miRNA, etc., que también jugará en el futuro un papel muy importante en el modelado de las rutas reguladoras implicadas en la regeneración (Friedländer *et al.*, 2009; González-Estévez *et al.*, 2009b; Lu *et al.*, 2009; Palakodeti, 2006; Qin *et al.*, 2012; Resch & Palakodeti, 2012; Sasidharan *et al.*, 2013; Thatcher & Patton, 2010).

La existencia de una pauta abierta de lectura (*Open Reading Frame*, ORF) en el genoma no implica necesariamente la existencia de un gen funcional. A pesar de los avances bioinformáticos, es todavía difícil predecir genes de manera precisa a partir de la información genómica (Guigó *et al.*, 2006; The ENCODE Project Consortium, 2007) o incluso transcriptómica (Steijger *et al.*, 2013). Esto es cierto en particular en el caso de genes pequeños, que pueden pasar desapercibidos, o genes sin homología a otros genes conocidos. Por tanto, la verificación del producto génico, esto es, la proteína, mediante técnicas proteómicas debe ser un importante primer paso en la anotación de un genoma (Mann & Pandey, 2000).

## Proteómica

Hasta ahora las planarias habían sido pobremente caracterizadas en términos de expresión proteica, seguramente en parte por la falta de herramientas para ello.

En nuestro trabajo encontramos cuatro grandes grupos de proteínas relacionadas con neoblastos:

1. Proteínas implicadas en la producción de energía y el metabolismo.
2. Reguladoras de la expresión génica y la transcripción.
3. Relacionadas con el desarrollo y la diferenciación.
4. Proteínas de respuesta al estrés y apoptosis.

Las dos primeras categorías responden a lo esperado en células que proliferan, como es el caso de las células madre. Las vías de desarrollo y diferenciación se

encuentran activas en células madre embrionarias, por ejemplo, mientras que los procesos de respuesta al estrés y la apoptosis son necesarios durante la remodelación del organismo.

De entre los genes detectados destaca *Smed-SmB*, publicado en un artículo propio (Fernández-Taboada *et al.*, 2010). Constituye el primer miembro de la superfamilia de proteínas LSm descrito en planaria, unas proteínas altamente conservadas implicadas en la maquinaria de splicing celular. Aunque *Smed-SmB* se expresa también en neuronas, se localiza en el núcleo y los cuerpos cromatoides de los neoblastos, y es necesario para mantener la capacidad proliferativa de estos. *Smed-SmB(RNAi)* provoca una rápida desorganización de los cuerpos cromatoides e impide la proliferación de los neoblastos. Todo ello conduce a una rápida reducción de la población de neoblastos, resultando en un fenotipo letal tanto en animales regenerantes como intactos. La mayoría del resto de genes pertenecen a dos grandes grupos, formados por la superfamilia Ras de *small GTPases*, *Rab-11B*, *Rab-39* y *Rac-1*, y las *heat shock proteins* 40, 60 y 70, además del factor de transcripción *Hunchback-like* y la quinasa *PrkC*. A excepción de las *heat-shock proteins* ninguno de estos genes ha sido vinculado antes con neoblastos. Todos estos genes se expresan en el blastema y/o el posblastema de animales regenerantes y, o bien desaparecen a medida que avanza la regeneración, o forman el típico patrón de expresión de neoblasto distribuido por todo el paréquima menos en la faringe y la región anterior a los ojos. Excepto en el caso de *Hunchback-like* y *Rab39*, la inhibición de estos genes a través de RNAi impide su regeneración y produce la muerte del animal. Finalmente un conjunto de otros seis genes obtenidos a partir de las trazas del genoma de planaria fueron validados mediante WISH, mostrando un patrón de expresión típico de neoblasto que desaparece en animales irradiados.

Estos resultados confirman la validez de una estrategia proteómica para encontrar nuevos genes de neoblasto que ayuden en la investigación de la biología del neoblasto. Tras nuestro estudio proteómico y movidos por la convicción de que la proteómica es complemento imprescindible de las técnicas genómicas, otros estudios han sido publicados más recientemente. Adamidi *et al.* (2011) y Önal *et al.* (2012) utilizaron una aproximación proteómica de tipo *shotgun* para validar miles de nuevos transcritos generando nuevas secuencias peptídicas.

Estudios sobre el recambio individual de proteínas a nivel de todo el organismo son ahora posibles con los avances técnicos en espectrometría de masas utilizando técnicas de marcaje metabólico con isótopos. Böser *et al.* (2013) han implementado en planarias la técnica de marcaje de aminoácidos con isótopos estables en cultivo celular *in vivo* (SILAC; Ong *et al.*, 2002) con el fin de identificar proteínas enriquecidas en sus células madre. Este método es capaz de detectar la abundancia de proteínas y cambios en sus niveles de forma cuantitativa con más precisión que los métodos sin marcaje. Aplicando este protocolo han identificado cientos de proteínas de células madre al comparar los proteomas de planarias irradiadas con planarias control y el de neoblastos aislados mediante FACS.

Puesto que la irradiación causa estrés celular y puede inducir cambios en la expresión génica no relacionados con las células madre, los datos fueron validados



mediante una aproximación independiente con neoblastos aislados por FACS (Hayashi *et al.*, 2006; Reddien *et al.*, 2005b). 838 proteínas estaban enriquecidas en X1, incluyendo muchas proteínas ya caracterizadas en células madre.

De acuerdo con nuestros resultados y otros estudios (Rouhana *et al.*, 2010, 2012), la anotación de Gene Ontology reveló que el 19 % de las proteínas estaban asociadas con unión al RNA, el 25 % con traducción, lo que sugiere que dada la especial importancia que los procesos postranscripcionales pueden tener en las células madre de planaria la expresión génica a nivel de transcrito puede no reflejar la abundancia real de proteínas.

Futuras aplicaciones de proteómica SILAC en planarias podrían incluir la cuantificación global de modificaciones postranscripcionales, como fosforilación, metilación o ubiquitinización durante la regeneración, o el análisis de interacciones proteína-proteína. Estos experimentos podrían proporcionar pistas importantes sobre mecanismos reguladores que no pueden ser capturadas por medio de análisis transcriptómicos.

Aunque la proteómica mediante SILAC en combinación con un modelo altamente plástico como la planaria se ha revelado como una herramienta potente para la identificación de proteínas de células madre conservadas, el análisis de su recambio y la caracterización de procesos basados en células madre, como la regeneración *in vivo*, es necesaria una revolución en las técnicas de análisis y cuantificación de proteína similar a la producida por la NGS en transcriptómica para avanzar en el estudio del proteoma a gran escala.

Además, es imprescindible realizar un fraccionamiento de los extractos proteicos por su localización en compartimentos celulares o por sus características físico-químicas, si se quiere extender el número de especies proteicas identificadas evitando el descubrimiento sistemático de solamente las moléculas más abundantes (Franco *et al.*, 2013).

Por último, a pesar de que la utilidad de las herramientas proteómicas ha sido probada fuera de toda duda en el estudio de la regeneración, varios organismos modelo todavía no han sido explorados con estas técnicas. En particular, aquellos que conciernen a algunos de los organismos con mayores capacidades de regenerar, como anfibios y equinodermos.

## Transcriptómica

Desde el desarrollo del método de secuenciación SAGE (Velculescu *et al.*, 1995), los sucesivos avances técnicos han hecho posible establecer los perfiles cuantitativos de expresión del mRNA de células y tejidos incluso en especies para las cuales se desconocen la mayoría de sus genes. Los métodos basados en tags, como SAGE o su versión mejorada de DGE, miden abundancias absolutas, proporcionando niveles de expresión digitales precisos. Aún así, aunque solo una parte del transcrito es analizado, impidiendo distinguir entre isoformas de *splicing* (Maier *et al.*, 2009).

En los últimos años se ha desarrollado la tecnología necesaria para llevar a cabo la secuenciación de transcriptomas completos, conocida como RNA-Seq, capaz de

producir millones de secuencias a un coste cada vez más asequible. Esta tecnología tiene varias ventajas, destacando la habilidad de detectar nuevos transcritos e isoformas. Algunos estudios han mostrado que la densidad de *reads* tiende a variar a lo largo del transcrito, lo que indica que esta técnica no está libre de sesgo. Este sesgo, como la selección preferente o la exclusión de determinadas secuencias, podría tener lugar durante el paso de ligación del adaptador, la amplificación por PCR o la propia secuenciación (Marioni *et al.*, 2008; Mortazavi *et al.*, 2008). Mejoras en la uniformización de la cobertura total (por ejemplo, mediante amplificación de muestras por MALBAC-PCR (Fu *et al.*, 2015; Huang *et al.*, 2015)) aumentarán la precisión y la fiabilidad de los análisis basados en RNA-Seq.

La disponibilidad de secuencias genómicas completas así como las tecnologías que permiten un análisis global del mRNA han incrementado enormemente nuestra capacidad para examinar los perfiles de expresión de células y tejidos, hasta el punto que el rápido crecimiento de estas técnicas durante la última década unido al estancamiento de las propias en la proteómica ha desplazado el interés por el estudio directo de las proteínas. En apenas tres años se han sido publicados hasta ocho transcriptomas de *S. mediterranea* (Adamidi *et al.*, 2011; Blythe *et al.*, 2010; Kao *et al.*, 2013; Labbé *et al.*, 2012; Resch *et al.*, 2012; Rouhana *et al.*, 2012; Sandmann *et al.*, 2011), incluyendo el nuestro (Abril *et al.*, 2010, Artículo 3). A corto plazo se seguirán produciendo nuevos transcriptomas y se integrarán para formar nuevos transcriptomas de referencia (como los publicados por Sandmann *et al.* (2011) o Kao *et al.* (2013)).

La cantidad de transcritos detectados por DGE es mucho mayor que la de proteínas obtenida por espectrometría de masas. La tabla 4 del Artículo 4 (Rodríguez-Esteban *et al.*, 2015) contiene una relación de los nuevos genes de neoblasto que han sido encontrados y validados mediante esta aproximación. Pese a su gran variedad, puede comprobarse que la mayoría corresponden a factores de transcripción, están implicados en embriogénesis, desarrollo o cáncer. Entre ellos, se han caracterizado genes relacionados con procesos de reparación del DNA, procesamiento del RNA, transporte nuclear y proliferación celular. Puede encontrarse una descripción más detallada de cada uno en el texto y en el documento 10 del material suplementario del Artículo 4 (Rodríguez-Esteban *et al.*, 2015). Los experimentos de WISH, RNAi e inmunohistoquímica con *Smed-meis-like*, miembro de la familia TALE-class caracterizada por la presencia de un dominio homeobox, demuestran que este gen es necesario específicamente para la regeneración anterior en planaria pero no para la posterior. Por otro lado, el factor nuclear Y (NF-Y) es un factor de transcripción formado por un complejo heterotrimérico. La caracterización de los tres genes que lo conforman sugiere que estaría implicado en los estadios tempranos de la diferenciación afectando al linaje epidérmico y posiblemente también al linaje neuronal.

DGE ha demostrado ser útil en planaria no solo para la identificación de nuevos genes de neoblasto. Este método ha sido aplicado también en un estudio sobre la vía de los EGFR, comparando animales control con animales privados mediante RNAi de uno de los intermediarios de esa vía, *Smed-egfr-3* (Artículo 5; Fraguas *et al.*, 2014).

### **Futuro del RNA-Seq: secuenciación de tercera generación**

Los sistemas de NGS actuales requieren un paso de amplificación por PCR de los fragmentos de DNA para emitir una señal luminosa lo suficientemente fuerte que pueda ser captada por una cámara de alta sensibilidad y poder asignar la base correspondiente. La amplificación por PCR ha sido una de las mayores revoluciones en la biología molecular, pero tiene algunos problemas inherentes asociados ya que introduce errores en las secuencias amplificadas y puede favorecer la amplificación de ciertas secuencias sobre otras, cambiando así la frecuencia relativa y la abundancia de los fragmentos de DNA amplificados. Idealmente, sería deseable poder determinar directamente la secuencia de cada molécula de DNA o RNA original, sin necesidad de amplificarla introduciendo potencialmente errores en su secuencia y un cierto *bias* en su cuantificación.

Los dispositivos de *reads* cortos de DNA (20-200 bases) ya han encontrado numerosas aplicaciones pero para secuenciación genómica y el análisis de variaciones estructurales genéticas, como variaciones del número de copias (*copy number variations*), translocaciones cromosómicas, inversiones, deleciones, inserciones y duplicaciones, sería una gran ayuda que la longitud del *read* en la moléculas de DNA original pudiera incrementarse hasta varios miles de bases. Idealmente, el objetivo sería determinar la secuencia del cromosoma completo a partir de una única molécula original de DNA.

De entre todas las tecnologías que se han estado desarrollando para secuenciar una única molécula de DNA con este objetivo, la tecnología de nanoporo (Clarke *et al.*, 2009; Steinbock & Radenovic, 2015) es la más prometedora. Esta aproximación se basa en detectar la modulación de la corriente iónica que se produce a medida que una molécula de DNA atraviesa un nanoporo artificial, revelando la características y parámetros (diámetro, longitud y conformación) de la molécula. Otra aproximación permite obtener secuencias de hasta 10Kb (PACBIO *single-molecule realtime cDNA reads*, SMRT; Flusberg *et al.*, 2010; Roberts *et al.*, 2013) y ha sido utilizada recientemente con éxito para la resolución de isoformas de *splicing* alternativo en humanos (Sharon *et al.*, 2013; Tilgner *et al.*, 2014) y en predicción de genes en organismos no modelo (Larsen *et al.*, 2014; Minoche *et al.*, 2015).

### **Correlación entre expresión de proteína y mRNA**

La expresión de un gen es aquel mecanismo por el cual la información que codifica en forma de DNA se convierte a proteína. En sentido estricto, por tanto, no debería utilizarse el término para referirse únicamente a su transcripción en forma de mRNA. A pesar de ello, suele hablarse habitualmente de niveles de expresión génica en los ensayos de RNA-Seq asumiendo que los cambios en los niveles de mRNA son paralelos a aquellos en los niveles de las proteínas que codifican. Sin embargo, biológicamente, entre los niveles de expresión de proteína y mRNA no tiene porque existir necesariamente una relación directa ni simple, debido a diversos mecanismos de regulación postranscripcional y postraduccional que se producen después de la

síntesis del mRNA.

Numerosos trabajos (la mayoría en cáncer y levadura) han reportado correlaciones bajas entre los niveles de proteína derivados de análisis con espectrometría de masas y los datos equivalentes de mRNA obtenidos tanto con *microarrays* (Greenbaum *et al.*, 2003, 2002; Gry *et al.*, 2009; Vogel *et al.*, 2010) como con SAGE (Greenbaum *et al.*, 2003, 2002; Gygi *et al.*, 1999a) y RNA-Seq (Fu *et al.*, 2009; Ning *et al.*, 2012). Ni siquiera en organismos procariontas mucho más simples como la bacteria *Escherichia coli* parece haber una correlación mejor (Taniguchi *et al.*, 2010). Otros estudios, en cambio, han encontrado una correlación positiva moderada entre los niveles de expresión de proteína y mRNA en organismos desde bacterias hasta mamíferos, con coeficientes de correlación de entre 0,2 y 0,5 (Fu *et al.*, 2009), o entre 0,36 y 0,76 (Maier *et al.*, 2009), muchos de los cuales han demostrado tener una clara relevancia biológica. En lo que sí parece haber acuerdo, es en que la tecnología de RNA-Seq proporciona una mejor estimación absoluta de los niveles de expresión frente a los *microarrays* cuando se comparan con los datos proteómicos (Fu *et al.*, 2009; Ning *et al.*, 2012). En cualquier caso, la principal conclusión de estos trabajos ha sido que a pesar de observar una correspondencia general en los niveles de mRNA y sus proteínas correspondientes, existe una considerable variación en esas correlaciones, especialmente para aquellos genes menos expresados (Gry *et al.*, 2009).

Son muchos los factores biológicos identificados que afectan a esta correlación además de factores técnicos y metodológicos. Varios de ellos afectan a la eficiencia de traducción, como la conformación (estructuras secundaria y terciaria) del mRNA, el *codon bias* o la cantidad y ocupación de los ribosomas. La síntesis de proteínas a partir de un cierto número de moléculas de mRNA puede verse estimulada o reprimida, desligando la transcripción y la traducción de manera continua o bajo ciertas condiciones.

Otro factor a tener en cuenta es el desfase temporal que puede existir entre la transcripción de un gen, su expresión en forma de proteína y la vida media de esta. Esto implica que una proteína puede continuar realizando su función después de que el mRNA a partir del cual se formó haya sido degradado o, por el contrario, el mRNA puede verse secuestrado un tiempo antes de empezar a traducirse como sucede en los cuerpos cromatoides de los neoblastos. El mayor factor postraduccional que influye en la correlación entre proteína y mRNA es la vida media individual de cada proteína. El tiempo de vida de una proteína en la célula depende de muchos factores: la estabilidad intrínseca de la proteína (por ejemplo, con presencia de dominios ricos en residuos PEST), el primer aminoácido amino-terminal, procesamientos postraduccionales como fosforilación, glicosilación o ubiquitinización, y la localización de la proteína. Las vidas medias de las proteínas son muy variables, desde unos pocos segundos a varios días. Estudios de correlación de varios parámetros muestran que el recambio de la proteína (*turnover*), estimado por la vida media de la proteína, es el parámetro biológico más importante que influye en la correlación entre proteína y mRNA (Wu *et al.*, 2008).

La expresión génica se compone de muchos pasos estocásticos que implican moléculas actuando a una concentración muy baja, lo que las hace muy difíciles de

monitorizar. Aunque el RNA-Seq ha mejorado la relación señal/ruido, no está completamente libre de ambigüedad. De todas formas, los métodos de cuantificación de transcritos son reproducibles y sensibles para rangos de varios órdenes de magnitud.

El error total en la caracterización cuantitativa de muestras complejas utilizando espectrometría de masas es difícil de evaluar. Varias fuentes técnicas y experimentales, así como de procesamiento de los datos pueden contribuir a la variabilidad inespecífica. A menudo se realizan pocas réplicas debido a que el análisis de los datos es muy costoso. El procesamiento de las muestras y la cuantificación de los péptidos digeridos pueden ser una fuente experimental de error. Diferencias en la separación cromatográfica, la ionización de los péptidos y en variaciones en el equipo de medida complican la comparación entre experimentos. El análisis de los datos de espectrometría de masas requiere muchos pasos, como normalización y diversos filtrados, detección de picos y cuantificación o la aplicación de modelos estadísticos. Un punto clave en la evaluación de datos proteómicos complejos es la elección del umbral en la identificación de los péptidos. Aumentar la puntuación mínima requerida reduce el número de falsos positivos a la vez que incrementa el de falsos negativos. El *bias* en las bases de datos de secuencias también es clave.

Los avances recientes en cuantificación de proteínas basadas en espectrometría de masas con menor error experimental y más precisión en la cuantificación, así como los métodos de secuenciación *high-throughput* para mRNA deberían ayudar en futuros análisis a dar una imagen más precisa de la correlación entre proteína y mRNA, y la influencia de los diferentes parámetros que la modifican. Análisis de muestras del mismo sistema biológico bajo diferentes condiciones, como respuesta al estrés celular o análisis en diferentes estadios del ciclo celular deberían también aportar información fisiológica específica derivada de los análisis de muestras biológicas complejas.

Por otra parte, aunque los valores de expresión de mRNA han demostrado ser útiles en un amplio rango de aplicaciones, incluyendo el diagnóstico y la clasificación de cánceres, estos resultados son, casi con certeza, solamente correlativos, y no causativos. Es más probable que sea la concentración de proteínas y sus interacciones la verdadera fuerza causativa en la célula. Las evidencias sugieren que los patrones de expresión de mRNA son necesarios pero insuficientes por si solos para la descripción cuantitativa de sistemas biológicos.

Hay que tener presente que el dogma central de la biología (enunciado en la introducción) no implica que un único gen resulte siempre en una sola proteína. A través de mecanismos como el *splicing* alternativo un único gen puede dar lugar a diferentes moléculas de mRNA o isoformas. Este mecanismo permite que a partir de un gen puedan formarse diferentes proteínas mediante la sustitución, la inserción o la delección de determinados fragmentos de su secuencia (Graveley, 2001). Una gran proporción de la complejidad del transcriptoma es debida al *splicing* de exones dentro de un único *locus* genómico. En humanos, por ejemplo, al menos entre el 60 y el 80 % de los genes muestran evidencia de *splicing* alternativo (Harrow *et al.*, 2012; Pervouchine *et al.*, 2012; The ENCODE Project Consortium, 2007), y hasta un 95 % en el caso de genes multiexónicos (Pan *et al.*, 2008). Por tanto, existe un único genoma pero múltiples transcriptomas y proteomas. Con la acumulación

de grandes cantidades de secuencias de DNA en las bases de datos durante la última década, se ha puesto de manifiesto que ni siquiera tener secuencias genómicas completas es suficiente para poder inferir de su función biológica. Además, si bien este dogma establece una secuencia de eventos unidireccional de gen a proteína, puede argumentarse también que son las proteínas quienes, a la postre, regulan la expresión de los propios genes.

Aunque desde el punto de vista de la estructura de los datos y el análisis de estos, existe paralelismo entre datos genómicos y transcriptómicos, la proteómica es todavía única en su constitución y los principios en los que se basa. La proteómica es, por tanto, complementaria a la genómica ya que se centra en el producto de los genes (Mann & Pandey, 2000). Combinar datos proteómicos con datos “ómicos” procedentes de experimentos transcriptómicos puede revelar interesantes facetas de las funciones celulares.

Por todo lo expuesto, y dado que la regulación postranscripcional es una característica importante de las células madre de planaria (Fernández-Taboada *et al.*, 2010; Rouhana *et al.*, 2010, 2012), no debería sorprendernos haber obtenido resultados diferentes entre las aproximaciones proteómica y transcriptómica. Esto reafirma nuestra convicción de que para conseguir una imagen completa debemos aproximarnos desde todos los ángulos.

## Perspectivas

Se ha hecho un considerable progreso en la comprensión de la biología de los neoblastos, su progenie y los fundamentos del proceso regenerativo. Tenemos ahora numerosas pistas sobre cómo se controlan los primeros acontecimientos en la diferenciación de los neoblastos y su mantenimiento. Sin embargo, todavía carecemos de descripciones mecanísticas detalladas. Estas verán la luz cuando los análisis transcriptómicos y funcionales (Abril *et al.*, 2010; Adamidi *et al.*, 2011; Blythe *et al.*, 2010; Böser *et al.*, 2013; Fernández-Taboada *et al.*, 2010; Kao *et al.*, 2013; Labbé *et al.*, 2012; Reddien *et al.*, 2005a; Resch *et al.*, 2012; Rouhana *et al.*, 2012; Sandmann *et al.*, 2011; Wang *et al.*, 2010) sean combinados con análisis bioquímicos y el desarrollo de aproximaciones de pérdida de expresión. Integrar toda la información en redes y modelos de regulación e interacción podría impulsar en el futuro la biología molecular sobre planarias.

Entender los mecanismos que mantienen y restablecen la estabilidad dinámica entre neoblastos y células diferenciadas con la forma correcta, el tamaño y las proporciones del cuerpo de la planaria constituye posiblemente el mayor desafío en la investigación de células madre en planaria. Un enlace crucial que falta entre morfogénesis y el sistema de células madre es el actualmente desconocido punto de intersección de las señales de formación de patrón con las de la determinación de progenitores.

Futuros trabajos se centrarán en los detalles mecanísticos que rodean la diferenciación de los neoblastos y cómo su progenie puede interpretar polaridad y formación

de patrón para diferenciarse en las estructuras correctas. Puesto que se observan también cambios tempranos en los patrones de expresión en animales irradiados, parece que las señales de polaridad y destino iniciales provienen de las células diferenciadas existentes y son independientes de los neoblastos (Gurley *et al.*, 2010). Concretamente, de células musculares subepidérmicas que enviarían señales que los neoblastos serían capaces de trazar (Witchley *et al.*, 2013). Si estas colas de polaridad son específicas de tejidos en animales con capacidad de regenerar es una cuestión importante. ¿Contienen los tejidos diferenciados de mamífero colas de polaridad capaces de coordinar correctamente la diferenciación de las células madre en los diferentes linajes? Si no, ¿pueden estas señales ser introducidas?

El grado desconocido de heterogeneidad en la población de neoblastos permanece como un cuello de botella conceptual y experimental. Se necesitan nuevas herramientas para definir subpoblaciones de neoblastos basados en criterios moleculares y funcionales. Establecer la fracción de neoblastos pluripotentes entre aquellos que expresan *smedwi-1*, el establecimiento de las condiciones para cultivos *in vitro*, o la localización de los neoblastos en su nicho, representan los siguientes pasos necesarios.

### **Single-cell sequencing en planaria**

La heterogeneidad celular es una propiedad aparente de los tejidos biológicos. Como hemos visto, el término “neoblasto” se refiere a una población de células y, puesto que las diferencias célula a célula existen en cualquier población, el criterio que designa a una célula particular como neoblasto es necesariamente impreciso. A pesar de que la población de neoblastos es heterogénea, las estrategias basadas en DGE o RNA-Seq a partir de homogenizados celulares parecen no ser todavía lo bastante resolutivas como para capturar genes diferencialmente expresados entre la población de neoblastos X1 (neoblastos en proliferación, S/G2/M) y X2 (interpretada como una mezcla de neoblastos en proliferación en fase G1 y progenie temprana o en diferenciación). Esto se debe seguramente a que alrededor de la mitad de las células en X2 son neoblastos que aún no están diferenciándose, lo cual hace más difícil capturar aquellos genes específicos de diferenciación que, además, pueden ser distintos en cada linaje.

Los análisis transcriptómicos convencionales requieren al menos de decenas de miles de células para proporcionar una visión general de la expresión génica, promediando la expresión de todas las células e ignorando la heterogeneidad celular de esa expresión. En el momento de escribir esta memoria ya es factible la aplicación de tecnologías de RNA-Seq que permiten obtener el perfil de expresión de cada célula individualmente (conocidas como *single-cell* RNA-Seq). Estas herramientas posibilitan por primera vez identificar fenotipos celulares distintos dentro de una población heterogénea. Es una tecnología joven y todavía en desarrollo, para la cual existen diferentes aproximaciones metodológicas. Entre ellas cabe citar MARS-Seq (Jaitin *et al.*, 2014) que, aplicando la misma idea de DGE, secuenciará únicamente el extremo 3' de cada molécula de mRNA para maximizar el número de transcritos

detectados por número de *reads*. Así, las recientes técnicas de *single-cell* RNA-Seq que son capaces de obtener los perfiles de expresión (Blakeley *et al.*, 2015; Grün *et al.*, 2015) o los genomas (Huang *et al.*, 2015) de células individuales proporcionan una resolución sin precedentes, lo que permite caracterizar una población célula a célula. Aunque existe un estudio a nivel de *single-cell* en planaria (van Wolfswinkel *et al.*, 2014), estas nuevas tecnologías no han sido utilizadas extensivamente a pesar de que serán la aproximación que posibilitará la caracterización de los distintos tipos de neoblastos y su progenie, la identificación de los distintos linajes celulares y facilitarán el estudio de su proceso de diferenciación.

## La planaria como organismo modelo en biomedicina

Aunque las células madre fueron vistas inicialmente como la panacea que sería capaz de curar una increíble variedad de enfermedades, la complejidad de las interacciones y de los sistemas reguladores, así como la variedad de tejidos y órganos en que estas células se pueden diferenciar, han impedido el éxito del trasplante directo en el lugar de la herida para restaurar los tejidos perdidos o dañados. Por esta razón, el estudio de cómo diferentes modelos animales utilizan y regulan las células madre o reorganizan las células que rodean la herida hacia la reparación del tejido y la regeneración, es probable que nos proporcione respuestas sobre porqué estos procesos no son activos en mamíferos. Por ejemplo, el conocimiento de porqué un proceso regenerativo tiene lugar en un organismo particular y no en otros, podría proporcionar nuevas vías para estimular la regeneración si las vías endógenas de regeneración no están disponibles.

En mamíferos, la pluripotencia se limita a embriones tempranos y es inducida y mantenida por un pequeño número de factores de transcripción fundamentales. Los adultos no cuentan con células madre pluripotentes pero se pueden derivar células pluripotentes de células madre embrionarias de embriones tempranos (Weissman, 2000) o a través de la reprogramación de células somáticas otra vez hacia un estado casi embrionario (Takahashi & Yamanaka, 2006). El descubrimiento de que sobreexpresando un cóctel de factores de transcripción se puede revertir células completamente diferenciadas a un estado pluripotente, ha puesto un gran énfasis en la importancia de los mecanismos transcripcionales involucrados en el establecimiento y mantenimiento de la pluripotencia. Las células madre pluripotentes, como una fuente *in vitro* potencialmente ilimitada de todos y cada uno de los tipos celulares, son de gran interés para la medicina regenerativa; aún quedando por resolver cuestiones fundamentales acerca de las bases mecánicas de la pluripotencia celular y la transición ordenada a la diferenciación de progenitores. Su actividad continuada, su abundancia en los tejidos adultos y su distancia evolutiva con los vertebrados hacen del neoblasto de planaria un sistema modelo necesario para el estudio *in vivo* de la pluripotencia de células madre.

Los neoblastos son células madre adultas pluripotentes remarcables porque las células madre adultas en otros organismos modelo están restringidas a un linaje y la pluripotencia somática existe solo de forma transitoria durante el desarrollo



embrionario temprano. Además, son la única fuente de nuevas células en planarias y se dividen continuamente para reemplazar todos los tipos celulares diferenciados. Las planarias, por tanto, existen en un estado dinámico estable entre un único tipo celular proliferante y múltiples células diferenciadas de vida corta. El mantenimiento de este estado estable constituye un problema universal en cualquier tejido que contiene células madre y es un fenómeno poco entendido. Su completo recambio en cuestión de semanas hace a las planarias un modelo único para el estudio de los mecanismos subyacentes y los principios conceptuales. El recambio dinámico caracteriza también el intestino de vertebrados, por ejemplo.

La distancia filogenética de los neoblastos respecto a las células madre embrionarias y las células madre pluripotentes inducidas de mamífero proporcionan una oportunidad única para investigar si la pluripotencia aparece como un mecanismo conservado evolutivamente o si, en lugar de eso, emerge a partir de interacciones generales entre procesos celulares como una propiedad del sistema. Por ejemplo, la formación de células madre pluripotentes inducidas normalmente requiere la expresión de factores de transcripción con dominio POU como Oct4 y Sox2, además de Nanog. Sin embargo, no se ha encontrado un homólogo de Nanog en planaria. Por ello, otro tema importante de la investigación en planaria es qué genes y vías conservados dirigen a los neoblastos a través de la historia regenerativa de la vida; sugiriendo que la regeneración, incluso en escenarios naturales de no regeneración, requerirá la correcta redistribución de redes génicas bien conocidas. Hasta ahora, la mayor parte de la investigación se ha centrado en genes y redes conservadas, mientras que los posibles papeles de nuevos genes específicos de planarias tienen que ser todavía investigados.

Para la regeneración en planaria, los neoblastos deben:

1. Tener una capacidad indefinida de autorenovación y una homeostasis de la población de neoblastos.
2. Mantener la pluripotencia y la producción de una progenie que pueda diferenciarse en todas las células y tejidos.
3. Interpretar correctamente las señales de polaridad y posicionamiento para asegurar la integración funcional de los nuevos tejidos con los viejos.

Estos son los mismos mecanismos que han de ser manipulados para que pueda tener éxito una terapia regenerativa basada en células madre. La pluripotencia inducida tiene un gran potencial para proporcionar una fuente de células madre para la medicina regenerativa (Deng, 2010; Takahashi & Yamanaka, 2006), pero utilizar estas células con seguridad para reparar o reemplazar tejidos requeriría entender cómo las señales en los tejidos existentes afectan la diferenciación de las células madre y cómo pueden ser manipuladas para facilitar la integración funcional. La planaria como sistema proporciona una oportunidad ideal para el estudio de estas señales.

El mantenimiento homeostático de los tejidos de la planaria y la regeneración después de una lesión requiere un control fino de la proliferación, la muerte celular

y los procesos de autofagia (González-Estévez *et al.*, 2007; Pellettieri & Alvarado, 2007; Pellettieri *et al.*, 2010). Algunos datos sugieren que las vías metabólicas que controlan la proliferación, el crecimiento y el decrecimiento están probablemente conservados, y que las planarias representan un excelente sistema modelo para estudiar los aspectos básicos de estas vías, muchas de las cuales están directamente relacionadas con la biología del cáncer. Se ha sugerido también que los neoblastos son un modelo potencial para entender el control de la proliferación de las células madre, lo cual es relevante también en cáncer (Oviedo & Beane, 2009; Oviedo *et al.*, 2008; Pearson & Sánchez Alvarado, 2008).

La comparación de los neoblastos con células embrionarias y células madre pluripotentes inducidas podría identificar componentes o dianas conservadas, ayudando así al entendimiento mecanístico de la pluripotencia. La posibilidad alternativa, esto es, la pluripotencia como un epifenómeno que no dependa explícitamente de circuitos moleculares conservados, sería igualmente interesante. En este caso, las comparaciones podrían ayudar a revelar los procesos celulares generales que dan lugar colectivamente a la pluripotencia como una propiedad emergente del sistema. Más allá, la contribución mecanística de los cuerpos cromatoides y los gránulos germinales a la pluripotencia permanecen como una importante área para la futura exploración, a pesar del hecho de que ni las células madre embrionarias ni las células madre pluripotentes inducidas parecen contener cantidades relevantes de RNPs. Es importante remarcar que tanto las células madre embrionarias como las células madre pluripotentes inducidas son productos artificiales de laboratorio. En vertebrados al menos, la pluripotencia somática *in vivo* solo existe transitoriamente durante los primeros estadios del desarrollo embrionario. Los neoblastos y las células germinales, por otro lado, han evolucionado como células pluripotentes permanentes. Su dependencia de las RNPs podría, por tanto, reflejar una función accesoria pero crítica requerida para el mantenimiento a largo plazo de la pluripotencia. El papel establecido de los gránulos germinales y los cuerpos cromatoides de salvaguardar la integridad del genoma encajaría con esta hipótesis. La eliminación de marcas epigenéticas, que permanece incompleta en células reprogramadas, podría ser también una posibilidad. En general, los neoblastos de planaria, como células pluripotentes naturales, complementan idealmente los sistemas *in vitro* de vertebrado hacia el objetivo médico de explotar la pluripotencia.

Dos estudios recientes han comparado los neoblastos con células madre de vertebrados. Önal *et al.* (2012) han examinado estadísticamente la abundancia relativa de homólogos de factores de pluripotencia de vertebrados y genes diana conocidos de Oct4/Nanog entre los transcriptomas y proteomas de neoblasto (Önal *et al.*, 2012). En ambos casos, los autores han encontrado enriquecimientos estadísticamente significativos, consistentes con una amplia conservación evolutiva de la pluripotencia. De manera similar, Labbé y colaboradores han reportado grados de solapamiento particularmente altos entre genes asociados con pluripotencia en vertebrados y genes específicamente sobreexpresados en neoblastos (Labbé *et al.*, 2012). Estos hallazgos proporcionan un esperanzador punto de inicio para la disección comparativa de la pluripotencia a través de distancias evolutivas largas.

Es alentador que lo que ya conocemos sobre los requerimientos moleculares para el mantenimiento de los neoblastos, pluripotencia y diferenciación muestra una conservación molecular remarcable con componentes requeridos para estos procesos en mamíferos y otros animales. Las células madre embrionarias de mamífero tienen los mismos determinantes moleculares básicos de pluripotencia encontrados en los neoblastos de planaria, sugiriendo que el mecanismo se ha conservado durante la evolución (Önal *et al.*, 2012).

En conclusión, las características claves de los neoblastos son su capacidad de autorenovarse de manera indefinida, su totipotencia y la habilidad de su progenie de interpretar las señales de diferenciación y polaridad para reemplazar correctamente las estructuras después de un daño en el tejido. La regeneración en planarias ofrece un paradigma para entender el control molecular y celular de la reparación y la regeneración de los tejidos animales, y puede proporcionar conocimientos valiosos para el uso seguro de células madre para reparar tejidos humanos dañados, enfermos o envejecidos con poca o ninguna capacidad regenerativa.

¿Podrán todos estos conocimientos ser utilizados para mejorar la salud y la longevidad de los seres humanos? Será posible comparar la regeneración de órganos en planarias con el desarrollo y la reparación de órganos en otros animales para comprender cómo genes conservados son utilizados de manera diferente en animales con y sin capacidad de regenerar. Seremos capaces de derivar beneficios para los humanos a partir de la investigación de fenómenos como la desdiferenciación y la regeneración en los animales más simples. El conocimiento obtenido acerca de los mecanismos moleculares asociados a los procesos celulares y la caracterización de las biomoléculas involucradas en la regeneración empezarán a revelar las diferencias en las capacidades regenerativas entre los metazoos. Estos estudios pueden también proporcionar información valiosa para acabar con enfermedades humanas, incluso las relacionadas con el envejecimiento. La investigación en células madre y medicina regenerativa ya ha mostrado grandes avances, especialmente en modelos murinos de enfermedades humanas, incluyendo la formación de varios tipos celulares especializados y la generación de órganos completos *in vitro* (Sánchez Alvarado & Yamanaka, 2014; Tanabe *et al.*, 2014) que culminarán probablemente en el desarrollo de estrategias biomiméticas hacia la regeneración de tejidos y órganos no predispuestos para ello (Franco *et al.*, 2013). Si se pueden controlar los mecanismos moleculares que definen la diferenciación y migración de las células en una matriz, podría llegar a obtenerse tejidos y órganos sintéticos mediante la impresión 3D con células (Huh *et al.*, 2011; Marga *et al.*, 2007; Murphy & Atala, 2014; van der Meer *et al.*, 2013).

### **La planaria como organismo modelo en cáncer**

La planaria está constituyéndose en un modelo molecular y genético emergente capaz de proporcionar nuevas revelaciones acerca de la regeneración en tejidos adultos, la biología del cáncer y otras enfermedades (Gentile *et al.*, 2010). Como se comentó en el Capítulo 1, se han encontrado en planaria homólogos de genes supresores tumorales humanos como *p53* y *PTEN*, y nosotros hemos visto que hasta

29 de los 42 genes de neoblasto encontrados gracias a la aproximación basada en DGE están relacionados con cáncer en humanos.

Una mejor comprensión de las fuerzas que controlan el crecimiento celular es esencial para desarrollar terapias efectivas en medicina regenerativa y cáncer. Históricamente, la literatura ha relacionado cáncer y regeneración, proponiendo la regeneración tanto como fuente de cáncer como método para inhibir la tumorigénesis, incluso definiendo el cáncer como un proceso regenerativo incompleto (Beachy *et al.*, 2004; Flier *et al.*, 1986; Riss *et al.*, 2006; Schäfer & Werner, 2008; Waddington, 1935).

El término regeneración implica una restauración bien coordinada de las células, los tejidos y los órganos que se han perdido física o funcionalmente. Este proceso de reparación debe conseguir el reconocimiento y la recapitulación de las estructuras perdidas, mientras que consigue simultáneamente la integración funcional entre los tejidos recién formados y los existentes. Además, la regeneración que implica proliferación celular (epimorfosis) requiere señales que regulen el ciclo celular, de manera que el número de células que entran en división sea el estrictamente necesario para llevar a cabo la regeneración y no mayor. Controlar el crecimiento celular es una tarea fundamental tanto en cáncer como en regeneración epimórfica. Las células involucradas deben ser dirigidas con precisión a las áreas donde son requeridas y una vez la regeneración se ha completado deben haber señales de terminación. De otra manera, la respuesta inicial continuaría de forma indefinida, conllevando consecuencias desastrosas para la homeostasis del organismo. Sin embargo, este proceso no siempre es infalible, como demuestran las crecientes evidencias que asocian regeneración con anomalías celulares relacionadas con cáncer (Beachy *et al.*, 2004; Schäfer & Werner, 2008).

Nuestro conocimiento actual sobre el cáncer y el crecimiento celular anormal ha derivado en mayoritariamente de los estudios en vertebrados, especialmente en mamíferos, pero se han descrito aberraciones del crecimiento similares en casi todos los grandes filos de invertebrados (Scharrer & Szabó Lochhead, 1950). A pesar de ser un fenómeno común, nuestro conocimiento sobre el camino evolutivo de sus mecanismos de regulación y los genes supresores tumorales implicados es actualmente muy restringido (Pearson & Sánchez Alvarado, 2008). La planaria tiene gran potencial como organismo modelo para contribuir significativamente en estas áreas. A diferencia de otros modelos de invertebrados, las planarias han sido utilizadas históricamente para testar diversos compuestos farmacológicos y carcinogénicos (Agata & Watanabe, 1999; Alvarado & Tsonis, 2006; Newmark & Sánchez Alvarado, 2002; Oviedo *et al.*, 2008; Pearson *et al.*, 2009; Reddien & Sánchez Alvarado, 2004; Rossi *et al.*, 2007; Saló, 2006; Stevens *et al.*, 2015). Los resultados de muchos estudios han concluido que los neoblastos guardan similitud funcional tanto con células madre adultas de mamífero como con células embrionarias tempranas; particularmente, los neoblastos responden a agentes carcinogénicos formando tumores benignos y malignos, así como teratomas (Oviedo & Beane, 2009). También observan que esta respuesta carcinogénica, tanto en vertebrados como en invertebrados, evoluciona hacia tumores en tejidos que no regeneran; mientras que en tejidos con capacidad

regenerativa simplemente presentan cierta hiperplasia.

Así, las planarias ofrecen un modelo simplificado para estudiar tumorigénesis químicamente inducida, tanto en la complejidad de un organismo completo como durante la regeneración de múltiples tejidos. La planaria es el modelo ideal para investigar conexiones entre cáncer y regeneración. Comparaciones de los patrones de expresión génica entre cáncer y regeneración han revelado importantes diferencias en muchas vías metabólicas, desde aquellas asociadas con hipoxia y el insulin-like growth factor-I, a genes regulando morfogénesis (CRYM, TCF21, CTGF, etc.) y glicólisis (como PGK1 y HK1) (Riss *et al.*, 2006). Se han identificado en planaria muchos genes candidatos y vías metabólicas asociados con cáncer (Oviedo *et al.*, 2008; Pearson & Sánchez Alvarado, 2008). Por tanto, las planarias son potencialmente el primer modelo de regeneración a gran escala para diseccionar los detalles metabólicos de regeneración y cáncer.

En un futuro cercano, será necesario estudiar modelos animales con grandes capacidades regenerativas así como los mecanismos del cáncer, utilizando el creciente número de recursos moleculares y genómicos que han potenciado nuestra habilidad para manipular el crecimiento celular y diseccionar la maquinaria implicada.

# Conclusiones



# Conclusiones

1. Se ha generado una librería de anticuerpos monoclonales con los que se ha realizado un *immunoscreening* de proteínas de membrana plasmática que ha permitido caracterizar dos subtipos de neoblastos que corresponden a una población inicial y a otra final de su progenie.
2. Se ha obtenido el primer perfil proteico de los neoblastos, lo que ha permitido caracterizar un conjunto de proteínas relacionadas con el metabolismo, el estrés y la apoptosis, además de la proliferación y la diferenciación de los neoblastos.
3. Se ha generado un transcriptoma de referencia de libre acceso de *Schmidtea mediterranea* que ha permitido identificar una gran diversidad de familias génicas, resultando una herramienta imprescindible en la caracterización molecular de esta especie como organismo modelo de regeneración.
4. La aplicación de un análisis de *Digital Gene Expression* sobre el transcriptoma de referencia ha demostrado ser un método eficiente para la cuantificación transcripcional y la identificación de nuevos genes de neoblasto.
5. La aproximación anterior sobre diferentes poblaciones de neoblastos, su progenie y células diferenciadas separadas por FACS, nos ha permitido caracterizar nuevos factores de transcripción y genes relacionados con el cáncer.
6. La heterogeneidad transcripcional de la población X2 requiere de un método de discriminación mas profundo, como un estudio de RNA-Seq a nivel de células individuales (*single-cell* RNA-Seq).





# **Bibliografía**



# Bibliografía

- ABOUBAKER, A. A. (2011). Planarian stem cells: a simple paradigm for regeneration. *Trends Cell Biol* **21**(5), 304–311.
- ABRIL, J. F., CEBRIÀ, F., RODRÍGUEZ ESTEBAN, G., HORN, T., FRAGUAS, S., CALVO, B., BARTSCHERER, K. & SALÓ, E. (2010). Smed454 dataset: unravelling the transcriptome of *Schmidtea mediterranea*. *BMC Genomics* **11**(1), 731.
- ADAMIDI, C., WANG, Y., GRUEN, D., MASTROBUONI, G., YOU, X., TOLLE, D., DODT, M., MACKOWIAK, S. D., GOGOL-DOERING, A., OENAL, P., RYBAK, A., ROSS, E., SÁNCHEZ ALVARADO, A., KEMPA, S., DIETERICH, C., RAJEWSKY, N. & W, C. (2011). De novo assembly and validation of planaria transcriptome by massive parallel sequencing and shotgun proteomics. *Genome Res* **21**(7), 1193–1200.
- ADELL, T., CEBRIÀ, F. & SALÓ, E. (2009a). Gradients in planarian regeneration and homeostasis. *Cold Spring Harb Perspect Biol* **2**(1), a000505.
- ADELL, T., SALO, E., BOUTROS, M. & BARTSCHERER, K. (2009b). Smed-evi/wntless is required for  $\beta$ -catenin-dependent and -independent processes during planarian regeneration. *Development* **136**(6), 905–910.
- AEBERSOLD, R. & MANN, M. (2003). Mass spectrometry-based proteomics. *Nature* **422**(6928), 198–207.
- AGATA, K., NAKAJIMA, E., FUNAYAMA, N., SHIBATA, N., SAITO, Y. & UMESONO, Y. (2006). Two different evolutionary origins of stem cell systems and their molecular basis. *Semin Cell Dev Biol* **17**(4), 503–509.
- AGATA, K. & WATANABE, K. (1999). Molecular and cellular aspects of planarian regeneration. *Semin Cell Dev Biol* **10**(4), 377–383.
- ALI, S. R., HIPPEMEYER, S., SAADAT, L. V., LUO, L., WEISSMAN, I. L. & ARDEHALI, R. (2014). Existing cardiomyocytes generate cardiomyocytes at a low rate after birth in mice. *Proc Natl Acad Sci U S A* **111**(24), 8850–8855.
- ALVARADO, A. S. & TSONIS, P. A. (2006). Bridging the regeneration gap: genetic insights from diverse animal models. *Nat Rev Genet* **7**(11), 873–884.

## Bibliografía

- ANSORGE, W. J. (2009). Next-generation DNA sequencing techniques. *N Biotechnol* **25**(4), 195–203.
- ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. (2000). Gene ontology: tool for the unification of biology. *Nat Genet* **25**(1), 25–29.
- BAGUÑÀ, J. (1974). Dramatic mitotic response in planarians after feeding, and a hypothesis for the control mechanism. *J Exp Zool* **190**(1), 117–122.
- BAGUÑÀ, J. (1976). Mitosis in the intact and regenerating planarian *Dugesia mediterranea*. *J Exp Zool* **195**(1), 65–79.
- BAGUÑÀ, J. (2012). The planarian neoblast: the rambling history of its origin and some current black boxes. *Int J Dev Biol* **56**(1-2-3), 19–37.
- BAGUÑÀ, J., CARRANZA, S., PALA, M., RIBERA, C., GIRIBET, G., ARNEDO, M. A., RIBAS, M. & RIUTORT, M. (1999). From morphology and karyology to molecules. new methods for taxonomical identification of asexual populations of freshwater planarians. a tribute to Professor Mario Benazzi. *Ital J Zool* **66**(3), 207–214.
- BAGUÑÀ, J. & ROMERO, R. (1981). Quantitative analysis of cell types during growth, degrowth and regeneration in the planarians *Dugesia mediterranea* and *Dugesia tigrina*. *Hydrobiologia* **84**(1), 181–194.
- BAGUÑÀ, J., ROMERO, R., SALÓ, E., COLLET, J., AULADELL, C., RIBAS, M., RIUTORT, M., GARCÍA-FERNÁNDEZ, J., BURGAYA, F. & BUENO, D. (1990). Growth, degrowth and regeneration as developmental phenomena in adult freshwater planarians. *Experimental Embryology in Aquatic Plants and Animals* **195**, 129–162.
- BAGUÑÀ, J., SALÓ, E. & AULADELL, C. (1989). Regeneration and pattern formation in planarians. iii. evidence that neoblasts are totipotent stem cells and the source of blastema cells. *Development* **107**, 77–86.
- BAHARVAND, H., FATHI, A., GOURABI, H., MOLLAMOHAMMADI, S. & SALEKDEH, G. H. (2008). Identification of mouse embryonic stem cell-associated proteins. *J Proteome Res* **7**(1), 412–423.
- BARDEEN, C. R. & BAETJER, F. H. (1904). The inhibitive action of the Roentgen rays on regeneration in planarians. *J Exp Zool* **1**(1), 191–195.
- BEACHY, P. A., KARHADKAR, S. S. & BERMAN, D. M. (2004). Tissue repair and stem cell renewal in carcinogenesis. *Nature* **432**(7015), 324–331.

- BELY, A. E. & NYBERG, K. G. (2010). Evolution of animal regeneration: re-emergence of a field. *Trends Ecol Evol* **25**(3), 161–170.
- BELY, A. E. & SIKES, J. M. (2010). Acoel and platyhelminth models for stem-cell research. *J Biol* **9**(2), 14.
- BERGMANN, O., BHARDWAJ, R. D., BERNARD, S., ZDUNEK, S., BARNABE-HEIDER, F., WALSH, S., ZUPICICH, J., ALKASS, K., BUCHHOLZ, B. A., DRUID, H. & ET AL (2009). Evidence for cardiomyocyte renewal in humans. *Science* **324**(5923), 98–102.
- BEST, J. B. & MORITA, M. (1982). Planarians as a model system for in vitro teratogenesis studies. *Teratog Carcinog Mutagen* **2**(3–4), 277–291.
- BLACKSTONE, N. W. & JASKER, B. D. (2003). Phylogenetic considerations of clonality, coloniality, and mode of germline development in animals. *J Exp Zool B Mol Dev Evol* **297B**(1), 35–47.
- BLAKELEY, P., FOGARTY, N. M. E., DEL VALLE, I., WAMAITHA, S. E., HU, T. X., ELDER, K., SNELL, P., CHRISTIE, L., ROBSON, P. & NIAKAN, K. K. (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* **Epub ahead of print**.
- BLYTHE, M. J., KAO, D., MALLA, S., ROWSELL, J., WILSON, R., EVANS, D., JOWETT, J., HALL, A., LEMAY, V., LAM, S. & ABOOBAKER, A. (2010). A dual platform approach to transcript discovery for the planarian *Schmidtea mediterranea* to establish RNAseq for stem cell and regeneration biology. *PLoS One* **5**(12), e15617.
- BONUCCELLI, L., ROSSI, L., LENA, A., SCARCELLI, V., RAINALDI, G., EVANGELISTA, M., IACOPETTI, P., GREMIGNI, V. & SALVETTI, A. (2010). An RbAp48-like gene regulates adult stem cells in planarians. *J Cell Sci* **123**(5), 690–698.
- BÖSER, A., DREXLER, H. A., REUTER, H., SCHMITZ, H., WU, G., SCHÖLER, H., GENTILE, L. & BARTSCHERER, K. (2013). SILAC proteomics of planarians identifies Ncoa5 as a conserved component of pluripotent stem cells. *Cell Rep* **5**(4), 1142–1155.
- CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**(1), 421.
- CANTAREL, B. L., KORF, I., ROBB, S. M., PARRA, G., ROSS, E., MOORE, B., HOLT, C., SANCHEZ ALVARADO, A. & YANDELL, M. (2007). MA-KER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**(1), 188–196.

## Bibliografia

- CEBRIÀ, F. (2005). Planarian homologs of *netrin* and *netrin receptor* are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. *Development* **132**(16), 3691–3703.
- CEBRIÀ, F. (2007). Regenerating the central nervous system: how easy for planarians! *Dev Genes Evol* **217**(11-12), 733–748.
- CEBRIÀ, F. (2008). Organization of the nervous system in the model planarian *Schmidtea mediterranea*: an immunocytochemical study. *Neurosci Res* **61**(4), 375–384.
- CEBRIÀ, F., GUO, T., JOPEK, J. & NEWMARK, P. A. (2007). Regeneration and maintenance of the planarian midline is regulated by a *slit* orthologue. *Dev Biol* **307**(2), 394–406.
- CEBRIÀ, F. & NEWMARK, P. A. (2007). Morphogenesis defects are associated with abnormal nervous system regeneration following *roboA* RNAi in planarians. *Development* **134**(5), 833–837.
- CHANDEBOIS, R. (1980). The dynamics of wound closure and its role in the programming of planarian regeneration. *Dev Growth Differ* **22**(4), 693–704.
- CHERA, S., GHILA, L., DOBRETZ, K., WENGER, Y., BAUER, C., BUZGARIU, W., MARTINOU, J.-C. & GALLIOT, B. (2009). Apoptotic cells provide an unexpected source of Wnt3 signaling to drive *Hydra* head regeneration. *Dev Cell* **17**(2), 279–289.
- CLARKE, J., WU, H.-C., JAYASINGHE, L., PATEL, A., REID, S. & BAYLEY, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* **4**(4), 265–270.
- COLLET, J. & BAGUÑÀ, J. (1987). Síntesi proteica i expressió gènica durant la regeneració a la planària *Dugesia (G) tigrina*. *Biologia del Desenvolupament* **5**, 175–186.
- COLLET, J. & BAGUÑÀ, J. (1993). Optimizing a method of protein extraction for two-dimensional electrophoretic separation of proteins from planarians (*Platyhelminthes, Turbellaria*). *Electrophoresis* **14**, 1054–1059.
- COLLINS, J. J., HOU, X., ROMANOVA, E. V., LAMBRUS, B. G., MILLER, C. M., SABERI, A., SWEEDLER, J. V. & NEWMARK, P. A. (2010). Genome-wide analyses reveal a role for peptide hormones in planarian germline development. *PLoS Biol* **8**(10), e1000509.
- CONTE, M., DERI, P., ISOLANI, M. E., MANNINI, L. & BATISTONI, R. (2009). A mortalin-like gene is crucial for planarian stem cell viability. *Developmental Biology* **334**(1), 109–118.

- COOMBS, K. M. (2011). Quantitative proteomics of complex mixtures. *Expert Rev Proteomics* **8**(5), 659–677.
- CRICK, F. (1970). Central dogma of molecular biology. *Nature* **227**(5258), 561–563.
- DE GODOY, L. M. F., OLSEN, J. V., COX, J., NIELSEN, M. L., HUBNER, N. C., FRÖHLICH, F., WALTHER, T. C. & MANN, M. (2008). Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**(7217), 1251–1254.
- DE ROBERTIS, E. M. (2010). Wnt signaling in axial patterning and regeneration: lessons from planaria. *Sci Signal* **3**(127), pe21.
- DENG, W. (2010). Induced pluripotent stem cells: paths to new medicines. *EMBO reports* **11**(3), 161–165.
- DOMON, B. & AEBERSOLD, R. (2006). Mass spectrometry and protein analysis. *Science* **312**(5771), 212–217.
- EDDY, E. M. (1975). Germ plasm and the differentiation of the germ cell line. *Int Rev Cytol* **43**, 229–80327.
- EISENHOFFER, G. T., KANG, H. & SÁNCHEZ ALVARADO, A. (2008). Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell Stem Cell* **3**(3), 327–339.
- EWEN-CAMPEN, B., SCHWAGER, E. E. & EXTAVOUR, C. G. M. (2009). The molecular machinery of germ line specification. *Mol Reprod Dev* **77**(1), 3–18.
- EXTAVOUR, C. G. M. (2007). Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. *Integr Comp Biol* **47**(5), 770–785.
- FELIX, D. A. & ABOOBAKER, A. A. (2010). The TALE class homeobox gene *Smed-prep* defines the anterior compartment for head regeneration. *PLoS Genet* **6**(4), e1000915.
- FERNÁNDEZ-TABOADA, E., MÖRITZ, S., ZEUSCHNER, D., STEHLING, M., SCHOLER, H. R., SALÓ, E. & GENTILE, L. (2010). Smed-SmB, a member of the LSm protein superfamily, is essential for chromatoid body organization and planarian stem cell proliferation. *Development* **137**(7), 1055–1065.
- FINN, R. D., BATEMAN, A., CLEMENTS, J., COGGILL, P., EBERHARDT, R. Y., EDDY, S. R., HEGER, A., HETHERINGTON, K., HOLM, L., MISTRY, J., SONNHAMMER, E. L., TATE, J. & PUNTA, M. (2013). Pfam: the protein families database. *Nucleic Acids Res* **42**(D1), D222–D230.



## Bibliografía

- FLIER, J. S., UNDERHILL, L. H. & DVORAK, H. F. (1986). Tumors: wounds that do not heal. *N Engl J Med* **315**(26), 1650–1659.
- FLUSBERG, B. A., WEBSTER, D. R., LEE, J. H., TRAVERS, K. J., OLIVARES, E. C., CLARK, T. A., KORLACH, J. & TURNER, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Meth* **7**(6), 461–465.
- FORSTHOEFEL, D. J. & NEWMARK, P. A. (2009). Emerging patterns in planarian regeneration. *Curr Opin Genet Dev* **19**(4), 412–420.
- FORSTHOEFEL, D. J., PARK, A. E. & NEWMARK, P. A. (2011). Stem cell-based growth, regeneration, and remodeling of the planarian intestine. *Dev Biol* **356**(2), 445–459.
- FRAGUAS, S., BARBERAN, S., IGLESIAS, M., RODRIGUEZ-ESTEBAN, G. & CEBRIA, F. (2014). *egr-4*, a target of EGFR signaling, is required for the formation of the brain primordia and head regeneration in planarians. *Development* **141**(9), 1835–1847.
- FRAGUAS, S., BARBERÁN, S. & CEBRIÀ, F. (2011). EGFR signaling regulates cell proliferation, differentiation and morphogenesis during planarian regeneration and homeostasis. *Dev Biol* **354**(1), 87–101.
- FRAGUAS, S., BARBERÁN, S., IBARRA, B., STÖGER, L. & CEBRIÀ, F. (2012). Regeneration of neuronal cell types in *Schmidtea mediterranea*: an immunohistochemical and expression study. *Int J Dev Biol* **56**(1-2-3), 143–153.
- FRANCO, C., SOARES, R., PIRES, E., KOCI, K., ALMEIDA, A. M., SANTOS, R. & COELHO, A. V. (2013). Understanding regeneration through proteomics. *Proteomics* **13**(3-4), 686–709.
- FRIEDLÄNDER, M. R., ADAMIDI, C., HAN, T., LEBEDEVA, S., ISENBARGER, T. A., HIRST, M., MARRA, M., NUSBAUM, C., LEE, W. L., JENKIN, J. C., SÁNCHEZ ALVARADO, A., KIM, J. K. & RAJEWSKY, N. (2009). High-resolution profiling and discovery of planarian small RNAs. *Proc Natl Acad Sci U S A* **106**(28), 11546–11551.
- FU, X., FU, N., GUO, S., YAN, Z., XU, Y., HU, H., MENZEL, C., CHEN, W., LI, Y., ZENG, R. & KHAITOVICH, P. (2009). Estimating accuracy of RNA-Seq and microarrays with proteomics. *BMC Genomics* **10**(1), 161.
- FU, Y., LI, C., LU, S., ZHOU, W., TANG, F., XIE, X. S. & HUANG, Y. (2015). Uniform and accurate single-cell sequencing based on emulsion whole-genome amplification. *Proc Natl Acad Sci U S A* **112**(38), 11923–11928.

- GENTILE, L., CEBRIÀ, F. & BARTSCHERER, K. (2010). The planarian flatworm: an in vivo model for stem cell biology and nervous system regeneration. *Dis Model Mech* **4**(1), 12–19.
- GONZÁLEZ-ESTÉVEZ, C., ARSENI, V., THAMBYRAJAH, R. S., FELIX, D. A. & ABOOBAKER, A. A. (2009a). Diverse miRNA spatial expression patterns suggest important roles in homeostasis and regeneration in planarians. *Int J Dev Biol* **53**(4), 493–505.
- GONZÁLEZ-ESTÉVEZ, C., ARSENI, V., THAMBYRAJAH, R. S., FELIX, D. A. & ABOOBAKER, A. A. (2009b). Diverse miRNA spatial expression patterns suggest important roles in homeostasis and regeneration in planarians. *Int J Dev Biol* **53**(4), 493–505.
- GONZÁLEZ-ESTÉVEZ, C., FELIX, D. A., ABOOBAKER, A. A. & SALÓ, E. (2007). Gtdap-1 promotes autophagy and is required for planarian remodeling during regeneration and starvation. *Proc Natl Acad Sci U S A* **104**(33), 13373–13378.
- GONZÁLEZ-ESTÉVEZ, C., FELIX, D. A., RODRÍGUEZ-ESTEBAN, G. & ABOOBAKER, A. A. (2012a). Decreased neoblast progeny and increased cell death during starvation-induced planarian degrowth. *Int J Dev Biol* **56**(1-2-3), 83–91.
- GONZÁLEZ-ESTÉVEZ, C., FELIX, D. A., SMITH, M. D., PAPS, J., MORLEY, S. J., JAMES, V., SHARP, T. V. & ABOOBAKER, A. A. (2012b). SMG-1 and mTORC1 act antagonistically to regulate response to injury and growth in planarians. *PLoS Genet* **8**(3), e1002619.
- GRAUMANN, J., HUBNER, N. C., KIM, J. B., KO, K., MOSER, M., KUMAR, C., COX, J., SCHOLER, H. & MANN, M. (2007). Stable isotope labeling by amino acids in cell culture (SILAC) and proteome quantitation of mouse embryonic stem cells to a depth of 5,111 proteins. *Mol Cell Proteomics* **7**(4), 672–683.
- GRAVELEY, B. R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet* **17**(2), 100–107.
- GREENBAUM, D., COLANGELO, C., WILLIAMS, K. & GERSTEIN, M. (2003). Comparing protein abundance and mrna expression levels on a genomic scale. *Genome Biol* **4**(9), 117.
- GREENBAUM, D., JANSEN, R. & GERSTEIN, M. (2002). Analysis of mRNA expression and protein abundance data: an approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts. *Bioinformatics* **18**(4), 585–596.

## Bibliografía

- GREMIGNI, V. (1988). Planarian regeneration: an overview of some cellular mechanisms. *Zool Sci* **5**, 1153–1163.
- GRY, M., RIMINI, R., STRÖMBERG, S., ASPLUND, A., PONTÉN, F., UHLÉN, M. & NILSSON, P. (2009). Correlations between RNA and protein expression profiles in 23 human cell lines. *BMC Genomics* **10**(1), 365.
- GRÜN, D., LYUBIMOVA, A., KESTER, L., WIEBRANDS, K., BASAK, O., SASAKI, N., CLEVERS, H. & VAN OUDENAARDEN, A. (2015). Single-cell messenger rna sequencing reveals rare intestinal cell types. *Nature* **Published online**.
- GUIGÓ, R., FLICEK, P., ABRIL, J. F., REYMOND, A., LAGARDE, J., DENOEUD, F., ANTONARAKIS, S., ASHBURNER, M., BAJIC, V. B., BIRNEY, E., CASTELO, R., EYRAS, E., UCLA, C., GINGERAS, T. R., HARROW, J., HUBBARD, T., LEWIS, S. E. & REESE, M. G. (2006). EGASP: the human ENCODE Genome Annotation Assessment Project. *Genome Biol* **7**(Suppl 1), S2.
- GUO, T., PETERS, A. H. & NEWMARK, P. A. (2006). A bruno-like gene is required for stem cell maintenance in planarians. *Dev Cell* **11**(2), 159–169.
- GURLEY, K. A., ELLIOTT, S. A., SIMAKOV, O., SCHMIDT, H. A., HOLSTEIN, T. W. & ALVARADO, A. S. (2010). Expression of secreted Wnt pathway components reveals unexpected complexity of the planarian amputation response. *Dev Biol* **347**(1), 24–39.
- GYGI, S., ROCHON, Y., FRANZA, B. R. & AEBERSOLD, R. (1999a). Correlation between protein and mRNA abundance in yeast. *Dev Genes Evol* **19**(3), 1720–1730.
- GYGI, S. P., RIST, B., GERBER, S. A., TURECEK, F., GELB, M. H. & AEBERSOLD, R. (1999b). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* **17**(10), 994–999.
- GÖRG, A., OBERMAIER, C., BOGUTH, G., HARDER, A., SCHEIBE, B., WILDGRUBER, R. & WEISS, W. (2000). The current state of two-dimensional electrophoresis with immobilized ph gradients. *Electrophoresis* **21**(6), 1037–1053.
- GÖRG, A., WEISS, W. & DUNN, M. J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics* **4**(12), 3665–3685.
- HANDBERG-THORSAGER, M. & SALÓ, E. (2007). The planarian nanos-like gene *Smednos* is expressed in germline and eye precursor cells during development and regeneration. *Dev Genes Evol* **217**(5), 403–411.
- HARROW, J., FRANKISH, A., GONZÁLEZ, J. M., TAPANARI, E., DIEKHANS, M., KOKOCINSKI, F., AKEN, B. L., BARRELL, D., ZADISSA, A., SEARLE, S., BARNES, I., BIGNELL, A., BOYCHENKO, V., HUNT, T., KAY, M.,

- MUKHERJEE, G., RAJAN, J., DESPACIO-REYES, G., SAUNDERS, G., STEWARD, C., HARTE, R., LIN, M., HOWALD, C., TANZER, A., DERRIEN, T., CHRAST, J., WALTERS, N., BALASUBRAMANIAN, S., PEI, B., TRESS, M., RODRÍGUEZ, J. M., EZKURDIA, I., VAN BAREN, J., BRENT, M., HAUSSELER, D., KELLIS, M., VALENCIA, A., REYMOND, A., GERSTEIN, M., GUIGÓ, R. & HUBBARD, T. J. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**(9), 1760–1774.
- HAY, E. D. & COWARD, S. J. (1975). Fine structure studies on the planarian, *Dugesia*: I. nature of the “neoblast” and other cell types in noninjured worms. *J Ultrastruct Res* **50**(1), 1–21.
- HAYASHI, T. & AGATA, K. (2012). A unique FACS method to isolate stem cells in planarian. *Methods Mol Biol* **879**, 29–37.
- HAYASHI, T., ASAMI, M., HIGUCHI, S., SHIBATA, N. & AGATA, K. (2006). Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Dev Growth Differ* **48**(6), 371–380.
- HAYASHI, T., SHIBATA, N., OKUMURA, R., KUDOME, T., NISHIMURA, O., TARUI, H. & AGATA, K. (2010). Single-cell gene profiling of planarian stem cells using fluorescent activated cell sorting and its “index sorting” function for stem cell research. *Dev Growth Differ* **52**(1), 131–144.
- HIGUCHI, S., HAYASHI, T., HORI, I., SHIBATA, N., SAKAMOTO, H. & AGATA, K. (2007). Characterization and categorization of fluorescence activated cell sorted planarian stem cells by ultrastructural analysis. *Dev Growth Differ* **49**(7), 571–581.
- HOFFROGGE, R., MIKKAT, S., SCHARF, C., BEYER, S., CHRISTOPH, H., PAHNKE, J., MIX, E., BERTH, M., UHRMACHER, A., ZUBRZYCKI, I. Z., MILJAN, E., VÖLKER, U. & ROLFS, A. (2006). 2-de proteome analysis of a proliferating and differentiating human neuronal stem cell line (ReNcell VM). *Proteomics* **6**(6), 1833–1847.
- HOHEISEL, J. D. (2006). Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* **7**(3), 200–210.
- HUANG, L., MA, F., CHAPMAN, A., LU, S. & XIE, X. S. (2015). Single-cell whole-genome amplification and sequencing: methodology and applications. *Annu Rev Genomics Hum Genet* **16**(1), 79–102.
- HUGHES, C. S., NUHN, A. A., POSTOVIT, L. M. & LAJOIE, G. A. (2011). Proteomics of human embryonic stem cells. *Proteomics* **11**(4), 675–690.
- HUH, D., HAMILTON, G. A. & INGBER, D. E. (2011). From 3D cell culture to organs-on-chips. *Trends Cell Biol* **21**(12), 745–754.

## Bibliografía

- ISHIZUKA, H., MAEZAWA, T., KAWAUCHI, J., NODONO, H., HIRAO, Y., NISHIMURA, O., NAKAGAWA, H., SEKII, K., TASAKA, K., TARUI, H., AGATA, K., HOSHI, M., KOBAYASHI, K., SAKAKIBARA, Y. & MATSUMOTO, M. (2007). The *Dugesia ryukyuensis* database as a molecular resource for studying switching of the reproductive system. *Zoolog Sci* **24**(1), 31–37.
- ISOLANI, M. E., ABRIL, J. F., SALÓ, E., DERI, P., BIANUCCI, A. M. & BATISTONI, R. (2013). Planarians as a model to assess in vivo the role of matrix metalloproteinase genes during homeostasis and regeneration. *PLoS One* **8**(2), e55649.
- JAITIN, D. A., KENIGSBERG, E., KEREN-SHAUL, H., ELEFANT, N., PAUL, F., ZARETSKY, I., MILDNER, A., COHEN, N., JUNG, S., TANAY, A. & AMIT, I. (2014). Massively parallel single-cell RNA-Seq for marker-free decomposition of tissues into cell types. *Science* **343**(6172), 776–779.
- JONES, C. E., BROWN, A. L. & BAUMANN, U. (2007). Estimating the annotation error rate of curated GO database sequence annotations. *BMC Bioinformatics* **8**(1), 170.
- KANEHISA, M., GOTO, S., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. (2013). Data, information, knowledge and principle: back to metabolism in kegg. *Nucleic Acids Res* **42**(D1), D199–D205.
- KANG, H. & ALVARADO, A. S. (2009). Flow cytometry methods for the study of cell-cycle parameters of planarian stem cells. *Dev Dyn* **238**(5), 1111–1117.
- KAO, D., FELIX, D. & ABOOBAKER, A. (2013). The planarian regeneration transcriptome reveals a shared but temporally shifted regulatory program between opposing head and tail scenarios. *BMC Genomics* **16**(14), 797.
- KING, R. S. & NEWMARK, P. A. (2013). In situ hybridization protocol for enhanced detection of gene expression in the planarian *Schmidtea mediterranea*. *BMC Dev Biol* **13**(1), 8.
- KOTAJA, N. (2006). Interplay of PIWI/Argonaute protein MIWI and kinesin KIF17b in chromatoid bodies of male germ cells. *J Cell Sci* **119**(13), 2819–2825.
- KUKURBA, K. R. & MONTGOMERY, S. B. (2015). RNA sequencing and analysis. *Cold Spring Harb Protoc* **2015**.
- KUMAR, A. & BROCKES, J. P. (2012). Nerve dependence in tissue, organ, and appendage regeneration. *Trends Neurosci* **35**(11), 691–699.
- KUMAR, C. & MANN, M. (2009). Bioinformatics analysis of mass spectrometry-based proteomics data sets. *FEBS Lett* **583**(11), 1703–1712.

- LABBÉ, R. M., IRIMIA, M., CURRIE, K. W., LIN, A., ZHU, S. J., BROWN, D. D., ROSS, E. J., VOISIN, V., BADER, G. D., BLENCOWE, B. J. & J, P. B. (2012). A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**(8), 1734–1745.
- LAPAN, S. W. & REDDIEN, P. W. (2011). *dlx* and *sp6-9* control optic cup regeneration in a prototypic eye. *PLoS Genet* **7**(8), e1002226.
- LARBI, N. B. & JEFFERIES, C. (2009). 2D-DIGE: comparative proteomics of cellular signalling pathways. *Methods Mol Biol* **517**, 105–132.
- LARSEN, P. A., HEILMAN, A. M. & YODER, A. D. (2014). The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms. *BMC Genomics* **15**(1), 720.
- LEUNG, Y. F. & CAVALIERI, D. (2003). Fundamentals of cDNA microarray data analysis. *Trends Genet* **19**(11), 649–659.
- LI, Q., YANG, H. & ZHONG, T. P. (2015). Regeneration across metazoan phylogeny: lessons from model organisms. *J Genet Genomics* **42**(2), 57–70.
- LU, Y.-C., SMIELEWSKA, M., PALAKODETI, D., LOVCI, M. T., AIGNER, S., YEO, G. W. & GRAVELEY, B. R. (2009). Deep sequencing identifies new and regulated microRNAs in *Schmidtea mediterranea*. *RNA* **15**(8), 1483–1491.
- LÓPEZ-OTÍN, C., BLASCO, M. A., PARTRIDGE, L., SERRANO, M. & KROEMER, G. (2013). The hallmarks of aging. *Cell* **153**(6), 1194–1217.
- MAIER, T., GÜELL, M. & SERRANO, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Lett* **583**(24), 3966–3973.
- MANN, M. & PANDEY, A. (2000). Proteomics to study genes and genomes. *Nature* **405**(6788), 837–846.
- MARGA, F., NEAGU, A., KOSZTIN, I. & FORGACS, G. (2007). Developmental biology and tissue engineering. *Birth Defects Res* **81**(4), 320–328.
- MARIONI, J. C., MASON, C. E., MANE, S. M., STEPHENS, M. & GILAD, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* **18**(9), 1509–1517.
- MAURER, M. H., FELDMANN JR, R. E., FÜTTERER, C. D., BUTLIN, J. & KUSCHINSKY, W. (2004). Comprehensive proteome expression profiling of undifferentiated versus differentiated neural stem cells from adult rat hippocampus. *Neurochem Res* **29**(6), 1129–1144.
- MILLER, I., CRAWFORD, J. & GIANAZZA, E. (2006). Protein stains for proteomic applications: which, when, why? *Proteomics* **6**(20), 5385–5408.

## Bibliografía

- MINOCHE, A. E., DOHM, J. C., SCHNEIDER, J., HOLTGRÄWE, D., VIEHÖVER, P., MONTFORT, M., ROSLEFF SÖRENSEN, T., WEISSHAAR, B. & HIMMELBAUER, H. (2015). Exploiting single-molecule transcript sequencing for eukaryotic gene prediction. *Genome Biol* **16**(1), 184.
- MITCHELL, A., CHANG, H.-Y., DAUGHERTY, L., FRASER, M., HUNTER, S., LOPEZ, R., MCANULLA, C., MCMENAMIN, C., NUKA, G., PESSEAT, S., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., RATO, C., YONG, S. Y., BATEMAN, A., PUNTA, M., ATTWOOD, T. K., SIGRIST, C. J., REDASCHI, N., RIVOIRE, C., XENARIOS, I., KAHN, D., GUYOT, D., BORK, P., LETUNIC, I., GOUGH, J., OATES, M., HAFT, D., HUANG, H., NATALE, D. A., WU, C. H., ORENGO, C., SILLITOE, I., MI, H., THOMAS, P. D. & FINN, R. D. (2014). The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* **43**(D1), D213–D221.
- MIYAOKA, Y. & MIYAJIMA, A. (2013). To divide or not to divide: revisiting liver regeneration. *Cell Div* **8**(1), 8.
- MOLINA, M. D., SALÓ, E. & CEBRIÀ, F. (2007). The BMP pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. *Dev Biol* **311**(1), 79–94.
- MORGAN, T. H. (1898). Experimental studies of the regeneration of *Planaria maculata*. *Archiv für Entwicklungsmechanik der Organismen* **7**(2–3), 364–397.
- MORGAN, T. H. (1901). *Regeneration*. Smithsonian Institution.
- MORITA, M. & BEST, J. B. (1984). Electron microscopic studies of planarian regeneration. iv. cell division of neoblasts in *Dugesia dorotocephala*. *J Exp Zool* **229**(3), 425–436.
- MORITZ, S., STÖCKLE, F., ORTMEIER, C., SCHMITZ, H., RODRÍGUEZ-ESTEBAN, G., KEY, G. & GENTILE, L. (2012). Heterogeneity of planarian stem cells in the S/G2/M phase. *Int J Dev Biol* **56**(1-2-3), 117–125.
- MORTAZAVI, A., WILLIAMS, B. A., McCUE, K., SCHAEFFER, L. & WOLD, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5**(7), 621–628.
- MOUTON, S., WILLEMS, M., BRAECKMAN, B. P., EGGER, B., LADURNER, P., SCHÄRER, L. & BORGONIE, G. (2009). The free-living flatworm *Macrostomum lignano*: a new model organism for ageing research. *Exp Gerontol* **44**(4), 243–249.
- MOUTON, S., WILLEMS, M., HOUTHOOFD, W., BERT, W. & BRAECKMAN, B. P. (2011). Lack of metabolic ageing in the long-lived flatworm *Schmidtea polychroa*. *Exp Gerontol* **46**(9), 755–761.

- MURPHY, S. V. & ATALA, A. (2014). 3D bioprinting of tissues and organs. *Nat Biotechnol* **32**(8), 773–785.
- NCBI GENOME DATABASE: SCHMIDTEA MEDITERRANEA (2014). <http://www.ncbi.nlm.nih.gov/genome/232> .
- NEWMARK, P., WANG, Y. & CHONG, T. (2008). Germ cell specification and regeneration in planarians. *Cold Spring Harb Symp Quant Biol* **73**(0), 573–581.
- NEWMARK, P. A. & SÁNCHEZ ALVARADO, A. (2000). Bromodeoxyuridine specifically labels the regenerative stem cells of planarians. *Dev Biol* **220**(2), 142–153.
- NEWMARK, P. A. & SÁNCHEZ ALVARADO, A. (2002). Not your father's planarian: a classic model enters the era of functional genomics. *Nat Rev Genet* **3**(3), 210–219.
- NING, K., FERMIN, D. & NESVIZHSHKII, A. I. (2012). Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. *J Proteome Res* **11**(4), 2261–2271.
- ONG, S.-E., BLAGOEV, B., KRATCHMAROVA, I., KRISTENSEN, D. B., STEEN, H., PANDEY, A. & MANN, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**(5), 376–386.
- ONG, S.-E. & MANN, M. (2005). Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **1**(5), 252–262.
- ORII, H., SAKURAI, T. & WATANABE, K. (2005). Distribution of the stem cells (neoblasts) in the planarian *Dugesia japonica*. *Dev Genes Evol* **215**(3), 143–157.
- ORII, H. & WATANABE, K. (2007). Bone morphogenetic protein is required for dorso-ventral patterning in the planarian *Dugesia japonica*. *Dev Growth Differ* **49**(4), 345–349.
- OVIEDO, N. J. & BEANE, W. S. (2009). Regeneration: the origin of cancer or a possible cure? *Semin Cell Dev Biol* **20**(5), 557–564.
- OVIEDO, N. J. & LEVIN, M. (2007). *smedinx-11* is a planarian stem cell gap junction gene required for regeneration and homeostasis. *Development* **134**(17), 3121–3131.
- OVIEDO, N. J., MOROKUMA, J., WALENTEK, P., KEMA, I. P., GU, M. B., AHN, J.-M., HWANG, J. S., GOJOBORI, T. & LEVIN, M. (2010). Long-range neural and gap junction protein-mediated cues control polarity during planarian regeneration. *Dev Biol* **339**(1), 188–199.



## Bibliografía

- OVIEDO, N. J., NEWMARK, P. A. & SÁNCHEZ ALVARADO, A. (2003). Allometric scaling and proportion regulation in the freshwater planarian *Schmidtea mediterranea*. *Dev Dyn* **226**(2), 326–333.
- OVIEDO, N. J., PEARSON, B. J., LEVIN, M. & SÁNCHEZ ALVARADO, A. (2008). Planarian PTEN homologs regulate stem cells and regeneration through TOR signaling. *Dis Model Mech* **1**(2-3), 131–143.
- PALAKODETI, D. (2006). MicroRNAs from the planarian *Schmidtea mediterranea*: a model system for stem cell biology. *RNA* **12**(9), 1640–1649.
- PALAKODETI, D., SMIELEWSKA, M., LU, Y.-C., YEO, G. W. & GRAVELEY, B. R. (2008). The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA* **14**(6), 1174–1186.
- PAN, F. C., BANKAITIS, E. D., BOYER, D., XU, X., VAN DE CASTEELE, M., MAGNUSON, M. A., HEIMBERG, H. & WRIGHT, C. V. E. (2013). Spatiotemporal patterns of multipotentiality in Ptf1a-expressing cells during pancreas organogenesis and injury-induced facultative restoration. *Development* **140**(4), 751–764.
- PAN, Q., SHAI, O., LEE, L. J., FREY, B. J. & BLENCOWE, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**(12), 1413–1415.
- PANDEY, A. & MANN, M. (2000). Proteomics to study genes and genomes. *Nature* **405**(6788), 837–846.
- PAPPIN, D. J., HOJRUP, P. & BLEASBY, A. J. (2013). Rapid identification of proteins by peptide-mass fingerprinting. *Curr Biol* **3**(6), 327–332.
- PARVINEN, M. (2005). The chromatoid body in spermatogenesis. *Int J Androl* **28**(4), 189–201.
- PATTON, W. F. & BEECHEM, J. M. (2002). Rainbow's end: the quest for multiplexed fluorescence quantitative analysis in proteomics. *Curr Opin Chem Biol* **6**(1), 63–69.
- PEARSON, B. & SÁNCHEZ ALVARADO, A. (2008). Regeneration, stem cells, and the evolution of tumor suppression. *Cold Spring Harb Symp Quant Biol* **73**(0), 565–572.
- PEARSON, B. J. & ALVARADO, A. S. (2009). A planarian p53 homolog regulates proliferation and self-renewal in adult stem cell lineages. *Development* **137**(2), 213–221.

- PEARSON, B. J., EISENHOFER, G. T., GURLEY, K. A., RINK, J. C., MILLER, D. E. & SÁNCHEZ ALVARADO, A. (2009). Formaldehyde-based whole-mount in situ hybridization method for planarians. *Dev Dyn* **238**(2), 443–450.
- PEIRIS, T. H., WECKERLE, F., OZAMOTO, E., RAMIREZ, D., DAVIDIAN, D., GARCIA-OJEDA, M. E. & OVIEDO, N. J. (2012). TOR signaling regulates planarian stem cells and controls localized and organismal growth. *J Cell Sci* **125**(7), 1657–1665.
- PELETTIERI, J. & ALVARADO, A. S. (2007). Cell turnover and adult tissue homeostasis: from humans to planarians. *Annu Rev Genet* **41**(1), 83–105.
- PELETTIERI, J., FITZGERALD, P., WATANABE, S., MANCUSO, J., GREEN, D. R. & SÁNCHEZ ALVARADO, A. (2010). Cell death and tissue remodeling in planarian regeneration. *Dev Biol* **338**(1), 76–85.
- PERVOUCHINE, D. D., KNOWLES, D. G. & GUIGO, R. (2012). Intron-centric estimation of alternative splicing from RNA-seq data. *Bioinformatics* **29**(2), 273–274.
- PETER, R., GSCHWENTNER, R., SCHURMANN, W., RIEGER, R. M. & LADURNER, P. (2003). The significance of stem cells in free-living flatworms: one common source for all cells in the adult. *J Appl Med* **2**, 21–35.
- PETERSEN, C. P. & REDDIEN, P. W. (2009). Wnt signaling and the polarity of the primary body axis. *Cell* **139**(6), 1056–1068.
- PETRALIA, R. S., MATTSO, M. P. & YAO, P. J. (2014). Aging and longevity in the simplest animals and the quest for immortality. *Ageing Res Rev* **16**(0), 66–82.
- QIN, Y.-F., ZHAO, J.-M., BAO, Z.-X., ZHU, Z.-Y., MAI, J., HUANG, Y.-B., LI, J.-B., CHEN, G., LU, P., CHEN, S.-J., SU, L. L., FANG, H. M., LU, J. K., ZHANG, Y. Z. & T, Z. S. (2012). Identification of small non-coding RNAs in the planarian *Dugesia japonica* via deep sequencing. *Genomics* **99**(5), 315–321.
- RANDO, T. A. (2006). Stem cells, ageing and the quest for immortality. *Nature* **441**(7097), 1080–1086.
- REDDIEN, P. W., BERMANGE, A. L., KICZA, A. M. & SANCHEZ ALVARADO, A. (2007). BMP signaling regulates the dorsal planarian midline and is needed for asymmetric regeneration. *Development* **134**(22), 4043–4051.
- REDDIEN, P. W., BERMANGE, A. L., MURFITT, K. J., JENNINGS, J. R. & SÁNCHEZ ALVARADO, A. (2005a). Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Dev Cell* **8**(5), 635–649.

## Bibliografía

- REDDIEN, P. W., OVIEDO, N. J., R, J. J., C, J. J. & ALVARADO, A. S. (2005b). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science* **310**(5752), 1327–1330.
- REDDIEN, P. W. & SÁNCHEZ ALVARADO, A. (2004). Fundamentals of planarian regeneration. *Annu Rev Cell Dev Biol* **20**(1), 725–757.
- RESCH, A. M. & PALAKODETI, D. (2012). Small RNA pathways in *Schmidtea mediterranea*. *Int J Dev Biol* **56**(1-2-3), 67–74.
- RESCH, A. M., PALAKODETI, D., LU, Y. C., HOROWITZ, M. & GRAVELEY, B. R. (2012). Transcriptome analysis reveals strain-specific and conserved stemness genes in *Schmidtea mediterranea*. *PLoS One* **7**(4), e34447.
- RINK, J. C. (2012). Stem cell systems and regeneration in planaria. *Dev Genes Evol* **223**(1-2), 67–84.
- RINK, J. C., VU, H. T.-K. & ALVARADO, A. S. (2011). The maintenance and regeneration of the planarian excretory system are regulated by EGFR signaling. *Development* **138**(17), 3769–3780.
- RISS, J., KHANNA, C., KOO, S., CHANDRAMOULI, G. V., YANG, H. H., HU, Y., KLEINER, D. E., ROSENWALD, A., SCHAEFER, C. F., BEN-SASSON, S. A., YANG, L., POWELL, J., KANE, D. W., STAR, R. A., APRELIKOVA, O., BAUER, K., VASSELLI, J. R., MARANCHIE, J. K., KOHN, K. W., BUETOW, K. H., LINEHAN, W. M., WEINSTEIN, J. N., LEE, M. P., KLAUSNER, R. D. & BARRETT, J. C. (2006). Cancers as wounds that do not heal: differences and similarities between renal regeneration/repair and renal cell carcinoma. *Cancer Res* **66**(14), 7216–7224.
- ROBB, S. M. C., GOTTING, K., ROSS, E. & SÁNCHEZ ALVARADO, A. (2015). SmedGD 2.0: the *Schmidtea mediterranea* genome database. *Genesis* **53**(8), 535–546.
- ROBERTS, R. J., CARNEIRO, M. O. & SCHATZ, M. C. (2013). The advantages of SMRT sequencing. *Genome Biol* **14**(6), 405.
- ROBERTS-GALBRAITH, R. H. & NEWMARK, P. A. (2015). On the organ trail: insights into organ regeneration in the planarian. *Curr Opin Genet Dev* **32**, 37–46.
- RODRÍGUEZ-ESTEBAN, G., GONZÁLEZ-SASTRE, A., ROJO-LAGUNA, J. I., SALÓ, E. & ABRIL, J. F. (2015). Digital gene expression approach over multiple RNA-Seq data sets to detect neoblast transcriptional changes in *Schmidtea mediterranea*. *BMC Genomics* **16**(1), 361.

- ROSSI, L., SALVETTI, A., LENA, A., BATISTONI, R., DERI, P., PUGLIESI, C., LORETI, E. & GREMIGNI, V. (2006). *DjPiwi-1*, a member of the *PAZ-Piwi* gene family, defines a subpopulation of planarian stem cells. *Dev Genes Evol* **216**(6), 335–346.
- ROSSI, L., SALVETTI, A., MARINCOLA, F. M., LENA, A., DERI, P., MANNINI, L., BATISTONI, R., WANG, E. & GREMIGNI, V. (2007). Deciphering the molecular machinery of stem cells: a look at the neoblast gene expression profile. *Genome Biol* **8**(4), R62.
- ROUHANA, L., SHIBATA, N., NISHIMURA, O. & AGATA, K. (2010). Different requirements for conserved post-transcriptional regulators in planarian regeneration and stem cell maintenance. *Dev Biol* **341**(2), 429–443.
- ROUHANA, L., VIEIRA, A. P., ROBERTS-GALBRAITH, R. H. & A, N. P. (2012). PRMT5 and the role of symmetrical dimethylarginine in chromatoid bodies of planarian stem cells. *Development* **139**(6), 1083–1094.
- SAFFMAN, E. E. & LASKO, P. (1999). Germline development in vertebrates and invertebrates. *Cell Mol Life Sci* **55**(8–9), 1141–63.
- SALVETTI, A. (2005). *DjPum*, a homologue of *Drosophila Pumilio*, is essential to planarian stem cell maintenance. *Development* **132**(8), 1863–1874.
- SALVETTI, A., ROSSI, L., BONUCCELLI, L., LENA, A., PUGLIESI, C., RAINALDI, G., EVANGELISTA, M. & GREMIGNI, V. (2009). Adult stem cell plasticity: neoblast repopulation in non-lethally irradiated planarians. *Dev Biol* **328**(2), 305–314.
- SALVETTI, A., ROSSI, L., DERI, P. & BATISTONI, R. (2000). An MCM2-related gene is expressed in proliferating cells of intact and regenerating planarians. *Dev Dyn* **218**(4), 603–614.
- SALÓ, E. (2006). The power of regeneration and the stem-cell kingdom: freshwater planarians (Platyhelminthes). *Bioessays* **28**(5), 546–559.
- SALÓ, E. & BAGUÑÀ, J. (1984). Regeneration and pattern formation in planarians. i. the pattern of mitosis in anterior and posterior regeneration in *Dugesia (G) tigrina*, and a new proposal for blastema formation. *J Embryol Exp Morphol* **83**, 63–80.
- SANDMANN, T., VOGG, M. C., OWLARN, S., BOUTROS, M. & BARTSCHE-RER, K. (2011). The head-regeneration transcriptome of the planarian *Schmidtea mediterranea*. *Genome Biol* **12**(8), R76.
- SASIDHARAN, V., LU, Y.-C., BANSAL, D., DASARI, P., PODUVAL, D., SESHASAYEE, A., RESCH, A. M., GRAVELEY, B. R. & PALAKODETI, D.

## Bibliografía

- (2013). Identification of neoblast- and regeneration-specific *miRNAs* in the planarian *Schmidtea mediterranea*. *RNA* **19**(10), 1394–1404.
- SATO, K., SHIBATA, N., ORII, H., AMIKURA, R., SAKURAI, T., AGATA, K., KOBAYASHI, S. & WATANABE, K. (2006). Identification and origin of the germline stem cells as revealed by the expression of nanos-related gene in planarians. *Dev Growth Differ* **48**(9), 615–628.
- SCADDEN, D. (2014). Nice neighborhood: emerging concepts of the stem cell niche. *Cell* **157**(1), 41–50.
- SCHARRER, B. & SZABÓ LOCHHEAD, M. (1950). Tumors in the invertebrates: a review. *Cancer Res* **10**, 403.
- SCHÄFER, M. & WERNER, S. (2008). Cancer as an overhealing wound: an old hypothesis revisited. *Nat Rev Mol Cell Biol* **9**(8), 628–638.
- SCIMONE, M. L., MEISEL, J. & REDDIEN, P. W. (2010). The Mi-2-like *Smed-CHD4* gene is required for stem cell differentiation in the planarian *Schmidtea mediterranea*. *Development* **137**(8), 1231–1241.
- SCIMONE, M. L., SRIVASTAVA, M., BELL, G. W. & REDDIEN, P. W. (2011). A regulatory program for excretory system regeneration in planarians. *Development* **138**(20), 4387–4398.
- SENYO, S. E., STEINHAUSER, M. L., PIZZIMENTI, C. L., YANG, V. K., CAI, L., WANG, M., WU, T.-D., GUERQUIN-KERN, J.-L., LECHENE, C. P. & LEE, R. T. (2012). Mammalian heart renewal by pre-existing cardiomyocytes. *Nature* **493**(7432), 433–436.
- SHARON, D., TILGNER, H., GRUBERT, F. & SNYDER, M. (2013). A single-molecule long-read survey of the human transcriptome. *Nat Biotechnol* **31**(11), 1009–1014.
- SHIBATA, N., HAYASHI, T., FUKUMURA, R., FUJII, J., KUDOME-TAKAMATSU, T., NISHIMURA, O., SANO, S., SON, F., SUZUKI, N., ARAKI, R., ABE, M. & AGATA, K. (2012). Comprehensive gene expression analyses in pluripotent stem cells of a planarian, *Dugesia japonica*. *Int J Dev Biol* **56**(1-2-3), 93–102.
- SHIBATA, N., ROUHANA, L. & AGATA, K. (2010). Cellular and molecular dissection of pluripotent adult somatic stem cells in planarians. *Dev Growth Differ* **52**(1), 27–41.
- SHIBATA, N., UMESONO, Y., ORII, H., SAKURAI, T., WATANABE, K. & AGATA, K. (1999). Expression of vasa(vas)-related genes in germline cells and totipotent somatic stem cells of planarians. *Dev Biol* **206**(1), 73–87.

- SOLANA, J. (2013). Closing the circle of germline and stem cells: the Primordial Stem Cell hypothesis. *EvoDevo* **4**(1), 2.
- SOLANA, J., KAO, D., MIHAYLOVA, Y., JABER-HIJAZI, F., MALLA, S., WILSON, R. & ABOOBAKER, A. (2012). Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNA-seq, RNA interference and irradiation approach. *Genome Biol* **13**(3), R19.
- SOLANA, J., LASKO, P. & ROMERO, R. (2009). *Spoltud-1* is a chromatoid body component required for planarian long-term stem cell self-renewal. *Dev Biol* **328**(2), 410–421.
- STASTNA, M., ABRAHAM, M. R. & VAN EYK, J. E. (2009). Cardiac stem/progenitor cells, secreted proteins, and proteomics. *FEBS Lett* **583**(11), 1800–1807.
- STEIJGER, T., ABRIL, J. F., ENGSTRÖM, P. G., KOKOCINSKI, F., THE RGASP CONSORTIUM, HUBBARD, T. J., GUIGÓ, R., HARROW, J. & BERTONE, P. (2013). Assessment of transcript reconstruction methods for rna-seq. *Nat Meth* **10**(12), 1177–1184.
- STEINBOCK, L. J. & RADENOVIC, A. (2015). The emergence of nanopores in next-generation sequencing. *Nanotechnology* **26**(7), 074003.
- STEVENS, A. S., PIROTTE, N., WOUTERS, A., ROTEN, A. V., BELLEGHEM, F. V., WILLEMS, M., CUYPERS, A., T, A. & K1, S. (2015). Redox-related mechanisms to rebalance cancer-deregulated cell growth the anticarcinogenic mechanisms of regeneration. *Curr Drug Targets* **Epub ahead of print**.
- STROME, S. & LEHMANN, R. (2007). Germ versus soma decisions: lessons from flies and worms. *Science* **316**(5823), 392–393.
- SÁNCHEZ ALVARADO, A. & NEWMARK, P. A. (1999). Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc Natl Acad Sci U S A* **96**(9), 5049–5054.
- SÁNCHEZ ALVARADO, A., NEWMARK, P. A., ROBB, S. M. & R, J. (2002). The *Schmidtea mediterranea* database as a molecular resource for studying platyhelminthes, stem cells and regeneration. *Development* **129**(24), 5659–5665.
- SÁNCHEZ ALVARADO, A. & YAMANAKA, S. (2014). Rethinking differentiation: stem cells, regeneration, and plasticity. *Cell* **157**(1), 110–119.
- TAKAHASHI, K. & YAMANAKA, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**(4), 663–676.

## Bibliografía

- TAKEDA, H., NISHIMURA, K. & AGATA, K. (2009). Planarians maintain a constant ratio of different cell types during changes in body size by using the stem cell system. *Zoolog Sci* **26**(12), 805–813.
- TAN, T. C. J., RAHMAN, R., JABER-HIJAZI, F., FELIX, D. A., CHEN, C., LOUIS, E. J. & ABOOBAKER, A. (2012). Telomere maintenance and telomerase activity are differentially regulated in asexual and sexual worms. *Proc Natl Acad Sci U S A* **109**(11), 4209–4214.
- TANABE, K., TAKAHASHI, K. & YAMANAKA, S. (2014). Induction of pluripotency by defined factors. *Proc Jpn Acad Ser B Phys Biol Sci* **90**(3), 83–96.
- TANAKA, E. & REDDIEN, P. (2011). The cellular basis for animal regeneration. *Dev Cell* **21**(1), 172–185.
- TANIGUCHI, Y., CHOI, P. J., LI, G. W., CHEN, H., BABU, M., HEARN, J., EMILI, A. & XIE, X. S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**(5991), 533–538.
- THATCHER, E. J. & PATTON, J. G. (2010). Small RNAs have a big impact on regeneration. *RNA Biol* **7**(3), 333–338.
- THE ENCODE PROJECT CONSORTIUM (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**(7146), 799–816.
- THE SCHMIDTEA MEDITERRANEA GENOME SEQUENCING PROJECT (2014). <http://genome.wustl.edu/genomes/detail/schmidtea-mediterranea>.
- TILGNER, H., GRUBERT, F., SHARON, D. & SNYDER, M. P. (2014). Defining a personal, allele-specific, and single-molecule long-read transcriptome. *Proc Natl Acad Sci U S A* **111**(27), 9869–9874.
- TIMMS, J. F. & CRAMER, R. (2008). Difference gel electrophoresis. *Proteomics* **8**(23-24), 4886–4897.
- TSONIS, P. A., MADHAVAN, M., CALL, M. K., GAINER, S., RICE, A. & DEL RIO-TSONIS, K. (2004). Effects of a CDK inhibitor on lens regeneration. *Wound Repair Regen* **12**(1), 24–29.
- UNLU, M., MORGAN, M. E. & MINDEN, J. S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* **18**(1), 2071–2077.
- VAN DER MEER, A. D., ORLOVA, V. V., TEN DIJKE, P., VAN DEN BERG, A. & MUMMERY, C. L. (2013). Three-dimensional co-cultures of human endothelial cells and embryonic stem cell-derived pericytes inside a microfluidic device. *Lab Chip* **13**(18), 3562.

- VAN WOLFSWINKEL, J., WAGNER, D. & REDDIEN, P. (2014). Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell* **15**(3), 326–339.
- VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. & KINZLER, K. W. (1995). Serial analysis of gene expression. *Science* **270**(5235), 484–487.
- VOGEL, C., DE SOUSA ABREU, R., KO, D., LE, S.-Y., SHAPIRO, B. A., BURNS, S. C., SANDHU, D., BOUTZ, D. R., MARCOTTE, E. M. & PENALVA, L. O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. *Mol Syst Biol* **6**, 400.
- WADDINGTON, C. H. (1935). Cancer and the theory of organisers. *Nature* **135**(3416), 606–608.
- WAGNER, D., HO, J. & REDDIEN, P. (2012). Genetic regulators of a pluripotent adult stem cell system in planarians identified by RNAi and clonal analysis. *Cell Stem Cell* **10**(3), 299–311.
- WAGNER, D. E., WANG, I. E. & REDDIEN, P. W. (2011). Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science* **332**(6031), 811–816.
- WANG, J., SAXE, J. P., TANAKA, T., CHUMA, S. & LIN, H. (2009). Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. *Curr Biol* **19**(8), 640–644.
- WANG, Y., STARY, J. M., WILHELM, J. E. & NEWMARK, P. A. (2010). A functional genomic screen in planarians identifies novel regulators of germ cell development. *Genes Dev* **24**(18), 2081–2092.
- WANG, Y., ZAYAS, R. M., GUO, T. & NEWMARK, P. A. (2007). *nanos* function is essential for development and regeneration of planarian germ cells. *Proc Natl Acad Sci U S A* **104**(14), 5901–5906.
- WEISSMAN, I. L. (2000). Stem cells: units of development, units of regeneration, and units in evolution. *Cell* **100**(1), 157–68.
- WENEMOSER, D. & REDDIEN, P. W. (2010). Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Dev Biol* **344**(2), 979–991.
- WIESE, S., REIDEGELD, K. A., MEYER, H. E. & WARSCHIED, B. (2007). Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7**(3), 340–350.



## Bibliografía

- WITCHLEY, J., MAYER, M., WAGNER, D., OWEN, J. & REDDIEN, P. (2013). Muscle cells provide instructions for planarian regeneration. *Cell Rep* **4**(4), 633–641.
- WOLFF, E. & DUBOIS, F. (1948). Sur la migration des cellules de regeneration chez les planaires. *Rev Suisse Zool* **55**, 218–227.
- WU, G., NIE, L. & ZHANG, W. (2008). Integrative analyses of posttranscriptional regulation in the yeast *Saccharomyces cerevisiae* using transcriptomic and proteomic data. *Curr Microbiol* **57**(1), 18–22.
- YAZAWA, S., UMESONO, Y., HAYASHI, T., TARUI, H. & AGATA, K. (2009). Planarian Hedgehog/Patched establishes anterior-posterior polarity by regulating Wnt signaling. *Proc Natl Acad Sci U S A* **106**(52), 22329–22334.
- YOSHIDA-KASHIKAWA, M., SHIBATA, N., TAKECHI, K. & AGATA, K. (2007). DjCBC-1, a conserved DEAD box RNA helicase of the RCK/p54/Me31B family, is a component of RNA-protein complexes in planarian stem cells and neurons. *Dev Dyn* **236**(12), 3436–3450.
- YU, J., CLIFTEN, P. F., JUEHNE, T. I., SINNWELL, T. M., SAWYER, C. S., SHARMA, M., LUTZ, A., TYCKSEN, E., JOHNSON, M. R., MINTON, M. R., KLOTZ, E. T., SCHRIEFER, A. E., YANG, W., HEINZ, M. E., CROSBY, S. D. & HEAD, R. D. (2015). Multi-platform assessment of transcriptional profiling technologies utilizing a precise probe mapping methodology. *BMC Genomics* **16**(1), 710.
- ZAYAS, R. M., HERNANDEZ, A., HABERMANN, B., WANG, Y., STARY, J. M. & NEWMARK, P. A. (2005). The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: analysis of ESTs from the hermaphroditic strain. *Proc Natl Acad Sci U S A* **102**(51), 18491–18496.
- ZENZMAIER, C., KOLLROSER, M., GESSLBAUER, B., JANDROSITZ, A., PREISEGGER, K.-H. & KUNGL, A. J. (2003). Preliminary 2-d chromatographic investigation of the human stem cell proteome. *Biochem Biophys Res Commun* **310**(2), 483–490.
- ÖNAL, P., GRÜN, D., ADAMIDI, C., RYBAK, A., SOLANA, J., MASTROBUONI, G., WANG, Y., RAHN, H.-P., CHEN, W., KEMPA, S., ZIEBOLD, U. & RAJEWSKY, N. (2012). Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *EMBO J* **31**(12), 2755–2769.

# **Anexos**



## Artículo 5

***egr-4, a target of EGFR signaling, is required for the formation of the brain primordia and head regeneration in planarians***

Susanna Fraguas, Sara Barberán, Marta Iglesias, **Gustavo Rodríguez-Esteban** and Francesc Cebrià.

**Development** (2014) 141:1835-1847



## Resumen

***egr-4*, un gen diana de la vía de señalización de los EGFR, es necesario para formar el primordio del cerebro y la regeneración de la cabeza en planarias.**

Debido a la gran diversidad y complejidad de fenotipos obtenidos en los diferentes *Smed-egfr* después de su silenciamiento (Fraguas *et al.*, 2011), se decidió realizar un análisis más detallado de uno de estos elementos de la vía de los EGFR, *Smed-egfr-3*, mediante nuevas tecnologías de secuenciación de alto rendimiento, concretamente *Digital Gene Expression* (DGE), con el objetivo de desvelar cómo regulan los diferentes procesos biológicos y saber a través de qué vías de señalización *downstream* los están regulando. En este artículo se caracteriza uno de los posibles genes diana de *Smed-egfr-3*, el gen *egr-4*, obtenido a partir de las librerías génicas resultantes de DGE.

El análisis del patrón de expresión de *egr-4*, un miembro de la familia de los genes "*early growth response*", mediante técnicas de hibridación *in situ*, revela que este gen se expresa principalmente en el sistema nervioso de las planarias y que se induce en el blastema tras pocas horas de producirse una herida. Aunque esta inducción inmediata parece independiente de la vías de los EGFR, pasa a ser dependiente de *Smed-egfr-3* a partir del segundo día de regeneración. Por otro lado, experimentos de irradiación que eliminan en pocos días los neoblastos no apuntan a una alteración de su expresión, de manera que podemos concluir que este gen se expresa en células diferenciadas.

A partir de los resultados obtenidos se sugiere una posible relación de los genes de la familia *egr* con la vía de señalización de los EGFR que regula la diferenciación celular durante la regeneración de las planarias y se postula una probable conexión entre la diferenciación del primordio del cerebro y la correcta progresión de la regeneración de la cabeza.



## RESEARCH ARTICLE

## STEM CELLS AND REGENERATION

# *egr-4*, a target of EGFR signaling, is required for the formation of the brain primordia and head regeneration in planarians

Susanna Fraguas, Sara Barberán, Marta Iglesias, Gustavo Rodríguez-Esteban and Francesc Cebrià\*

**ABSTRACT**

During the regeneration of freshwater planarians, polarity and patterning programs play essential roles in determining whether a head or a tail regenerates at anterior or posterior-facing wounds. This decision is made very soon after amputation. The pivotal role of the Wnt/ $\beta$ -catenin and Hh signaling pathways in re-establishing anterior-posterior (AP) polarity has been well documented. However, the mechanisms that control the growth and differentiation of the blastema in accordance with its AP identity are less well understood. Previous studies have described a role of *Smed-egfr-3*, a planarian epidermal growth factor receptor, in blastema growth and differentiation. Here, we identify *Smed-egr-4*, a zinc-finger transcription factor belonging to the early growth response gene family, as a putative downstream target of *Smed-egfr-3*. *Smed-egr-4* is mainly expressed in the central nervous system and its silencing inhibits anterior regeneration without affecting the regeneration of posterior regions. Single and combinatorial RNA interference to target different elements of the Wnt/ $\beta$ -catenin pathway, together with expression analysis of brain- and anterior-specific markers, revealed that *Smed-egr-4*: (1) is expressed in two phases – an early *Smed-egfr-3*-independent phase and a late *Smed-egfr-3*-dependent phase; (2) is necessary for the differentiation of the brain primordia in the early stages of regeneration; and (3) that it appears to antagonize the activity of the Wnt/ $\beta$ -catenin pathway to allow head regeneration. These results suggest that a conserved EGFR/*egr* pathway plays an important role in cell differentiation during planarian regeneration and indicate an association between early brain differentiation and the proper progression of head regeneration.

**KEY WORDS:** Planarian, Early growth response genes, Patterning

**INTRODUCTION**

After almost any type of amputation, freshwater planarians are capable of regenerating missing regions with the appropriate polarity. Once polarity is established, the appropriate morphogenetic and patterning programs must direct the differentiation of those blastemas into either anterior or posterior regions and their corresponding tissues and organs. Thus, cephalic ganglia develop *de novo* in decapitated planarians. The instrumental role of Wnt/ $\beta$ -catenin signaling in defining head versus tail identity during planarian regeneration is well documented (Gurley et al., 2008; Iglesias et al., 2008; Petersen and Reddien, 2008). However, less is known about the genes and/or pathways that mediate

the patterning, growth and differentiation of blastemas after the re-establishment of polarity (Felix and Aboobaker, 2010).

Conserved signaling pathways play important roles in morphogenesis in all animals. One such pathway is the epidermal growth factor receptor (EGFR) pathway, which regulates multiple biological processes, including cell proliferation, differentiation, apoptosis and cell survival. We have previously shown that silencing of *Smed-egfr-3*, a planarian homologue of epidermal growth factor receptors, impairs regeneration and blastema growth in these organisms, probably by disrupting cellular differentiation (Fraguas et al., 2011). Similar effects have been reported after inhibition of ERK (MAPK extracellular signal-related kinase; Tasaki et al., 2011), a gene whose expression is regulated by EGFR signaling in many organisms. In the present study, we conducted digital gene expression (DGE) analyses to identify putative target genes of *Smed-egfr-3* in order to better characterize the function of the EGFR signaling pathway during planarian regeneration. One of the isolated candidate genes that was downregulated after *egfr-3* RNAi was *Smed-egr-4* (Wenemoser et al., 2012), henceforth *egr-4*, a member of the early growth response (*egr*) gene family. *egr* genes were first characterized by their induction by nerve growth factor (Milbrandt, 1987) and other mitogens (Sukhatme et al., 1987), and are implicated in the regulation of multiple cellular processes (Calogero et al., 2004; Cole et al., 1989; Dussmann et al., 2011; Shafarenko et al., 2005; Sukhatme et al., 1988). Members of this family of zinc-finger transcription factors are considered immediate-early genes; they are rapidly induced by many environmental signals, including growth factors, hormones and neurotransmitters (Thiel and Cibelli, 2002), and are rapidly and transiently upregulated by a variety of signaling pathways, including the EGFR and MAPK/ERK signaling pathways (Aggeli et al., 2010; Cabodi et al., 2008; Kaufmann and Thiel, 2001; Ludwig et al., 2011b; Mayer et al., 2009; Tsai et al., 2000). EGFR signaling in mice stimulates the expression of *egr2* to promote cell proliferation during bone formation (Chandra et al., 2013), while EGF induces Egr-1 expression in endothelial cells (Tsai et al., 2000).

We found that *egr-4* was mainly expressed in the central nervous system (CNS) and was rapidly induced after different types of injury. Although early *egr-4* expression after injury was independent of EGFR signaling, it became *Smed-egfr-3* dependent from the second day of regeneration. Functional analyses based on RNA interference (RNAi) revealed that *egr-4* was required for head regeneration but not for the regeneration of posterior regions. *egr-4* silencing significantly impaired the formation of anterior blastemas; these animals exhibited either small mispatterned cephalic ganglia or a total absence of new brain tissue. The differentiation of other cell types normally present in the anterior region was also altered in *egr-4*(RNAi) animals. However, the early expression of polarity determinants required for the re-establishment of anterior polarity was unaffected. Simultaneous silencing of *egr-4* and different elements of the Wnt/ $\beta$ -catenin pathway revealed that *egr-4* is required for the differentiation of the brain primordia. Moreover, the results of these experiments suggest

Departament de Genètica de la Universitat de Barcelona and Institut de Biomedicina de la Universitat de Barcelona (IBUB), Avenida Diagonal 643, Edifici Prevoosti planta 1, Barcelona 08028, Spain.

\*Author for correspondence (fcebrias@ub.edu)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Received 17 July 2013; Accepted 3 March 2014



that *egr-4* promotes head regeneration by antagonizing the inhibitory action of the Wnt/ $\beta$ -catenin pathway. Taken together, these findings identify *egr-4* as one of the first known genes necessary for the differentiation of the brain primordia during planarian regeneration. In view of these findings, we discuss how the failure to differentiate proper brain primordia may lead to the blockade of blastema growth and head regeneration.

## RESULTS

### DGE analyses identify *Smed-egr-4* as a putative target of *Smed-egr-3*

We have previously demonstrated that silencing of *Smed-egr-3*, a planarian homologue of the epidermal growth factor receptor (EGFR) family, blocks regeneration, probably by disrupting cellular differentiation (Fraguas et al., 2011). To better characterize the functions of the EGFR pathway during planarian regeneration, we constructed and sequenced DGE (Digital Gene Expression) libraries to compare the transcriptomic profiles of control and *egr-3(RNAi)* regenerating animals (supplementary material Table S1). Genes predicted to undergo up- or downregulation after *Smed-egr-3* RNAi on days 1 and 3 post-amputation are shown in supplementary material Table S2. After applying the cut-off criteria, 680 genes were upregulated on day 1 and 2403 on day 3 in *Smed-egr-3(RNAi)* animals. A total of 655 genes were downregulated on day 1 and 800 on day 3 in *Smed-egr-3(RNAi)* animals (supplementary material Fig. S1). All genes that were up- and downregulated on day 3 post-amputation (2949 and 921, respectively) were annotated and assigned Gene Ontology (GO) categories (supplementary material Fig. S2). Because EGFR signaling regulates multiple downstream targets and pathways, the annotated genes fell into a wide range of GO categories. Overrepresented categories included 'signal transduction', 'cell differentiation', 'response to stress', 'catabolic process', 'nucleotide binding' and 'protein kinase activity' (supplementary material Fig. S2). To identify putative target genes of *Smed-egr-3* that may be linked to the phenotypic defects observed after its silencing (Fraguas et al., 2011), we selected a group of genes for further characterization, based either on their proposed involvement in the EGFR pathway in other models or their role in processes such as neurogenesis, cell proliferation and differentiation, cancer and tumorigenesis, apoptosis and cell survival, and inflammation and the immune response (supplementary material Fig. S3).

Among the candidate targets that exhibited interesting RNAi phenotypes and merited further characterization was a gene displaying sequence similarity to the *early growth response (egr)* gene family of transcription factors (Contig\_2669\_90e), a well-known family of genes that are regulated by EGFR signaling in other models (Cabodi et al., 2008; Kaufmann and Thiel, 2001; Mayer et al., 2009; Tsai et al., 2000). Although several members of this family have been described in planarians (Onal et al., 2012; Sandmann et al., 2011; Wenemoser et al., 2012; Wagner et al., 2012), Contig\_2669\_90e was the only *egr* gene downregulated after *Smed-egr-3* RNAi. Sequence analyses revealed that this gene corresponded to the previously annotated *egr-4* (Wenemoser et al., 2012) and phylogenetic analyses confirmed that *egr-4* belongs to this gene family (supplementary material Fig. S4).

Whole-mount *in situ* hybridization in intact planarians revealed *egr-4* expression mainly in the cephalic ganglia, with weak expression in the ventral nerve cord (vnc) and the mesenchyme (Fig. 1A). We have previously reported *Smed-egr-3* expression in neoblasts and in the CNS (Fraguas et al., 2011). Here, we better characterized the expression of *Smed-egr-3* by fluorescent *in situ* hybridization (FISH) (Fig. 1B). Although we did not succeed in performing double FISH

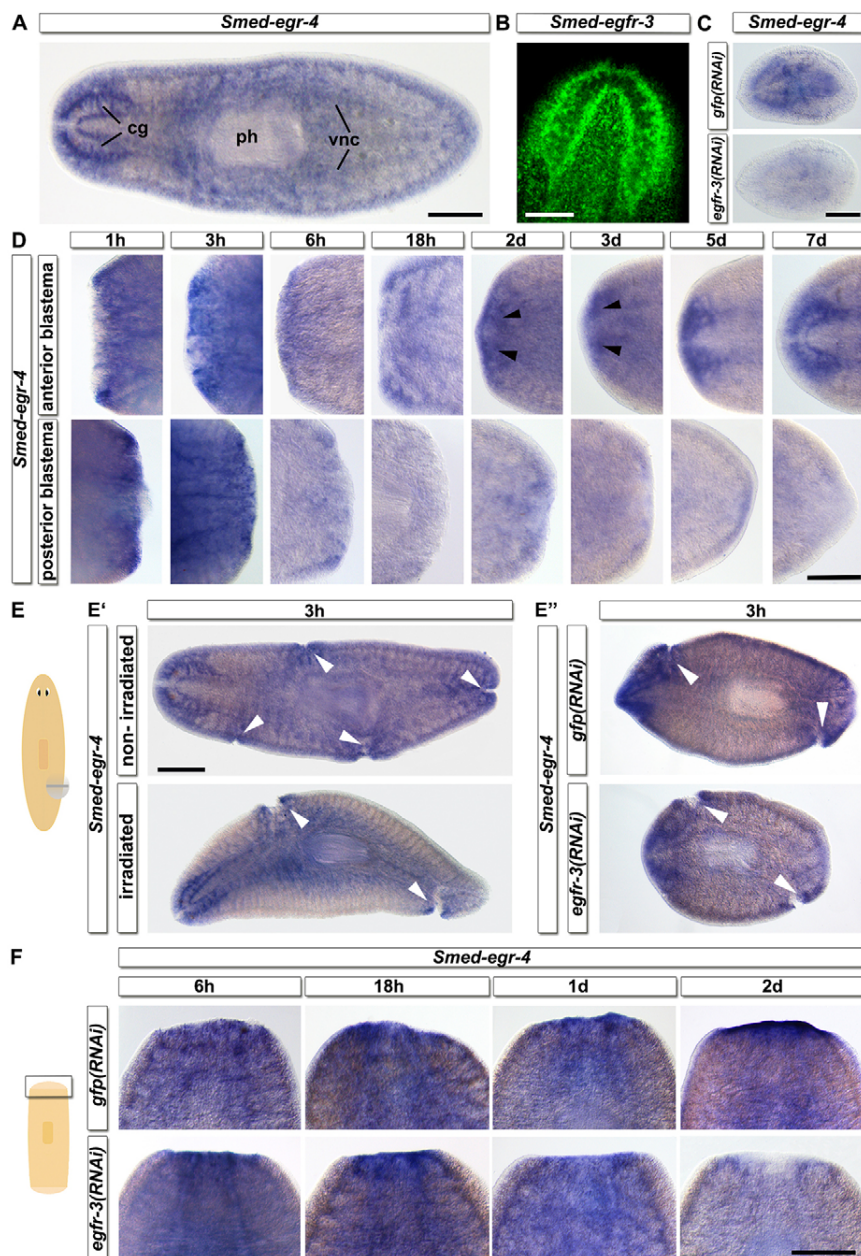
of *egr-4* and *Smed-egr-3*, their similar expression patterns (compare Fig. 1A with 1B) suggest that these genes are probably co-expressed in the CNS. Moreover, *egr-4* expression in the cephalic ganglia was strongly reduced after *Smed-egr-3* silencing, further supporting the DGE data (Fig. 1C). To investigate whether *egr-4* expression in the mesenchyme corresponded to neoblasts we performed *in situ* hybridization on irradiated animals at different time points (supplementary material Fig. S5). Whereas *Smed-egr-3* expression in the mesenchyme disappears 1 day after irradiation (Fraguas et al., 2011), no change in the pattern of *egr-4* expression was observed, even at 7 days after neoblast elimination (supplementary material Fig. S5), suggesting an absence of *egr-4* expression in stem cells. During anterior regeneration, *egr-4* was upregulated in the wound region from day 1 of regeneration (Fig. 1D). As regeneration progressed, *egr-4* expression became restricted mainly to the differentiating cephalic ganglia (arrowheads in Fig. 1D). By contrast, no upregulation of *egr-4* was observed during posterior regeneration at later time points (Fig. 1D). However, strong expression at earlier stages (1 h and 3 h post-amputation) was observed in both anterior and posterior blastemas. *egr-4* has been recently categorized as a W1 (wound-induced class 1) gene (Wenemoser et al., 2012). *In situ* hybridization revealed rapid upregulation of *egr-4* after a small incision (Fig. 1E), in agreement with previous reports for other planarian *egr* genes (Wenemoser et al., 2012; Sandmann et al., 2011). This strong upregulation of *egr-4* was not dependent on neoblasts (Fig. 1E') or on *Smed-egr-3* (Fig. 1E''), as no differences were observed in irradiated or *Smed-egr-3(RNAi)* animals.

Finally, to better characterize the dependence of *egr-4* expression on *Smed-egr-3* during regeneration we carried out a detailed time-course analysis of *egr-4* expression after *Smed-egr-3(RNAi)* (Fig. 1F). Normal *egr-4* expression in the wound region was observed up to day 1 of regeneration. However, by day 2, *egr-4* expression was downregulated (Fig. 1F), suggesting that *egr-4* expression becomes *Smed-egr-3* dependent once its expression becomes restricted to the regenerating cephalic ganglia.

### *egr-4* is necessary for proper head regeneration and neoblast differentiation

We conducted RNAi experiments to characterize the role of *egr-4* in regeneration (see Materials and Methods). Both trunk and tail fragments regenerating a new head exhibited blastemas that were either very small or abnormally differentiated. After 8 days of regeneration, about half of the anteriorly regenerating trunks ( $n=18/34$ ) formed normal-sized blastemas but with cyclopic eyes, and very few ( $n=3/34$ ) regenerated small blastemas without eyes (Fig. 2A). By contrast, at the same time point, most tails ( $n=30/34$ ) failed to properly regenerate a head. Of these 30 tails, some developed very small blastemas ( $n=12/30$ ), whereas others appeared to only heal the wound ( $n=18/30$ ) (Fig. 2A). These results indicate that the severity of *egr-4* RNAi varied along the anterior-posterior (AP) axis. Posterior regeneration proceeded normally (Fig. 2A), indicating that *egr-4* RNAi specifically blocked anterior regeneration.

Although *egr-4* was not expressed in neoblasts, we investigated whether these defects in regeneration were due to a non-cell-autonomous effect of *egr-4* RNAi on neoblast maintenance and/or proliferation. However, no differences in neoblast number or distribution were observed after immunostaining and *in situ* hybridization for the neoblast-specific markers anti-SMEDWI-1 antibody and *Smed-histone-2B*, respectively (Fig. 2B) (Reddien et al., 2005; Guo et al., 2006; Solana et al., 2012). Thus, neoblasts were normally found below the wound epithelium on day 3. Eight



**Fig. 1. *egr-4* expression pattern in intact and regenerating planarians.** (A) In intact animals, *egr-4* was expressed in the cephalic ganglia (cg), ventral nerve cord (vnc) and mesenchyme. (B) *Smed-egr-3* expression in the cephalic ganglia. (C) *egr-4* expression was downregulated after *Smed-egr-3* silencing. These samples correspond to regenerating head fragments after 10 days of regeneration. (D) Expression of *egr-4* in anteriorly and posteriorly regenerating bipolar trunks. Arrowheads indicate the brain primordia. (E) *egr-4* was rapidly upregulated after injury. Small incisions induced *egr-4* expression after 3 h (arrowheads). This early expression was not dependent on either neoblasts (E') or *Smed-egr-3* (E''). (F) *Smed-egr-3* silencing resulted in the downregulation of *egr-4* expression from day 2 of regeneration. (A,C-E) Anterior towards the left. (B,F) Anterior at the top. Scale bars: 300  $\mu$ m in A,D,F; 200  $\mu$ m in B,C,E.

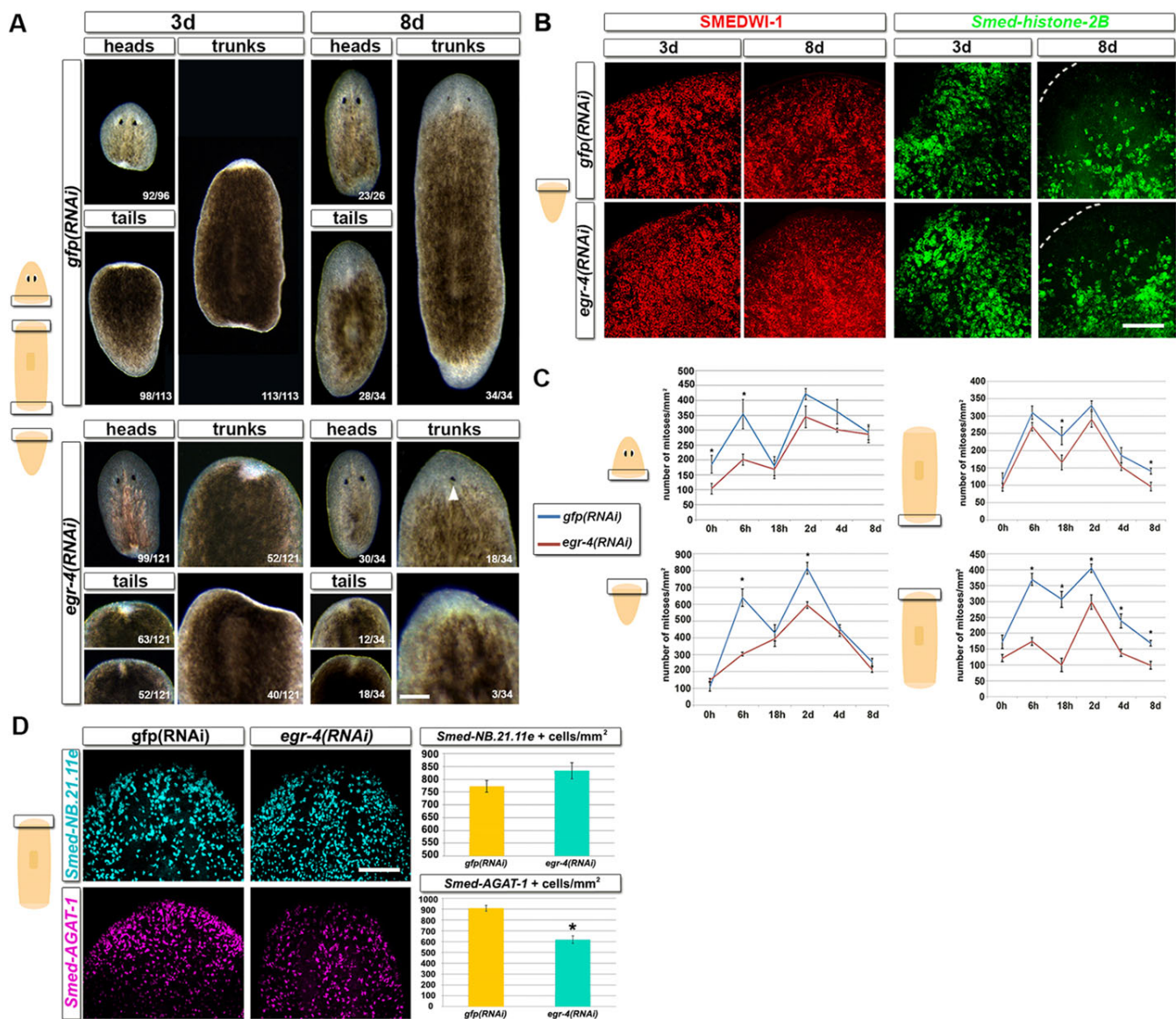
days after amputation, few *Smed-histone-2B*-positive cells were detected in the new head region of controls. Similarly, despite impaired blastema growth and regeneration, few *Smed-histone-2B*-expressing cells were observed after *egr-4(RNAi)*, indicating that neoblast dynamics (in terms of distribution) was unaffected.

We next used an anti-phospho-histone H3 antibody to characterize the mitotic response after *egr-4(RNAi)* in both anterior and posterior regenerating fragments (Fig. 2C). In most cases, the first mitotic peak, which is associated with injury and occurs 6 h post-amputation, was significantly attenuated when compared with controls. Interestingly, the second mitotic peak at 48 h post-amputation, which is associated with tissue loss (Wenemoser and Reddien, 2010), was significantly attenuated in anterior, but not posterior, blastemas. However, despite this decrease in the rate of proliferation, a significant number of mitoses were observed at all stages after *egr-4* RNAi (supplementary material Fig. S6), suggesting that the severe impairment of blastema growth and regeneration in these animals was not exclusively due to

defective mitosis. Finally, we characterized neoblast progeny at different stages of differentiation using the lineage markers *Smed-NB21.11e* and *Smed-AGAT-1*, which are specific to early and late neoblast postmitotic progeny, respectively (Eisenhoffer et al., 2008). Although the early neoblast progeny were unaffected, the number of late progeny was significantly decreased after *egr-4* RNAi (Fig. 2D). Taken together, these results indicate that *egr-4* RNAi impairs anterior regeneration probably by affecting cell differentiation rather than neoblast pool maintenance or proliferation.

#### ***egr-4* RNAi impairs tissue and organ differentiation**

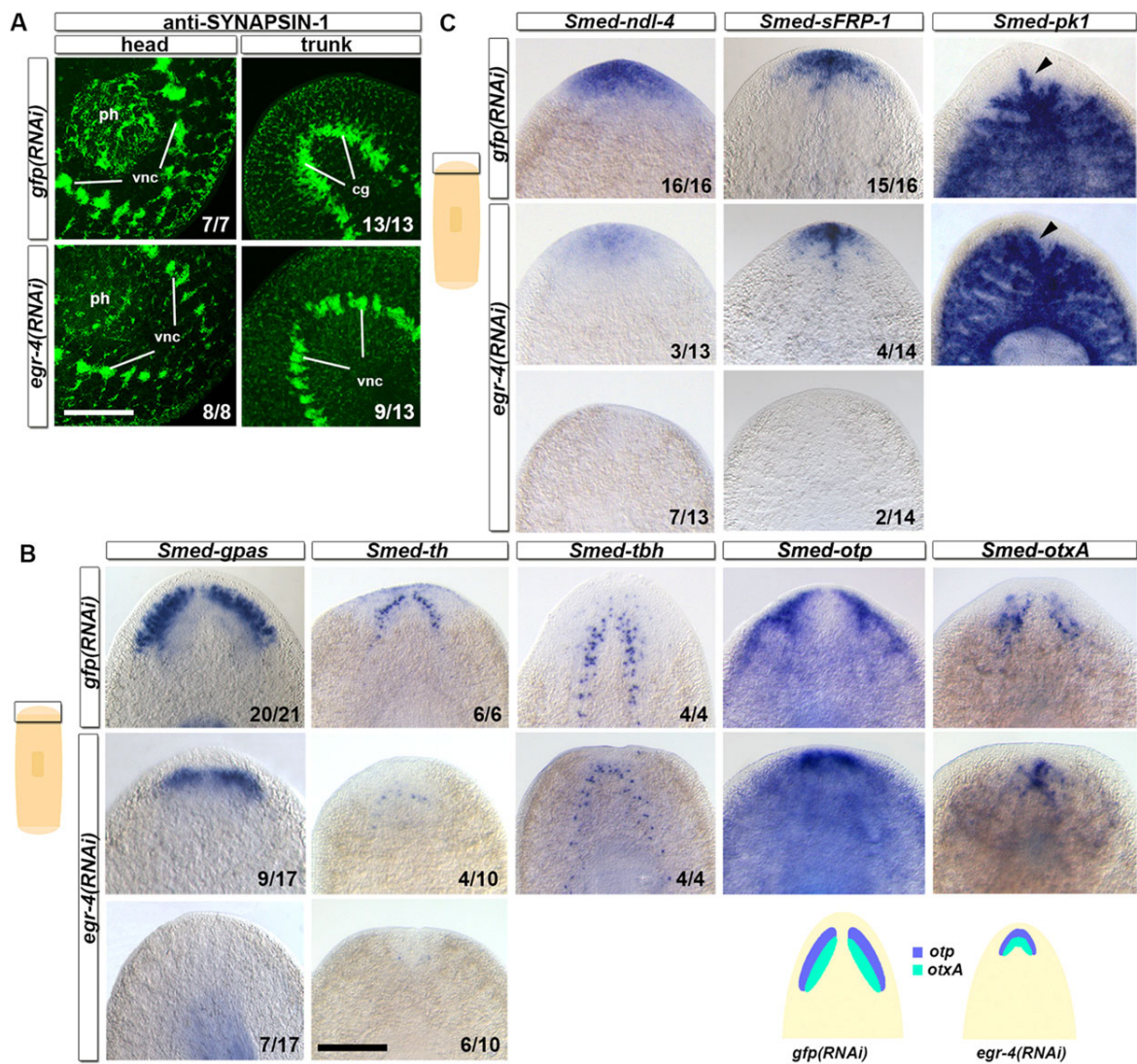
To further characterize these defects in regeneration, we first used the pan-neural marker anti-SYNORF-1 (Cebrià, 2008) to study CNS regeneration. Both control and *egr-4(RNAi)* head fragments regenerated normal pharynges and ventral nerve cords grew into the newly developed tails (Fig. 3A). By contrast, in anteriorly regenerating trunks, *egr-4* RNAi generally resulted in the formation



**Fig. 2. Effects of *egr-4* RNAi on head regeneration and neoblast dynamics.** (A) Effects of *egr-4* RNAi after 3 and 8 days (d) of regeneration in head, bipolar trunk and tail fragments. Although heads regenerated normally, most tail fragments were unable to regenerate a proper head after *egr-4* silencing. Anteriorly regenerating *egr-4*(RNAi) trunks exhibited a milder phenotype when compared with corresponding tail fragments, displaying defects mainly in head morphogenesis, as evidenced by the development of cyclopic eyes (arrowhead). (B) The impairment of head regeneration was not due to neoblast loss, as demonstrated by using anti-SMEDWI-1 immunostaining and *Smed-histone-2B* *in situ* hybridization. Dashed lines indicate the border of the anterior head. (C) Quantification of mitotic cells after *egr-4* RNAi, detected by anti-phospho-histone H3 immunostaining in head, bipolar trunk and tail fragments at different time points. Values represent the mean  $\pm$  s.e.m. (Student's *t*-test, \**P*<0.05) of an average of 10 samples per time point and amputation level. (D) Although early neoblast progeny (*Smed-NB.21.11e*-positive cells) were unaffected by *egr-4* RNAi, a significant decrease in the number of late neoblast progeny (*Smed-AGAT-1*-positive cells) was observed. Samples correspond to trunk sections after 8 days of regeneration. Values represent the mean  $\pm$  s.e.m. (Student's *t*-test, \**P*<0.05) of an average of 10 samples per time point. All panels are oriented with the anterior towards the top. Scale bars: 200  $\mu$ m for *egr-4* (RNAi) trunks and 300  $\mu$ m for all other panels in A; 200  $\mu$ m in B,D.

of blastemas with truncated ventral nerve cords and little cephalic ganglia differentiation, whereas normal cephalic ganglia regeneration was observed in controls (Fig. 3A). We next used additional neural (Fig. 3B) and anterior (Fig. 3C) markers to better characterize the *egr-4*(RNAi) phenotype. After 5-7 days of regeneration, most *egr-4*(RNAi) trunk fragments displayed significantly smaller or near-absent cephalic ganglia after *in situ* hybridization for *Smed-gpas*, a gene specific to the brain lateral branches (Iglesias et al., 2011). Moreover, the expression of markers of distinct neuronal populations such as *Smed-tbh* (octopaminergic neurons; Fraguas et al., 2012; Nishimura

et al., 2008) and *Smed-th* (dopaminergic neurons; Fraguas et al., 2011; Nishimura et al., 2007) was significantly reduced (Fig. 3B). These markers revealed that *egr-4* (RNAi) animals developed small and aberrant cephalic ganglia. To study the patterning of these cephalic ganglia, we analyzed expression of the *otd/Otx* family gene *Smed-otxA* and the homeobox-containing gene *orthopedia Smed-otp* (Umesono et al., 1997, 1999; Iglesias et al., 2011) (Fig. 3B). *otp* is normally expressed in the most lateral region of the cephalic ganglia, whereas *otxA* is expressed more medially (Umesono et al., 1999; Iglesias et al., 2011). Although *egr-4*(RNAi) animals often displayed



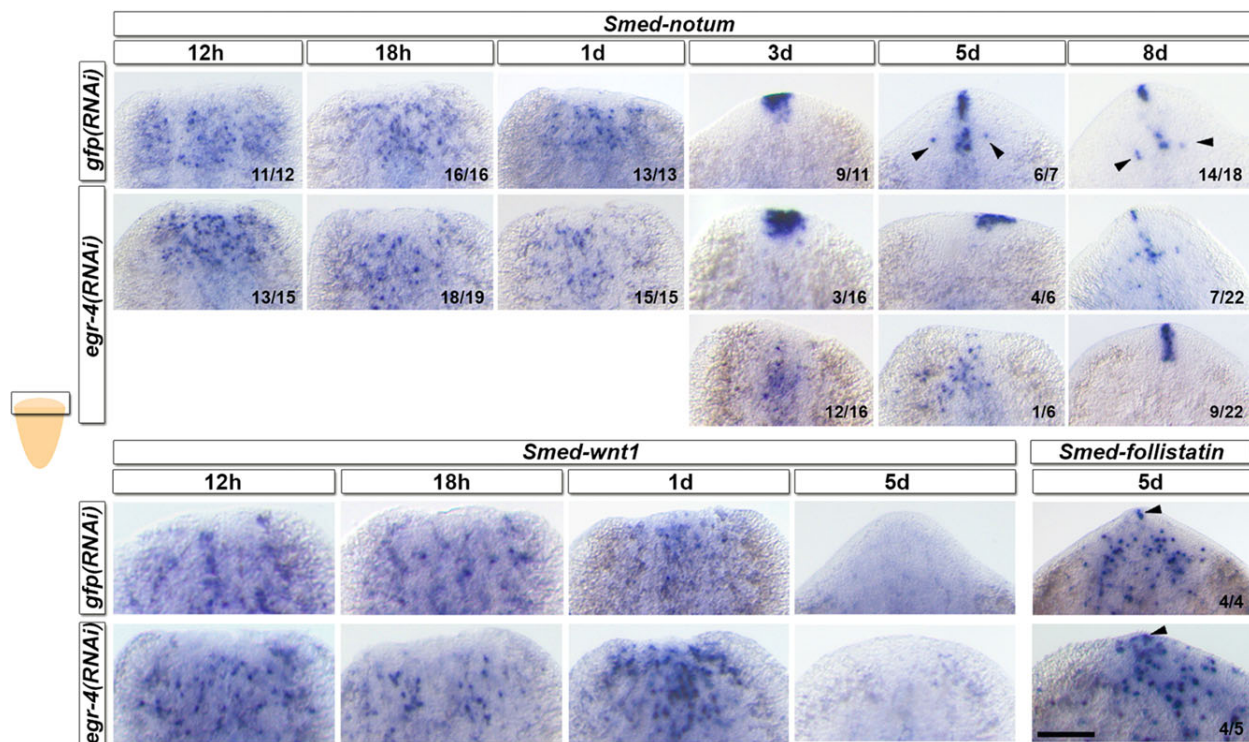
**Fig. 3. Loss of brain and anterior markers after *egr-4* RNAi.** (A) Whole-mount immunostaining with anti-SYNORF1 in head fragments revealed correct pharynges with typical nervous plexus and ventral nerve cords in both control and *egr-4(RNAi)* animals. The majority of anteriorly regenerating *egr-4(RNAi)* trunks exhibited truncated ventral nerve cords with no cephalic ganglia differentiation, in contrast to the normal regeneration seen in control animals. (B) The expression of several markers of specific brain subpopulations was reduced in *egr-4(RNAi)* animals. Most of these animals regenerated small cephalic ganglia instead of normal bilateral ganglia, as evidenced by the expression of *Smed-gpas*, *Smed-th* and *Smed-tbh*. In some cases, the expression of these markers was completely abolished. However, the patterning of those small abnormal cephalic ganglia appeared not to be affected after *in situ* hybridization with the mediolateral patterning genes *Smed-otp* and *Smed-otxA*. (C) Expression of the anterior marker *ndl-4* was also significantly attenuated. By contrast, the expression of the anterior marker *sFRP-1* was normal in most *egr-4(RNAi)* animals ( $n=8/14$ ). Finally, *egr-4(RNAi)* animals failed to correctly regenerate the anterior gut branch (arrowheads). cg, cephalic ganglia; vnc, ventral nerve cords; ph, pharynx. In A all panels are oriented with the anterior towards the top left. In B and C all panels are oriented with the anterior towards the top. Samples correspond to trunks after 5-8 days of regeneration. Scale bars: 150  $\mu$ m in head and 200  $\mu$ m in trunk fragments in A; 300  $\mu$ m in B,C.

abnormal cephalic ganglia, these distinct domains along the mediolateral axis were clearly distinguishable (Fig. 3B). Taken together, our data indicate that *egr-4* RNAi results in the development of small but well-patterned cephalic ganglia.

In addition to the aforementioned CNS defects, *egr-4(RNAi)* animals failed to properly regenerate the anterior gut branch into the blastema (Fig. 3C). Moreover, the expression of the anterior marker *Smed-ndl-4*, a FGF-receptor-like protein of the *nou-darake* family (Cebrià et al., 2002a; Rink et al., 2009), was either significantly reduced or completely absent in the majority of *egr-4(RNAi)* animals (Fig. 3C). The expression of *Smed-sFRP-1*, another anterior marker (Gurley et al., 2008; Petersen and Reddien, 2008), was not as

strongly affected. Most *egr-4(RNAi)* animals ( $n=8/14$ ) exhibited normal *Smed-sFRP-1* expression; some ( $n=4/14$ ) showed a very slight decrease (probably due to the smaller size of the regenerated anterior region) and the marker was completely absent in only 2 of the 14 animals (Fig. 3C). These results indicate that not all genes previously proposed as candidate anterior patterning genes (*ndl-4* and *sFRP-1*; Chen et al., 2013) respond equally to *egr-4* RNAi.

Taken together, our data confirm that *egr-4* silencing impairs the proper differentiation of several cell types and organs (CNS and gut) during anterior regeneration. However, no defects were observed after *egr-4* silencing in intact non-regenerating planarians; all CNS- and anterior-specific markers were normally expressed (supplementary



**Fig. 4. The early expression of AP polarity determinants is unaffected by *egr-4* silencing.** *In situ* hybridization with *Smed-notum*, *Smed-wnt1* and *Smed-follistatin* after *egr-4* RNAi in regenerating tail fragments at different time points. Although *egr-4* silencing had no effect on *Smed-notum* expression at early time points, a delay in the temporal expression of this gene was observed beginning on day 3. Arrowheads indicate *Smed-notum*-positive bilateral cells adjacent to photoreceptors. The dynamics of *Smed-wnt1* expression were unaffected by *egr-4* RNAi. After 5 days of regeneration, normal *Smed-follistatin* expression was detected in the anterior-most tip of the regenerating head (arrowheads). All panels are oriented with anterior towards the top. Scale bar: 200  $\mu$ m.

material Fig. S7A). Moreover, neoblast proliferation was unaltered in intact *egr-4(RNAi)* animals when compared with corresponding controls (supplementary material Fig. S7B).

#### ***egr-4* RNAi disrupts head regeneration without impairing the establishment of anterior polarity**

To determine whether the impaired head regeneration observed after *egr-4* RNAi was due to defects in the establishment of anterior polarity, we analyzed the expression of the anterior polarity determinants *Smed-notum*, *Smed-wnt1* and *Smed-follistatin* (Adell et al., 2009; Gurley et al., 2010; Petersen and Reddien, 2009, 2011; Yazawa et al., 2009; Roberts-Galbraith and Newmark, 2013). Control animals exhibited three typical phases of *Smed-notum* expression (Fig. 4). At 12 h post-amputation, *Smed-notum* expression was strongly upregulated in the wound region, giving rise to a dotted pattern of discrete cells. Between 18 h and 1 day after amputation, the number of *Smed-notum*-positive cells decreased (12 h to 1 day, phase I). After 3 days of regeneration, *Smed-notum*-positive cells coalesced at the tip of the anterior blastema (phase II). Finally, between 5 and 8 days post-amputation, the expression of *Smed-notum* became restricted to a small number of cells at the tip of the new head and along the midline, and to two bilateral groups of cells (arrowheads in Fig. 4; phase III), a pattern similar to that observed in intact planarians. No differences in the pattern of *Smed-notum* expression were observed between *egr-4(RNAi)* animals and controls up to 3 days post-amputation (Fig. 4). From this stage on, however, the progression of *Smed-notum* expression appeared to be delayed or arrested in *egr-4(RNAi)* animals (Fig. 4); after 3 days of regeneration, the clustering of *Smed-notum*-positive cells at the tip of the blastema seen

in control animals was observed in only 3 out of 16 *egr-4(RNAi)* animals. By day 5 of regeneration, this clustering of *Smed-notum*-positive cells was observed in most *egr-4(RNAi)* animals. Normal *Smed-notum* expression was detected in several *egr-4(RNAi)* animals ( $n=7/22$ ) by day 8; in others ( $n=9/22$ ) the pattern of *Smed-notum*-positive cells was somewhere between that of phase II and III, while the remaining *egr-4(RNAi)* animals displayed no *notum* expression ( $n=6/22$ ).

We observed no significant differences in the expression pattern of *Smed-wnt1* between controls and *egr-4(RNAi)* animals (Fig. 4). Between 12 h and 1 day post-amputation, a dotted distribution of discrete *Smed-wnt1*-expressing cells was observed in the anterior wound. By 2 days post-amputation this anterior expression had been completely lost in both controls and *egr-4(RNAi)* animals, indicating that the lack of head regeneration after *egr-4* silencing was not due to blastema posteriorization (Fig. 4). *Smed-wnt1* expression in the tip of the tail was normal in all these samples (data not shown).

*Smed-follistatin* and *Smed-notum* co-expression was recently demonstrated in planarians in a small cluster of cells at the tip of the head (Roberts-Galbraith and Newmark, 2013). Functional analyses suggest that *Smed-follistatin* inhibits the Activin/ActR-1/Smad2/3 signaling pathway, which represses anterior regeneration. It has been proposed that *Smed-follistatin* and *Smed-notum* cooperate to promote anterior identity, thus acting as an anterior signaling center (Roberts-Galbraith and Newmark, 2013). We found that *Smed-follistatin* expression at the tip of the regenerating head re-appeared after *egr-4* RNAi, although slightly later than in controls, as also observed for *Smed-notum* (data not shown). However, after 5 days

of regeneration most animals displayed normal *Smed-follistatin* expression at their anterior-most tips (Fig. 4).

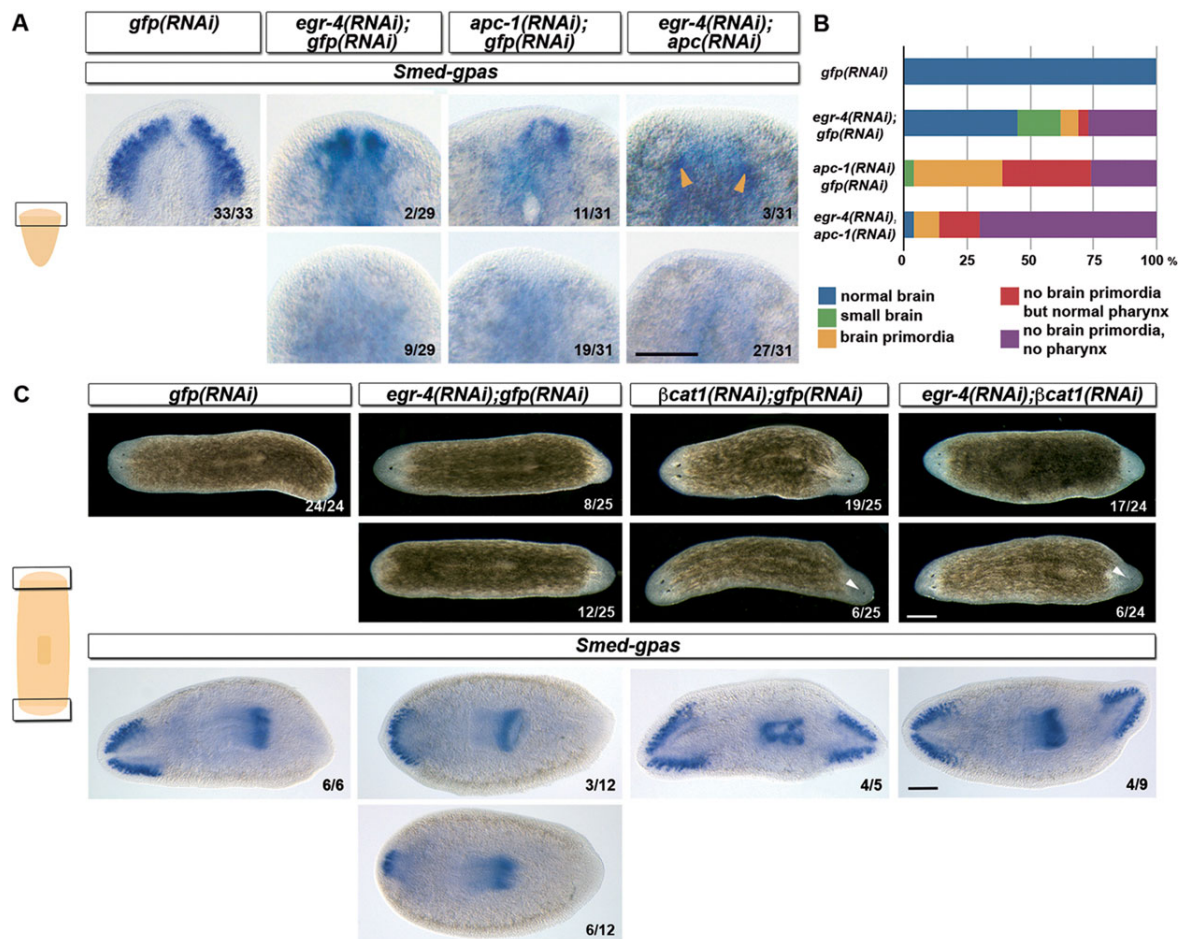
Overall, these results suggest that the polarity determinants *Smed-notum*, *Smed-wnt1* and *Smed-follistatin* are normally expressed in the early regenerative stages during which AP polarity is established (Petersen and Reddien, 2009, 2011) and that *egr-4* silencing impairs normal head regeneration without disrupting the establishment of anterior polarity.

#### ***egr-4* is required for the formation of the brain primordia**

As *egr-4* was highly expressed in the mature and differentiating cephalic ganglia from very early stages of regeneration, we investigated its role in the regeneration of the cephalic ganglia by simultaneously silencing *egr-4* and *Smed-apc-1*, an inhibitor of the Wnt/ $\beta$ -catenin pathway. *apc-1* silencing results in the regeneration of a tail, rather than a head, from anterior wounds (Gurley et al., 2008). Remarkably, despite this polarity reversal, two small neuronal clusters known as 'brain primordia' differentiate within

these posteriorized anterior blastemas (Evans et al., 2011; Iglesias et al., 2011). These brain primordia are equivalent to what it has also been termed 'brain rudiments' and correspond to the initial neuronal clusters that will differentiate into the new cephalic ganglia (Cebrià et al., 2002b; Kobayashi et al., 2007; Agata and Umesono, 2008).

As previously shown, *egr-4* silencing inhibited the differentiation of normal cephalic ganglia (Fig. 5A). After *apc-1* RNAi, approximately one-third of the tails regenerated small brain primordia with a similar morphology to that described previously (Evans et al., 2011; Iglesias et al., 2011). By contrast, no brain tissues were detected in most of the double *egr-4(RNAi);apc-1(RNAi)* regenerating tails, although in very few cases the differentiation of very reduced brain primordia was observed (Fig. 5A,B). Although *in situ* hybridization revealed no brain primordia in over half of the *apc-1(RNAi)* tails, differentiation of a pharynx was observed in most cases, indicating that the regenerative process had been somehow moved forward. By contrast, in *egr-4(RNAi);apc-1(RNAi)* animals, a greater number of tails failed to regenerate any brain primordia, when compared with either



**Fig. 5. *egr-4* is required for the differentiation of the brain primordia and the effects of its silencing are reversed by  $\beta$ cat1 RNAi.** (A) *In situ* hybridization with the brain marker *Smed-gpas* in tail fragments after 7-9 days of regeneration. After *egr-4;apc-1* RNAi, most animals failed to differentiate any brain or developed very small brain primordia (arrowheads) when compared with *apc-1(RNAi)* samples. (B) Quantification of the different phenotypes obtained after individual and simultaneous silencing of *egr-4* and *apc-1*. In all cases in which a normal brain (blue) or some brain tissue (yellow and green) differentiated, a pharynx was also observed. (C) Double *egr-4;beta-cat1* RNAi rescued *egr-4(RNAi)* phenotypes. In live images of bipolar trunks after 7 days of regeneration following two rounds of RNAi and amputation, *egr-4* knockdown resulted in animals with cyclopic eyes or small blastemas without eyes.  $\beta$ cat1(RNAi) animals displayed the typical anteriorization of the posterior blastema; some samples displayed cyclopia in the posterior head (arrowhead). Double *egr-4(RNAi);beta-cat1(RNAi)* animals displayed the same phenotype as  $\beta$ cat1(RNAi) animals, with a similar percentage of cyclopic posterior heads (arrowhead). *In situ* hybridization with *Smed-gpas* revealed that most *egr-4(RNAi)* animals regenerated smaller, mispatterned brains when compared with controls. All  $\beta$ cat1(RNAi) and *egr-4(RNAi);beta-cat1(RNAi)* animals regenerated normal anterior brains and most regenerated normal posterior brains. All panels are oriented with the anterior towards the left. Scale bars: 200  $\mu$ m in A; 300  $\mu$ m in C.

single *apc-1* or *egr-4* RNAi counterparts, and most of these failed to develop a normal pharynx (Fig. 5B). These results indicate that the differentiation of the brain primordia after *apc-1* RNAi requires *egr-4*.

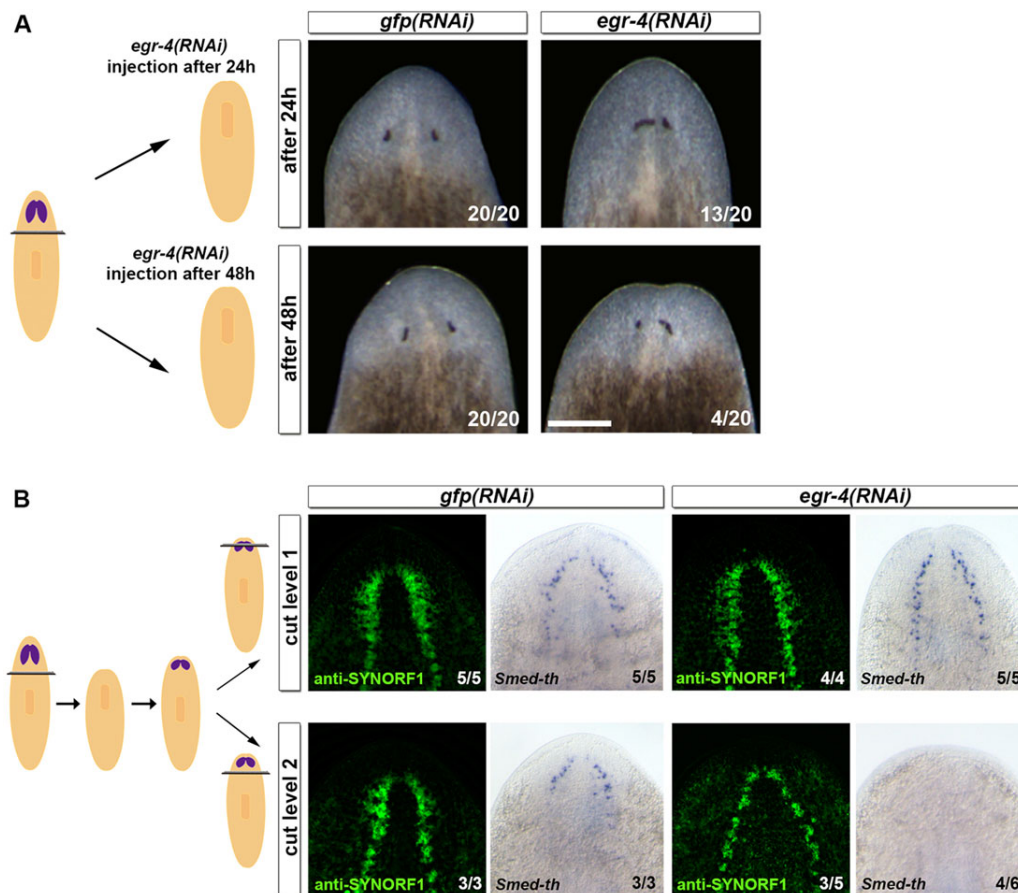
Given that *egr-4* RNAi inhibits only anterior regeneration and the Wnt/ $\beta$ -catenin pathway mediates the specification of head versus tail regeneration, we further investigated the relationship between *egr-4* and the Wnt/ $\beta$ -catenin pathway suggested by our data. Although most single *egr-4(RNAi)*-treated bipolar trunks displayed either smaller anterior blastemas with cyclopia or without eyes, treatment with double *egr-4(RNAi); $\beta$ cat1(RNAi)* resulted in two-headed animals with normal anterior heads and two bilateral eyes, as also observed after  *$\beta$ cat1* RNAi alone (Fig. 5C). The ability of  *$\beta$ cat1* RNAi to rescue head regeneration in *egr-4(RNAi)* animals suggests that the *egr-4(RNAi)* phenotype requires  *$\beta$ cat1*, and indicates that in normal physiological situations *egr-4* may antagonize  $\beta$ -catenin activity to allow head regeneration.

#### ***egr-4(RNAi)*-induced defects in regeneration depend on the timing of silencing and are rescued by the presence of brain tissue**

We next sought to delimit the period during which *egr-4* is required for head regeneration. *egr-4* was silenced by dsRNA injections 24 h and 48 h after amputation. Most of the animals injected 48 h post-amputation developed normal heads and only small number

displayed defects in eye morphogenesis (Fig. 6A). By contrast, most of the animals injected 24 h post-amputation exhibited defects in the regenerated eyes (including ectopic and fused eyes; Fig. 6A). However, all of these animals regenerated normal-sized heads and brains (data not shown), in contrast to the severe impairment in head regeneration observed when *egr-4* was silenced prior to amputation (Figs 2-5). Although it is unclear exactly how long after injection RNAi begins to exert its effect, these results suggest that *egr-4* function is necessary during the first 2-3 days post-amputation, after which it may not be essential for proper head regeneration.

As this time window coincides with the development of the brain primordia (Cebrià et al., 2002b), we speculated that no phenotype is observed when *egr-4* is silenced 48 h post-amputation because the brain primordia are already formed at this stage. We thus investigated whether *egr-4* RNAi impaired head regeneration even when some brain tissue remained after amputation. Animals were injected with *egr-4* dsRNA, amputated and allowed to regenerate. As expected, no defects in head regeneration were observed after this first round of amputation (see Materials and Methods). These animals were then re-injected (beginning 5 days post-amputation) and re-amputated at two different levels: level 1, immediately posterior to the newly regenerated eyes, leaving some of the regenerated brain tissue in the stump; and level 2, which involved decapitation and the removal of all brain tissue (Fig. 6B). Animals



**Fig. 6. The effects of *egr-4* RNAi depend on the time of the silencing and the presence of brain tissue.** (A) Injection of the first dose of *egr-4* dsRNA after 24 h of regeneration resulted in defects in the newly regenerated eyes of most animals. By contrast, most animals regenerated normally when the first injection was delivered 48 h after amputation. (B) After the first round of injection and amputation, regenerating animals were re-injected and re-amputated, leaving a portion of the newly regenerated small brain (level 1) or removing all brain tissue (level 2). Analysis of the expression of the CNS markers anti-SYNORF1 and *Smed-th* revealed that the presence of brain tissue rescued head regeneration after *egr-4* RNAi. Scale bar: 200  $\mu$ m.

amputated at level 1 regenerated normally and differentiated normal brains as evidenced by anti-SYNORF-1 immunostaining and *in situ* hybridization with *Smed-th* (Fig. 6B). However, most of the decapitated fragments from level 2 amputations failed to regenerate proper heads or cephalic ganglia (Fig. 6B). Similarly, *egr-4(RNAi)* animals that were amputated sagittally along the midline regenerated properly (supplementary material Fig. S8). These results suggest that the presence of brain tissue is sufficient to rescue the blockade of head regeneration following *egr-4* RNAi.

## DISCUSSION

### ***egr-4*, a downstream target of EGFR signaling, regulates cell differentiation in anterior blastemas**

EGFR signaling plays important roles in many biological processes by activating or inhibiting many downstream pathways, including PI3K/AKT, MAPK (mitogen-activated protein kinase), PLC $\gamma$  and JAK/STAT (signal transducer and activator of transcription) (Haley and Gullick, 2009). Previous reports suggest an important role for the EGFR signaling pathway in cell differentiation during planarian regeneration and homeostasis (Fraguas et al., 2011; Rink et al., 2011). *Smed-egfr-1* regulates eye-pigment cell differentiation; *Smed-egfr-3* seems to be necessary for blastema growth and cell differentiation (Fraguas et al., 2011); and *Smed-egfr-5* is required for flame cell maintenance and regeneration, and for guiding branch extension in protonephridia (Rink et al., 2011). These observations are consistent with the demonstrated role of the EGFR signaling pathway in cell differentiation in other organisms (Jones et al., 2009; Lejard et al., 2011; Harris and Horvitz, 2011); for example, EGFR is required for the differentiation of mammary epithelial cells (Mukhopadhyay et al., 2013) and human neural progenitors (Lemcke and Kuznetsov, 2013). Here, we identify *egr-4* as a putative target of *Smed-egfr-3* (Fig. 1) that is required for anterior regeneration, likely through its regulation of neoblast differentiation.

Although *Smed-egfr-3* is expressed in neoblasts and in the CNS of planarians (Fraguas et al., 2011; Fig. 1B), we found that *egr-4* was mainly expressed in the CNS and in the mesenchyme in irradiation-insensitive cells, suggesting that *egr-4* and *Smed-egfr-3* are co-expressed in the CNS (compare Fig. 1A and 1B). Indeed, silencing of *Smed-egfr-3* resulted in marked downregulation of *egr-4* expression in the cephalic ganglia (Fig. 1C).

Like other planarian *egr* homologues, *egr-4* expression was rapidly and locally upregulated after any small incision or during regeneration. Remarkably, this effect was independent of *Smed-egfr-3* (Fig. 1E',F). However, after 2 days of regeneration, the expression of *egr-4* in the blastema became *Smed-egfr-3* dependent (Fig. 1F). These results suggest that *egr-4* is expressed in two distinct phases during regeneration, an early, *Smed-egfr-3*-independent, phase, and a subsequent phase that is controlled by EGFR signaling. These findings are supported by previous studies demonstrating that the expression of *egr* genes is regulated by the EGFR pathway in different organisms and contexts (Kaufmann and Thiel, 2002; Tsai et al., 2000; Lindzen et al., 2012).

The silencing of *egr-4* impaired head regeneration, resulting in the development of either extremely reduced blastemas or small blastemas with aberrant photoreceptors and cephalic ganglia (Figs 2 and 3). Planarian regeneration is dependent on pluripotent stem cells known as neoblasts. Specific anti-SMEDWI-1 and *Smed-histone-2B* markers revealed no differences between control and *egr-4(RNAi)* animals (Fig. 2B). We did, however, observe differences in neoblast mitotic activity between these two groups. In *egr-4(RNAi)* animals, the first, wound-related mitotic peak (at 6 h) was markedly attenuated in anterior stumps and in posterior

stumps from head sections, when compared with controls. This decrease may reflect a role of early *egr-4* expression in this initial proliferative response. Interestingly, the second mitotic peak (associated with tissue loss and appearing at 48 h) was significantly attenuated only in anterior blastemas. This decrease may be a consequence, at least in part, of the marked attenuation of the first mitotic peak. However, because the decrease in the second peak was specific to anterior regeneration, we cannot rule out a potential role of later *egr-4* expression in regulating neoblast proliferation (directly or indirectly) at this stage. A normal bimodal proliferative response was observed after *egr-4* RNAi. Together with the slight attenuation of neoblast proliferation seen during posterior regeneration and the presence of a large number of mitotic cells in anterior blastemas at all stages (supplementary material Fig. S6), these findings suggest that the severe phenotypes observed after *egr-4* silencing are not solely due to defects in proliferation. Although *egr-4* silencing had no effect on early neoblast differentiation (normal anti-SMEDWI-1 and *Smed-NB21.11e* expression was observed), it resulted in a significant decrease in late neoblast progeny (Fig. 2D), suggesting that *egr-4* RNAi impairs late neoblast differentiation.

Taken together, these results suggest that *egr-4* is expressed in two distinct phases during regeneration. Early *egr-4* expression may participate in the initial (injury-induced) proliferative response, whereas the late expression, mediated by EGFR signaling, appears to regulate cell differentiation during anterior regeneration. These data provide the first evidence of a conserved EGFR/*egr* pathway in planarians. Further experiments will be necessary to better understand the relationship between *egr-4* and *Smed-egfr-3*, and to elucidate their specific roles during anterior and posterior regeneration.

### ***egr-4* is required for early differentiation of the cephalic ganglia downstream of polarity determinants**

*egr* genes have been implicated in cell proliferation, differentiation, inflammation, apoptosis, wound healing and liver regeneration (Thiel and Cibelli, 2002; Dussmann et al., 2011). Several *egr* genes have been identified in planarians (Sandmann et al., 2011; Wagner et al., 2012; Wenemoser et al., 2012).

Many of these genes, including *egr-4*, are upregulated immediately after injury or amputation and are barely expressed (if at all) in intact planarians (Sandmann et al., 2011; Wenemoser et al., 2012). To date, no functional characterization of the injury-induced expression of planarian *egr* genes has been reported. However, given the decrease reported here in the first mitotic peak after *egr-4* RNAi and the demonstrated role of several *egr* genes in wound healing, tissue fibrosis and inflammatory responses in other models (Schmidt et al., 2008; Wu et al., 2009; Chen et al., 2006), it seems plausible that early *egr* genes expression in planarians is involved in the initial stages of regeneration.

*egr-4* was also expressed in the mature and differentiating CNS (Fig. 1). Given the conserved function of *egr* genes in neural development (reviewed by O'Donovan et al., 1999; Pérez-Cadahía et al., 2011) and the severe phenotypes observed in the regenerating CNS after *egr-4* RNAi, we investigated whether this gene is required for the regeneration of cephalic ganglia in planarians. The results of double RNAi of *egr-4* and *Smed-apc-1* (Fig. 5) suggest that *egr-4* is necessary for the early differentiation of the brain primordia. However, *egr-4* was not required for the maintenance of the CNS either in intact planarians (supplementary material Fig. S7) or in regenerating sections in which brain tissue remained after amputation (Fig. 6). These findings identify *egr-4* as the first gene



known to be involved in the initial stages of neural regeneration in planarians. This role in neural differentiation is in agreement with functions attributed to *egr* genes in other organisms. For example, *egr-1* regulates astrocyte growth through via activation of the ERK signaling cascade triggered by EGF receptors (Mayer et al., 2009; Biesiada et al., 1996). In cultured PC12 cells, p35 is essential for neurite outgrowth and is induced via the *egr-1*-mediated activation of ERK (Harada et al., 2001). Similarly, *egr-1* regulates neurite extension during development in *Xenopus* (Anelli et al., 2013), and ERK-dependent *egr-4* expression is required for the maturation of GABAergic neurons (Ludwig et al., 2011a).

The silencing of different genes required for the respecification of anterior polarity or pole formation produces phenotypes similar to those described here after *egr-4* RNAi (Petersen and Reddien, 2009, 2011; Almuedo-Castillo et al., 2012; Roberts-Galbraith and Newmark, 2013; Blassberg et al., 2013; Chen et al., 2013). It is thus possible that the regeneration defects observed after *egr-4* RNAi are due to dysregulation of the respecification of polarity, given that double *egr-4(RNAi)/β-catenin(RNAi)* rescues proper head regeneration. However, although the knockdown of polarity or pole determinants completely silences *Smed-notum* (Roberts-Galbraith and Newmark, 2013; Blassberg et al., 2013; Chen et al., 2013), the expression of *Smed-notum* and *Smed-follistatin*, genes that are required for anterior polarity (Roberts-Galbraith and Newmark, 2013; Petersen and Reddien, 2011), is upregulated and largely maintained in regenerating *egr-4(RNAi)* animals. After normal upregulation of *Smed-notum* in *egr-4(RNAi)* animals during the first 24 h of regeneration (neoblast-independent upregulation; Chen et al., 2013), the recovery of the normal expression pattern in the days that follow is delayed when compared with controls. This delay may be associated with the attenuation of the second mitotic peak induced by *egr-4(RNAi)*. By 5-8 days of regeneration, most animals exhibited *Smed-notum* expression at the tip of the blastema. Moreover, *egr-4(RNAi)* did not affect the normal expression of *Smed-follistatin* at the blastema tip. We cannot rule out the possibility that the delay in restoring *Smed-notum* and *Smed-follistatin* expression at the tip of the blastema in *egr-4* knockdowns in turn influences polarity respecification (and/or maintenance), thus giving rise to the observed defects. Nonetheless, our findings suggest that the impaired head regeneration observed after *egr-4* RNAi is caused by disruption of *egr-4*-mediated differentiation of the cephalic ganglia.

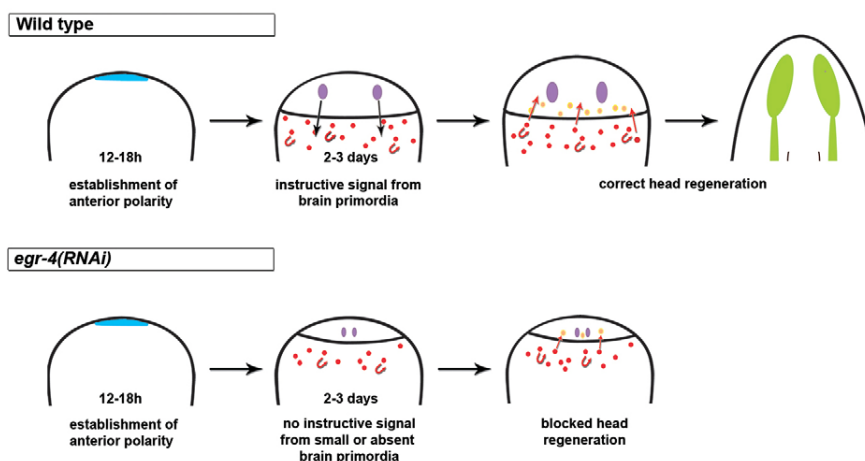
The results of the double *egr-4(RNAi)/β-catenin(RNAi)* suggest that the impairment of head and cephalic ganglia regeneration after *egr-4*

silencing is  $\beta$ -catenin dependent. Thus, in wild-type planarians, *egr-4* may inhibit  $\beta$ -catenin activity to allow head regeneration. In support of this hypothesis, defects in brain regeneration were more severe in *egr-4(RNAi);apc-1(RNAi)* animals, in which  $\beta$ -catenin activity is augmented. These results suggest that the effect of *egr-4* on cephalic ganglia regeneration is at least partially mediated by antagonism of the Wnt/ $\beta$ -catenin pathway. Although the Wnt/ $\beta$ -catenin pathway has been primarily associated with the establishment of AP polarity (Iglesias et al., 2008; Gurley et al., 2008; Petersen and Reddien, 2008), other studies suggest that this pathway may regulate the differentiation of cephalic ganglia independently of axial polarity (Iglesias et al., 2011; Blassberg et al., 2013).

Overall, our data point to *egr-4* as a unique factor that plays a key role in the differentiation of the brain primordia by antagonizing  $\beta$ -catenin function downstream of the polarity determinants *Smed-notum* and *Smed-wnt1*. Further experiments will be required to unravel the exact relationship between *egr-4* and Wnt/ $\beta$ -catenin signaling.

### The role of the cephalic ganglia in head regeneration

If *egr-4* is required for the early development of the cephalic ganglia, how does the inhibition of CNS differentiation block head regeneration? Previous studies have shown that the inhibition of *Smed-egr-3* (Fraguas et al., 2011) and ERK signaling (Tasaki et al., 2011) yields phenotypes similar to those obtained after *egr-4* RNAi: blastema growth is severely affected due to impaired cell differentiation. Tasaki and co-workers (2011) have suggested that ERK activation within the blastema is necessary for the early differentiation of an initial cohort of cells that is subsequently required to induce neoblast proliferation in the stump, and is probably necessary for the migration of neoblast progeny into the blastema, ultimately promoting blastema growth. We propose that this cohort of cells that is necessary to maintain blastema growth might be the initial brain primordia, the differentiation of which is dependent on *egr-4*. Under normal physiological situations, the brain primordia could emit some form of signal to promote neoblast proliferation in the stump and/or the migration of neoblast progeny cells into the blastema, where they subsequently differentiate. However, this mechanism may be dysregulated in *egr-4(RNAi)* animals, in which differentiation of the brain primordia is impaired (Fig. 7). Future experiments will be required to investigate this hypothesis. Further suggesting that brain cells are required for blastema growth, regenerating *egr-4(RNAi)* planarians in which some brain tissues was retained following amputation (Fig. 6; supplementary material Fig. S8) were capable of regenerating



**Fig. 7. Proposed model illustrating the requirement of brain primordia for head regeneration in planarians.** In wild type the brain primordia would send some signal to the stump to promote the proliferation, migration or differentiation of the neoblasts to allow blastema growth and head regeneration. In the absence of a proper brain primordia after *egr-4* RNAi, the lack of such putative inducing signal would explain the inhibition of head regeneration. Blue, anterior pole; red dots, proliferating neoblasts; orange dots, differentiating neoblasts; purple dots, brain primordia; green, mature CNS.

normal heads with proper brains. Given that EGFR/ERK signaling regulates *egr* genes in several other contexts (Mayer et al., 2009; Harada et al., 2001; Tarcic et al., 2012; Mukhopadhyay et al., 2013), planarians provide a unique opportunity with which to study the function of this conserved pathway during regeneration.

A recent study proposed that during planarian regeneration neoblasts follow a default program, triggered by ERK activation, to differentiate into head tissues (Umesono et al., 2013). These authors suggest that in posterior regions the Wnt/ $\beta$ -catenin pathway inhibits this ERK activation, resulting in regeneration of a tail. It will thus be of interest to analyze the relationship between *egr-4* and ERK in future studies to determine to what extent *egr-4* mediates the differentiation of head tissues after ERK activation.

Finally, the requirement of the brain primordia for blastema growth supports an evolutionarily conserved role of the nervous system in animal regeneration (Kumar et al., 2007; Kumar and Brockes, 2012; Miljkovic-Licina et al., 2007; Singer and Craven, 1948). Although several planarian studies have proposed a role of the nervous system in regeneration (Baguña et al., 1989; Cebrià and Newmark, 2007; Oviedo et al., 2010; Stéphan-Dubois and Lender, 1956), the underlying molecular basis remains unknown. Accordingly, *egr-4* may be the first identified gene linking the differentiation of the brain primordia with head regeneration in planarians.

## MATERIALS AND METHODS

### Animals

*Schmidtea mediterranea* from the BCN-10 clonal line were used in all experiments. Planarians were starved for at least 1 week prior to experiments. Genes and RNAi experiments were named using the nomenclature proposed by Reddien et al. (2008).

### Construction, sequencing and analysis of the DGE libraries

For the DGE experiment, total RNA from control (GFP) and *Smed-egr-3* (*RNAi*) planarians was extracted 1 day and 3 days after amputation using TRIzol (Invitrogen). Libraries were generated by the Skuldtech transcriptomic service (Montpellier, France). Sequence tag preparation was performed using Illumina's Digital Gene Expression Tag Profiling Kit according to the manufacturer's instructions (version 2.1B). Cluster generation was performed after applying 4 pM of each sample to the individual lanes of the Illumina 1G flowcell according to the manufacturer's instructions. Image analysis and base calling were performed using the Illumina Pipeline, from which sequence tags were obtained after purity filtering. This was followed by sorting and counting the unique tags.

Four libraries were obtained from control and *Smed-egr-3* (*RNAi*) animals on days 1 and 3 post-amputation (henceforth referred to as C1, C3, R1 and R3) (supplementary material Table S1). For comparison between libraries, tags with a sum of occurrences below 10 in the two compared sets or with a  $P > 0.005$  were discarded.  $P$ -value determination was performed as previously described (Piquemal et al., 2002). Tags were mapped using the short sequence mapping tools SeqMap (Jiang and Wong, 2008) and MPscan (Rivals et al., 2009) against the Smed454\_90edb transcriptome dataset (Abril et al., 2010). Annotation of the sequences and assignment of Gene Ontology (GO) categories was performed using the Blast2GO suite (Conesa et al., 2005).

### RNAi experiments

Double-stranded RNAs (dsRNA) were delivered into the planarian digestive system for three consecutive days as previously described (Sánchez-Alvarado and Newmark, 1999). All controls were injected with GFP dsRNA. After one round of dsRNA injections and amputation all *egr-4* (*RNAi*) planarians regenerated well. Defective phenotypes were identified after two rounds of injections and amputation. To increase the penetrance of the phenotypes the animals were injected with dsRNA for 3 consecutive days, cultured un-cut for 5 more days, re-injected for 3 consecutive days and then amputated pre- and post-pharyngeally. All

experiments were performed following this protocol unless otherwise specified. In all figures showing a quantification of the RNAi phenotypes when the fractions do not add up to 100% the missing animals are normal, unless specified in the main text.

### In situ hybridization

Gene expression analysis was carried out by whole-mount *in situ* hybridization, as previously described (Molina et al., 2007; Umesono et al., 1997; Cebrià et al., 2007; Pearson et al., 2009). All samples were observed through a Leica MZ16F stereomicroscope and images from representative organisms of each experiment were captured with a ProgRes C3 camera from Jenoptik. Confocal laser scanning microscopy was performed with a Leica SP2. A lethal dose of 100 Gy was used in irradiation experiments, and animals were fixed and hybridized at the indicated time points.

### Immunohistochemistry

Immunostaining was carried out as described previously (Cebrià and Newmark, 2005). The following antibodies were used: anti-SYNORF-1, a monoclonal antibody specific for synapsin, which was used as a pan-neural marker (Cebrià, 2008) (1:10; Developmental Studies Hybridoma Bank); anti-SMEDWI-1 (1:1500; Guo et al., 2006; März et al., 2013); and anti-phospho-histone H3 (H3P), which was used to detect mitotic cells (1:300; Cell Signaling Technology). Alexa 488-conjugated goat anti-mouse (1:400) and Alexa 568-conjugated goat anti-rabbit (1:1000; Molecular Probes) were used as secondary antibodies. Confocal laser scanning microscopy was performed using a Leica TCS-SPE and a Leica SP2. Confocal stacks from representative organisms in each experimental condition are shown.

### Acknowledgements

The authors thank Chema Martín-Durán for his help with the phylogenetic analyses and Josep Francesc Abril for providing the bioinformatics resources for analysis of the DGE data. We also thank Emili Saló and Teresa Adell for helpful comments and discussions. We thank Kerstin Bartscherer for the anti-SMEDWI-1 antibody. We thank Owen Howard for advice on English style.

### Competing interests

The authors declare no competing financial interests.

### Author contributions

F.C., S.F., M.I. and S.B. conceived and designed the experiments. S.F., S.B. and M.I. performed the experiments. G.R.-E. and F.C. analyzed the DGE data. S.F. and F.C. wrote the paper.

### Funding

S.F. is supported by an FPI fellowship (MICINN, Spain). S.B. is supported by a FI fellowship (Generalitat de Catalunya). This work was supported by Ministerio de Economía y Competitividad (Spain) [BFU2008-00710, BFU2011-22510 and BFU2012-31701] and by Generalitat de Catalunya [2009-SGR-1018] to F.C. Deposited in PMC for immediate release.

### Supplementary material

Supplementary material available online at <http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.101345/-DC1>

### References

- Abril, J. F., Cebrià, F., Rodríguez-Esteban, G., Horn, T., Fraguas, S., Calvo, B., Bartscherer, K. and Saló, E. (2010). Smed454 dataset: unravelling the transcriptome of *Schmidtea mediterranea*. *BMC Genomics* **11**, 731.
- Adell, T., Saló, E., Boutros, M. and Bartscherer, K. (2009). Smed-Evi/Wntless is required for beta-catenin-dependent and -independent processes during planarian regeneration. *Development* **136**, 905-910.
- Agata, K. and Umesono, Y. (2008). Brain regeneration from pluripotent stem cells in planarian. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 2071-2078.
- Aggeli, I. K., Beis, I. and Gaitanaki, C. (2010). ERKs and JNKs mediate hydrogen peroxide-induced Egr-1 expression and nuclear accumulation in H9c2 cells. *Physiol. Res.* **59**, 443-454.
- Almuedo-Castillo, M., Sureda-Gómez, M. and Adell, T. (2012). Wnt signaling in planarians: new answers to old questions. *Int. J. Dev. Biol.* **56**, 53-65.
- Anelli, T., Cardarelli, S., Ori, M., Nardi, I., Biagioni, S. and Poiana, G. (2013). 5-Hydroxytryptamine 1A and 2B serotonin receptors in neurite outgrowth: involvement of early growth response protein 1. *Dev. Neurosci.* **35**, 450-460.

- Baguña, J., Saló, E. and Romero, R. (1989). Effects of activators and antagonists of the neuropeptides substance P and substance K on cell proliferation in planarians. *Int. J. Dev. Biol.* **33**, 261-266.
- Biesiada, E., Razandi, M. and Levin, E. R. (1996). Egr-1 activates basic fibroblast growth factor transcription. Mechanistic implications for astrocyte proliferation. *J. Biol. Chem.* **271**, 18576-18581.
- Blassberg, R. A., Felix, D. A., Tejada-Romero, B. and Aboobaker, A. A. (2013). PBX/extradenticle is required to re-establish axial structures and polarity during planarian regeneration. *Development* **140**, 730-739.
- Cabodi, S., Morello, V., Masi, A., Cicchi, R., Broggio, C., Distefano, P., Brunelli, E., Silengo, L., Pavone, F., Arcangeli, A. et al. (2008). Convergence of integrins and EGF receptor signaling via PI3K/Akt/FoxO pathway in early gene egr-1 expression. *J. Cell. Physiol.* **218**, 294-303.
- Calogero, A., Lombardi, V., De Gregorio, G., Porcellini, A., Ucci, S., Arcella, A., Caruso, R., Gagliardi, F. M., Gulino, A., Lanzetta, G. et al. (2004). Inhibition of cell growth by EGR-1 in human primary cultures from malignant glioma. *Cancer Cell Int.* **4**, 1.
- Cebrià, F. (2008). Organization of the nervous system in the model planarian *Schmidtea mediterranea*: an immunocytochemical study. *Neurosci. Res.* **61**, 375-384.
- Cebrià, F. and Newmark, P. A. (2005). Planarian homologs of netrin and netrin receptor are required for proper regeneration of the central nervous system and the maintenance of nervous system architecture. *Development* **132**, 3691-3703.
- Cebrià, F. and Newmark, P. A. (2007). Morphogenesis defects are associated with abnormal nervous system regeneration following roboA RNAi in planarians. *Development* **134**, 833-837.
- Cebrià, F., Kobayashi, C., Umesono, Y., Nakazawa, M., Mineta, K., Ikeo, K., Gojobori, T., Itoh, M., Taira, M., Sánchez Alvarado, A. et al. (2002a). FGFR-related gene *nou-darake* restricts brain tissues to the head region of planarians. *Nature* **419**, 620-624.
- Cebrià, F., Nakazawa, M., Mineta, K., Ikeo, K., Gojobori, T. and Agata, K. (2002b). Dissecting planarian central nervous system regeneration by the expression of neural-specific genes. *Dev. Growth Differ.* **44**, 135-146.
- Cebrià, F., Guo, T., Jopek, J. and Newmark, P. A. (2007). Regeneration and maintenance of the planarian midline is regulated by a slit orthologue. *Dev. Biol.* **307**, 394-406.
- Chandra, A., Lan, S., Zhu, J., Siclari, V. A. and Qin, L. (2013). Epidermal growth factor receptor (EGFR) signaling promotes proliferation and survival in osteoprogenitors by increasing early growth response 2 (EGR2) expression. *J. Biol. Chem.* **288**, 20488-20498.
- Chen, S.-J., Ning, H., Ishida, W., Sodin-Semrl, S., Takagawa, S., Mori, Y. and Varga, J. (2006). The early-immediate gene EGR-1 is induced by transforming growth factor-beta and mediates stimulation of collagen gene expression. *J. Biol. Chem.* **281**, 21183-21197.
- Chen, C.-C. G., Wang, I. E. and Reddien, P. W. (2013). *pbx* is required for pole and eye regeneration in planarians. *Development* **140**, 719-729.
- Cole, A. J., Saffen, D. W., Baraban, J. M. and Worley, P. F. (1989). Rapid increase of an immediate early gene messenger RNA in hippocampal neurons by synaptic NMDA receptor activation. *Nature* **340**, 474-476.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M. and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674-3676.
- Dussmann, P., Pagel, J. I., Vogel, S., Magnusson, T., Zimmermann, R., Wagner, E., Schaper, W., Ogris, M. and Deindl, E. (2011). Live in vivo imaging of Egr-1 promoter activity during neonatal development, liver regeneration and wound healing. *BMC Dev. Biol.* **11**, 28.
- Eisenhoffer, G. T., Kang, H. and Sánchez Alvarado, A. (2008). Molecular analysis of stem cells and their descendants during cell turnover and regeneration in the planarian *Schmidtea mediterranea*. *Cell Stem Cell* **3**, 327-339.
- Evans, D. J., Owlarn, S., Tejada-Romero, B., Chen, C. and Aboobaker, A. A. (2011). Combining classical and molecular approaches elaborates on the complexity of mechanisms underpinning anterior regeneration. *PLoS ONE* **6**, e27927.
- Felix, D. A. and Aboobaker, A. A. (2010). The TALE class homeobox gene *Smed-prep* defines the anterior compartment for head regeneration. *PLoS Genet.* **6**, e1000915.
- Fraguas, S., Barberán, S. and Cebrià, F. (2011). EGFR signaling regulates cell proliferation, differentiation and morphogenesis during planarian regeneration and homeostasis. *Dev. Biol.* **354**, 87-101.
- Fraguas, S., Barberán, S., Ibarra, B., Stöger, L. and Cebrià, F. (2012). Regeneration of neuronal cell types in *Schmidtea mediterranea*: an immunohistochemical and expression study. *Int. J. Dev. Biol.* **56**, 143-153.
- Guo, T., Peters, A. H. F. M. and Newmark, P. A. (2006). A Bruno-like gene is required for stem cell maintenance in planarians. *Dev. Cell* **11**, 159-169.
- Gurley, K. A., Rink, J. C. and Sánchez Alvarado, A. (2008). Beta-catenin defines head versus tail identity during planarian regeneration and homeostasis. *Science* **319**, 323-327.
- Gurley, K. A., Elliott, S. A., Simakov, O., Schmidt, H. A., Holstein, T. W. and Sánchez Alvarado, A. (2010). Expression of secreted Wnt pathway components reveals unexpected complexity of the planarian amputation response. *Dev. Biol.* **347**, 24-39.
- Haley, J. D. and Gullick, W. J. (2009). EGFR Signaling Networks in Cancer Therapy. *New York, USA*: Humana Press.
- Harada, T., Morooka, T., Ogawa, S. and Nishida, E. (2001). ERK induces p35, a neuron-specific activator of Cdk5, through induction of Egr1. *Nat. Cell Biol.* **3**, 453-459.
- Harris, D. T. and Horvitz, H. R. (2011). MAB-10/NAB acts with LIN-29/EGR to regulate terminal differentiation and the transition from larva to adult in *C. elegans*. *Development* **138**, 4051-4062.
- Iglesias, M., Gomez-Skarmeta, J. L., Saló, E. and Adell, T. (2008). Silencing of *Smed-beta-catenin1* generates radial-like hypercephalized planarians. *Development* **135**, 1215-1221.
- Iglesias, M., Almuedo-Castillo, M., Aboobaker, A. A. and Saló, E. (2011). Early planarian brain regeneration is independent of blastema polarity mediated by the Wnt/ $\beta$ -catenin pathway. *Dev. Biol.* **358**, 68-78.
- Jiang, H. and Wong, W. H. (2008). SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* **24**, 2395-2396.
- Jones, J. E., Wang, L., Kropf, P. L., Duan, R. and Johnson, D. E. (2009). Src family kinase gene targets during myeloid differentiation: identification of the *EGR-1* gene as a direct target. *Leukemia* **23**, 1933-1935.
- Kaufmann, K. and Thiel, G. (2001). Epidermal growth factor and platelet-derived growth factor induce expression of Egr-1, a zinc finger transcription factor, in human malignant glioma cells. *J. Neuro. Sci.* **189**, 83-91.
- Kaufmann, K. and Thiel, G. (2002). Epidermal growth factor and thrombin induced proliferation of immortalized human keratinocytes is coupled to the synthesis of Egr-1, a zinc finger transcriptional regulator. *J. Cell. Biochem.* **85**, 381-391.
- Kobayashi, C., Saito, Y., Ogawa, K. and Agata, K. (2007). Wnt signaling is required for antero-posterior patterning of the planarian brain. *Dev. Biol.* **306**, 714-724.
- Kumar, A. and Brockes, J. P. (2012). Nerve dependence in tissue, organ, and appendage regeneration. *Trends Neurosci.* **35**, 691-699.
- Kumar, A., Godwin, J. W., Gates, P. B., Garza-Garcia, A. A. and Brockes, J. P. (2007). Molecular basis for the nerve dependence of limb regeneration in an adult vertebrate. *Science* **318**, 772-777.
- Lejard, V., Blais, F., Guerquin, M.-J., Bonnet, A., Bonnin, M.-A., Havis, E., Malbouyres, M., Bidaud, C. B., Maro, G., Gilardi-Hebenstreit, P. et al. (2011). EGR1 and EGR2 involvement in vertebrate tendon differentiation. *J. Biol. Chem.* **286**, 5855-5867.
- Lemcke, H. and Kuznetsov, S. A. (2013). Involvement of connexin43 in the EGF/EGFR signalling during self-renewal and differentiation of neural progenitor cells. *Cell. Signal.* **25**, 2676-2684.
- Lindzen, M., Carvalho, S., Starr, A., Ben-Chetrit, N., Pradeep, C.-R., Köstler, W. J., Rabinkov, A., Lavi, S., Bacus, S. S. and Yarden, Y. (2012). A recombinant decoy comprising EGFR and ErbB-4 inhibits tumor growth and metastasis. *Oncogene* **31**, 3505-3515.
- Ludwig, A., Uvarov, P., Soni, S., Thomas-Crusells, J., Airaksinen, M. S. and Rivera, C. (2011a). Early growth response 4 mediates BDNF induction of potassium chloride cotransporter 2 transcription. *J. Neurosci.* **31**, 644-649.
- Ludwig, A., Uvarov, P., Pellegrino, C., Thomas-Crusells, J., Schuchmann, S., Saarma, M., Airaksinen, M. S. and Rivera, C. (2011b). Neurturin evokes MAPK-dependent upregulation of Egr4 and KCC2 in developing neurons. *Neural Plast.* **2011**, 1-8.
- März, M., Seebeck, F. and Bartscherer, K. (2013). A Pitx transcription factor controls the establishment and maintenance of the serotonergic lineage in planarians. *Development* **140**, 4499-4509.
- Mayer, S. I., Rössler, O. G., Endo, T., Charnay, P. and Thiel, G. (2009). Epidermal-growth-factor-induced proliferation of astrocytes requires Egr transcription factors. *J. Cell Sci.* **122**, 3340-3350.
- Milbrandt, J. (1987). A nerve growth factor-induced gene encodes a possible transcriptional regulatory factor. *Science* **238**, 797-799.
- Miljkovic-Licina, M., Chera, S., Ghila, L. and Galliot, B. (2007). Head regeneration in wild-type hydra requires de novo neurogenesis. *Development* **134**, 1191-1201.
- Molina, M. D., Saló, E. and Cebrià, F. (2007). The BMP pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. *Dev. Biol.* **311**, 79-94.
- Mukhopadhyay, C., Zhao, X., Maroni, D., Band, V. and Naramura, M. (2013). Distinct effects of EGFR ligands on human mammary epithelial cell differentiation. *PLoS ONE* **8**, e75907.
- Nishimura, K., Kitamura, Y., Inoue, T., Umesono, Y., Sano, S., Yoshimoto, K., Inden, M., Takata, K., Taniguchi, T., Shimohama, S. et al. (2007). Reconstruction of dopaminergic neural network and locomotion function in planarian regenerates. *Dev. Neurobiol.* **67**, 1059-1078.
- Nishimura, K., Kitamura, Y., Inoue, T., Umesono, Y., Yoshimoto, K., Taniguchi, T. and Agata, K. (2008). Characterization of tyramine beta-hydroxylase in planarian *Dugesia japonica*: cloning and expression. *Neurochem. Int.* **53**, 184-192.
- O'Donovan, K. J., Tourtellotte, W. G., Millbrandt, J. and Baraban, J. M. (1999). The EGR family of transcription-regulatory factors: progress at the interface of molecular and systems neuroscience. *Trends Neurosci.* **22**, 167-173.

- Önal, P., Grün, D., Adamidi, C., Rybak, A., Solana, J., Mastrobuoni, G., Wang, Y., Rahn, H.-P., Chen, W., Kempa, S. et al. (2012). Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *EMBO J.* **31**, 2755-2769.
- Oviedo, N. J., Morokuma, J., Walentek, P., Kema, I. P., Gu, M. B., Ahn, J.-M., Hwang, J. S., Gojoberi, T. and Levin, M. (2010). Long-range neural and gap junction protein-mediated cues control polarity during planarian regeneration. *Dev. Biol.* **339**, 188-199.
- Pearson, B. J., Eisenhoffer, G. T., Gurley, K. A., Rink, J. C., Miller, D. E. and Sánchez Alvarado, A. (2009). Formaldehyde-based whole-mount in situ hybridization method for planarians. *Dev. Dyn.* **238**, 443-450.
- Pérez-Cadahía, B., Drobic, B. and Davie, J. R. (2011). Activation and function of immediate-early genes in the nervous system. *Biochem. Cell Biol.* **89**, 61-73.
- Petersen, C. P. and Reddien, P. W. (2008). Smed-betacatenin-1 is required for anteroposterior blastema polarity in planarian regeneration. *Science* **319**, 327-330.
- Petersen, C. P. and Reddien, P. W. (2009). A wound-induced Wnt expression program controls planarian regeneration polarity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17061-17066.
- Petersen, C. P. and Reddien, P. W. (2011). Polarized notum activation at wounds inhibits Wnt function to promote planarian head regeneration. *Science* **332**, 852-855.
- Piquemal, D., Commes, T., Manchon, L., Lejeune, M., Ferraz, C., Pugnère, D., Demaille, J., Elalouf, J.-M. and Marti, J. (2002). Transcriptome analysis of monocytic leukemia cell differentiation. *Genomics* **80**, 361-371.
- Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R. and Sánchez Alvarado, A. (2005). Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Dev. Cell* **8**, 635-649.
- Reddien, P. W., Newmark, P. A. and Sánchez-Alvarado, A. (2008). Gene nomenclature guidelines for the planarian *Schmidtea mediterranea*. *Dev. Dyn.* **237**, 3099-3101.
- Rink, J. C., Gurley, K. A., Elliott, S. A. and Sánchez Alvarado, A. (2009). Planarian Hh signaling regulates regeneration polarity and links Hh pathway evolution to cilia. *Science* **326**, 1406-1410.
- Rink, J. C., Vu, H. T.-K. and Sánchez Alvarado, A. (2011). The maintenance and regeneration of the planarian excretory system are regulated by EGFR signaling. *Development* **138**, 3769-3780.
- Rivals, E., Salmela, L., Kiiskinen, P., Kalsi, P. and Tarhio, J. (2009). MPSCAN: fast localisation of multiple reads in genomes. *Lect. Notes Bioinform.* **5724**, 246-260.
- Roberts-Galbraith, R. H. and Newmark, P. A. (2013). Follistatin antagonizes activin signaling and acts with notum to direct planarian head regeneration. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1363-1368.
- Sánchez-Alvarado, A. and Newmark, P. A. (1999). Double-stranded RNA specifically disrupts gene expression during planarian regeneration. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5049-5054.
- Sandmann, T., Vogg, M. C., Owlarn, S., Boutros, M. and Bartscherer, K. (2011). The head-regeneration transcriptome of the planarian *Schmidtea mediterranea*. *Genome Biol.* **12**, R76.
- Schmidt, J., Stoffels, B., Moore, B. A., Chanthaphavong, R. S., Mazie, A. R., Buchholz, B. M. and Bauer, A. J. (2008). Proinflammatory role of leukocyte-derived Egr-1 in the development of murine postoperative ileus. *Gastroenterology* **135**, 926-936.
- Shafarenko, M., Liebermann, D. A. and Hoffman, B. (2005). Egr-1 abrogates the block imparted by c-Myc on terminal M1 myeloid differentiation. *Blood* **106**, 871-878.
- Singer, M. and Craven, L. (1948). The growth and morphogenesis of the regenerating forelimb of adult Triturus following denervation at various stages of development. *J. Exp. Zool.* **108**, 279-308.
- Solana, J., Kao, D., Mihaylova, Y., Jaber-Hijazi, F., Malla, S., Wilson, R. and Aboobaker, A. (2012). Defining the molecular profile of planarian pluripotent stem cells using a combinatorial RNA-seq, RNA interference and irradiation approach. *Genome Biol.* **13**, R19.
- Stéphan-Dubois, F. and Lender, T. (1956). Corrélatons humorales dans la régénération de planaires paludicoles. *Ann. des Sc. Nat. Zool.* **18**, 223-230.
- Sukhatme, V. P., Kartha, S., Toback, F. G., Taub, R., Hoover, R. G. and Tsai-Morris, C. H. (1987). A novel early growth response gene rapidly induced by fibroblast, epithelial cell and lymphocyte mitogens. *Oncogene Res.* **1**, 343-355.
- Sukhatme, V. P., Cao, X., Chang, L. C., Tsai-Morris, C.-H., Stamenkovich, D., Ferreira, P. C. P., Cohen, D. R., Edwards, S. A., Shows, T. B., Curran, T. et al. (1988). A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization. *Cell* **53**, 37-43.
- Tarcic, G., Avraham, R., Pines, G., Amit, I., Shay, T., Lu, Y., Zwang, Y., Katz, M., Ben-Chetrit, N., Jacob-Hirsch, J. et al. (2012). EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF. *FASEB J.* **26**, 1582-1592.
- Tasaki, J., Shibata, N., Nishimura, O., Itomi, K., Tabata, Y., Son, F., Suzuki, N., Araki, R., Abe, M., Agata, K. et al. (2011). ERK signaling controls blastema cell differentiation during planarian regeneration. *Development* **138**, 2417-2427.
- Thiel, G. and Cibelli, G. (2002). Regulation of Life and Death by the zinc finger transcription factor egr-1. *J. Cell. Physiol.* **193**, 287-292.
- Tsai, J. C., Liu, L., Guan, J. and Aird, W. C. (2000). The egr-1 gene is induced by epidermal growth factor in ECV304 cells and primary endothelial cells. *Am. J. Physiol. Cell Physiol.* **279**, 1414-1424.
- Umesono, Y., Watanabe, K. and Agata, K. (1997). A planarian orthopedia homolog is specifically expressed in the branch region of both the mature and regenerating brain. *Dev. Growth Differ.* **9**, 723-727.
- Umesono, Y., Watanabe, K. and Agata, K. (1999). Distinct structural domains in the planarian brain defined by the expression of evolutionarily conserved homeobox genes. *Dev. Genes Evol.* **209**, 31-39.
- Umesono, Y., Tasaki, J., Nishimura, Y., Hrouda, M., Kawaguchi, E., Yazawa, S., Nishimura, O., Hosoda, K., Inoue, T. and Agata, K. (2013). The molecular logic for planarian regeneration along the anterior-posterior axis. *Nature* **500**, 73-76.
- Wagner, D. E., Ho, J. J. and Reddien, P. W. (2012). Genetic regulators of a pluripotent adult stem cell system in planarians identified by RNAi and clonal analysis. *Cell Stem Cell* **10**, 299-311.
- Wenemoser, D. and Reddien, P. W. (2010). Planarian regeneration involves distinct stem cell responses to wounds and tissue absence. *Dev. Biol.* **344**, 979-991.
- Wenemoser, D., Lapan, S. W., Wilkinson, A. W., Bell, G. W. and Reddien, P. W. (2012). A molecular wound response program associated with regeneration initiation in planarians. *Genes Dev.* **26**, 988-1002.
- Wu, M., Melichian, D. S., de la Garza, M., Gruner, K., Bhattacharyya, S., Barr, L., Nair, A., Shahrara, S., Sporn, P. H. S., Mustoe, T. A. et al. (2009). Essential roles for early growth response transcription factor Egr-1 in tissue fibrosis and wound healing. *Am. J. Pathol.* **175**, 1041-1055.
- Yazawa, S., Umesono, Y., Hayashi, T., Tarui, H. and Agata, K. (2009). Planarian Hedgehog/Patched establishes anterior-posterior polarity by regulating Wnt signaling. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22329-22334.



# **Agradecimientos**



¡Muchas gracias a todos!





# Notas







