

71-7  
FACH-FC  
Biotecnología  
Z83  
C.1

FACULTAD DE CIENCIAS  
UNIVERSIDAD DE CHILE

IDENTIFICACIÓN Y CARACTERIZACIÓN DE GENES  
DIFERENCIALMENTE EXPRESADOS DURANTE LA  
GASTRULACIÓN DE *Drosophila melanogaster*.

Memoria de Título  
Entregada a la  
Universidad de Chile  
En cumplimiento parcial de los requisitos  
para optar al Título de  
Ingeniero en Biotecnología Molecular



Por  
Alejandro Antonio Zúñiga Prado  
Julio, 2004  
Santiago - Chile

Director de Tesis: Dra. Verónica Cambiazo Ayala.  
Laboratorio de Bioinformática y Expresión Génica  
INTA – Universidad de Chile



**“IDENTIFICACIÓN Y CARACTERIZACIÓN DE GENES  
DIFERENCIALMENTE EXPRESADOS DURANTE LA  
GASTRULACION DE *Drosophila melanogaster*”**

Memoria de Título entregada a la Universidad de Chile en cumplimiento parcial de los requisitos para optar al Título de Ingeniero en Biotecnología Molecular.

**ALEJANDRO ANTONIO ZUÑIGA PRADO**

**Directora de Memoria de Título**

Dra. Verónica Cambiazo Ayala

Firma manuscrita en azul de Verónica Cambiazo Ayala, sobre una línea horizontal.

**Comisión de Evaluación de la Memoria de Título**



**Dr. Miguel Allende Connelly**

Presidente Comisión

Dra. Verónica Palma Alvarado

Firma manuscrita en azul de Miguel Allende Connelly, sobre una línea horizontal.

Firma manuscrita en azul de Verónica Palma Alvarado, sobre una línea horizontal.

Santiago de Chile, \_\_\_\_\_ 2004



*A mis padres Eustaquio y Antonieta  
Quienes impulsaron mis primeros pasos, me mostraron el mundo  
y comenzaron mi historia.*

*A mi hermano Rodrigo, por ser mi guía y modelo*

*A María Inés y Tamara, por su alegría, cariño y entusiasmo*

*A Andreita, mi futura esposa, por su amor incondicional  
por su apoyo constante durante este trabajo  
por convertirse en el pilar fundamental de mi vida.*

## AGRADECIMIENTOS



Mas que dar gracias, debo declararme en bancarrota, por que estoy en deuda:

Con verónica, por creer en mí y darme la oportunidad de aprender, por sus palabras de aliento, por plantearme nuevos desafíos y nuevas metas, por impulsarme a ir cada vez mas lejos. Por ser un modelo de profesionalismo y pasión por la ciencia. Por enseñarme que el trabajo y el esfuerzo valen mas que el talento.

Con Mauricio, por su apoyo, su entusiasmo y por estimular mi intelecto. Por estar siempre dispuesto a razonar y discutir los porqués y los alcances de cada nuevo experimento y cada nuevo análisis.

Con Mauricio, Cristián y Hans, Miriam y Talía, los mas veteranos, los incondicionales que estuvieron conmigo desde el principio, los compañeros que me recibieron y me enseñaron.

Con Luis y Rodrigo, los más fieles, con los que siempre pude contar.

Con Marcos y Carlos, quienes nunca dudaron de mi.

Con Juan, Igor y Felipe, por su amistad y camaradería.

Con Pablo y Patricio, sus aportes han sido fundamentales.

A todos ustedes les debo lo que soy en ciencias y lo que soy es esta tesis.

Esta tesis se llevo a cabo bajo el financiamiento del proyecto FONDECYT 1010693.

# ÍNDICE DE CONTENIDOS



AGRADECIMIENTOS.....	iii
ÍNDICE DE CONTENIDOS.....	iv
ÍNDICE DE TABLAS.....	viii
ÍNDICE DE FIGURAS.....	ix
LISTA DE ABREVIATURAS.....	xi
RESUMEN.....	xiii
ABSTRACT.....	xv
INTRODUCCIÓN.....	1
1. Contribuciones del modelo <i>Drosophila melanogaster</i> al conocimiento biológico.....	1
2. Desarrollo embrionario de <i>Drosophila</i> : morfología y genética.....	3
3. Estudios a nivel genómico en organismos modelo.....	8
4. Hipótesis.....	14
5. Objetivo general.....	14
6. Objetivos específicos.....	14
MATERIALES Y MÉTODOS.....	16
1. MATERIALES.....	16
1.1. Cepas de <i>Drosophila melanogaster</i> .....	16
1.2. Cepas bacterianas.....	16
1.3. Vector.....	16



1.4. Partidores y adaptadores.....	17
1.5. Bases de datos y herramientas bioinformáticas.....	19
2. MÉTODOS.....	20
2.1. Colecta de embriones.....	20
2.2. Procedimientos de biología molecular.....	21
2.2.1. Electroforesis.....	21
2.2.2. Extracción de RNA total.....	21
2.2.3. Síntesis de cDNA.....	21
2.2.4. Purificación de DNA plasmidial a gran escala.....	23
2.2.5. PCR en colonias.....	24
2.2.6. PCR a gran escala.....	24
2.2.7. PCR cuantitativo en tiempo real.....	25
2.3. Sustracción.....	26
2.3.1. Hibridación sustractiva por supresión.....	26
2.3.2. Verificación de la eficiencia de la sustracción.....	27
2.3.3. Northern blots virtuales.....	27
2.4. Construcción de la genoteca.....	28
2.4.1. Ligación.....	28
2.4.2. Transformación.....	28
2.4.3. Selección de transformantes y construcción de la genoteca.....	29
2.4.5. Análisis de la presencia de insertos en los clones.....	29
2.5. Identificación y caracterización de genes.....	29
2.5.1. Obtención de secuencias.....	29
2.5.2. Análisis de secuencias.....	30
2.6. Hibridación en matrices.....	31

2.6.1. Matrices de hibridación.....	31
2.6.2. Obtención de sondas.....	32
2.6.3. Marcación de sondas.....	32
2.6.4. Hibridación con sondas radioactivas.....	34
2.7. Análisis de datos.....	34
2.7.1. Obtención de imágenes.....	34
2.7.2. Filtración, normalización y análisis de los datos.....	35
2.8. Hibridaciones <i>in situ</i> .....	38
 RESULTADOS.....	 40
 1. Obtención de un cDNA enriquecido en transcritos diferencialmente expresados en gástrula.....	 40
1.1. Sustracción de secuencias comunes.....	40
1.2. Eficiencia de la sustracción.....	42
2. Construcción y caracterización de una genoteca específica de gastrulación.....	45
2.1. Construcción de la genoteca de sustracción.....	45
2.2. Verificación de insertos clonados.....	45
2.3. Caracterización de la genoteca de sustracción.....	47
2.3.1. Estimación del porcentaje de genes diferencialmente expresados.....	47
2.3.2. Análisis de Northern blots virtuales.....	50
3. Identificación y caracterización de los genes aislados.....	51
3.1. Obtención y análisis de las secuencias.....	51
3.2. Identificación de los clones aislados.....	52



3.3. Caracterización de los genes aislados.....	57
3.4. Composición funcional de la genoteca.....	61
4. Análisis de expresión diferencial.....	64
4.1. Optimización de condiciones experimentales.....	64
4.2. Perfiles de expresión temporal.....	69
4.2.1. Obtención de sondas que representan los distintos estados del desarrollo embrionario.....	70
4.2.2. Medición de las abundancias relativas de genes en 5 intervalos de tiempo.....	71
4.2.3. Obtención de perfiles de expresión temporal.....	75
4.4. Expresión espacio temporal.....	85
 DISCUSIÓN.....	 87
1. Características de la metodología utilizada.....	88
2. Composición de la genoteca.....	91
3. Caracterización de los genes aislados.....	94
4. Genes de importancia en el desarrollo.....	96
5. Expresión temporal.....	100
 CONCLUSIONES.....	 109
 REFERENCIAS BIBLIOGRÁFICAS.....	 111

## ÍNDICE DE TABLAS



Tabla 1. Clasificación de los clones secuenciados.....	54
Tabla 2. Ortólogos de los genes no caracterizados.....	54
Tabla 3. Comparación de genes aislados en nuestra genoteca con aquellos identificados en otros experimentos de expresión génica.....	56
Tabla 4. Genes cuya expresión cambia significativamente entre gástrula y blastodermo sincicial.....	69
Tabla 5. Perfiles de expresión de genes que codifican proteínas.....	76
Tabla 6. Perfiles de expresión de clones no anotados y RNAs no traducidos.....	82

## ÍNDICE DE FIGURAS



Figura 1. Estados del desarrollo de <i>Drosophila melanogaster</i> .....	5
Figura 2. Productos de la hibridación sustractiva por supresión.....	42
Figura 3. Verificación de la eficiencia de la sustracción.....	44
Figura 4. Eficiencia de la transformación.....	46
Figura 5. Caracterización de la genoteca.....	48
Figura 6. Expresión diferencial de genes en el estado de gástrula.....	49
Figura 7. Northern blots virtuales.....	51
Figura 8. Frecuencia de genes con funciones moleculares asignadas.....	58
Figura 9. Frecuencia de genes con procesos biológicos asignados.....	59
Figura 10. Frecuencia de genes con componentes celulares asignados.....	60
Figura 11. Representatividad de funciones moleculares en la genoteca respecto al genoma.....	62
Figura 12. Representatividad de procesos biológicos en la genoteca respecto al genoma.....	63
Figura 13. Recuperación de señales obtenidas al sembrar DNA plasmidial o productos de PCR.....	66
Figura 14. Reproducibilidad entre membranas.....	67
Figura 15. Verificación de las sondas estado-específicas.....	71
Figura 16. Hibridación en matrices con cDNAs de embriones en diferentes estados de desarrollo.....	73
Figura 17. Cambio en la abundancia relativa de los transcritos en las distintas sondas.....	74
Figura 18. Perfiles de expresión promedio.....	80

Figura 19. Perfiles de expresión de grupos funcionales.....81

Figura 20. Agrupamiento de genes de acuerdo a sus perfiles de expresión.....84

Figura 21. Expresión espacio-temporal de los genes CG11212 y CG6234.....86

Figura 22. Grupos funcionales encontrados en la genoteca.....97



## LISTA DE ABREVIATURAS

BCIP	fosfato de 5-bromo-4-cloro-3-indol
BDGP	Berkeley Drosophila Genome Project
BLAST	Basic Local Alignment Search Tool
Bs	blastodermo sincicial
cDNA	DNA complementario
CDS	secuencia codificante
col	colaboradores
dATP	desoxiadenina
dCTP	desoxicitidina
dGTP	desoxiguanina
DNA	ácido desoxiribonucleico
dNTPs	desoxinucleotidos trifosfatos
DTT	ditiotreitól
dTTP	desoxitimidina
dUTP	desoxiuracilo
EBI	European Bioinformatics Institute
EST	expressed sequence tag
g	gramo
G	gástrula
GO	Gene Ontology
HS	solucion de hibridación
IPTG	isopropiltiogalactósido
Kb	kilo base

kDa	kilo dalton
L	litro
LB	medio de cultivo Luria broth
M	molaridad
ME	EGTA 10% en metanol
mRNA	RNA mensajero
NBT	nitro blue tetrazolium
pb	pares de bases
PBS	tampón fosfato salino
PCR	reacción en cadena de la polimerasa
PM	estandar de peso molecular
PP	paraformaldehído al 4% en PBS
RNA	ácido ribonucleico
rpm	revoluciones por minuto
RT	transcripción reversa
SDS	dodecil sulfato de sodio
SMART	Switch Mechanism At the 5' end of RNA Templates
TE	Tris - EDTA
Tris	Tris-(hidroximetil)-aminometano
U	unidad internacional de enzima, $\mu$ moles de sustrato procesado por minuto
UTR	region no traducida
UV	ultra violeta

## RESUMEN

Como un primer paso en la identificación de genes que participan en los procesos morfogénéticos de la gastrulación de *Drosophila melanogaster* en esta memoria de título se realizó una hibridación sustractiva por supresión entre cDNA de embriones en estado de gástrula y cDNA de embriones en estado de blastodermo sincicial con el objetivo de remover los transcritos maternos. El resultado de esta sustracción fue clonado generando una genoteca estado-específica enriquecida en secuencias que se expresan diferencialmente durante la gastrulación. Los análisis realizados mediante ensayos de hibridación en membranas y Northern blots virtuales indican que el 64% de los genes aislados aumenta su expresión al menos dos veces en gástrula respecto a blastodermo sincicial.

La secuenciación de 588 clones aislados permitió identificar 283 genes únicos, 138 de los cuales no han sido caracterizados. Con el objetivo de reconocer grupos funcionales, la composición de genes de la genoteca fue analizada tanto a través de la información que entrega la literatura como mediante la utilización de la base de datos de Gene Ontology. Un análisis comparativo de la composición de secuencias de la genoteca con aquella del genoma de *Drosophila*, reveló que la genoteca está constituida por un subconjunto de genes que operan en morfogénesis y especificación de patrones, además de migración y diferenciación celular en desmedro de procesos necesarios para mantener el estado estable de las células.

Durante el desarrollo de esta memoria se implementó la metodología de los *macroarrays* de cDNA, esto es la hibridación de sondas complejas en membranas de *nylon* que contenían secuencias de cDNA los clones aislados. Además, se diseñó un protocolo de análisis que permite reconocer cambios significativos en la expresión

temporal de los genes analizados. Utilizando esta metodología se realizaron experimentos de hibridación en membranas con el objetivo de obtener los perfiles de expresión de los genes identificados en la genoteca durante 5 intervalos de tiempo cubriendo los estados 2 al 12 del desarrollo embrionario de *Drosophila*. Estos análisis permitieron distinguir grupos de genes de expresión transitoria y otros de expresión sostenida durante los estados analizados y revelaron que 38 genes alcanzan su máximo de expresión durante la celularización del embrión, mientras que 17 presentan máximos de expresión durante la gastrulación y 98 en estados posteriores del desarrollo embrionario. Se analizaron con mayor detalle los 17 genes que aumentan de manera transitoria su expresión durante la gastrulación, ellos codifican para proteínas previamente caracterizadas o presuntas proteínas con funciones en la regulación de la dinámica del citoesqueleto, señalización intracelular y tráfico de membranas.

Este trabajo aporta una gran cantidad de información relacionada con la expresión temporal de los genes aislados, muchos de ellos no caracterizados, que combinada con la información contenida en otras bases de datos, tales como Gene Ontology o InterPro, nos permite proyectar estos resultados hacia la formulación de posibles redes de interacción génica que controlan los procesos de gastrulación. La información entregada es relevante en temas como genómica y biología molecular y aporta un conjunto de genes que son candidatos interesantes para estudios funcionales posteriores orientados a evaluar su participación en procesos particulares del desarrollo embrionario.

## ABSTRACT

As a first step towards the identification of genes that participate in the morphogenetic process of *Drosophila melanogaster* gastrulation, in this work we applied a suppression subtractive hybridization procedure using gastrula cDNA as tester and sincitial blastoderm cDNA as driver to remove maternal transcripts. The subtracted cDNA fraction was cloned to generate a stage-specific library that whose enriched in sequences differentially expressed during gastrulation. Analysis of macroarray and virtual Northern blots hybridization assays indicated that the level of expression of 64% of the isolated genes was at least two fold higher in gastrula than in sincitial blastoderm.

After sequencing 588 clones from our library, we identified 283 individual genes, 138 of them has not been characterized before. In order to recognize functional groups, the composition of genes in the library was analyzed using information from the literature and from the Gene Ontology data base. A comparative analysis between sequence composition of the library and the entire genome of *Drosophila* reveals that the library was mainly composed by genes that function in morphogenesis and patterning events as well as cell differentiation and migration. Genes that are necessary to maintain the steady state of cells were under-represented in our library.

In this work we implemented the cDNA macroarray technology, in doing so, we hybridized nylon membranes that contained the cDNA sequence of isolated clones with complex cDNA probes. In addition, we apply a protocol for data analysis that allowed us to recognize significant changes in the temporal expression of these genes. Using this methodology, we obtained the expression profiles of the isolated genes along 5 time points between stage 2 to 12 of *Drosophila* embryonic development. These analyses allowed us to distinguish groups of genes with transient and sustained expression

during the stages analyzed, revealing 38 genes that showed their highest level of expression during cellularization, while 17 genes showed their highest level of expression during gastrulation and 98 in later stages of embryogenesis. The group of 17 genes that displayed transient peaks of expression during gastrulation was analyzed with more detail. They code for known and putative proteins that function as regulators of cytoskeleton dynamics, components of intracellular signalling pathways and membrane trafficking.

This work contributes with the identification of temporal expression patterns of several genes, including a number of uncharacterized genes differentially expressed in gastrulation. This data along with the information available from databases such as Gene Ontology and InerPro, would permit us to formulate putative gene interaction networks that control gastrulation processes. The work presented here is relevant to the fields of genomics and molecular biology and provides a set of candidate genes for further functional studies that evaluate their role in particular process during embryonic development.

## INTRODUCCIÓN

### 1. Contribuciones del modelo *Drosophila melanogaster* al conocimiento biológico.

En 1910 T. H. Morgan escogió a *Drosophila* para estudiar la herencia, comenzando con una mosca mutante que tenía los ojos blancos. Él y tres de sus estudiantes, A. H. Sturtevant, C. B. Bridges y H. J. Muller en los siguientes 5 años propusieron la entonces revolucionaria teoría cromosómica de la herencia. Utilizando una metodología tan básica, como cruzar mutantes y contar descendientes, Morgan alcanzó logros notables que lo llevaron a ganar el premio Nóbel en 1933. En 1913, Sturtevant construyó el primer mapa genético mostrando que los genes estaban dispuestos en orden lineal. Muller por su parte, introdujo el uso de cromosomas balanceadores, que portan inversiones que impiden la recombinación y permiten mantener mutaciones letales en la forma de heterocigotos.

El mapeo físico de los genes tiene sus raíces en el descubrimiento de los cromosomas politénicos de las glándulas salivales de la mosca *Bibio hortulanus*. En 1934, T. S. Painter hizo los primeros dibujos de los cromosomas politénicos en *Drosophila*, los que incluían la localización cromosómica de varios genes. Entre 1935 y 1938, Bridges publicó mapas de estos cromosomas con tal exactitud que aún hoy son usados. Ahora sabemos que los genes fueron ubicados con un margen de error en intervalos de menos de 100 kb.

En 1927 Muller demostró que la radiación ionizante causa mutaciones, y en conjunto con rayos X provoca re-ordenamientos cromosómicos, con estos trabajos, recibió el premio Nóbel en 1946. D. L. Lindsley y L. Sandler en 1970 usaron sistemáticamente estos métodos para generar grupos ordenados de duplicaciones y

deleciones de 500 kb en los principales autosomas. Este trabajo fue la génesis del concepto de la búsqueda a escala genómica de perturbaciones genotípicas en los metazoos.

D. S. Hogness en 1972 propuso clonar insertos de gran tamaño para construir mapas físicos que permitieran un estudio más detallado de la estructura cromosómica. Para 1975, fueron fabricadas genotecas de clones que contenían el genoma completo y utilizadas para la búsqueda de secuencias específicas. Hacia 1978, segmentos de DNA cromosómicos que correspondían a mas de 200 kb de DNA geonómico fueron clonados en el bacteriófago *lambda* mediante la técnica de *chromosome walking*. En 1979, utilizando una inversión cromosómica fue posible localizar al complejo Bithorax de genes homeobox y realizar la primera clonación posicional de un gen, *ultrabithorax*.

En 1980, C. Nüsslein-Volhard y E. Wieschaus implementaron una estrategia sistemática de mutación y búsqueda de fenotipos letales que afectaban la cutícula de la larva con el objetivo de identificar todos los genes involucrados en procesos fundamentales del desarrollo. La caracterización molecular de los genes encontrados reveló que estos codifican para los componentes de las principales cascadas de transducción de señales responsables de que las células adquieran valores de posición en el embrión y sean destinadas a formar parte de tejidos particulares (Nüsslein-Volhard, 1996; Wieschaus, 1996), Ellos compartieron el premio Nobel en 1995 (revisado en Rubin y Lewis, 2000).

La importancia que adquirió *Drosophila* como organismo modelo tras 90 años de investigación la convirtió en un buen candidato para estar entre los primeros genomas secuenciados, permitiendo obtener la porción eucromática de este en marzo del 2000 (Adams y col., 2000). Sin embargo, la obtención de la secuencia es solo el punto de partida para un análisis más amplio del genoma, en donde se hace necesario

anotar correctamente sus genes, interpretar la secuencia genómica desde el punto de vista de las secuencias expresadas, y predecir las estructuras de sus RNAs y productos proteicos. Las funciones de estos productos y la manera en las cuales están controladas tanto su expresión como su actividad deben ser dilucidadas, una tarea ardua que requiere de una amplia variedad de métodos experimentales y computacionales (Celniker y Rubin, 2003).

Los primeros estudios a nivel genómico se llevaron a cabo hace tres décadas, cuando se estimó que en *Drosophila* se producían en cada etapa del desarrollo entre 5000 a 7000 especies distintas de RNA poliadenilado y que la composición de este RNA cambiaba durante el desarrollo (Levy y Manning, 1981). Estudios posteriores se enfocaron en la detección y caracterización de genes cuya acción es relevante durante el desarrollo. En la actualidad, es posible reunir estos dos enfoques en uno solo, mediante la tecnología de hibridación en matrices. Así, es posible realizar estudios a gran escala de un importante número de genes, sin perder la identidad de estos, en situaciones en que los procesos que regulan la expresión de genes son altamente dinámicos, como es el caso del desarrollo embrionario de *Drosophila*.

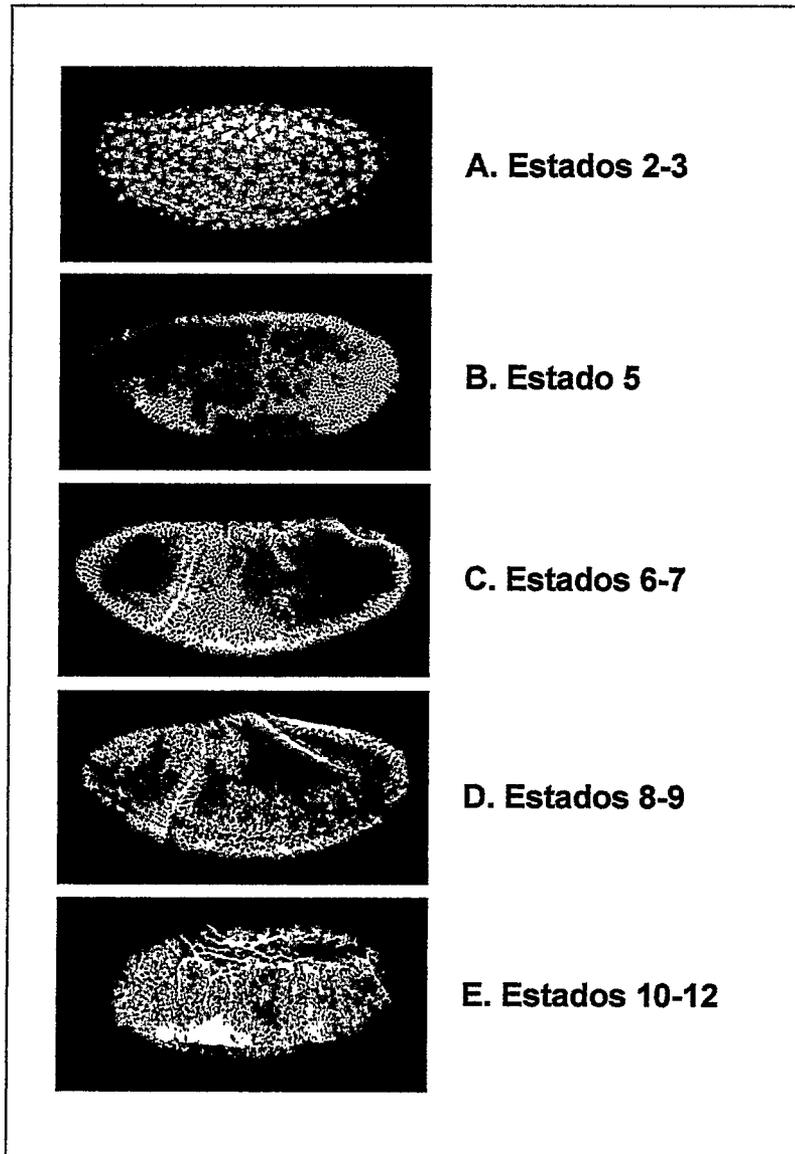
## **2. Desarrollo embrionario de *Drosophila*: morfología y genética.**

Durante el desarrollo, las células del embrión deben cumplir dos tareas principales: deben ser programadas para formar una parte específica del cuerpo y además deben sufrir cambios de forma, posición y patrones de expresión génica que conducirán a este destino. El embrión resultante tendrá diferentes tipos celulares arreglados en un complejo patrón espacial.

En *Drosophila*, un sincicio o saco lleno de núcleos pasa a ser un embrión con simetría bien definida a lo largo de los ejes dorso-ventral y antero-posterior. Estos ejes

están determinados por mensajeros maternos presentes en el ovocito que codifican proteínas distribuidas en gradientes, las que orquestarán adecuadamente este sofisticado ordenamiento, regulando la transcripción y la traducción de las proteínas cigóticas. Las estructuras anteriores están determinadas por el producto de los genes *bicoid* y *hunchback* mientras que las estructuras posteriores lo están por los de *nanos* y *caudal* (Klinger, 1990). El establecimiento de un gradiente de la proteína Dorsal, mediante su traslocación diferencial al núcleo mediada por el receptor Toll, define el eje dorso-ventral del embrión (Roth y col., 1989).

Durante la primera etapa de la morfogénesis, el cigoto presenta un clivaje superficial cuya principal característica es la división de los núcleos previa a la formación de membranas celulares. Hasta la octava división mitótica, los 256 núcleos se encuentran ubicados en el centro del huevo. Durante la décima división mitótica, los núcleos alcanzan la periferia del embrión, donde se dividen cuatro veces más. En este punto, el embrión es llamado blastodermo sincicial (estados 2-3, figura 1 A). Aunque los 6000 núcleos, morfológicamente idénticos, comparten un citoplasma común, este no es homogéneo ya que cada núcleo es acompañado por sus propias proteínas de citoesqueleto y su porción de citoplasma. Luego del ciclo 13, la membrana plasmática se pliega, gracias a una reorganización dramática del citoesqueleto, envolviendo cada núcleo y dando origen al blastodermo celular (estado 5, figura 1 B). Al comienzo de la celularización se forma una red de actina-miosina hexagonal que provee un marco estructural y proporciona fuerza contráctil para la invaginación sincronizada de las membranas. Esta red deriva en estructuras anulares compuestas por filamentos de actina y miosina II. Al final de la celularización, la contracción de estos anillos resulta en la formación de una membrana basal, definiendo el tejido epitelial del blastodermo celular.



**Figura 1. Estados del desarrollo de *Drosophila melanogaster*.** Inmunofluorescencia de montajes completos de embriones en lo distintos estados del desarrollo utilizando un anticuerpo anti-tubulina para visualizar las células. La figura muestra embriones en estado 2-3 (A), estado 5 (B), Estados 6-7 (C), estados 8-9 (D) y estados 10-12 (E). Fotografías cortesía Dra. Verónica Cambiazo.

En esta etapa se requiere por primera vez de la transcripción de los genes del cigoto, entre ellos los genes *nullo* y *serendipity α*, son esenciales para la iniciación, estabilización y reorganización de los microfilamentos que conducirán la invaginación de la membrana plasmática durante la celularización del embrión (Postner y Wieschaus, 1994; Ibensouda y col., 1993; Scheweisguth y col, 1990).

La gastrulación (estado 6, figura 1 C) comienza luego de la 14<sup>ava</sup> división nuclear y es el proceso por el cual las células del blastodermo se diferencian en tres capas germinales claramente distinguibles: mesodermo, endodermo y ectodermo (Leptin, 1995). Esta diferenciación es consecuencia, principalmente, de dos tipos de movimientos morfogénéticos, la invaginación del mesodermo y endodermo y la elongación del ectodermo a lo largo del extremo posterior del embrión. Además, las células del mesodermo sufren una transición desde un tipo celular epitelial a uno mesenquimático.

Los primeros movimientos morfogénéticos que ocurren en la gastrulación incluyen la internalización del mesodermo y parte del endodermo que consta de 1000 células a lo largo de la línea media ventral, generando así el surco ventral. Las células polares forman la placa dorsal, que se mueve hacia el polo anterior a través del eje dorsal y se forma el surco cefálico en el tercio anterior del embrión, por la invaginación transitoria de un grupo de células. Estos movimientos están acompañados por cambios en la forma de las células, las que sucesivamente aumentan de tamaño a lo largo del eje apical-basolateral, experimentan constricciones apicales y acortamientos adquiriendo una forma trapezoide. Por su parte, la banda germinal que originará el ectodermo, se despliega desde la zona ventral hacia el extremo posterior en un movimiento llamado extensión de la banda germinal. Su extremo posterior se pliega originando la invaginación del proctodeo, justo detrás de la cabeza. Para acomodar la

banda germinal en extensión, el epitelio dorsal forma profundos pliegues dando lugar a una delgada capa celular, la amnioserosa.

Mutaciones de efecto materno, como *dorsal* o *toll* eliminan las células ventrales y extienden los dominios laterales y dorsales (Nüsslein-Volhard y col., 1980). En embriones de madres mutantes, las células ventrales no cambian de forma y el surco ventral no se forma. Dorsal activa a los factores de transcripción Twist (helix-loop-helix) y Snail (dedos de Zinc) que se expresan tempranamente en el embrión, específicamente en las células ventrales (Thisse y col., 1988; Boulay y col., 1987). Mutaciones en ambos genes eliminan la formación del surco ventral y prolongan los dominios laterales sobre esta zona. Aunque estos factores de transcripción son relevantes en la diferenciación del mesodermo, los mecanismos de activación y represión de blancos transcripcionales son más complejos de lo que se pensó en un principio (Casal y Leptin, 1996).

El final de la gastrulación (estados 8-9, figura 1 D) está marcado por la invaginación del primordio del endodermo por medio de las invaginaciones del intestino medio anterior y la invaginación del intestino medio posterior (Poulson, 1950). La señal de posición para esta invaginación es provista por la madre y activa un par de genes *gap*, *huckebein* y *tailless* (Weigel y col., 1990; Casanova, 1990). Presumiblemente, *huckebein* especifica el destino de la mayoría de las células posteriores que se invaginan y *tailless* el de las más anteriores. Evidencias morfológicas y genéticas indican que tanto la invaginación del surco ventral como la invaginación del intestino medio son generadas por los mismos mecanismos celulares (Parks y Wieschaus, 1991; Sweeton y col., 1991) ya que presentan comportamientos celulares equivalentes y ambos procesos requieren de la expresión de *concertina* y *folded gastrulation* para lograr la constricción coordinada de sus células (Sweeton y col., 1991; Campos Ortega

y Hartenstein, 1985). Luego de la gastrulación (estados 10-12, figura 1 E) se forma el estomodeo, se delinear los segmentos y la banda germinal se retrae. Este es un periodo principalmente de crecimiento (Campos Ortega y Hartenstein, 1985).

La mayoría de los genes mencionados hasta ahora, claves en los procesos morfogénéticos que hemos revisado, han sido descubiertos mediante la genética clásica, basada en el análisis de fenotipos mutantes. Sin embargo, la estrategia genética presenta algunas limitaciones, ya que mutantes con fenotipos parcialmente penetrantes son ignorados, y genes cuya función puede ser compensada por otros o por vías alternativas son difícilmente detectados. Estas dos categorías de genes parecen representar a la mayoría de los genes de *Drosophila*, dado que dos tercios de ellos no son requeridos para su viabilidad (Miklos y Rubin, 1996). Tomando en cuenta lo anterior, en los últimos años se han aplicado protocolos alternativos al análisis genético que permiten identificar grupos de genes sobre la base de cambios en sus perfiles de expresión y asociarlos con procesos particulares del desarrollo embrionario de *Drosophila* (Furlong y col., 2001; Arbeitman y col., 2002, Stathopoulos y col., 2002).

### **3. Estudios a nivel genómico en organismos modelo.**

Actualmente, el secuenciamiento y anotación de los genomas de los principales organismos modelo ha evidenciado un nuevo desafío: multitud de genes sin función. La llamada genómica funcional, apoyada principalmente en la tecnología de hibridación en matrices, ha proporcionado un análisis descriptivo de los genes mediante la obtención de sus perfiles de expresión bajo distintas condiciones fisiológicas.

A continuación mencionaré algunos de los trabajos que han aplicado herramientas de análisis genómico al estudio del desarrollo embrionario de algunos de

los principales organismos modelos, con especial énfasis en aquellos que utilizaron hibridación en matrices.

Las células del mesénquima primario del erizo de mar (*Strongylocentrotus purpuratus*) han sido un importante modelo experimental para analizar el comportamiento celular y la morfogénesis durante la gastrulación, caracterizados principalmente por una conversión epitelio-mesénquima y por procesos de migración celular. Los eventos celulares han sido descritos en detalle y algunos genes involucrados en los mecanismos moleculares responsables de estos comportamientos han sido caracterizados y clonados, pero muchos otros son aún desconocidos. Con el objetivo de entender estos mecanismos, Zhu y col. (2001) llevaron a cabo un análisis a gran escala de los mRNA expresados por estas células durante la gastrulación. Su estrategia experimental se basó en la obtención de RNA de células cultivadas *in vitro* que muestran el fenotipo de células del mesénquima primario. Este RNA fue utilizado para la síntesis de cDNA, el que fue clonado y secuenciado. Con ello obtuvieron una genoteca de 8000 clones que codifican para cerca de 1200 proteínas. Además, examinaron los patrones espacio-temporales de expresión de 20 clones por medio de hibridaciones *in situ*, estos ensayos que revelaron que 8 de estos clones se expresan específicamente durante la gastrulación. Este es el primer estudio a gran escala en erizo que busca la identificación de nuevos genes involucrados en el desarrollo. Entre los 771 genes identificados se encuentran genes que codifican para proteínas involucradas en ciclo, crecimiento y muerte celular, transducción de señales, migración, adhesión celular y citoesqueleto, matriz extracelular, metabolismo y regulación de la transcripción. Además analizaron las proteínas presentes mediante electroforesis bidimensional en gel, notando que existe un cambio notable en la composición del proteoma al iniciarse la gastrulación y que la mayoría de las proteínas que inician su

síntesis en este estado persisten en el tiempo. Este estudio identifica un gran número de potenciales reguladores de la morfogénesis en erizo, los que son candidatos atractivos para estudios posteriores. Finalmente los autores concluyen, luego de analizar los genes encontrados, que las vías de regulación de la especificación del destino de las células del mesénquima primario deben estar ligadas a la activación específica de moléculas efectoras río abajo que ejecutan el notable programa de la morfogénesis de estas células.

Estudios a nivel genómico en *Caenorhabditis elegans* fueron realizados por Baugh y col. (2003). Incorporando la tecnología de hibridación en matrices de vidrio, pudieron determinar que el 68% de un total de 8890 transcritos detectados, que corresponden a cerca de la mitad de los marcos de lectura abierta del genoma, cambian su abundancia entre el estado de una y 190 células (183 minutos posteriores al primer clivaje,  $P < 0,05$ ). Estas mediciones fueron validadas al analizar los patrones de expresión de 30 genes que habían sido previamente caracterizados. La utilización de breves intervalos de tiempo entre cada extracción de RNA permitieron resolver con detalle la dinámica del transcriptoma, esto es, la abundancia de los transcritos presentes en un estado específico. El adecuado tratamiento estadístico de los datos (análisis multivariado entre intervalos de tiempo) hizo posible la determinación de un cambio drástico en el estado de cuatro células reflejando la degradación de la mayoría de los transcritos maternos y la síntesis de los transcritos cigóticos. El cambio en los niveles de transcripción se mantiene hasta el inicio de la gastrulación, en donde se alcanza un estado estable, lo que reflejaría el control transcripcional por parte del embrión. Un análisis gráfico de los perfiles de expresión de los genes que más cambian en el transcurso de tiempo examinado permitió concluir que los genes son regulados en escalas de tiempo cortas con picos de expresión transitorios, incluso se

pueden circunscribir a solo un ciclo celular, indicando que las decisiones tomadas por la célula en esta etapa son rápidas. Finalmente, un análisis de clasificación por agrupamiento jerárquico permitió concluir que los genes están corregulados y que la composición de estos grupos esta relacionada con la función de los genes debido al enriquecimiento de ciertas categorías funcionales en los grupos recuperados.

Enfoques similares han sido aplicados al desarrollo de *Drosophila melanogaster*. Casal y Leptin (1996), asumiendo que los genes que determinan la diferenciación de un tejido particular se expresan específicamente en ese tejido, realizaron hibridaciones sustractivas entre cDNA de embriones silvestres y embriones mutantes para generar una genoteca enriquecida en secuencias presentes en el mesodermo, pudiendo identificar 13 nuevos genes que están regulados por *twist* y *snail* según resultados de hibridaciones *in situ*.

Mutantes que afectan la expresión de *twist* fueron utilizados por Furlong y col. (2001) para desarrollar experimentos de hibridación en matrices de vidrio sembradas con 5.081 genes únicos de *Drosophila* con el objetivo de estudiar sus perfiles de expresión durante el desarrollo del mesodermo. 130 genes fueron identificados mediante una disminución de sus abundancias relativas al comparar sus perfiles de expresión entre embriones mutantes para *twist* y embriones silvestres, de varias edades. De acuerdo a sus perfiles de expresión fueron incluidos en dos grupos: blancos transcripcionales de *twist* (genes de expresión temprana) o blancos transcripcionales de alguno de los genes activados por *twist* (genes de expresión tardía). Estudios más detallados de uno de estos genes, *gleeful*, determinaron que este gen es esencial en el desarrollo de músculo, ya que células silvestres precursoras de tejido nervioso expresaron marcadores musculares al inducir una expresión ectópica de *gleeful*.

Un análisis similar fue desarrollado por Stathopoulos y col. (2002) utilizando mutantes que afectan la distribución de Dorsal en el núcleo. Experimentos de hibridación en matrices de vidrio conteniendo la totalidad de los genes de *Drosophila* permitieron identificar 40 nuevos blancos transcripcionales de Dorsal que codifican proteínas involucradas en la comunicación célula-célula, re-organización del citoesqueleto y factores de transcripción. La transcripción de estos genes se activa a diferentes concentraciones nucleares de Dorsal, permitiendo distinguir genes involucrados en desarrollo de mesodermo (alta concentración), neuroectodermo (concentración media) y ectodermo dorsal (bajas concentraciones). Los resultados de los experimentos de hibridación en matrices fueron validados para algunos de estos genes mediante la distribución espacial observada por hibridaciones *in situ* en embriones en celularización. Este trabajo incluye análisis bioinformáticos de las secuencias reguladoras de los genes identificados, encontrándose motivos para la unión de Dorsal con distinta afinidad, en cada una de estas tres clases de genes.

Arbeitman y col., (2002) examinaron los perfiles de expresión de un tercio de los genes de *Drosophila* durante el ciclo de vida completo de este organismo. Los resultados son similares a los encontrados en *C. elegans* pudiendo clasificar los genes en categorías según sus perfiles de expresión. Uno de los grupos destacados está compuesto por 534 genes de expresión temprana, encontrándose un máximo de expresión en el estado de blastodermo celular. El diseño experimental permite detectar cascadas de regulación previamente caracterizadas como la activación de *dmef2* por *twist* y la consecuente activación de genes de diferenciación muscular. Los análisis de agrupamiento jerárquico asociando genes que comparten perfiles de expresión revela que, para la mayoría de los grupos encontrados, cada grupo contiene genes que están relacionados funcionalmente. Por lo tanto, basados en la similitud de los perfiles de

expresión, los autores asignan al 53 % de los genes funciones tentativas en diferentes procesos celulares, por ejemplo: diferenciación de músculo, línea germinal del macho o línea germinal de la hembra.

Estos trabajos revelan que es posible aplicar estas tecnologías a problemáticas de la biología del desarrollo y que estudios a gran escala, a nivel genómico, son una alternativa válida para encontrar nuevos genes involucrados en los procesos celulares que se están llevando a cabo en el embrión.

En esta memoria, he querido aplicar este nuevo enfoque utilizando las tecnologías de hibridación en matrices y aprovechar el desarrollo concomitante de la Bioinformática, para estudiar genes de expresión temprana en el embrión, principalmente aquellos de expresión cigótica, que participen en los procesos de morfogénesis durante las etapas tempranas del desarrollo de *Drosophila*. Bajo el supuesto de que los genes que cumplen funciones estructurales o reguladoras durante la gastrulación se encuentran transcripcionalmente activos, en las células que formaran las tres capas germinales, he aplicado un procedimiento de hibridación sustractiva para aislar genes diferencialmente expresados durante los estados iniciales del desarrollo embrionario. De esta forma y utilizando cDNAs de embriones de estados 6-7 (gástrulas) como *tester* y cDNAs de embriones de estados 2-3 (blastodermos sinciciales) como *driver*, generé una genoteca de cDNAs sustraídos que se encuentra enriquecida en transcritos diferencialmente expresados al inicio de la gastrulación. La caracterización de estos genes a través del análisis espacio-temporal de sus patrones de expresión y finalmente la implementación de ensayos funcionales apropiados permitirá reconocer su función durante la gastrulación.

#### **4. Hipótesis.**

Genes cuyos niveles de expresión aumentan durante la gastrulación contribuyen a establecer el fenotipo del embrión en este estado del desarrollo.

#### **5. Objetivo general.**

Determinar los cambios transcripcionales asociados a estados precisos del desarrollo embrionario en *Drosophila melanogaster*

#### **6. Objetivos específicos.**

1). Implementar un procedimiento que permita la obtención de una genoteca estado-específica de *Drosophila* y el análisis simultáneo de la expresión de múltiples genes.

Con el propósito de obtener esta población de genes diferencialmente expresados, se realizó un ensayo de hibridación sustractiva por supresión que generó una población de cDNAs enriquecidos en secuencias expresadas durante la gastrulación. Los cDNAs fueron clonados con el propósito de generar una genoteca estado-específica. Esta genoteca fue caracterizada mediante ensayos de hibridación en matrices y análisis de Northern blots virtuales.

2). Identificar genes asociados con el proceso de gastrulación a través de la secuenciación de una fracción de los clones obtenidos y examinar la similitud de estas secuencias con los genes anotados.

Los clones generados fueron secuenciados y las secuencias alineadas contra las bases de datos del proyecto Genoma de *Drosophila* (BDGP, Berkeley *Drosophila* Genome Project). Los genes encontrados fueron caracterizados sobre la base de sus

términos GO (Gene Ontology), búsqueda de genes ortólogos en otras especies e información disponible de patrones de expresión espacio temporal.

3). Caracterizar los genes identificados sobre la base de sus cambios de expresión temporal durante el desarrollo de *Drosophila melanogaster*.

Para lograr este objetivo, los clones fueron sembrados en membranas e hibridados con poblaciones de cDNAs representativos de embriones de 5 intervalos de tiempo que cubren los primeros 12 estados de desarrollo embrionario. Además, dos de los genes aislados fueron examinados mediante hibridaciones *in situ* con el objetivo de establecer sus patrones de expresión espacio-temporal.

## MATERIALES Y MÉTODOS

### 1. MATERIALES.

#### 1.1. Cepas de *Drosophila melanogaster*.

Durante el desarrollo de esta tesis, la cepa Canton S fue utilizada como cepa silvestre. Esta cepa fue obtenida en Bloomington *Drosophila* Stocks Center (Indiana).

#### 1.2. Cepas bacterianas.

Se utilizó la cepa JM109 de *Escherichia coli* para la construcción de la genoteca estado específica y para el clonamiento de los genes usados como controles.

#### 1.3. Vector.

Todos los insertos de DNA fueron ligados al vector pGEM-T Easy (Promega) para su clonamiento. Este vector, de alto número de copias, contiene los promotores SP6 y T7 flanqueando una región de clonamiento con múltiples sitios de restricción. Esta región está contenida dentro de la región codificante del péptido  $\alpha$  de la enzima  $\beta$ -galactosidasa. La inactivación de este péptido mediante la inserción de un fragmento de DNA en esta región, permite seleccionar las colonias que portan inserto al crecer las colonias transformadas en medio conteniendo X-Gal, sustrato de la  $\beta$ -galactosidasa. Las colonias bacterianas del color blanco indican la inactivación de la enzima. La selección de los transformantes se lleva a cabo gracias a que el vector porta la secuencia codificante de la enzima  $\beta$ -lactamasa, la que confiere resistencia a ampicilina.

#### 1.4. Partidores y adaptadores.

El procedimiento de hibridación sustractiva se realizó con los siguientes partidores y adaptadores provistos por el fabricante (BD Biosciences)

Adaptador 1:

5'-CTAATACGACTCATCTCGGGCTCGAGCGGCCGCCCGGGCAGGT-3'  
3'-GGCCCGTCCA-5'

Adaptador 2R:

5'-CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGT-3'  
3'-GCGCCGGCTCCA-5'

Primer 1:

5'-CTAATACGACTCACTATAGGGC-3'

Nested PCR primer 1:

5'-TCGAGCGGCCGCCCGGGCAGGT-3'

Nested PCR primer 2R:

5'-AGCGTGGTCGCGGCCGAGGT-3'

La síntesis de cDNA de doble hebra utilizando el método SMART™ utiliza los siguientes partidores y oligonucleótidos provistos por el fabricante (BD Biosciences):

BD SMART II™ A Oligonucleotide:

5'-AAGCAGTGGTATCAACGCAGAGTACGCGGG-3'

3' BD SAMRT™ CDS Primer II A:

5'-AAGCAGTGGTATCAACGCAGAGTACT<sub>(30)</sub>VN-3 (N=A,C, G, o T; V=A, G, o C)

5' PCR Primer II A:

5'-AAGCAGTGGTATCAACGCAGAGT-3'

Los siguientes partidores fueron utilizados en las reacciones de PCR realizadas en esta tesis:

SP6: 5'-ATTAGGTGACACTATAGAAT-3'

T7: 5'-TAATACGACTCACTATAGGG-3'

*tubulina* sense: 5'-GCGTGTGAACACTTCCAAT-3

*tubulina* antisense: 5'-TTGCGATTGGAAGCGTAAAC-3'

*g3pdh* sense: 5'-ACCACAGTCCATGCCATCAC-3'

*g3pdh* antisense: 5'-TCCACCACCCTGTTGCTGTA-3'

*actina* sense: 5'-CACCGGTATCGTTCTGGACT-3'

*actina* antisense: 5'-CTCGTAGGACTTCTCCAACG-3'

*rp49* sense: 5'-AAGCTAGCCCAACCTG-3'

*rp49* antisense: 5'-TCTTGAGAACGCAGGC-3'

*twist* sense: 5'-CTGGAGGACACAGACTACAA-3'

*twist* antisense: 5'-CTCGTCGGGAAAATCGAAAGT-3'

*snail* sense: 5'-TTCGTGGAGGAGCGTCTGC-3'

*snail* antisense: 5'-GCGGCGTCATCTGCTGGTA-3'

*fog* sense: 5'-CCCAATTGTCTGCTGGCTGT-3'

*fog* antisense: 5'-CGTCGCCGTAGTCATAGTCT-3'

Las mediciones de abundancia realizadas mediante PCR cuantitativo en tiempo real utilizaron los siguientes partidores:

*dorsal* sense: 5'-CGCCATATCG TCCTCATCCC-3'

*dorsal* antisense: 5'-CTGCCATCCTGGTGGTCATT-3'

*ser α* sense: 5'-GCTTTGACGGCTTTGTG-3'

*ser α* antisense: 5'-TGTGCGGATGAAGTGGA-3'

*tinman* sense: 5'-ATCGTCGTCCCTATCGC-3'

*tinman* antisense: 5'-CACCTGGCTGTTATCCTTT-3'

### 1.5. Bases de datos y herramientas bioinformáticas.

Se efectuaron alineamientos contra las secuencias depositadas en las bases de datos de BDGP (Berkeley Drosophila Genome Project). Las bases de datos consultadas fueron descargadas del sitio <http://www.flybase.org/.bin/fbidbatch.html> y fueron las siguientes: secuencias codificantes (CDS), transcritos completos (CDS más UTR), Secuencias de eucromatina y heterocromatina y genes anotados.

La búsqueda de dominios proteicos fue realizada consultando las bases de datos de InterPro vía web (<http://www.ebi.ac.uk/interpro/>).

La búsqueda de ortólogos fue realizada mediante BLASTP consultando las bases de datos de proteínas de *H. Sapiens*, *M. musculus* y *C. elegans* del EBI (European Bioinformatics Institute) desde <http://www.ebi.ac.uk/proteome/index.html>.

Todos los datos reunidos por estos análisis, además de los datos de expresión obtenidos en esta tesis se encuentran en el sitio web <http://200.89.72.98/>

[~pmoreno/flyAtInta/web/login.php](http://200.89.72.98/~pmoreno/flyAtInta/web/login.php) en una base de datos diseñada como parte de la práctica profesional de Pablo Moreno (estudiante de Ingeniería Civil en Biotecnología).

## 2. MÉTODOS.

### 2.1. Colecta de embriones.

Los embriones fueron colectados sobre placas conteniendo agar en jugo de manzana al 2%, cubiertas con una capa de ácido acético y levadura para estimular la oviposición. Para la extracción de RNA, los embriones fueron colectados en tritón 0,1% en ringer (KCl 182 mM, NaCl 46 mM, CaCl<sub>2</sub> 3 mM, Tris-HCl pH 7.2 10 mM) preparado con agua libre de nucleasas, decorionizados con hipoclorito de sodio 50% en ringer-tritón por 5 minutos y lavados en ringer-tritón. Los embriones seleccionados fueron congelados en nitrógeno líquido y mantenidos a -80°C. Para la caracterización morfológica de los distintos estados del desarrollo embrionario, se utilizó un criterio visual de acuerdo con la información detallada en Campos-Ortega y Hartenstein (1985).

Para los ensayos de hibridación *in situ*, embriones de todos los estados fueron colectados en tritón 0,05% en PBS (PBTritón), decorionizados con hipoclorito de sodio 50% en PBTritón por 5 minutos, lavados 3 veces en PBTritón y fijados en una solución 2:1 heptano: paraformaldehído al 4% en PBS (PP) e incubados 15 minutos con agitación. Las dos fases formadas fueron eliminadas y reemplazadas con una solución 1:2 heptano:metanol y agitados en *vortex* por 1 minuto. Se eliminaron ambas fases concentrándose las membranas vitelinas en la fase superior. Los embriones fueron lavados 3 veces con metanol para luego incubar sucesivamente en: EGTA al 10% en metanol (ME), ME:PP 7:3, 1:1 y 3:7, por 5 minutos cada vez. Los embriones fueron incubados en PP por 20 minutos y lavados con PBS. En este punto los embriones fueron guardados deshidratándolos en sucesivos lavados con etanol 30, 50 y 70% y mantenidos a -20°C hasta su utilización.

## **2.2. Procedimientos de biología molecular.**

### **2.2.1. Electroforesis.**

Todas las electroforesis de DNA se llevaron a cabo en geles de agarosa al 1% en buffer TAE (Tris, ácido acético, EDTA) visualizado con bromuro de etidio (Asufel, 1997). Para las electroforesis de RNA, se usó buffer TAE preparado con agua libre de nucleasas.

### **2.2.2. Extracción de RNA total.**

El RNA total fue extraído desde embriones homogeneizados en RNAwiz™ (Ambion). 0,2 volúmenes de cloroformo fueron agregados a esta mezcla, sometiéndola a agitación y centrifugación y rescatando el RNA en la fase acuosa superior, el que fue precipitado con isopropanol mediante centrifugación. El precipitado fue lavado con etanol 75% y precipitado por centrifugación. Finalmente el RNA total fue resuspendido en agua libre de nucleasas. Se obtuvo un rendimiento promedio de 100 ng de RNA por embrión. La integridad del RNA fue verificada por electroforesis en gel de agarosa. La calidad se evaluó mediante la síntesis de cDNA y la amplificación de distintos genes por PCR a partir de este cDNA.

### **2.2.3. Síntesis de cDNA.**

El RNA fue usado como sustrato para la síntesis de cDNA de hebra simple mediante transcripción reversa en una reacción estándar, utilizando 200 U de la enzima superscript II (Superscript II RNase H<sup>-</sup> Reverse Transcriptase, Invitrogen) y 0,5 µg de partidor oligo dT (15 nucleótidos). La calidad del cDNA resultante fue verificada

mediante su utilización como sustrato para la amplificación de los genes *tubulina*, *actina* y *fog*.

El cDNA de doble hebra fue sintetizado usando el método SMART (SMART™ PCR cDNA síntesis kit, Clontech®) que permite amplificar el cDNA resultante manteniendo las abundancias relativas de los genes (Chenchik y cols, 1998). El primer paso consiste en la síntesis de cDNA de hebra simple a partir de 2 µg de RNA total utilizando 1,2 µM del partidor oligo(dT) modificado, 3' BD SMART™ CDS Primer II A, 1,2 µM de BD SMART II™ A Oligonucleotide, que permite el efecto de cambio de templado (Template switching) y 100 U de la enzima BD PowerScript™ Reverse Transcriptase. La reacción lleva además 1 mM de cada dNTP, 2 mM de DTT y buffer de la enzima (BD Biosciences) y se incuba a 42°C por 1 hora. Cuando la enzima alcanza el extremo 5' del mRNA, la actividad transferasa terminal agrega unos pocos nucleótidos adicionales, principalmente dCTP al extremo 3' del cDNA. El oligonucleótido SMART II, que contiene una cola de poli(G) se aparea con los nucleótidos agregados al cDNA creando un templado extendido. La enzima entonces cambia de templado y continúa transcribiendo hasta el extremo del oligonucleótido. Todos los cDNAs resultantes tienen en sus extremos 5' y 3' una secuencia que será reconocida por el partidor PCR Primer II A en las subsecuentes reacciones de PCR. Este procedimiento permite obtener cDNAs de hebra simple de largo completo. La amplificación de este cDNA se lleva cabo por una reacción de LD PCR (Long Distance PCR), utilizando un termociclador PTC-100 (MJ Research). Para ello, 1 µL del cDNA de hebra simple recién sintetizado fue adicionado a una reacción que contiene 10 µL de 10 X BD advantage 2 PCR buffer, 2,2 mM de cada dNTP, 0,24 µM del partidor PCR Primer II A y 2 µL de 50X BD advantage 2 polymerase mix (BD Biosciences). La reacción comienza con una desnaturalización a 94°C por 1 minuto, seguida de 17

ciclos con pasos de 95°C por 15 segundos, 65°C por 30 segundos y 68°C por 6 minutos. Durante este procedimiento, fue necesario optimizar el número de ciclos de reacción durante los cuales los productos se encuentran en fase exponencial de amplificación. Para esto, se realizó un ensayo que consistió en tomar alícuotas de 15  $\mu\text{L}$  de las reacciones de PCR luego de 15, 18, 21 y 24 ciclos de amplificación y fraccionarlas en un gel de agarosa al 1%. A medida que aumenta el número de ciclos, la intensidad de las bandas correspondientes a los cDNAs más abundantes en la muestra, también aumenta, hasta un punto máximo, el que en nuestro caso fue alcanzado luego de 18 ciclos de amplificación. En este punto los cDNAs correspondientes a las especies de mayor abundancia han finalizado su crecimiento exponencial. Esta amplificación fue repetida varias veces en éste y otros experimentos con idénticos resultados, obteniendo un promedio de 500 ng/ $\mu\text{L}$  de cDNA SMART. La calidad del cDNA fue verificada mediante la amplificación de genes constitutivos y estado-específicos. Este cDNA SMART fue usado para realizar la sustracción, para la fabricación de sondas complejas representativas de cada estado embrionario y para la medición de abundancias relativas a través de PCR cuantitativo en tiempo real.

#### **2.2.4. Purificación de DNA plasmidial a gran escala.**

El DNA plasmidial de 96 clones fue aislado simultáneamente por lisis alcalina, Para ello, las bacterias fueron inoculadas en placas de 96 pocillos conteniendo 800  $\mu\text{L}$  de medio LB 2X, ampicilina y crecidas 24 horas con agitación suave (100 rpm). Las placas fueron centrifugadas 10 minutos a 2000 rpm, el sobrenadante fue descartado y las bacterias resuspendidas en 80  $\mu\text{L}$  de solución de resuspensión (Glucosa 30 mM, Tris HCl 15 mM, EDTA- $\text{Na}_2$  30 mM, RNAsa I 60  $\mu\text{g}/\text{mL}$ ). 80  $\mu\text{L}$  de solución de lisis (NaOH 0,2 N, SDS 1%) fueron agregados a las mezclas e incubadas 2 minutos a

temperatura ambiente luego de lo cual se adicionaron 80  $\mu$ L de solución de neutralización (acetato de potasio 3,6 M, ácido acético 14%) y fueron agitadas en vortex suave por 2 minutos. El homogeneizado fue filtrado al vacío (manifold, Millipore) a través de una placa filtro para retener el desecho celular (Multiscreen-DV Millipore 0.65  $\mu$ m Hydrophilic, Low Protein Binding Durapore Membrane). El filtrado fue recibido en una placa (Multiscreen-FB Millipore 1.0  $\mu$ m Glass Fiber Type B Filter) conteniendo 150  $\mu$ L de solución de unión (KI 6.1 M). El DNA fue lavado con etanol 80% y eluído con 50  $\mu$ L de buffer TE (Tris 10 mM, EDTA 1 mM) sobre placas de 96 pocillos mediante centrifugación.

#### **2.2.5. PCR en colonias.**

Las bacterias fueron crecidas en placas de 96 pocillos conteniendo 200  $\mu$ L de medio LB con ampicilina 50  $\mu$ g/mL a 37°C por 22 horas a 100 rpm. Alícuotas de 5  $\mu$ L de este cultivo fueron agregados a 33  $\mu$ L de agua des-ionizada y fueron calentados a 95°C por 15 minutos, luego de los cuales se les adicionó el resto de los componentes: buffer mix 1X (Fermentas), dNTPs 0,25 mM, MgCl<sub>2</sub> 2 mM, partidor SP6 y partidor T7 1 pM y 2,5 U de Taq DNA polimerasa (Fermentas). El programa consta de 35 ciclos que incluyen desnaturalización a 94°C por 30 segundos, alineamiento a 46°C por 30 segundos y extensión a 68°C por 2 minutos e incorpora una extensión final a 68°C por 15 minutos.

#### **2.2.6. PCR a gran escala.**

Estas reacciones fueron efectuadas en placas de 96 pocillos (Termowell Costar 6511, Corning Incorporated) usando como sustrato para la amplificación el DNA plasmidial aislado como se describió en 2.2.4 y el termociclador PTC-100 (MJ

Research). La reacción (50  $\mu$ L) consistió en: 0,2  $\mu$ M de cada uno de los partidores SP6 y T7, 5 U de Taq DNA polimerasa (Invitrogen), 0,2 mM de cada dNTP, 1 mM de  $MgCl_2$  y buffer de la enzima. Las amplificaciones fueron analizadas mediante electroforesis en gel de agarosa al 1%.

### **2.2.7. PCR cuantitativo en tiempo real.**

El análisis de abundancia relativa de cada gen marcador en todas las muestras de cDNA SMART de embriones de distintas edades se realizó mediante PCR en tiempo real utilizando el kit Light Cycler FastStart DNA Master SYBR Green I (Roche). Cada reacción de amplificación fue llevada a cabo en capilares utilizando 50 ng de cDNA, 0,5  $\mu$ M de cada partidor, 1,5 mM de  $MgCl_2$ , y 2  $\mu$ L de LC-FastStart DNA Master SYBR Green I, 10x (Roche) en 20  $\mu$ L de reacción. Las reacciones de PCR fueron realizadas en el Light Cycler Instrument (Roche) y el programa de amplificación fue diseñado utilizando el Light Cycler software v. 3.5 (Roche). Este programa incluye una etapa de activación a 95°C por 10 minutos y una etapa de amplificación que consta 35 ciclos de un paso de 95°C por 3 segundos, alineamiento a la temperatura adecuada para cada par de partidores por 15 segundos y un paso a 72°C por 30 segundos. Los productos de cada amplificación fueron analizados mediante electroforesis con el propósito de descartar amplificaciones inespecíficas. La medida de la abundancia de cada transcrito en la muestra fue realizada a través de una interpolación desde curvas de calibración construidas a partir de diluciones seriadas de DNA plasmidial o fragmentos de PCR de concentración conocida. Una vez determinada la abundancia de cada gen, esta fue normalizada respecto a la abundancia de *actina* en cada muestra, con el propósito de poder comparar entre distintas muestras.

## 2.3. Sustracción.

### 2.3.1. Hibridación sustractiva por supresión.

Esta técnica fue realizada de acuerdo con el protocolo recomendado por el fabricante (Clontech PCR-Select™ cDNA Subtraction Kit, Clontech) descrito además por Diatchenko y cols (1996). Brevemente, esta técnica consta de la separación del cDNA *tester* en dos poblaciones, que se ligan independientemente a los adaptadores 1 y 2R. Luego cada población es desnaturalizada e hibridada independientemente con un exceso de cDNA *driver*, que no tiene adaptadores ligados. El resultado de estas dos hibridaciones se mezcla y se hibrida con un exceso cDNA *driver*, esta vez sin desnaturalizar. Este cDNA resultante es amplificado en una reacción de PCR, llamada PCR primario usando el partidor 1 y luego en una segunda reacción, llamada PCR secundario usando los partidores *Nested* PCR Primer 1 y *Nested* PCR primer 2R. Estos partidores se aparean con las secuencias de los adaptadores ligadas a cada fragmento de cDNA. Esto da como resultado una población de cDNA enriquecido en secuencias cuya abundancia relativa es mayor en el *tester* que en el *driver*. El cDNA usado como *tester* corresponde a cDNA SMART de embriones en estado de gástrula (estados 6 y 7) y el cDNA *driver*, el estado de referencia, a cDNA SMART de embriones en estado de blastodermo sincicial (estados 2 y 3). Simultáneamente se realizó una sustracción reversa utilizando cDNA SMART de embriones de estado 2-3 como *tester* y cDNA SMART de embriones 6-7 como *driver*. Se efectuó una tercera sustracción que es un control positivo del procedimiento usando como *tester* una mezcla de cDNA de músculo esquelético y cDNA del fago ØX174 y como *driver* el cDNA de músculo esquelético.

### **2.3.2. Verificación de la eficiencia de la sustracción.**

Para este fin se realizó una cinética de amplificación de *tubulina*, *g3pdh* y *twist* a partir de cDNA sustraído y no sustraído. La reacción lleva los partidores necesarios (10  $\mu$ M), dNTPs (10mM), 10X PCR buffer, Taq DNA polimerasa (Fermentas) y agua hasta completar un volumen de 50  $\mu$ L. El programa consta de desnaturalización por 30 segundos a 94°C, alineamiento de los partidores por 30 segundos a una temperatura optima para cada par y una extensión de dos minutos a 68°C. La reacción consta de 35 ciclos, extrayendo 5  $\mu$ L de la mezcla cada cinco ciclos a partir del ciclo 15, las cuales fueron sometidas a un análisis electroforético. Una segunda verificación se llevó a cabo midiendo la abundancia del transcrito de *actina*, tanto en muestras sustraídas como no sustraídas mediante PCR cuantitativo en tiempo real.

### **2.3.3. Northern blots virtuales.**

El cDNA SMART obtenido a partir de embriones en estados 2-3 y 6-7 fue sometido a electroforesis y transferido a membranas de nylon de acuerdo con el método de Southern descrito por Ausubel (1997). El DNA fue transferido a membranas de nylon (Gene Screen Plus, BioRad) por capilaridad por 16 horas. Finalmente, las membranas fueron lavadas en una solución de 2X SSC, secadas, y el DNA fue unido covalentemente a la membrana mediante irradiación de luz UV.

## **2.4. Construcción de la genoteca.**

### **2.4.1. Ligación.**

300  $\mu\text{L}$  de producto de PCR secundario fueron purificados mediante el Kit DNA Wizard Clean Up (Promega). El producto de la purificación fue usado para la reacción de ligación al vector pGEM-T Easy (Promega). La reacción contiene Buffer Ligasa 2X, vector, DNA Ligasa y agua hasta completar un volumen de 10  $\mu\text{L}$ . La reacción fue incubada toda la noche a 4°C.

### **2.4.2. Transformación.**

Alícuotas de 2  $\mu\text{L}$  de la reacción de ligación fueron desnaturalizadas a 70°C por 10 minutos para inactivar la enzima, y mantenidas en hielo. Estas alícuotas fueron agregadas a 50  $\mu\text{L}$  de células competentes *E. coli* JM109, mezcladas suavemente y la mezcla fue mantenida en hielo por 30 minutos al cabo de los cuales se aplicó un golpe de calor en un baño a 42°C por 45 segundos. Luego de incubar las células en hielo por 2 minutos, se adicionó 950  $\mu\text{L}$  de medio SOC ( $\text{MgCl}_2$  10 mM,  $\text{MgSO}_4$  10 mM, Glucosa 0.4% en medio SOB estéril que a su vez contiene Triptona 2%, extracto de levadura 0.5%, NaCl 10 mM y KCl 2.5 mM) y estas fueron incubadas 1 hora a 37°C con agitación suave. Las bacterias fueron sembradas sobre placas conteniendo LB, agar 1,5% y ampicilina sobre las que se vierten 100  $\mu\text{L}$  de IPTG 0.1 M y 20  $\mu\text{L}$  de X-gal 50  $\mu\text{g}/\text{mL}$  e incubadas por 16-18 horas a 37°C.

### **2.4.3. Selección de transformantes y construcción de la genoteca.**

Los transformantes que llevan insertos de cDNA fueron seleccionados por su resistencia a ampicilina y por su incapacidad para degradar el X-gal, lo que originan el crecimiento de colonias blancas. Estas fueron sembradas en placas de 96 pocillos (Orange Scientific) conteniendo 200  $\mu$ L de medio LB, agar 1,5% y 50  $\mu$ g/mL de ampicilina. Los clones fueron replicados en placas de 96 pocillos (Orange Scientific) conteniendo LB, glicerol 30% y 50  $\mu$ g/mL ampicilina para almacenarlos a  $-80^{\circ}\text{C}$ .

### **2.4.5. Análisis de la presencia de insertos en los clones.**

Las verificaciones de la presencia de un inserto en cada clon se llevaron a cabo mediante la digestión del DNA plasmidial de 10 clones con la enzima EcoRI y mediante la amplificación por PCR en colonias de 96 clones. Todas las colonias probadas tenían inserto.

## **2.5. Identificación y caracterización de genes.**

### **2.5.1. Obtención de secuencias.**

Replicas de 7 placas congeladas de la genoteca, correspondientes a 652 clones, fueron secuenciadas en la empresa Agencourt Bioscience Corporation (Beverly MA). Durante este proceso, el DNA plasmidial de cada clon fue purificado utilizando el método CosMCPrep<sup>TM</sup>, desarrollado por dicha empresa. Este método se basa en el sistema SPRI (Solid Phase Reversible Immobilisation) para la purificación del DNA plasmidial ([http://www.agencourt.com/products/spri\\_reagents/cosmcprep/](http://www.agencourt.com/products/spri_reagents/cosmcprep/)). Para las reacciones de secuenciación del DNA se utilizó el partidador SP6 y los *kits*

BigDye Terminator v.3.1 (Applied Biosystems) y CleanSEQ (Agencourt). La secuenciación se realizó en un equipo ABI PRISM 3700/3730xl, las secuencias fueron leídas en varios formatos utilizando el *software* Phred. Se obtuvieron lecturas de 1000 pb en promedio.

### **2.5.2. Análisis de secuencias.**

Las secuencias obtenidas se analizaron mediante Batch BLASTN contra la base de datos versión 3.2 descargada desde el sitio del proyecto genoma de *Drosophila* (BDGP). Se utilizó la base de datos correspondiente a secuencias de transcritos, que incorporan las secuencias de los CDS (coding sequences) y los UTR de cada gen (untranslated region). De las 638 secuencias obtenidas, solo 490 poseen un valor  $E < 10^{-11}$ . Este valor indica la probabilidad de encontrar un alineamiento igual al azar en  $10^{-11}$  bases de datos de igual tamaño. Clones con valores mayores de E fueron alineados contra las bases de datos de euromatina y heterocromatina y de genes anotados.

Los resultados de los alineamientos incluyen su asociación con las bases de datos de Gene Ontology ([www.geneontology.org](http://www.geneontology.org), Gene Ontology Consortium, 2001) los cuales fueron usados para análisis posteriores que permitieron la caracterización de genes sobre la base de Función Molecular, Proceso Biológico y Componente Celular. 79 genes que no se encuentran anotados en Gene Ontology fueron analizados mediante InterPro con el fin de encontrar dominios proteicos conocidos. Este análisis permitió asignar uno o más términos de Gene Ontology (GO) a 16 genes adicionales gracias a que algunos dominios proteicos encontrados poseen un término GO asignado.

## 2.6. Hibridación en matrices.

### 2.6.1. Matrices de hibridación.

Durante esta tesis se diseñaron dos tipos de matrices de hibridación: 1) Las matrices Tipo I, utilizadas para los ensayos de caracterización de la genoteca, corresponden a membranas de nylon (Gene Screen Plus, BioRad) de 8 por 12 cms que fueron sembradas alternativamente con DNA plasmidial o productos de PCR, provenientes de los clones aislados. El DNA fue sembrado en tetraplicado usando un sembrador manual de 96 pins metálicos (VP380 Colony Copier, V&P Scientific, Inc.). La cantidad de DNA sembrado en cada punto fue cuantificada por espectrofotometría y varía en los distintos eventos de sembrado entre 6 y 60 ng por punto de siembra. 2) Las matrices Tipo II, utilizadas para los ensayos de expresión diferencial, corresponden a membranas de nylon (Gene Screen Plus, BioRad) de 8 por 12 cms que fueron sembradas con DNA plasmidial aislado a gran escala o productos del PCR en colonias. El DNA fue sembrado mediante el sembrador automático ChipWriter Compact Arrayer (BioRad). Cada punto de siembra tiene en promedio 4 ng de DNA, con un diámetro de 600  $\mu\text{m}$  manteniendo una distancia de 1500  $\mu\text{m}$  entre ellos. En ambos tipos de matrices se sembraron DNA plasmidial o productos de PCR correspondientes a los siguientes controles: 1) Controles positivos: los genes *twist*, *snail* y *fog*, que son genes de expresión cigótica (Leptin, 1999). 2) genes de expresión constitutiva: *actina*, *tubulina* y *rp49*, que permiten la comparación entre membranas, y 3) controles negativos: DNA plasmidial sin inserto y solvente.

Todas las membranas fueron tratadas con solución de desnaturalización (NaOH 0.5 M, NaCl 1.5 M) 7 minutos sin agitación y 2 veces con solución de neutralización (NaCl 1.5 M, Tris 0.5 M, EDTA 1 mM) 3 minutos con agitación. Las membranas fueron

secadas y tratadas con luz UV por 45 segundos. Todas las membranas fueron reutilizadas al menos una vez luego de lavarlas 3 veces con una solución de SDS 0.1% a 100°C.

### **2.6.2. Obtención de sondas.**

Los ensayos de hibridación fueron realizados con sondas simples y complejas. Para cada sonda simple el clon seleccionado fue crecido en medio LB y ampicilina por 16 horas a 37°C con agitación. El DNA plasmidial fue extraído mediante lisis alcalina y el inserto fue aislado mediante digestión con EcoRI, o amplificado por PCR utilizando los partidores SP6 y T7. El cDNA así obtenido fue purificado a partir de geles de agarosa utilizando el método Concert Rapid Gel Extraction System (Life Technologies). Los experimentos sobre membranas Tipo I fueron llevados a cabo hibridando sondas complejas correspondientes a cDNA sustraído, el que fue digerido con la enzima de restricción Rsa I con el objetivo de remover los adaptadores. Para esta reacción se utilizaron 25 µg de cDNA y 33 U de enzima, los que fueron incubados a 37°C por 12 horas. Los experimentos sobre membranas Tipo II fueron llevados a cabo hibridando sondas complejas correspondientes a cDNA SMART de los estados a comparar.

### **2.6.3. Marcación de sondas.**

Las sondas simples fueron marcadas utilizando el kit Random Primers DNA Labeling system (Invitrogen). 200 ng de producto de PCR fueron marcados en una reacción que contiene 15 µL de Random Primers Buffer Mixture, 0,01 mM de dATP, dGTP y dTTP, 30 mCi de [<sup>32</sup>P]dCTP y 3 U de Polimerasa Klenow. La reacción fue

incubada a 25°C por 90 minutos. Las sondas así sintetizadas fueron purificadas con el Kit Wizard DNA Clean Up (Promega), desnaturalizadas y utilizadas de inmediato.

Las sondas complejas de cDNA SMART (400 ng) fueron marcadas radioactivamente siguiendo el protocolo de *random priming* descrito anteriormente. Alternativamente se utilizaron otros dos protocolos de marcación de sondas complejas:

- 1) Incorporación del nucleótido radiactivo al momento de sintetizar el cDNA a partir de RNA total. Para esto se realizaron dos reacciones con 3 y 10 µg de RNA total usando como partidor 1 y 2 µg de poli(dT) respectivamente. Cada reacción lleva además 0,5 mM de dATP, dGTP y dTTP y 30 mCi de [<sup>32</sup>P]dCTP. La mezcla fue desnaturalizada a 70°C por 10 minutos y la temperatura bajada gradualmente 5°C cada un minuto hasta alcanzar 42°C. Transcurridos 10 minutos de incubación fueron agregados 1 mM de DTT, buffer de la enzima y 200 U de la enzima superscript II (Superscript II RNase H<sup>-</sup> Reverse Transcriptase, Invitrogen). Cada reacción fue incubada a 42°C por 2 horas. Posteriormente, las reacciones fueron inactivadas incubándolas a 70°C por 15 minutos y el RNA fue degradado agregando 1 µL de SDS 1%, 1 µL de EDTA 0,5 M y 3 µL de NaOH 3M. Cada reacción fue incubada a 68°C por 30 minutos y luego a temperatura ambiente por 15 minutos. Finalmente se agregaron 20 µL de Tris·HCl pH 8,0 y 6 µL de HCl 1 N. Las sondas fueron purificadas con el Kit Wizard DNA Clean Up (Promega), desnaturalizadas y utilizadas de inmediato (Bernard y col., 1996);
- 2) Marcación del cDNA de doble hebra no amplificado mediante el método de *random priming*. Para esto, 4 µg de cDNA de hebra simple fueron usados para una reacción de síntesis de la segunda hebra utilizando 0,8 mM de cada dNTP y la mezcla Second Strand Enzyme Cocktail (24 U de DNA polimerasa I, 1,5 U de RNasa H y 4,8 U de DNA ligasa, Clontech). La reacción fue incubada a 16°C por dos horas. Posteriormente, el DNA de doble hebra fue purificado por precipitación con acetato de sodio e isopropanol. 400 µg

de este cDNA de doble hebra no amplificado fueron marcados radioactivamente con el protocolo de *random priming* descrito anteriormente.

#### **2.6.4. Hibridación con sondas radioactivas.**

Las membranas fueron pre-hibridadas a 42°C por 2 horas con agitación suave, en botellas conteniendo 5 mL de solución de hibridación (SSC 5X, solución de Denhardt 5X, SDS 1% y Formamida 50%) y 50 µg/mL de DNA de esperma de Salmón. Posteriormente la sonda fue desnaturalizada y agregada a la solución de hibridación. El periodo de hibridación consta de 16 horas a 42°C con la misma agitación. Tanto para los Northern blots virtuales como las membranas Tipo I, las membranas fueron lavadas una vez con solución I (SSC 2X) por 10 minutos a 25°C y solución II (SSC 2X, SDS 0,1%) por 10 minutos a 25°C, dos veces con la solución III (SSC 1X, SDS 0,1%) por 15 minutos a 65°C y una vez con la solución IV (SSC 0,1X) por 15 minutos a 25°C. Las membranas fueron puestas en bolsas plásticas transparentes selladas enfrentando placas fotográficas Kodak a -80°C por tiempos variables (24-268 horas).

En el caso de las membranas Tipo II, las membranas fueron sometidas a 4 lavados con una solución SSC 2X, SDS 1% por 15 minutos a 65°C. Luego fueron selladas en bolsas de plástico transparente para enfrentarlas a pantallas de exposición (Imaging screen-K, BioRad).

#### **2.7. Análisis de datos.**

##### **2.7.1. Obtención de imágenes.**

La señal asociada a cada punto de siembra presente en membranas Tipo I, fue capturada en una placa fotográfica (Kodak). La información digitalizada de esta señal

es representada en una imagen en formato TIFF obtenida a través del *scanner* AGFA Snap Scan 1212p. Los valores crudos de intensidad de hibridación para cada punto de siembra fueron obtenidos utilizando el Software Kodak 1D, el cual proporciona una grilla con celdas de tamaño fijo. Los valores de Intensidad neta fueron obtenidos restando a cada valor crudo el valor de ruido asociado a la membrana que corresponde al promedio de los valores de ruido obtenidos en celdas vacías distribuidas en diferentes puntos de ésta.

La señal asociada a cada punto de siembra presente en membranas Tipo II fue capturada en una pantalla de BaFBr:Eu<sup>+2</sup> luego de tiempos variables de exposición, entre 2 y 96 horas. La información digitalizada de esta señal es representada en una imagen en formato TIFF obtenida utilizando el Phosphoimager Personal FX (BioRad) a una resolución de 50 µm por píxel. Los valores crudos de intensidad de hibridación para cada punto de siembra fueron obtenidos con el software VersArray Analyzer versión 4.5.1.46 (BioRad), el cual utiliza una grilla de tamaño de celda variable para cuantificar la densidad de píxeles (intensidad cruda). Los valores de intensidad neta fueron obtenidos restando a cada valor crudo el valor de ruido local asociado a cada celda.

### **2.7.2. Filtración, normalización y análisis de los datos.**

En las membranas Tipo I, el valor de intensidad neta asociado a cada gen fue calculado promediando las señales de su tetraplicado. Si un valor incrementó la desviación estándar, este valor fue eliminado y el promedio fue recalculado con los valores restantes. Los valores fueron normalizados respecto a la señal de hibridación de la *actina*, obteniéndose finalmente valores que corresponden a la razón entre el promedio de intensidades del tetraplicado y el promedio de intensidades de la *actina* en

su membrana. Este proceso de normalización permite la comparación entre membranas.

El valor de intensidad neta asociado a cada gen en las membranas Tipo II corresponde al promedio del duplicado. Fue calculado el coeficiente de variación (Dudoit y col., 2000) definido como la razón entre la desviación estándar y el valor promedio de los duplicados considerando solo aquellos valores mayores que cero. Aquellos valores cuyo coeficiente de variación fue igual o mayor que 0.5 fueron descartados de los análisis posteriores.

Se evaluó la calidad de cada punto de siembra según el procedimiento descrito por Wang y col., (2001), calculando un índice llamado  $q_{com}$  que corresponde al promedio de cuatro índices que evalúan la calidad de la señal obtenida de acuerdo con:

1) El tamaño del punto,  $q_{size}$ .

$$q_{size} = \exp \left( \frac{-|A - A_0|}{A_0} \right)$$

Donde A es el área del punto y  $A_0$  es el área promedio de todos los puntos. Este índice penaliza puntos muy intensos de tamaño grande o muy pequeño que pueden deberse a hibridación inespecífica.

2) La relación señal ruido,  $q_{sig-noise}$ .

$$q_{sig-noise} = 1 - [bkg_1 / (sig + bkg_1)] = sig / (sig + bkg_1)$$

Donde sig es la señal medida y  $bkg_1$  el ruido local. Este índice cuantifica cuán bien se puede resolver una señal verdadera del ruido del sistema.

3) La variabilidad del ruido local,  $q_{bkg1}$ .

$$q_{bkg1} = f_1 / CV_{bkg}$$

Donde  $CV_{bkg}$  es el coeficiente de variación del ruido local y  $f_1$  es una constante de normalización que satisface  $\max(q_{bkg1}) = 1$ . Este índice refleja la variabilidad del ruido local y es un indicador de contaminantes en el área medida.

4) La variabilidad del ruido local respecto al ruido de la membrana,  $q_{bkg2}$ .

$$q_{bkg2} = f_2 \{ 1 - [ bkg_1 / ( bkg_1 - bkg_0 ) ] \}$$

Donde  $bkg_0$  es el promedio global del ruido y  $f_2$  es una constante de normalización que satisface  $\max(q_{bkg2}) = 1$ . Este índice distingue regiones problemáticas dentro de la membrana ya que estas tendrán un promedio de ruido mas alto que el promedio de la membrana.

Todos aquellos duplicados en los que alguno de los puntos de siembra tuvo un  $q_{com}$  menor a 0.6 fueron descartados de los subsiguientes análisis. Con el propósito de comparar las señales de hibridación recuperadas entre distintas membranas, los datos fueron normalizados utilizando la señal promedio del gen *rp49* en cada membrana, cuya expresión es constitutiva durante el periodo de tiempo estudiado. Los cambios de expresión fueron calculados como la razón entre la abundancia relativa en una membrana respecto a la abundancia medida en la membrana de referencia que corresponde a la membrana hibridada con sonda representativa de los estados 2-3. En algunos experimentos, se calculó el estadístico t (test de t utilizando el software SYSTAT v. 10.0) con el objetivo de detectar los genes que presentan cambios significativos de expresión entre los estados comparados.

Paralelamente se realizó un análisis de agrupamiento de genes que presentan perfiles de expresión similares. Para ello fue calculada la razón de cambio para cada gen entre cada estado y un estado de referencia, en este caso, embriones en gastrulación (estados 6-7) con el objetivo de incluir en el análisis aquellos genes cuyo valor de intensidad neta normalizada en el estado de blastodermo sincicial fue cero

pero presentaron señales positivas en los estados posteriores. Los datos fueron analizados mediante un software de Cluster jerárquico (Eisen y col., 1998), utilizando como parámetro las distancias promedio y como criterio de verificación la correlación de Pearson. El software Treeview (Eisen y col., 1998) permitió visualizar los resultados mediante un dendrograma asociado a un mapa de colores que describe la expresión relativa de los genes en una escala de verde a rojo y que permite identificar grupos de genes por inspección visual.

## **2.8. Hibridaciones *in situ*.**

Hibridaciones *in situ* de los clones seleccionados se llevaron a cabo basadas en el protocolo descrito por Lehmann y Tautz (1994). Brevemente, 1 µg de cDNA fue obtenido mediante digestión con EcoRI desde DNA plasmidial y marcado mediante *random priming* utilizando 0,25 mM de dATP, dCTP, dGTP, 0,13 mM de dTTP y 0,07 mM de Digixigenin-11-dUTP alkali-labile además de 4 U de Polimerasa Klenow (Dig-High Prime, Roche). La reacción fue incubada a 37°C por 12 horas. La sonda así sintetizada fue purificada por precipitación con acetato de sodio e isopropanol y guardada a -20°C hasta su utilización.

Embriones de todos los estados previamente colectados fueron permeabilizados incubándolos con proteinasa K. Los embriones fueron re-fijados en PP y lavados nuevamente con PBT. Luego fueron pre-hibridados a 45°C en HS (formamida 50%, SSC 5X, heparina 50 µg/mL, tween 20 0,1%) conteniendo DNA de esperma de salmón. Pasado este tiempo la sonda desnaturalizada fue agregada a la solución de hibridación. Los embriones fueron hibridados toda la noche a 45°C con agitación suave (40 rpm). Luego los embriones fueron lavados sucesivamente en soluciones de concentraciones decrecientes de HS en PBT.

Los embriones fueron incubados con anticuerpo anti-digoxigenina acoplado a fosfatasa alcalina (dilución 1:2000) por 5 horas a temperatura ambiente, lavados con PBT y luego con una solución de 100 mM NaCl, 50 mM MgCl<sub>2</sub>, 100 mM Tris pH 9,5, Tween 20 0,1 %. Finalmente fueron incubados en la solución anterior conteniendo 0.16 mg/mL de BCIP y 0.36 mg/mL de NBT por 10-60 minutos en oscuridad. La reacción fue detenida lavando con una solución de EDTA 20 mM en PBT, los embriones fueron montados en portaobjetos usando el medio de montaje basado en Xileno Cytoseal XYL y examinados en un microscopio Carl Zeiss modelo Axiovert 25.

## RESULTADOS

### 1. Obtención de un cDNA enriquecido en transcritos diferencialmente expresados en gástrula.

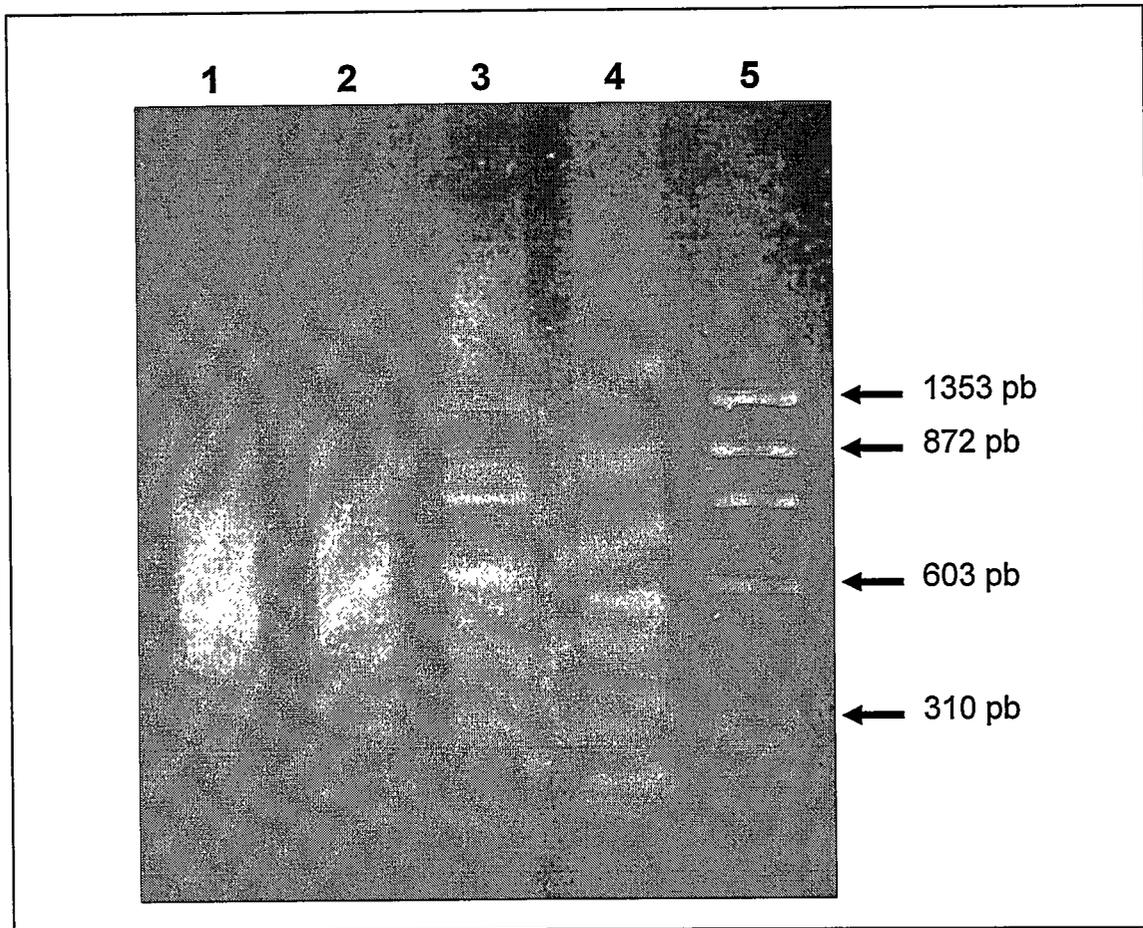
#### 1.1. Sustracción de secuencias comunes.

La obtención de secuencias diferencialmente expresadas en el estado de gástrula, con respecto al estado de blastodermo sincicial, se efectuó mediante un ensayo de hibridación sustractiva por supresión. El cDNA de doble hebra, sintetizado mediante la utilización del método SMART, a partir de RNA total de embriones en estados 6 y 7 del desarrollo (Campos-Ortega y Hartenstein, 1985) fue hibridado con cDNA SMART de embriones en estado de blastodermo sincicial, estados 2 y 3. Con esto se logró, sustraer aquellas secuencias comunes a ambas poblaciones de cDNA, que en este caso deberían corresponder principalmente a transcritos maternos, es decir, aquellos que estaban presentes en el ovocito antes de la fecundación. Considerando que la transcripción de los genes cigóticos se inicia en el estado 5, el resultado esperado de esta sustracción es la obtención de una población de cDNAs enriquecida en secuencias expresadas por el cigoto, preferentemente entre los estados 5-7 del desarrollo. Además, se efectuó la sustracción inversa, es decir, cDNA de blastodermo sincicial sustraído con cDNA de gástrula. Esta sustracción se realizó con el objetivo de sintetizar sonda para la verificación de la expresión diferencial de los clones aislados en la genoteca (sección III.c).

En paralelo se realizó un control del método, sustrayendo cDNA de músculo esquelético de ratón desde una mezcla de cDNA de músculo esquelético de ratón y

cDNA del fago ØX174 y enriqueciendo el cDNA resultante en secuencias pertenecientes al fago.

La figura 2 (A) muestra el patrón electroforético de los cDNAs sustraídos tanto de embriones en estados 6-7 como del control (carriles 1 y 3). Al comparar ambos cDNAs con sus controles negativos, es decir, cDNAs que no fueron sometidos a la sustracción (carriles 2 y 4), podemos ver que en el caso de los cDNAs de gástrula, no existe gran diferencia en el patrón electroforético, ambos muestran fragmentos con una distribución de tamaños entre 300 y 800 pb (carriles 1 y 2). No fue posible distinguir bandas individuales en el cDNA sustraído, lo que indica una alta complejidad de esta muestra, consistente en la presencia de numerosos fragmentos de cDNA de largo variable. Sin embargo, al comparar el patrón electroforético del control sustraído (carril 3) con el patrón del fago lambda (carril 5), vemos que coinciden claramente los cuatro fragmentos de mayor tamaño correspondientes a 1353, 1078, 872 y 603 pb, a diferencia del control no sustraído (carril 4) en el cual las dos bandas que se observan con mayor claridad no coinciden con el patrón electroforético del cDNA del fago. Este resultado nos indica que la reacción control de la sustracción se llevó a cabo satisfactoriamente y por lo tanto es esperable que el cDNA de gástrula se encuentre también sustraído de manera correcta. Resultados similares fueron obtenidos para la sustracción inversa.



**Figura 2. Productos de la hibridación sustractiva por supresión.** Electroforesis en un gel de agarosa al 1% de los productos de la hibridación sustractiva por supresión: cDNA de gástrula fue sustraído con cDNA de blastodermo sincicial (carril 1); cDNA de gástrula sin sustraer (carril 2); cDNA del fago ØX174 + cDNA de músculo esquelético sustraída con cDNA de músculo esquelético (carril 3) y cDNA de músculo esquelético sin sustraer (carril 4). En el carril 5 se observa el patrón electroforético del DNA del fago ØX174.

## 1.2. Eficiencia de la sustracción.

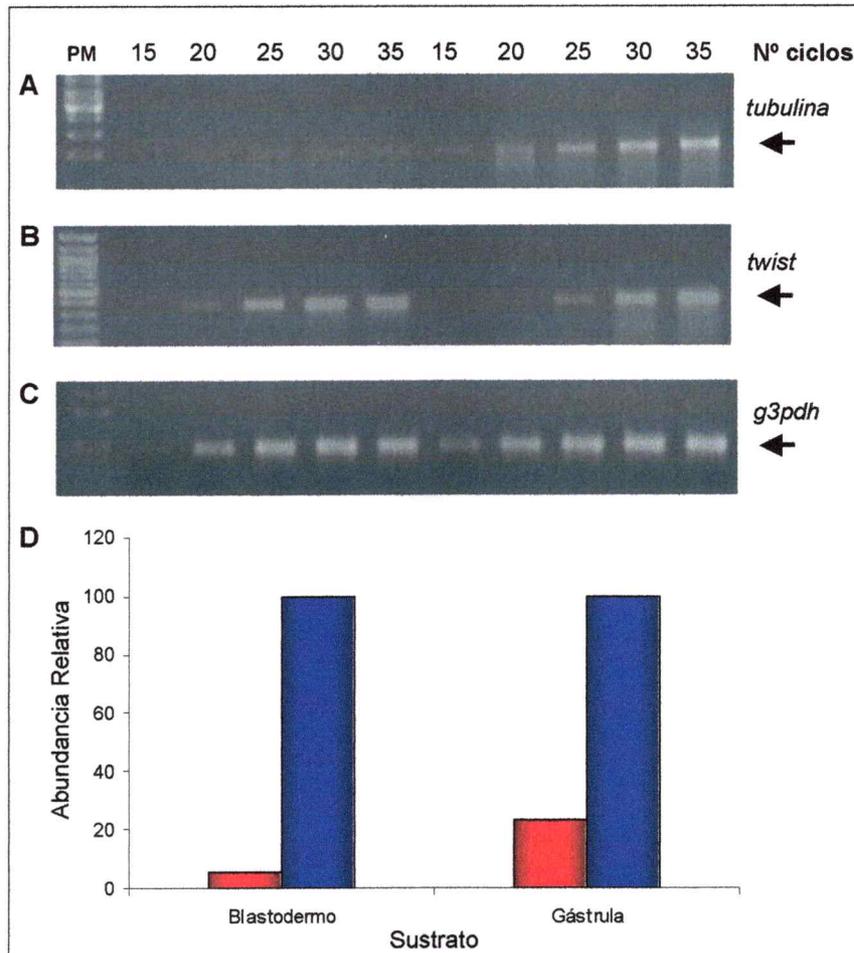
Con el objetivo de verificar la eficiencia de la sustracción realizamos una cinética de amplificación en la cual comparamos la abundancia de transcritos correspondientes a genes de expresión constitutiva y estado-específica en los cDNAs, sustraído y sin sustraer provenientes de gástrula (experimental) y músculo esquelético (control). En el caso del cDNA experimental se realizó una cinética de amplificación del transcrito del gen de expresión constitutiva *tubulina*, usando como sustrato iguales

cantidades (50 ng) del cDNA de gástrula sustraído (figura 3 A, izquierda) y del cDNA de gástrula sin sustraer (figura 3 A, derecha) y la amplificación del transcrito del gen *twist*, específico de gástrula, desde cDNA sustraído (figura 3 B, izquierda) y cDNA sin sustraer (figura 3 B, derecha). Paralelamente, se analizó la amplificación del transcrito del gen *g3pdh*, presente en el cDNA músculo esquelético y no en el fago ØX174, usando como sustrato una alícuota del cDNA control sustraído (figura 3 C, izquierda) y sin sustraer (figura 3 C, derecha).

Al comparar la cinética de amplificación de *tubulina*, se aprecia que hasta el ciclo 35 la cantidad de transcrito amplificado desde el cDNA sustraído no es significativa, en cambio cuando el cDNA sin sustraer fue usado como sustrato, pudimos detectar el producto de amplificación desde el ciclo 20 y en el ciclo 35 este fue 6.13 veces más abundante que el medido en el ciclo 35 de la muestra sustraída. Este resultado indica que como se espera para un gen de expresión constitutiva su transcrito se encuentre en menor abundancia en el cDNA sustraído comparado con el cDNA sin sustraer. Lo mismo se puede decir, para otro gen de expresión constitutiva como lo es el gen *g3pdh*, ya que este fue detectado en el ciclo 20 al usar el cDNA sustraído como sustrato y en el ciclo 15 al usar cDNA sin sustraer. Por otra parte, el transcrito del gen *twist* es detectado en el ciclo 20 en el cDNA sustraído y solo en el ciclo 25 en el cDNA sin sustraer, indicando un enriquecimiento en el cDNA sustraído de secuencias que se expresan preferentemente durante la gastrulación.

Estos datos fueron verificados mediante la medición de las abundancias del transcrito de otro gen constitutivo, *actina*, por PCR cuantitativo en tiempo real, en muestras de cDNA de gástrula sustraído y cDNA de blastodermo sincicial sustraído (sustracción inversa), y en sus respectivos cDNAs controles. El transcrito de *actina* es

al menos 5 veces más abundante en las muestras correspondientes a los cDNAs controles respecto de los cDNAs sustraídos (figura 3 D).



**Figura 3. Verificación de la eficiencia de la sustracción. A)** Amplificación del transcrito de *tubulina* desde cDNA de gástrula sustraído (izquierda) o cDNA de gástrula sin sustraer (derecha). **B)** Amplificación del transcrito de *twist* desde cDNA de gástrula sustraído (izquierda) o cDNA de gástrula sin sustraer (derecha). **C)** Amplificación del transcrito del gen *g3pdh* desde la mezcla de cDNA control sustraída (izquierda) y no sustraída (derecha). Alícuotas de 5  $\mu$ L de reacciones de PCR fueron analizadas mediante electroforesis en geles de agarosa al 1% entre los ciclos 15 al 35 de amplificación. PM = 1Kb DNA ladder. **D)** Cuantificación de la abundancia relativa del transcrito de actina mediante PCR cuantitativo en tiempo real. El transcrito fue amplificado utilizando como sustrato los cDNAs sustraídos (barras rojas) o los cDNAs sin sustraer (barras azules). Las reacciones de PCR fueron llevadas a cabo en el Light Cycler Instrument (Roche) Para cada estado, la concentración fue determinada a partir de una curva de calibración realizada con un vector que contiene *actina*. Este valor fue dividido por el máximo valor alcanzado en alguna de las amplificaciones generando un valor = 1 para el cDNA en que se encuentra mas expresado.

## **2. Construcción y caracterización de una genoteca específica de gastrulación.**

### **2.1. Construcción de la genoteca de sustracción.**

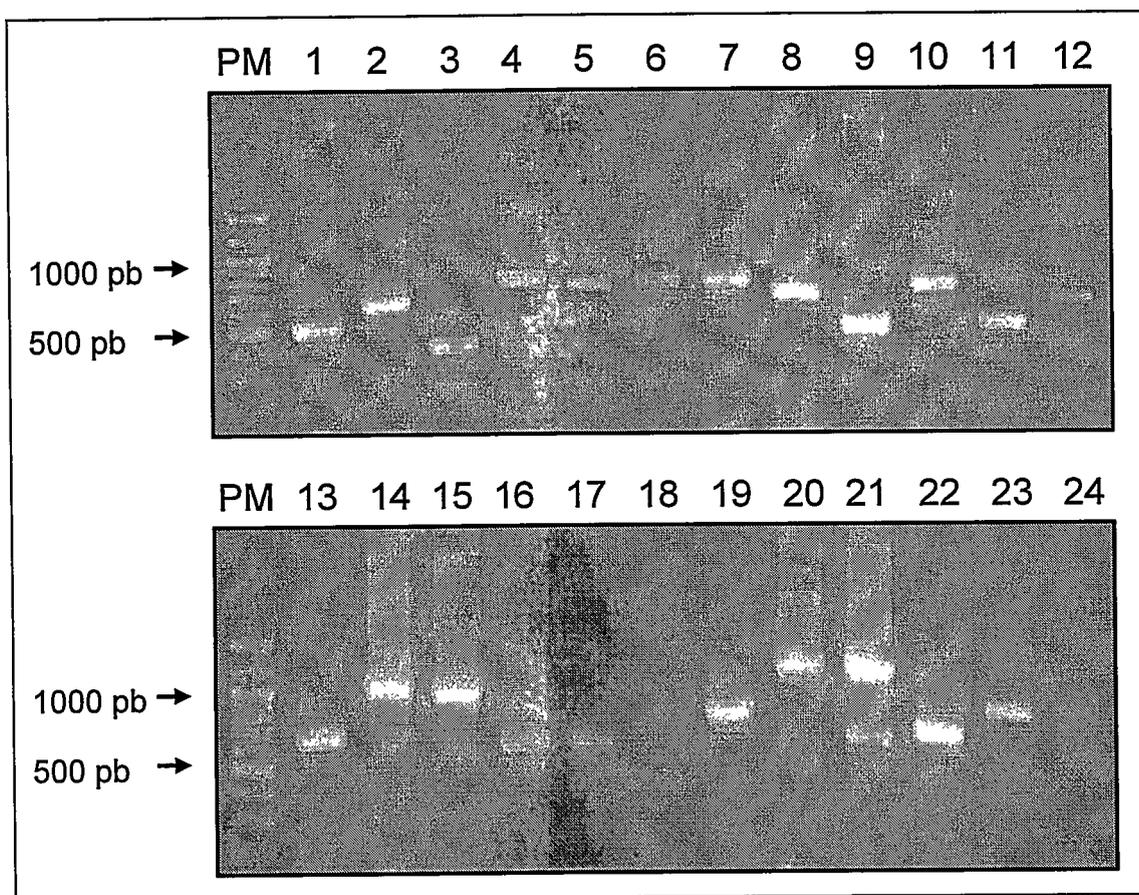
El conjunto de cDNAs obtenidos a partir de la hibridación sustractiva por supresión fue purificado, ligado al vector pGEM-T Easy (Promega) y utilizado para la transformación de bacterias competentes *E. coli*. Los transformantes positivos fueron seleccionados manualmente y sembrados en placas de 96 pocillos conteniendo medio LB, agar y ampicilina. Finalmente, obtuvimos 1428 clones distribuidos en 15 placas. Los clones fueron replicados en medio líquido y congelados con glicerol a  $-80^{\circ}\text{C}$ , para su posterior utilización.

La purificación de los productos de la sustracción aumentó en un orden de magnitud la eficiencia de transformación ya que en un experimento anterior al clonar el cDNA sin purificar, se obtuvo solamente 150 clones transformantes. La clonación de los productos purificados se realizó en tres ocasiones obteniéndose números comparables de transformantes en cada una de ellas, este resultado permite decir que, bajo nuestras condiciones experimentales, la eficiencia de este evento de clonación es reproducible.

### **2.2. Verificación de insertos clonados.**

Con el propósito de verificar si las colonias aisladas poseían un inserto de cDNA, se seleccionó una muestra al azar de 136 colonias, correspondiendo al 9.5% de la genoteca. Los insertos fueron amplificados mediante PCR desde 5  $\mu\text{L}$  de cultivo bacteriano utilizando partidores universales (T7 y SP6) obteniéndose productos de amplificación en todos los casos. La figura 4 muestra un patrón electroforético representativo de estas reacciones de PCR. Pese a que existe amplificación

inespecífica, debido a la baja temperatura de alineamiento de los partidores usados (42 °C), se observa para cada caso la presencia de una banda principal, cuyo tamaño varía entre 500 a 1000 pares de bases. Estos tamaños se deben a que los cDNA fueron digeridos con Rsa I, procedimiento necesario para la sustracción. Los resultados de este experimento nos permitieron inferir que todos los clones presentes en la genoteca poseen inserto de cDNA.



**Figura 4. Eficiencia de la transformación.** Electroforesis en geles de agarosa al 1% de insertos de cDNA amplificados por PCR usando los partidores SP6 y T7. La figura muestra 24 de 136 clones escogidos al azar, los que corresponden al 9.5% del total de clones presentes en la genoteca. En todos los casos probados se observa amplificación mayoritaria de un fragmento.

### **2.3. Caracterización de la genoteca de sustracción.**

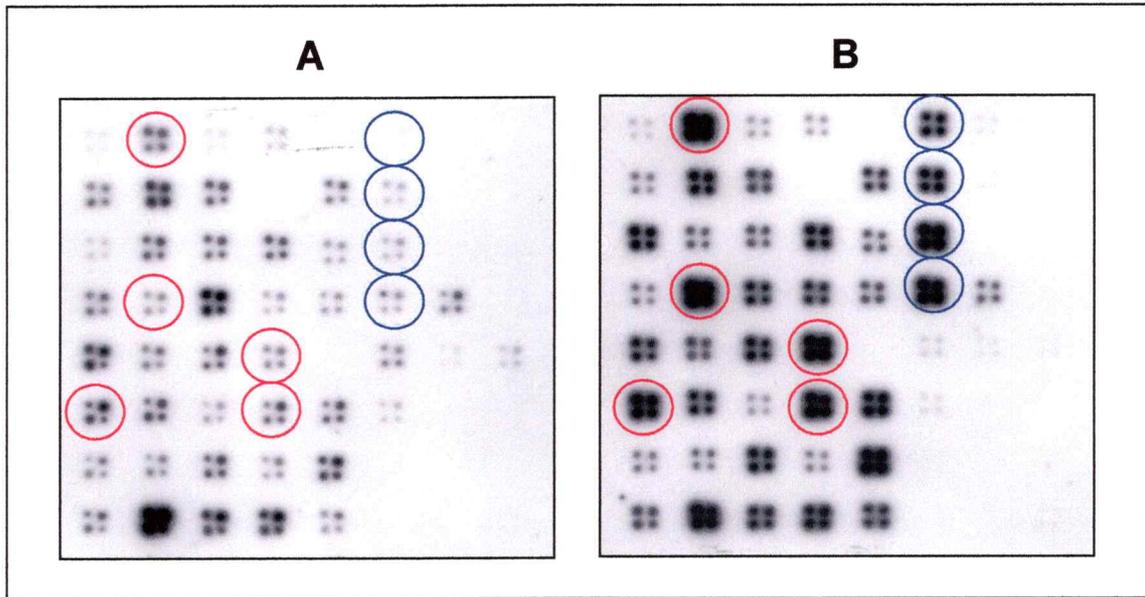
Es necesario estimar que proporción de genes presentes en la genoteca corresponde efectivamente a genes diferencialmente expresados en gástrula y cual corresponde a falsos positivos de la sustracción. Para ello fueron realizados ensayos de hibridación en matrices de nylon que contenían DNA plasmidial o fragmentos de PCR de los clones aislados y ensayos de hibridación en Northern blots virtuales.

#### **2.3.1. Estimación del porcentaje de genes diferencialmente expresados.**

Se diseñaron matrices de hibridación Tipo I sembrando productos de PCR y DNA plasmidial de 136 clones escogidos al azar con el objetivo de estimar la presencia de genes diferencialmente expresados en el estado de gástrula. Para ello, las matrices fueron hibridadas con sondas complejas marcadas con [<sup>32</sup>P]dCTP sintetizadas a partir del cDNA de gástrula sustraído con cDNA de blastodermo sincicial y del cDNA obtenido en la sustracción inversa. La ventaja de usar sondas sintetizadas a partir de cDNA sustraído consiste en aumentar el grado de sensibilidad del ensayo, dado que aquellos transcritos poco abundantes, al utilizar sondas no sustraídas, generan señales débiles que son difíciles de distinguir del ruido asociado a este ensayo de hibridación. La figura 5 muestra una imagen representativa de los resultados obtenidos en este ensayo en donde los círculos rojos señalan aquellos clones que presentan una señal de hibridación más intensa al ser hibridados con el cDNA de gástrula sustraído. En la misma figura los círculos azules indican controles positivos, que corresponden a transcritos de expresión diferencial en gástrula; de arriba hacia abajo encontramos los genes CG10035, CG4922, CG13427 y CG13333 (González-Agüero y col., 2004).

El análisis de las señales de hibridación obtenidas se realizó calculando el logaritmo natural de la intensidad de hibridación de cada clon (promedio de 4 puntos de

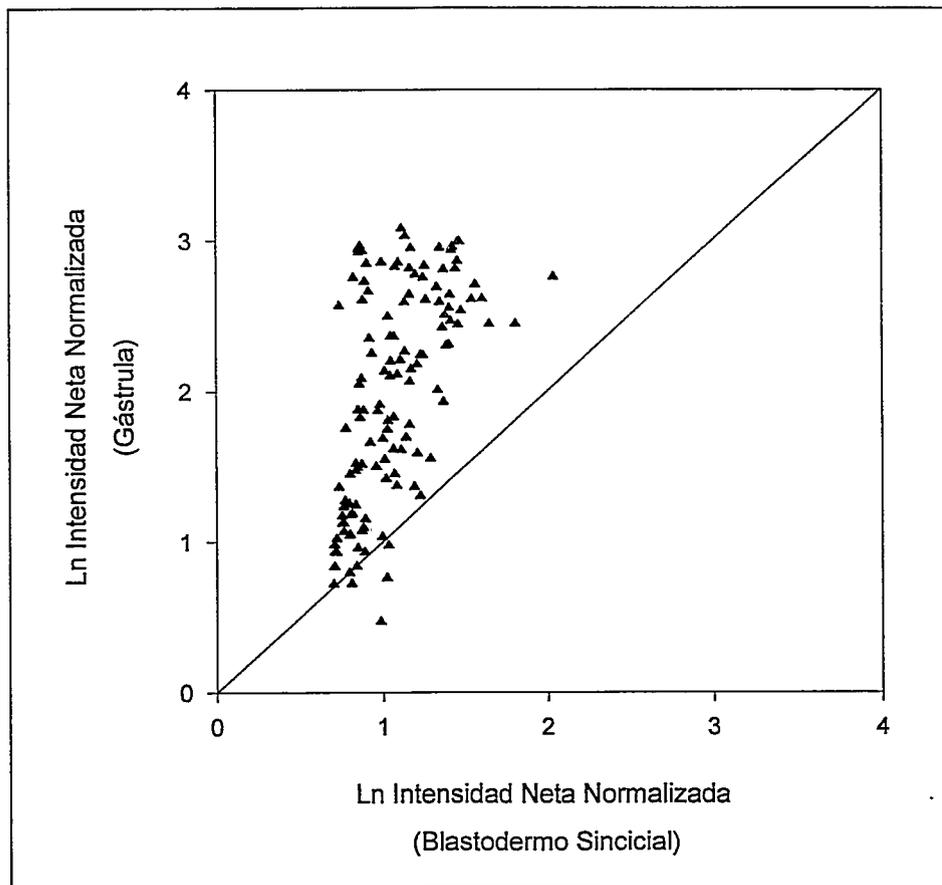
hibridación en la membrana). Estos valores fueron normalizados utilizando las señales de hibridación del gen constitutivo *actina*, lo que nos permitió comparar estas señales en ambas membranas. La razón entre los valores normalizados para cada hibridación, cDNA de gástrula sustraído/cDNA blastodermo sincicial sustraído, son entonces una medida de la abundancia relativa del transcrito



**Figura 5. Caracterización de la genoteca.** Matrices de hibridación tipo I sembradas en cuadruplicado con productos de PCR obtenidos partir de 50 clones seleccionados al azar fueron hibridadas con sondas complejas correspondientes a cDNA de blastodermo sincicial sustraído con cDNA de gástrula (A) y cDNA de gástrula sustraído con cDNA de blastodermo sincicial (B). Los círculos rojos denotan genes cuya intensidad de hibridación es mayor en las matrices hibridadas con cDNA de gástrula sustraído. Los círculos azules señalan genes conocidos de expresión cigótica utilizados como controles positivos.

correspondiente en cada sonda. En la figura 6 se han graficado los logaritmos de las señales de intensidad obtenidas para la hibridación de estos 136 clones con ambas sondas. Para aquellos clones cuya abundancia no cambia en los estados comparados, la razón entre las señales de intensidad de hibridación es 1 (en la línea de identidad). Aquellos clones cuya abundancia relativa es mayor en gástrula que en blastodermo presentan razones de intensidad de hibridación mayores a 1 (sobre la línea de

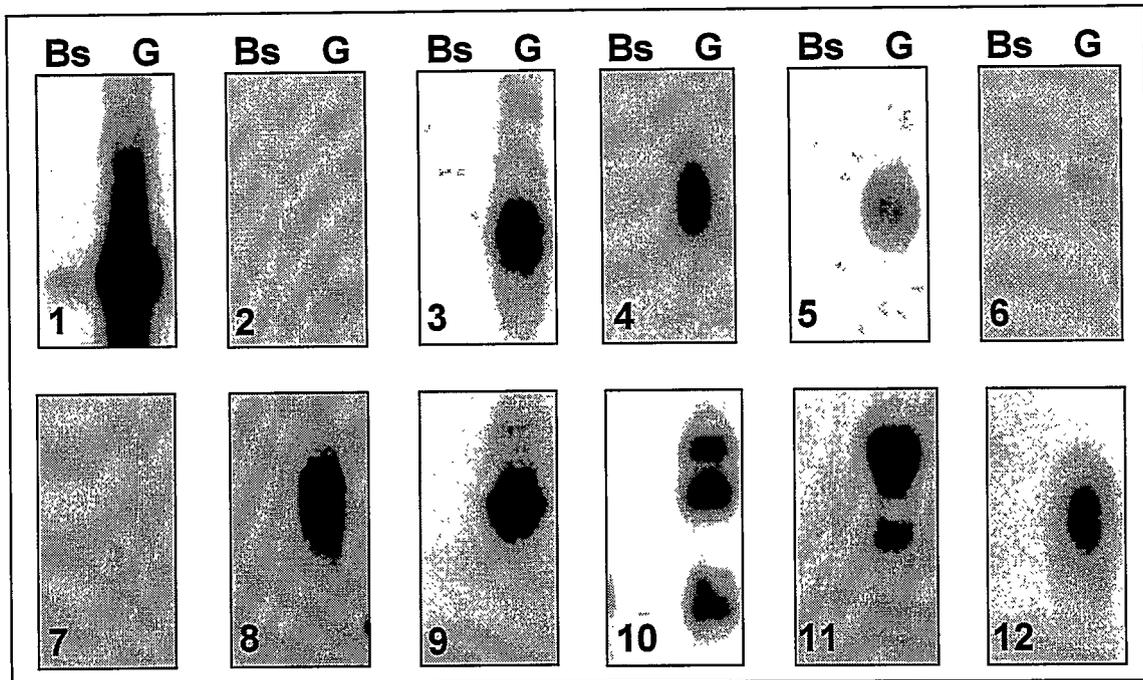
identidad). De los 136 clones, el 64% presentan razones mayores o igual a dos, es decir, la abundancia relativa del correspondiente transcrito es el doble o más, en el estado de gástrula. El restante 36% fueron considerados como falsos positivos de la sustracción. Este resultado permite estimar que dos tercios de la genoteca esta enriquecida en secuencias que se expresan diferencialmente en el estado de gástrula.



**Figura 6. Expresión diferencial de genes en el estado de gástrula.** Los valores de intensidad para cada punto de siembra sobre las matrices tipo I fueron obtenidos utilizando el software 1D (Kodak). El valor de ruido calculado para cada membrana fue restado a los correspondientes valores de intensidad cruda y éstos fueron normalizados utilizando los datos de intensidad obtenidos en cada membrana para la expresión de un gen constitutivo (*actina*). El gráfico muestra el logaritmo natural de las intensidades de hibridación del cDNA de gástrula sustraído con blastodermo contra el logaritmo natural de las intensidades de hibridación del cDNA de blastodermo sustraído con gástrula. El valor de intensidad de aquellos clones que hibridan en proporción similar con transcritos presentes en ambos cDNAs se ubica en la línea de identidad trazada. Sin embargo, aquellos clones expresados preferentemente en gástrula se desvían de este comportamiento ubicándose sobre esta línea. Esto se cumple para el 64% de los clones sembrados.

### **2.3.2. Análisis de Northern blots virtuales.**

Los resultados de las hibridaciones en matrices fueron sometidos a una prueba de verificación mediante ensayos de Northern blots virtuales. Para esto, se sintetizaron sondas marcadas con [<sup>32</sup>P]dCTP de 12 clones cuyas razones de intensidad G/Bs fueron mayores a 4. La figura 7 muestra los resultados de este ensayo, la señal neta en cada hibridación fue obtenida restando el ruido a la señal cruda. Estos valores fueron normalizados respecto a la hibridación de actina con el objetivo de comparar las abundancias de cada clon en blastodermo sincicial y en gástrula. La razón de intensidad G/Bs en estos Northern blots fue calculada indicando que los clones 1, 2, 3 y 4 aumentan su expresión en 1.66, 3.9, 24.3 y 26 veces en gástrula respecto a blastodermo sincicial. El resto de los clones (5-12) presentan una señal de hibridación indetectable en blastodermo sincicial, por lo que podemos suponer que la expresión de estos transcritos se activa durante la gastrulación. Este ensayo confirma los resultados de la hibridación en matrices. Podemos decir que con la aplicación de la hibridación sustractiva por supresión aislamos mensajeros cuya expresión aumenta en el estado de gástrula.



**Figura 7. Northern blots virtuales.** La expresión diferencial de 12 clones fue verificada mediante Northern blots virtuales, en los cuales la membrana de nylon contiene cDNA sintetizado a partir de RNA total de embriones en estado de blastodermo sincicial (Bs) o en estado de gástrula (G). En todos los casos, la intensidad de hibridación de las sondas con el cDNA de gástrula fue mayor concordando con los experimentos de hibridación en matrices tipo I.

### 3. Identificación y caracterización de los genes aislados.

#### 3.1. Obtención y análisis de las secuencias.

De acuerdo con nuestras capacidades, 648 clones seleccionados al azar fueron enviados a la empresa Agencourt Bioscience Corporation (Beverly MA) para su secuenciación.

Se obtuvieron las secuencias de 632 clones con un largo promedio de 1000 pb. Las secuencias obtenidas en formato Fasta incluyen fragmentos del vector en el cual se hizo la clonación. Para efectos de búsquedas de homologías por alineamiento y otras tareas ligadas al análisis de secuencia, es necesario remover de la secuencia leída las regiones pertenecientes al vector. Con este fin y como parte de la práctica

profesional de Pablo Moreno (estudiante de Ingeniería Civil en Biotecnología) se implementó una serie de rutinas en el lenguaje de programación Perl que permiten obtener las secuencias de cada inserto sin vector. Además, se eliminaron secuencias de mala calidad (valor phred 20) y aquellas que, por problemas durante el proceso de ligación, contenían fragmentos de varios cDNAs diferentes, quedando 588 secuencias.

### **3.2. Identificación de los clones aislados.**

En primer término se realizó un BLASTN (Altschul y col., 1990) masivo de las secuencias curadas de la genoteca contra las bases de datos conformadas por las regiones codificantes (CDS) y regiones no traducidas (UTRs) de los transcritos del genoma completo de *Drosophila*, disponibles en el Release 3.2 de BDGP (Tabla 1). Con el propósito de resumir los 588 resultados generados por BLAST, se escribió una rutina en Perl que permite extraer los mejores alineamientos e información sobre ellos. Estos resultados fueron analizados en forma manual e identificamos 490 cDNAs (83%) que presentaban buenos alineamientos con transcritos de *Drosophila*, nuestro punto de corte ( $E < 10^{-11}$ ) fue determinado de manera empírica para incluir secuencias con alto porcentaje de cobertura y un alto porcentaje de identidad. De estos, 88 genes, es decir, el 32.8% estaban representados por más de un clon en la genoteca, lo que resultó en la identificación de 268 genes únicos. Las secuencias de los restantes 98 cDNAs fueron comparadas, mediante BLASTN, con la base de datos que contiene las secuencias de eucromatina de *Drosophila* y encontramos que 15 de nuestras secuencias (2.55%) correspondían a genes que codifican para mensajeros no-traducidos mientras que 63 cDNAs (10.7%) alineaban con regiones intergénicas o intrones en BDGP y 16 con elementos repetitivos o de transposición (2.72%). Estos resultados indican que existen secuencias dentro del genoma que probablemente

codifican proteínas funcionales que los algoritmos de anotación computacional no han podido detectar y de los cuales no se tiene ninguna información.

Los 268 genes únicos que codifican proteínas pueden ser clasificados en tres categorías: 1) genes no caracterizados: aquellos anotados por BDGP mediante algoritmos de anotación como Genie o GeneScan; 2) genes parcialmente caracterizados: aquellos que presentan homología de secuencias con genes descritos en otros organismos o han sido aislados en experimentos de expresión génica a gran escala, y 3) genes caracterizados: aquellos que han sido identificados en estudios fenotípicos, y en algunos casos, mediante ensayos funcionales.

En el caso de los genes no caracterizados, se realizó una búsqueda de ortólogos en otros organismos. Para ello se obtuvieron las secuencias de las proteínas codificadas por estos genes y se realizó un alineamiento contra bases de datos de proteínas de *H. sapiens*, *M. musculus* y *C. elegans*. La Tabla 2 resume el resultado de estos alineamientos, en donde encontramos proteínas ortólogas para 27 genes con un porcentaje de identidad superior al 50%. Aunque algunos presentan ortólogos en varios organismos, la tabla resume el mejor alineamiento para cada secuencia, en este caso, los ortólogos con mayor porcentaje de identidad se encuentran en *H. sapiens* y *C. elegans*.

**Tabla 1. Clasificación de los clones secuenciados.** Resultados del alineamiento contra las bases de datos de BDGP ([www.flybase.net](http://www.flybase.net)).

Características	Nº	Porcentaje
<i>Clones secuenciados correctamente</i>	588	100,0
Genes anotados en BDGP	521	88,6
Genes que codifican proteínas	490	83,3
Genes que codifican RNAs no traducidos	15	2,6
Retrotransposones	16	2,7
Secuencias sin anotación	67	10,7
<i>Genes que codifican proteínas</i>		
Secuencias únicas	268	100,0
Caracterizados	107	39,9
Parcialmente caracterizados	23	8,6
No caracterizados	138	51,5

**Tabla 2. Ortólogos de los genes no caracterizados.** Las proteínas codificadas por los genes no caracterizados fueron alineadas contra las bases de datos de proteínas de InterPro. La tabla indica la especie, el porcentaje de identidad, y una descripción del mejor ortólogo encontrado. Están incluidas aquellas proteínas con más de un 50% de identidad.

CG <sup>1</sup>	Organismo	% de identidad	descripción
CG8415	<i>H. sapiens</i>	90.91	40S ribosomal protein S23
CG13045	<i>H. sapiens</i>	88.89	Hypothetical protein KIAA1399 (Fragment)
CG18591	<i>H. sapiens</i>	75.82	Small nuclear ribonucleoprotein E
CG5684	<i>H. sapiens</i>	74.81	CCR4-NOT transcription complex, subunit 7
CG9924	<i>H. sapiens</i>	74.60	Speckle-type POZ protein
CG5277	<i>H. sapiens</i>	73.08	TGF beta-inducible nuclear protein 1
CG7748	<i>H. sapiens</i>	70.13	Source of immunodominant MHC-associated peptides
CG17489	<i>H. sapiens</i>	67.22	MSTP030
CG33096	<i>H. sapiens</i>	65.48	Similar to CGI-67 protein (Fragment)
CG14641	<i>H. sapiens</i>	64.87	Hypothetical protein
CG1515	<i>H. sapiens</i>	61.73	SNARE protein Ykt6
CG12581	<i>C. elegans</i>	59.39	Hypothetical protein R148.5
CG18375	<i>H. sapiens</i>	59.07	Apoptosis stimulating of p53 protein 1
CG4897	<i>C. elegans</i>	58.55	60S ribosomal protein L7
CG13454	<i>H. sapiens</i>	57.89	L-myc-2 protein
CG14939	<i>H. sapiens</i>	57.75	Hypothetical protein (Fragment)
CG11597	<i>H. sapiens</i>	57.49	Serine/threonine protein phosphatase 4
CG10417	<i>H. sapiens</i>	56.89	Protein phosphatase 2C gamma isoform
CG3402	<i>C. elegans</i>	56.64	Hypothetical 13.5 kDa protein C45G9.7 in chromosome III
CG9629	<i>C. elegans</i>	56.50	Putative aldehyde dehydrogenase family 7 member A1 homolog
CG4111	<i>H. sapiens</i>	56.20	60S ribosomal protein L35

CG6684	<i>C. elegans</i>	54.87	40S ribosomal protein S25
CG11901	<i>H. sapiens</i>	53.55	Elongation factor 1-gamma
CG6196	<i>H. sapiens</i>	53.50	Human full-length cDNA clone CS0DK012YE11
CG1906	<i>H. sapiens</i>	52.80	Protein phosphatase 2C beta isoform
CG32850	<i>H. sapiens</i>	52.60	RING finger protein 11
CG4759	<i>H. sapiens</i>	52.59	60S ribosomal protein L27

1: Nomenclatura de identificación de cada gen usada por BDGP que significa Computed Gene.

A continuación y con el objetivo de analizar la composición de nuestra genoteca, y de compararla con trabajos previos enfocados en el desarrollo embrionario de *Drosophila*, se efectuó una búsqueda en la literatura de los genes aislados en los siguientes estudios: 1) análisis de hibridación sustractiva de secuencias expresadas en mesodermo (Casal y Leptin, 1996); 2) experimentos en matrices de hibridación que comparan la expresión de 5081 genes en mutantes con pérdida y ganancia de función del gen *twist* (Furlong y col., 2001); 3) matrices de hibridación que comparan la expresión de 13.500 genes en mutantes con pérdida y ganancia de función del gen *dorsal* (Stathopoulos y col., 2002); 4) matrices de hibridación sembradas con 4.800 genes de *Drosophila*, e hibridadas con sondas obtenidas de 66 intervalos de tiempo durante su ciclo de vida (Arbeitman y col., 2002). La Tabla 3 muestra una lista de los 42 genes aislados en la genoteca, tanto caracterizados como no caracterizados, que han sido descritos en los estudios mencionados, estos genes mostraron ser expresados en el cigoto y en algunos casos, poseer alguna función durante el desarrollo temprano de *Drosophila*. La primera columna muestra la nomenclatura utilizada por BDGP donde cada gen es identificado por el prefijo, CG (Computed Gene) seguido de un número. La segunda columna muestra el símbolo de cada gen, que corresponde a una abreviación del nombre completo, que se muestra en la tercera columna. Finalmente la cuarta columna indica el estudio en que el gen fue encontrado. Dado que estos trabajos están enfocados principalmente en la búsqueda de genes

involucrados en el desarrollo, este resultado valida el procedimiento experimental utilizado.

**Tabla 3. Comparación de genes aislados en nuestra genoteca con aquellos identificados en otros experimentos de expresión génica.** Los genes listados se encuentran en la fracción secuenciada de la genoteca de sustracción y han sido reportados con anterioridad en los siguientes estudios: Casal y Leptin, 1996 (1); Furlong y col., 2001 (2); Stathopoulos y col., 2002 (3); Arbeitman y col., 2002 (4).

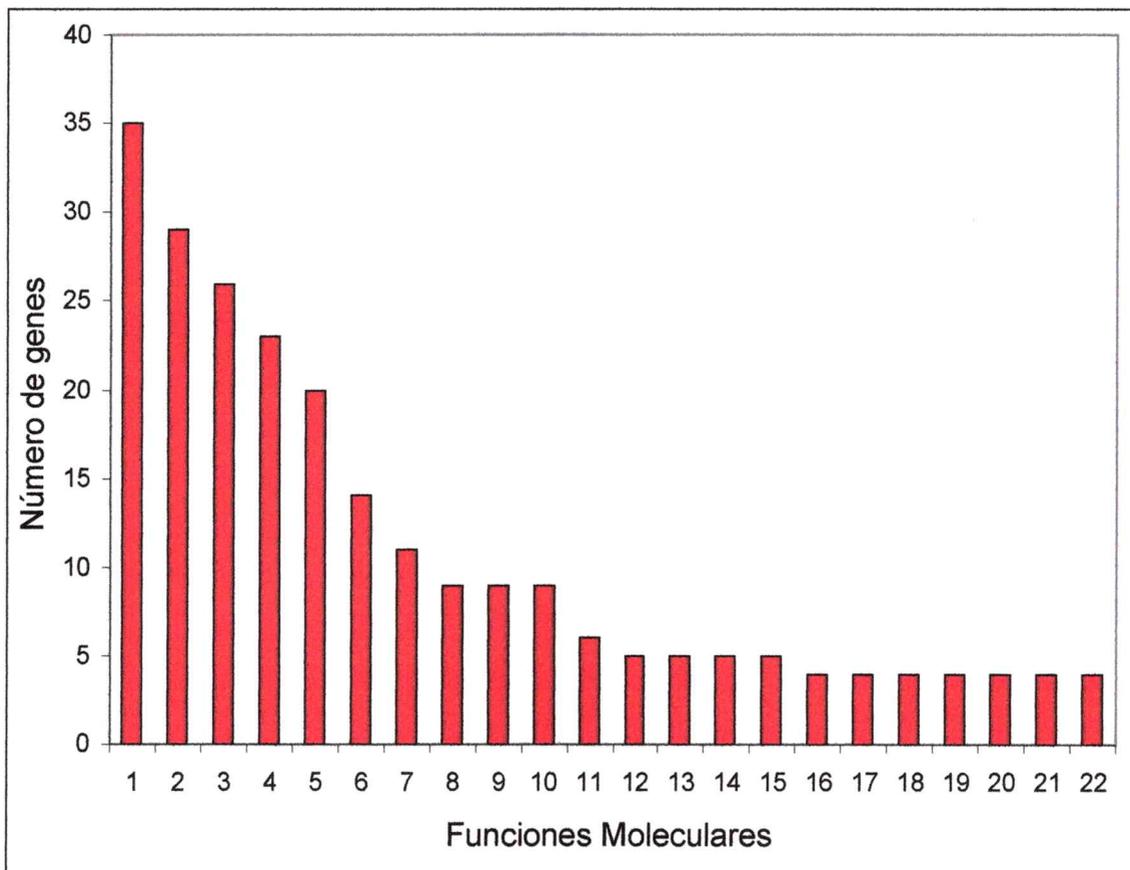
CG	Símbolo	Nombre completo	Referencia
CG4500	CG4500		1, 2, 3, 4
CG18485	stumps		1, 2, 3
CG10619	tup	tailup	2, 4
CG15319	nej	nejire	2, 4
CG32372	CG32372		1
CG33200	ventrally-expressed-protein-D		1
CG40138	CG40138		2
CG8776	CG8776		2
CG12919	egr	eiger	3
CG7642	ry	rosy	3
CG1007	emc	extra macrochaetae	4
CG10218	Mhcl	Myosin heavy chain-like	4
CG1028	Antp	Antennapedia	4
CG10417	CG10417		4
CG11212	Ptr	Ptc-related	4
CG11280	trn	tartan	4
CG11605	jbug	jitterbug	4
CG1225	RhoGEF3		4
CG12369	Lac	Lachesin	4
CG14026	tkv	thickveins	4
CG15009	ImpL2	Ecdysone-inducible gene L2	4
CG15095	I(2)08717		4
CG15658	CG15658		4
CG1677	CG1677		4
CG17921	HmgZ	HMG protein Z	4
CG2198	Ama	Amalgam	4
CG2835	G-salpa60A	G protein salpha 60A	4
CG2916	septin 05		4
CG3851	odd	odd skipped	4
CG4123	Mipp1	Multiple inositol polyphosphate phosphatase 1	4
CG5002	CG5002		4
CG5059	CG5059		4
CG5170	Dp1	Dodeca-satellite-binding protein 1	4
CG5888	CG5888		4
CG6234	CG6234		4
CG7748	OstStt3	Oligosaccharyl transferase 3	4
CG8654	CG8654		4
CG8846	Thor	Thor	4
CG8960	CG8960		4
CG9188	sip2	septin interacting protein 2	4
CG9629	CG9629		4
CG9704	Nrt	Neurotactin	4

### **3.3. Caracterización de los genes aislados.**

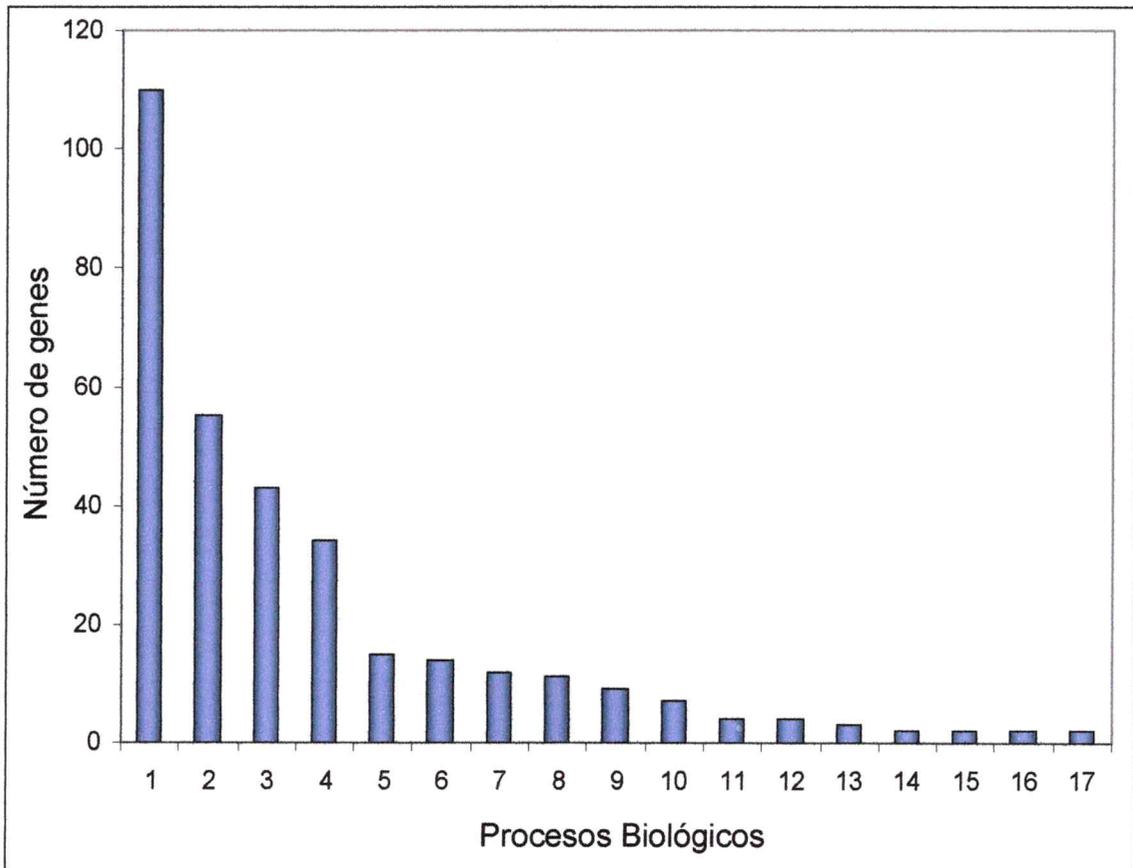
Con el objetivo de caracterizar los 268 genes únicos se utilizó la base de datos de Gene Ontology, la que proporciona un vocabulario controlado, preciso y ordenado de manera jerárquica que permite describir los productos génicos de algunos eucariontes sobre la base de tres atributos:

1. Función Molecular, definida como las tareas realizadas por un producto génico individual.
2. Proceso Biológico, definido como los propósitos biológicos amplios efectuados por conjuntos ordenados de funciones moleculares.
3. Componente celular, definido como la localización, estructura subcelular o complejo macromolecular donde actúa un producto génico individual.

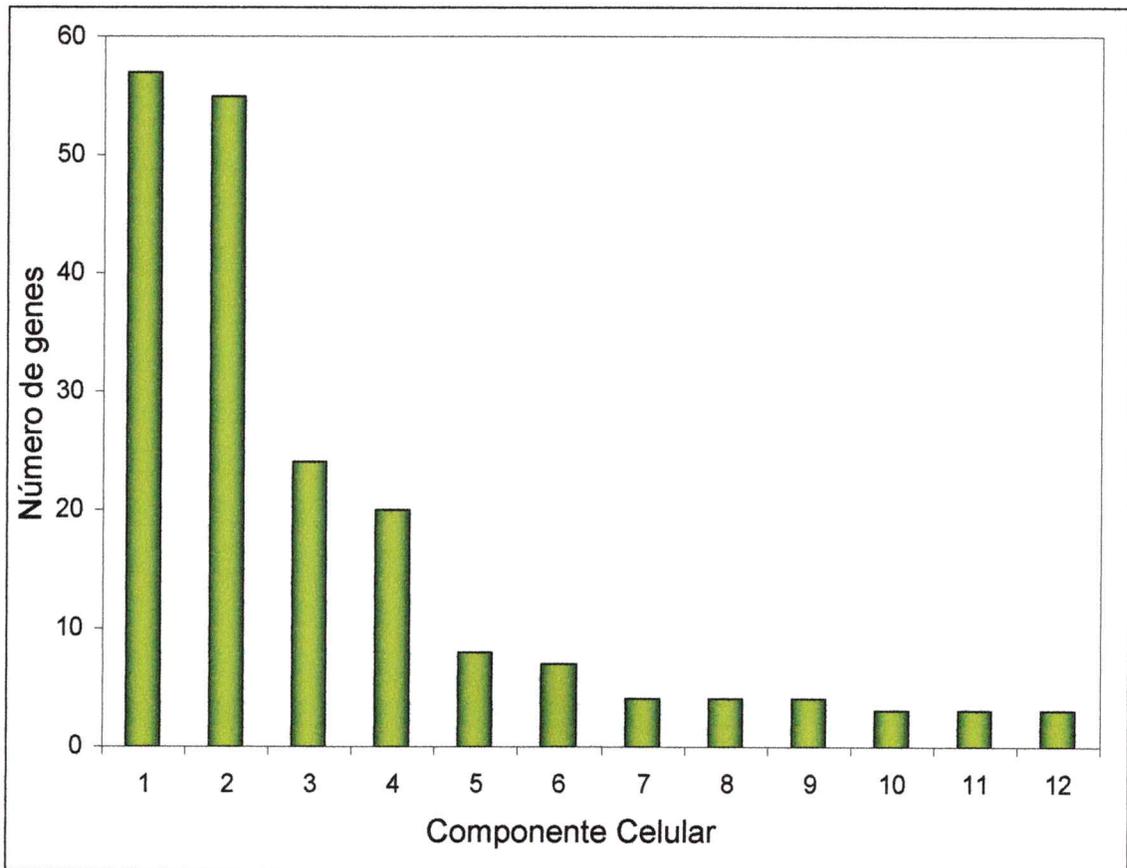
De esta manera cada una de las tres ontologías permite de forma independiente asignar propiedades a los genes de acuerdo con las funciones moleculares que poseen (figura 8), procesos biológicos en los cuales participan (figura 9) y los componentes celulares en los cuales se encuentran (figura 10). En cada una de las figuras se ha graficado el número de genes con un término GO asignado contra los términos GO más representados en la genoteca.



**Figura 8. Frecuencia de genes con funciones moleculares asignadas.** El gráfico muestra el número de genes a cuyos productos se les atribuyen las siguientes funciones moleculares: 1, Unión a ácidos nucleicos. 2, Hidrolasa. 3, Transferasa. 4, Unión a proteínas. 5, Quinasa. 6, Factor de transcripción de la RNA polimerasa II. 7, Constituyente estructural del ribosoma. 8, Receptor. 9, Transportador. 10, Oxidoreductasa. 11, Transportador de iones. 12, Factor de transcripción. 13, Represor transcripcional. 14, Ligasa. 15, Unión a iones metálicos. 16, Unión a Nucleótidos. 17, Helicasa. 18, Inhibidor de enzimas. 19, Receptor de proteínas de señalización. 20, Constituyente estructural del citoesqueleto. 21, Transportador de electrones. 22, Factor de traducción. El 63,43% de los genes aislados posee al menos una función molecular asignada. Este análisis se realizó sobre el segundo estrato de esta ontología.



**Figura 9. Frecuencia de genes con procesos biológicos asignados.** El gráfico muestra el número de genes cuyos productos participan en los siguientes procesos biológicos: 1, Metabolismo. 2, Crecimiento Celular. 3, Comunicación celular. 4, Morfogénesis. 5, Reproducción. 6, Especificación de patrones. 7, Respuesta a Estímulos. 8, Desarrollo embrionario. 9, Motilidad celular. 10, Diferenciación celular. 11, Respuesta a estrés. 12, Secreción, 13, Desarrollo larval. 14, Celularización. 15, Desarrollo post-embrionario. 16, Muerte. 17, Regulación de la expresión génica. El 55.22% de los genes aislados posee al menos un proceso biológico asignado. Este análisis se realizó sobre el segundo estrato de esta ontología.

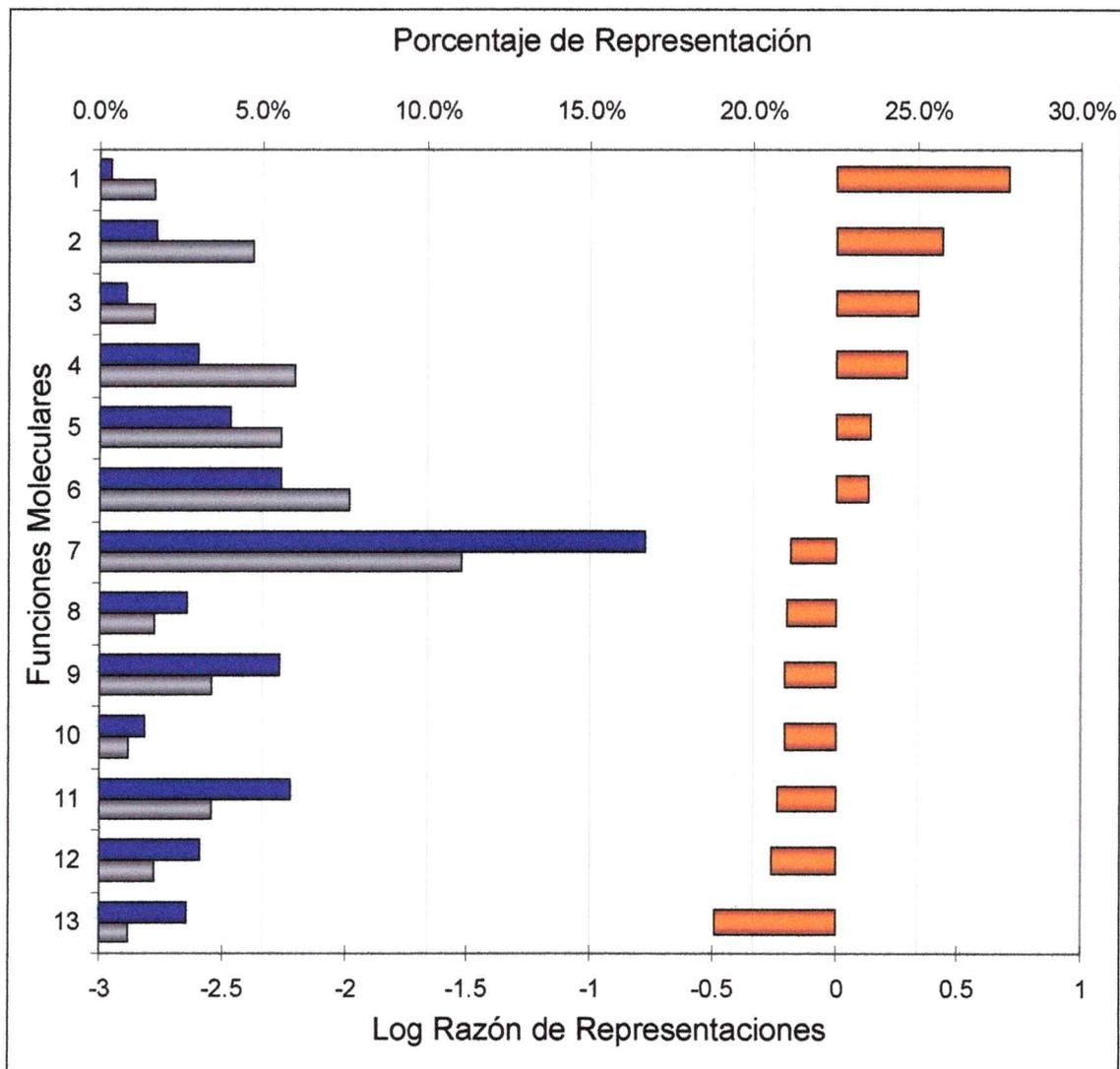


**Figura 10. Frecuencia de genes con componentes celulares asignados.** El gráfico muestra el número de genes cuyos productos son encontrados en los siguientes componentes celulares: 1, Núcleo. 2, Citoplasma. 3, Membrana plasmática. 4, Complejo ribonucleoproteína. 5, Integral a membrana. 6, Cromosoma. 7, Membrana mitocondrial. 8, Sistema de endomembrana. 9, Membrana interna. 10, Complejo de la ligasa de ubiquitina. 11, Exosoma. 12, Periferia celular. El 45.52% de los genes posee al menos un componente celular asignado. Este análisis se realizó sobre el tercer estrato de esta ontología.

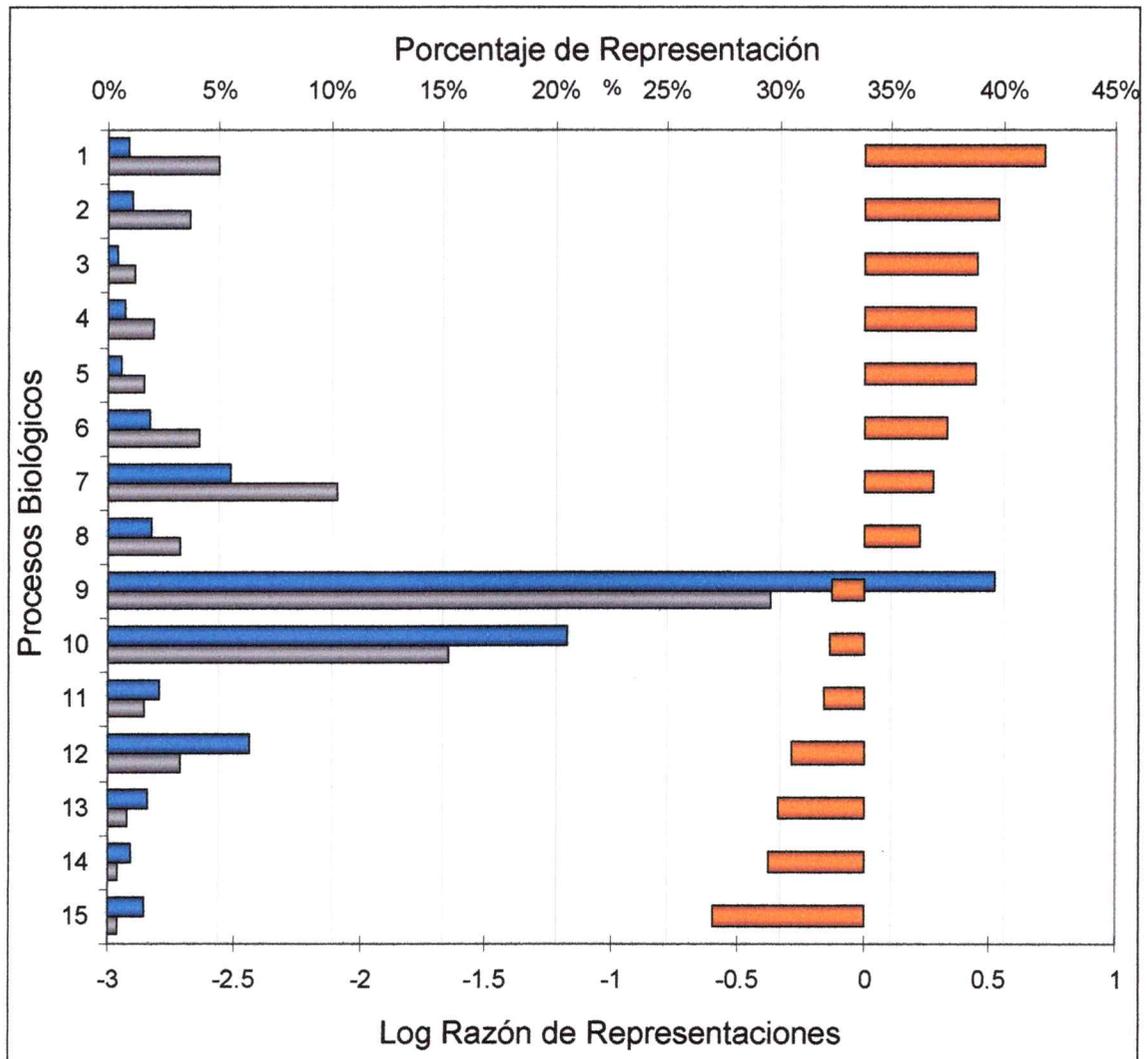
### 3.4. Composición funcional de la genoteca.

Con el propósito de comparar la composición de la genoteca de sustracción, en términos de las funciones moleculares y procesos biológicos presentes en ella respecto a la composición del genoma completo de *Drosophila*, la representatividad de cada uno de los términos GO fue contabilizada mediante un software desarrollado para este propósito en nuestro laboratorio (Moreno y col., manuscrito en preparación), y expresada en porcentaje de representatividad respecto al total de términos GO asignados, tanto para la genoteca como para el genoma completo. La figura 11 muestra aquellos términos GO correspondientes a Función Molecular que más cambian su representatividad en la genoteca respecto al genoma. A la derecha de la figura se encuentra graficado el logaritmo en base 10 de la razón entre el porcentaje encontrado para la genoteca y el porcentaje encontrado para el genoma. De esta manera, aquellas funciones moleculares que se encuentran más representadas en la genoteca, tendrán un valor de logaritmo positivo. Este análisis revela que en la genoteca de sustracción se encuentran más representadas funciones moleculares que participan en las maquinarias transcripcional y traduccional y menos representadas funciones moleculares que corresponden a cascadas de señalización.

Un análisis similar respecto a los procesos biológicos en los cuales participan los productos génicos codificados por los genes de la genoteca se muestra en la figura 12. En este caso, los procesos biológicos más representados son aquellos que tienen lugar en los estados iniciales del desarrollo. Los procesos menos representados son aquellos relacionados con el metabolismo, crecimiento y mantención celular y la respuesta a estímulos.



**Figura 11. Representatividad de funciones moleculares en la genoteca respecto al genoma.** El porcentaje de aparición de un término GO indica el porcentaje de proteínas que tienen el término asignado en la genoteca de sustracción (barras grises) y en la totalidad de los genes de *Drosophila* (barras azules). Los términos GO que más cambian de representatividad en esta comparación corresponden a: 1, represor transcripcional; 2, Constituyente estructural del ribosoma; 3, factor de traducción con unión a ácidos nucleicos; 4, factor de transcripción de la RNA polimerasa II; 5, Quinasa; 6, unión a proteínas; Mientras que los términos 7, hidrolasa, 8, Proteínas receptora de señales; 9, receptor; 10, regulador de GTPasa; 11, oxidoreductasa; 12, constituyente estructural del citoesqueleto; 13, receptor, permanecen invariantes o se encuentran menos representados en la genoteca. El gráfico incluye el logaritmo en base 10 de la razón entre el porcentaje obtenido para un término GO (barras naranjas) en la genoteca y el genoma. Para este análisis se utilizó el tercer estrato de clasificación.



**Figura 12. Representatividad de procesos biológicos en la genoteca respecto al genoma.** El porcentaje de aparición de un término GO indica el porcentaje de proteínas que tienen el término asignado en la genoteca de sustracción (barras grises) y en la totalidad de los genes de *Drosophila* (barras azules). Los términos GO que mas cambian de representatividad en esta comparación corresponden a: 1, especificación de patrones; 2, desarrollo embrionario; 3, desarrollo de la larva; 4, diferenciación celular; 5, secreción; 6, reproducción. Mientras que los términos 7, morfogénesis; 8, motilidad celular; 9, metabolismo; 10, crecimiento celular y/o mantenimiento; 11, respuesta a estrés; 12, respuesta a estímulos externos; muerte; 14, respuesta a estímulo endógeno; 15, muerte celular, permanecen invariantes o se encuentran menos representados en la genoteca. Para este análisis se utilizó el tercer estrato de clasificación. El gráfico incluye el logaritmo en base 10 de la razón entre el porcentaje obtenido para un término GO (barras naranjas) en la genoteca y el genoma.

#### 4. Análisis de expresión diferencial.

##### 4.1. Optimización de condiciones experimentales.

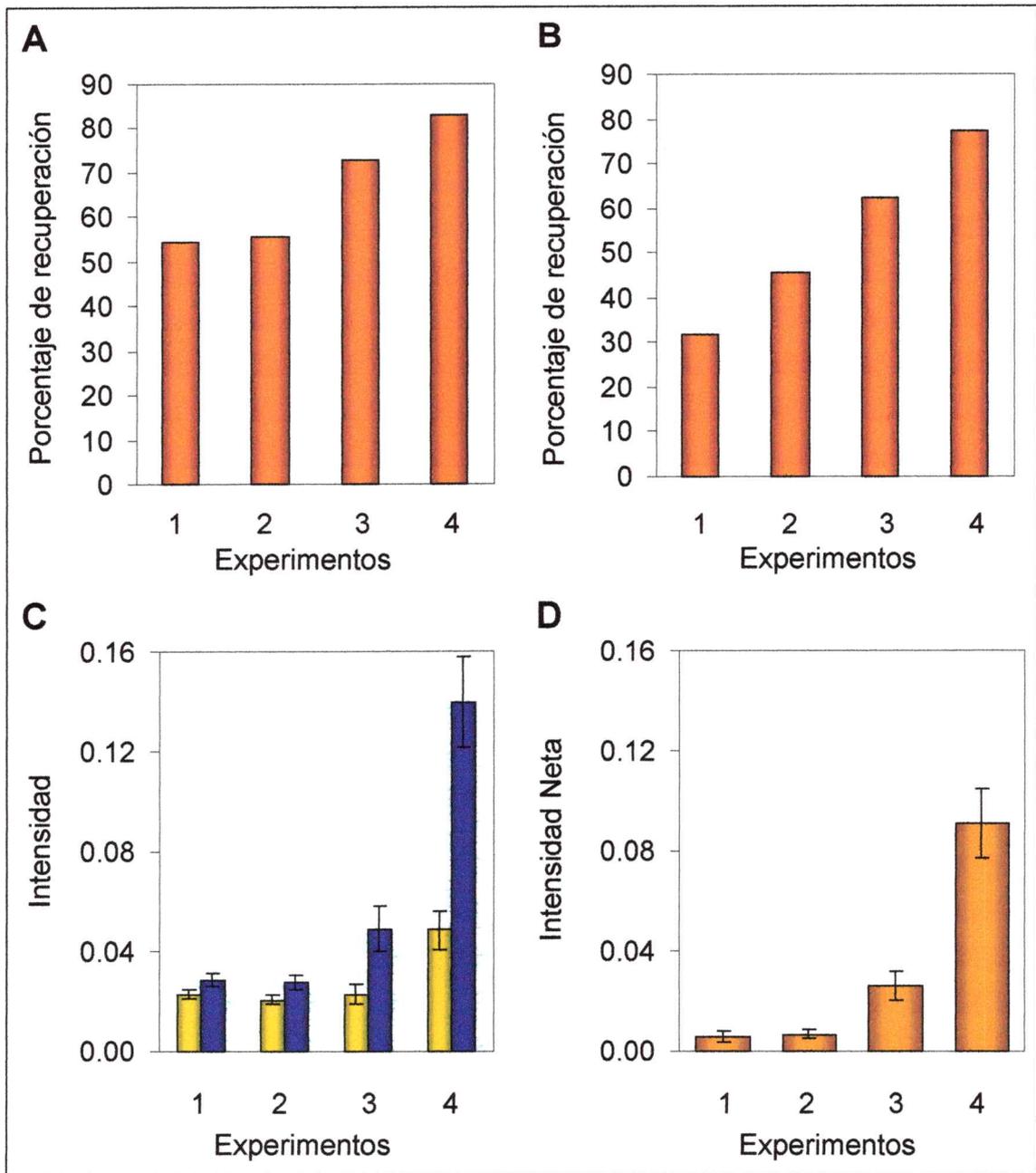
Con el objetivo de optimizar la cantidad y calidad de los datos obtenidos a partir de ensayos de hibridación en matrices, determinada por parámetros como reproducibilidad del evento de sembrado y cantidad de clones en cada membrana, se diseñaron matrices tipo II utilizando un sembrador automático.

Primero se evaluaron tres métodos para la obtención de sonda marcada: 1) la incorporación del nucleótido radiactivo al momento de sintetizar el cDNA a partir de RNA total, 2) la marcación del cDNA de doble hebra no amplificado mediante el método de *random priming*, y 3) la marcación del cDNA de doble hebra SMART mediante el método de *random priming*. Este último método reportó los mejores resultados ya que la actividad específica de esta sonda fue de 0.49  $\mu\text{Ci}/\text{ng}$  comparada con los 0.013  $\mu\text{Ci}/\text{ng}$  del primer método y los 0.26  $\mu\text{Ci}/\text{ng}$  del segundo, lo que redujo de manera significativa el tiempo de exposición aumentando la calidad de la imagen obtenida.

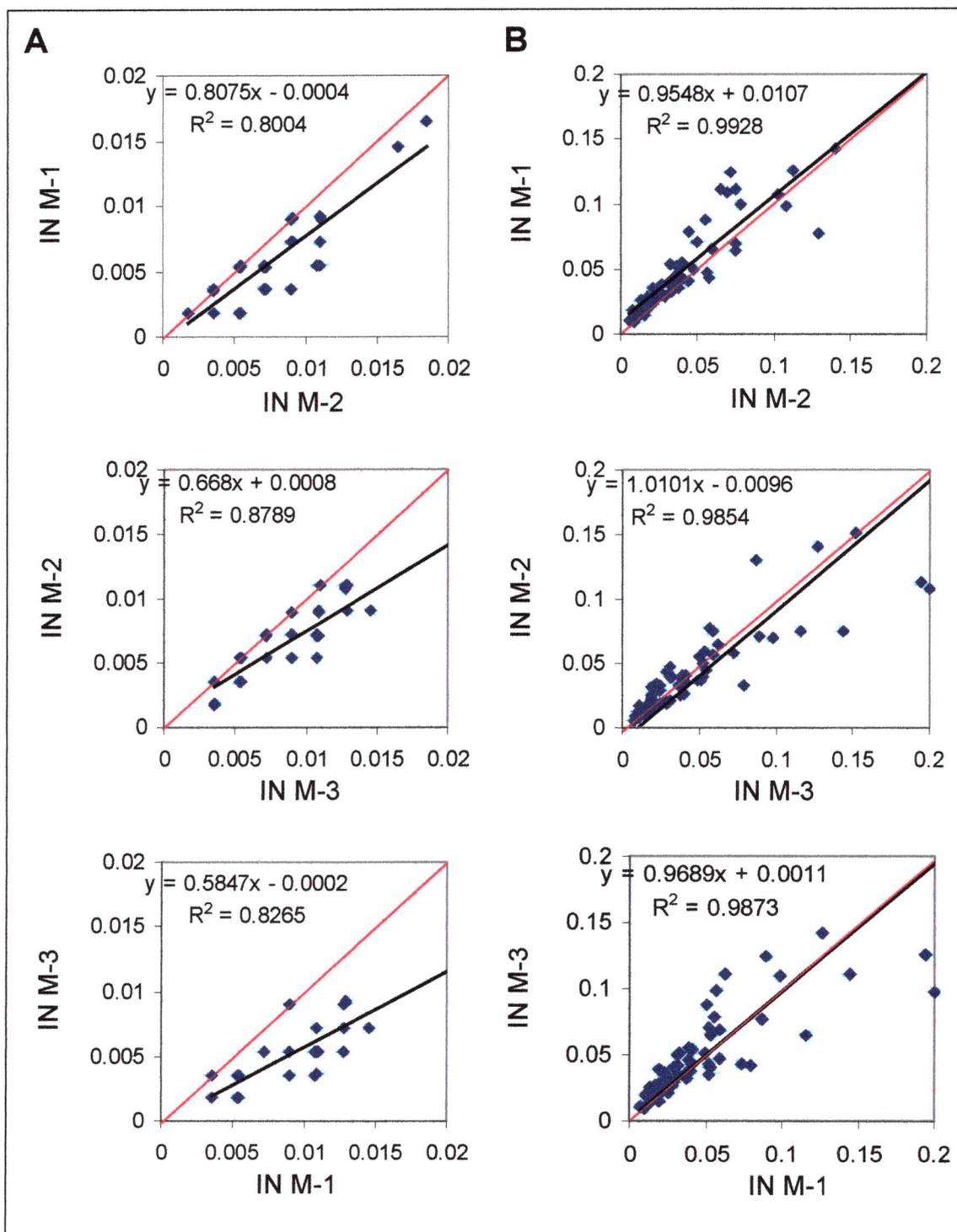
Posteriormente se evaluó la calidad de la señal que se obtiene al sembrar DNA plasmidial o producto de PCR sin purificar. Para descartar algún efecto producido por el origen de la sonda utilizada en este ensayo, se prepararon dos sondas distintas correspondientes a cDNA de doble hebra de blastodermo sincicial y cDNA de doble hebra de gástrula, las que fueron hibridadas con tres membranas sembradas con ambos tipos de DNA. Los datos de intensidad obtenidos para cada punto de siembra fueron analizados considerando dos atributos de la señal recuperada: 1) valor umbral sobre el cual la señal es confiable, y 2) índice de calidad de cada triplicado. El valor umbral particular para cada triplicado fue definido como el promedio del ruido local para

las tres replicas sembradas, más el doble de la desviación estándar de este ruido. Todo punto cuyo promedio de señal cruda fuera menor a su valor umbral fue descartado. El segundo filtro consta del calculo del índice de calidad dado por el coeficiente de variación, que corresponde a la razón entre la desviación estándar del promedio de las señales netas (señal cruda – ruido) de cada replica y este promedio (Dudoit y col., 2000). Aquellos clones cuyo coeficiente de variación fue mayor a 0.5 fueron eliminados. La figura 13 A muestra el porcentaje de recuperación que se obtiene después de aplicar el primer filtro, el valor umbral y la figura 13 B, el porcentaje de recuperación después de aplicar el segundo filtro utilizando dos sondas distintas. La figura 13 C muestra la relación entre la señal y el ruido en cada experimento que dará origen a la señal neta mostrada en la figura 13 D.

La figura 14 muestra las curvas de tendencia para las tres combinaciones de pares de membranas sembradas con DNA plasmidial y productos de PCR, todas hibridadas con cDNA obtenido de gástrula. Las curvas muestran el grado de agregación de los datos filtrados, indicado por el valor de  $r^2$  y la pendiente de la recta para cada caso. El análisis de estos indicadores nos dice que la reproducibilidad entre los triplicados es mayor al sembrar productos de PCR, ya que el valor de las pendientes varía entre 0.95 y 1.01 y el índice de correlación varía entre 0.98 y 0.99. Al sembrar DNA plasmidial en cambio, las pendientes varían entre 0.58 y 0.80 y los índices de correlación caen a valores entre 0.80 y 0.87. Este comportamiento se refleja gráficamente al comparar la línea de tendencia en cada caso (línea negra) y la línea de identidad (línea roja) que simboliza una correlación perfecta indicando una reproducibilidad óptima. Estos ensayos permitieron determinar el tipo de DNA a sembrar en los siguientes experimentos.



**Figura 13. Recuperación de señales obtenidas al sembrar DNA plasmidial o productos de PCR.** (A) y (B) muestran los porcentajes de recuperación de datos en cada experimento al aplicar el primer y segundo filtro en forma sucesiva. (C) muestra la relación señal ruido que origina la señal neta graficada en (D) luego de eliminar los datos de mala calidad. Los datos fueron obtenidos a partir de experimentos de hibridación en membranas tipo II: 1) DNA plasmidial hibridado con Bs; 2) DNA plasmidial hibridado con G; 3) productos de PCR hibridado con Bs y 4) productos de PCR hibridado con G.



**Figura 14. Reproducibilidad entre membranas.** Intensidades netas versus intensidades netas entre la membrana 1 y la membrana 2 (IN M-1 v/s IN M-2), entre la membrana 2 y la membrana 3 (IN M-2 v/s IN M-3) y entre la membrana 3 y la membrana 1 (IN M-3 v/s IN M-1) en matrices tipo II sembradas con DNA plasmidial (A) o con productos de PCR (B). Todas las membranas fueron hibridadas con cDNA de embriones en estados 6-7. En cada caso se indica la ecuación de la recta y el índice de correlación ( $r^2$ ).

El paso siguiente en esta optimización fue evaluar las matrices de hibridación tipo II respecto de su capacidad para resolver expresión diferencial de transcritos en muestras complejas de cDNA. Para esto, sembramos en sextuplicado en la misma membrana, productos de PCR de 192 clones más una serie de controles positivos y negativos (ver materiales y métodos). Estos clones fueron sembrados en duplicado y el patrón completo de siembra fue repetido tres veces en la misma membrana, las que fueron consideradas replicas experimentales o del ensayo de hibridación del evento de sembrado. Las membranas fueron hibridadas con sondas elaboradas desde cDNA SMART correspondientes a los estados de gástrula y blastodermo sincicial. Las señales de hibridación fueron adquiridas mediante el software Versarray versión 4.5.1.46 (BioRad), el cual utiliza una grilla de tamaño de celda variable para cuantificar la densidad de píxeles (intensidad cruda) y el ruido local. Se calculó el coeficiente de variación de las señales obtenidas para cada uno de los sextuplicados; los datos con coeficiente de variación menor a 0.5 fueron eliminados. Aquellos clones que pasaron el filtro de los coeficientes de variación fueron sometidos a una prueba estadística t con el propósito de evaluar si los cambios observados en la abundancia relativa de los transcritos eran significativos. El cambio en los niveles de expresión de cada transcrito fue calculado como la razón entre la señal promedio obtenida para cada clon en la hibridación con cDNA de gástrula y la señal promedio obtenida en la hibridación con cDNA de blastodermo sincicial. El resultado de este análisis revela que existen 22 genes (Tabla 4) que cambian significativamente su expresión ( $p < 0.001$ ) en gástrula con respecto a blastodermo sincicial. Las razones de cambio calculadas para estos 22 genes fluctúan entre 1.69 y 10.8 indicando que el cambio en la abundancia relativa de un transcrito en los estados probados, aún si es menor a dos veces, es detectable y

consistente ya que se repite en las 6 replicas con una baja probabilidad de que este cambio se deba al azar.

**Tabla 4. Genes cuya expresión cambia significativamente entre gástrula y blastodermo sincicial.** Los clones de estos genes fueron sembrados en sextuplicado en cada una de las membranas e hibridados con sondas correspondientes a cDNA de doble hebra perteneciente a blastodermo sincicial y gástrula, marcado mediante *random priming*. Los datos obtenidos fueron sometidos a una prueba estadística t donde n=6. La razón gástrula/blastodermo sincicial (G/Bs) indica el cambio en el nivel de expresión de cada transcrito ( $p < 0.001$ ).

CG	Símbolo	Nombre completo	razón G/Bs
CG6684	<i>RpS25</i>	<i>Ribosomal protein S25</i>	10.81
CG4440	CG4440		7.22
CG10035	CG10035		6.06
CG10758	<i>drl</i>	<i>Derailed</i>	4.93
CG16983	<i>skpA</i>		4.37
CG9704	<i>Nrt</i>	<i>Neurotactin</i>	3.80
CG10594	<i>Cyp307a1</i>		3.19
CG12487	CG12487		3.13
CG4800	CG4800		3.01
CG17117	<i>hth</i>	<i>Homothorax</i>	2.96
CG3340	<i>Kr</i>	<i>Kruppel</i>	2.91
CG10711	CG10711		2.89
CG11280	<i>trn</i>	<i>Tartan</i>	2.60
CG10218	<i>Mhcl</i>	<i>Myosin heavy chain-like</i>	2.53
CG5026	CG5026		2.49
CG9277	<i>betaTub56D</i>	<i>beta-Tubulin at 56D</i>	2.43
CG7004	<i>fwd</i>	<i>four wheel drive</i>	2.12
CG17216	<i>KP78b</i>		1.97
CG1740	<i>Ntf-2</i>		1.90
CG10307	CG10307		1.88
CG31607	CG31607		1.71
CG4500	CG4500		1.69

#### 4.2. Perfiles de expresión temporal.

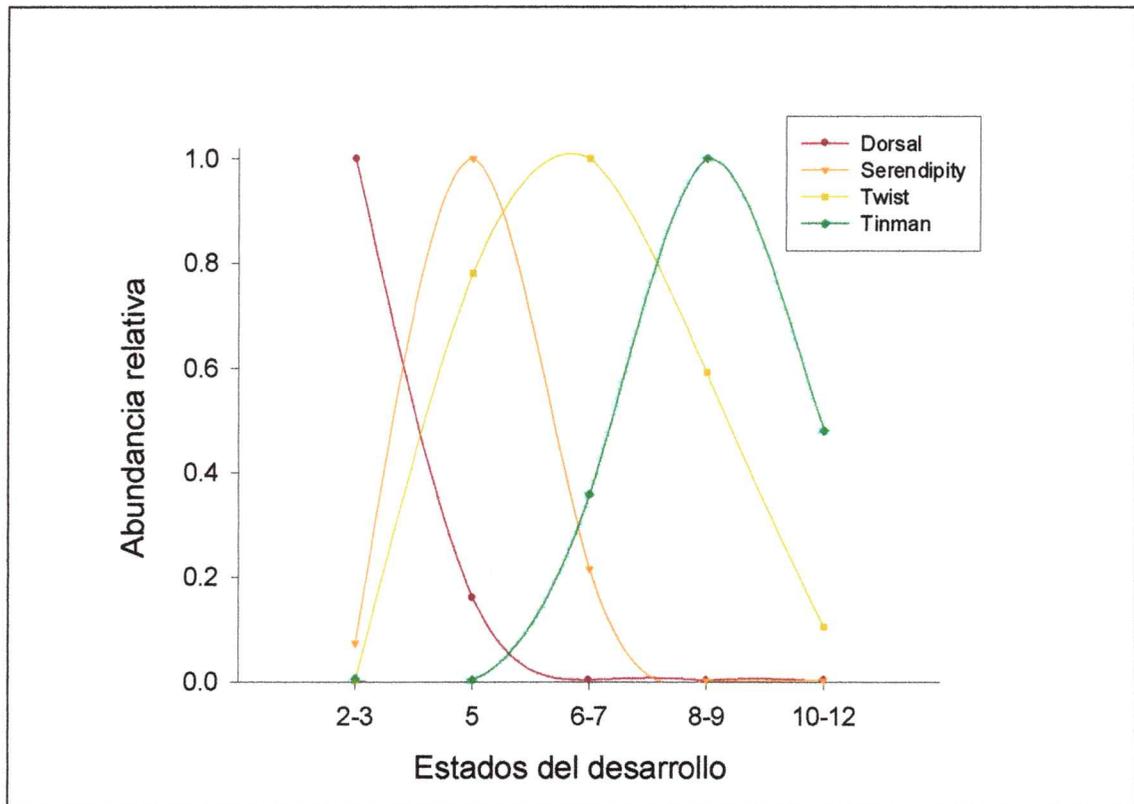
Sabemos que los genes aislados aumentan su abundancia relativa en el estado de gástrula respecto al estado de blastodermo, sin embargo no sabemos como se comportan en las etapas restantes del desarrollo embrionario. Con el objetivo de estudiar este comportamiento, se midieron las abundancias relativas de todos los clones secuenciados mediante experimentos de hibridación sobre matrices Tipo II

utilizando como sonda 5 poblaciones de cDNAs representativas de distintos intervalos de tiempo, que comprenden los estados 2 y 12 del desarrollo.

#### **4.2.1. Obtención de sondas que representan los distintos estados del desarrollo embrionario.**

Las sondas utilizadas corresponden a poblaciones de cDNA total de doble hebra sintetizado mediante el método SMART, a partir de RNA total extraído de embriones de estados 2-3, estado 5, estado 6-7, estado 8-9 y estado 10-12, los cuales fueron escogidos manualmente mediante una cuidadosa inspección visual. Con el objetivo de verificar lo acertado de la selección de los embriones, se midió la abundancia relativa de genes marcadores para cada estado mediante PCR cuantitativo en tiempo real. La abundancia de cada gen en cada muestra fue medida y normalizada respecto a la cuantificación de *actina*, un gen constitutivo. En la figura 15 vemos que cada gen presenta un valor máximo en su abundancia relativa en los estados esperados. Por ejemplo, el gen de expresión materna *dorsal* presenta un máximo en su abundancia relativa en los estados 2-3, luego de los cuales, su abundancia baja notablemente. El gen *serendipity*  $\alpha$  esta involucrado en la celularización y su abundancia relativa es mínima en los estados 2-3, aumenta hasta alcanzar un máximo en embriones de estado 5, cuando ocurre la celularización, y luego disminuye. El mismo comportamiento presentan los otros genes marcadores: *twist*, un factor de transcripción involucrado en la diferenciación de mesodermo que alcanza su máxima abundancia en embriones de estado 6-7, durante la gastrulación y *tinman*, que al igual que *twist*, es un factor de transcripción involucrado en mesodermo, pero su expresión es más tardía, en embriones de estados 8-9. Finalmente, esta figura nos indica que la selección visual de los embriones se realizó de manera correcta y que la sonda

utilizada probablemente contiene transcritos representativos de los intervalos de tiempo seleccionados.

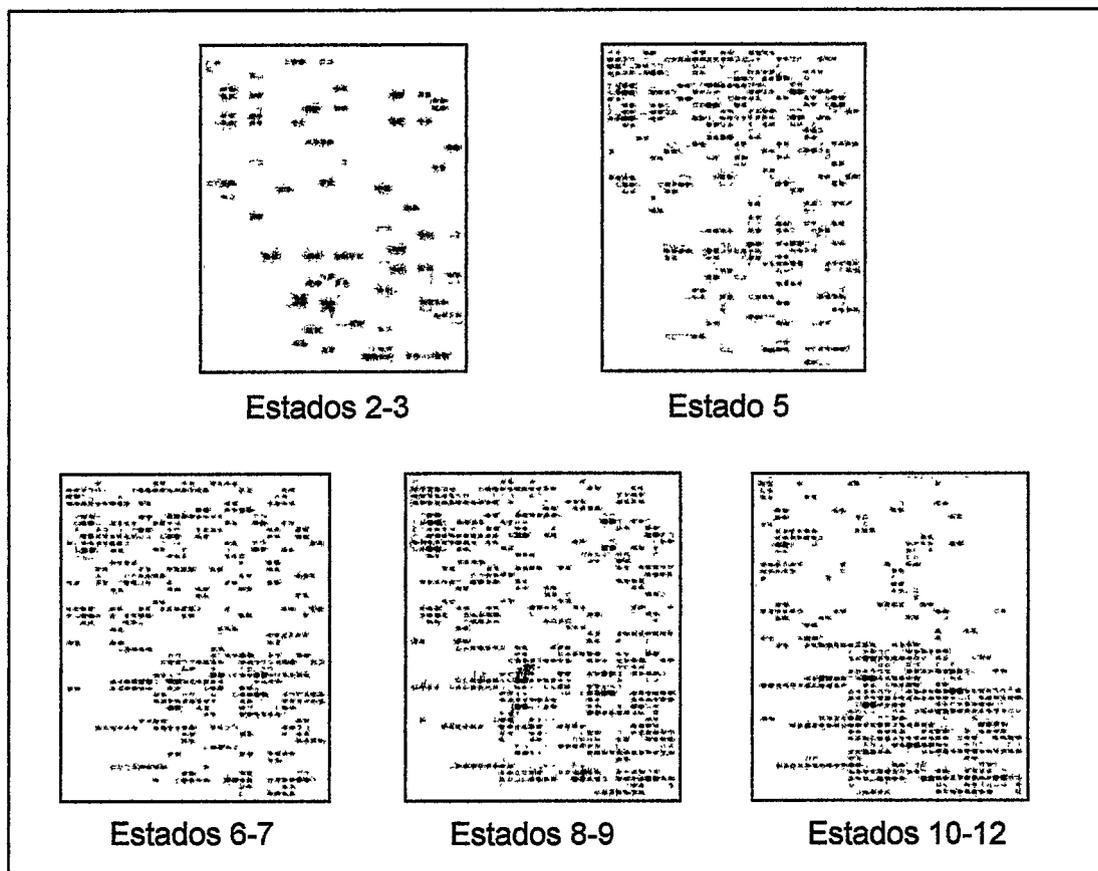


**Figura 15. Verificación de las sondas estado-específicas.** Cuantificación de la abundancia relativa de genes marcadores de 4 estados del desarrollo mediante PCR en tiempo real. Cada gen fue amplificado utilizando como sustrato los cDNAs de doble hebra que corresponden a cada sonda utilizada posteriormente en los experimentos de hibridación. Las reacciones de PCR fueron llevadas a cabo en el Light Cycler Instrument (Roche). Las concentraciones obtenidas para cada fragmento en la muestra original fueron normalizadas respecto del gen *actina*, cuya expresión no cambia durante los estados analizados. Para cada gen probado, el valor normalizado medido fue dividido por el máximo valor alcanzado en alguna de las 5 amplificaciones generando un valor de 1 para el estado en que se encuentra más expresado y valores menores a 1 en los estados restantes.

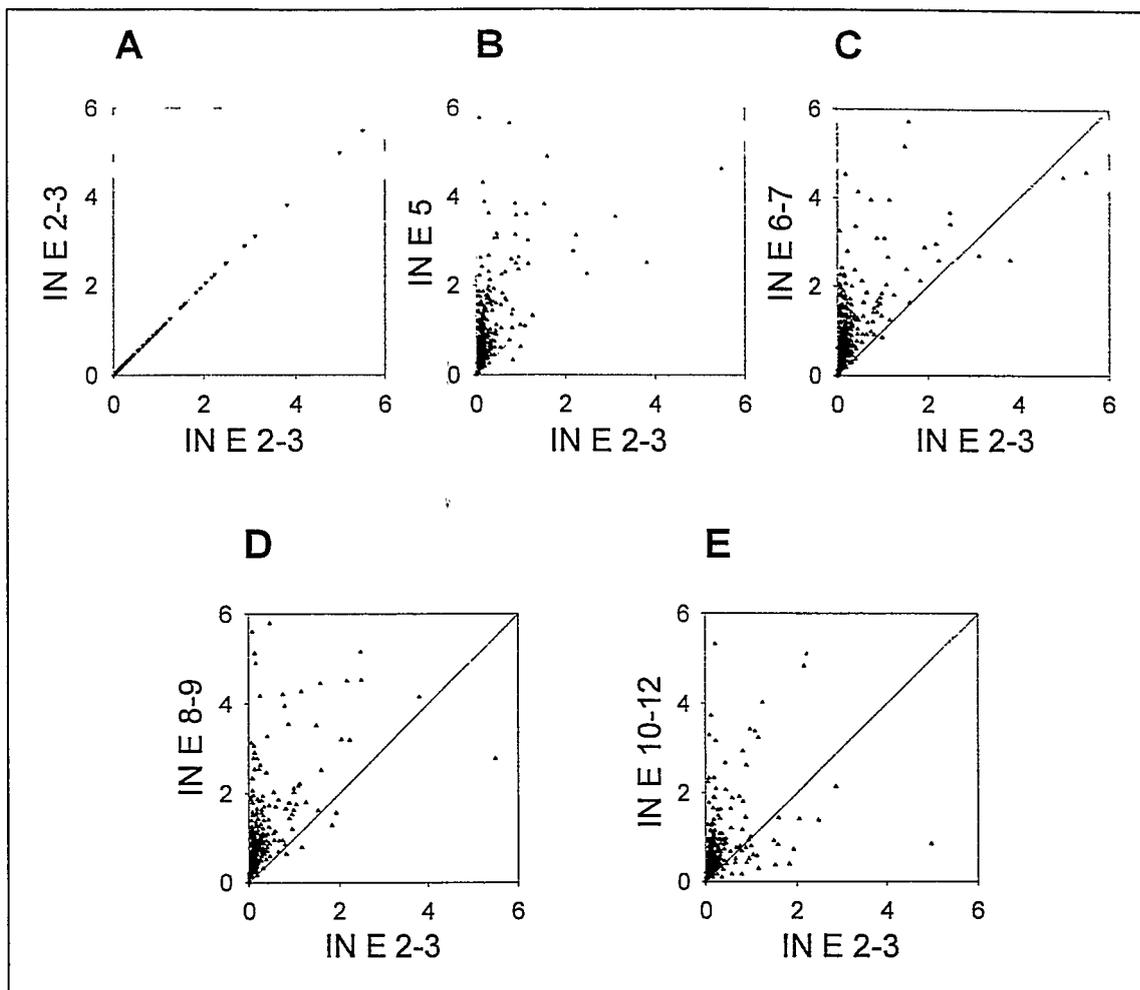
#### 4.2.2. Medición de las abundancias relativas de genes en 5 intervalos de tiempo.

Las sondas caracterizadas en el párrafo precedente fueron utilizadas para hibridar matrices sembradas con todos los clones secuenciados. Una inspección visual de las imágenes nos revela un cambio en el patrón de hibridación, ya que en

embriones de estado 5 existe un aumento en las abundancias relativas de la mayoría de los genes respecto a embriones de estados 2-3 (figura 16). Este aumento se mantiene hasta los estados 8-9, decayendo para muchos clones en embriones de estados 10-12. Este comportamiento se refleja más claramente al graficar los datos de intensidad neta normalizada extraídos desde cada una de las membranas en función de los datos extraídos desde la membrana de referencia, es decir, aquella hibridada con cDNA de estados 2-3, blastodermo sincicial (figura 17). Cuando graficamos los datos extraídos desde la referencia sobre si mismos, vemos que ellos caen sobre la línea de identidad, ya que la razón de cada valor sobre si mismo es 1 (figura 17 A). Sin embargo, al graficar los datos correspondientes al estado 5 en función de la referencia (figura 17 B), casi la totalidad de los puntos caen sobre la línea de identidad, lo que significa que la razón entre los datos es mayor que 1, revelando un aumento en la abundancia relativa de la mayoría de los transcritos. Este aumento en las abundancias relativas se mantiene hasta los estados 8-9 (figuras 17 C y D). En los estados 10-12 la abundancia relativa de un grupo de transcritos disminuye incluso cayendo por valores bajo la referencia, esto es evidente al observar la población de puntos que caen bajo la línea de identidad (figura 17 E).



**Figura 16. Hibridación en matrices con cDNAs de embriones en diferentes estados de desarrollo.** Las imágenes obtenidas con el Phosphorimager Personal FX (BioRad) a una resolución de 50  $\mu\text{m}$  por píxel, corresponden a las matrices de hibridación Tipo II sembradas en duplicado con productos de PCR de todos los clones secuenciados e hibridadas con sondas radioactivas sintetizadas a partir del RNA total extraído de embriones correspondientes a estados 2-3, estado 5, estados 6-7, estados 8-9 y estados 10-12 del desarrollo. Se sembraron además controles positivos y negativos.



**Figura 17. Cambio en la abundancia relativa de los transcritos presentes en las distintas sondas.** Los valores de intensidad para cada punto de siembra sobre las matrices mostradas en la figura 16 fueron obtenidos utilizando el software VersArray Analyzer versión 4.5.1.46 (BioRad) el cual resta el ruido local asociado a cada punto obteniendo el valor de intensidad neta. Los valores correspondientes a cada membrana fueron normalizados utilizando los valores de intensidad obtenidos para la expresión de un gen constitutivo (*rp49*). Los valores graficados corresponden al logaritmo en base dos de la Intensidad neta normalizada obtenida en cada hibridación (IN), correspondiendo a una medida de la abundancia del transcrito en la sonda usada. Los gráficos muestran las abundancias de los transcritos obtenidos en los estados 2-3 (A), estado 5 (B), estados 6-7 (C), estados 8-9 (D), estados 10-12 (E) relativas a la hibridación de referencia (estado 2-3). Si un transcrito es igual de abundante en el experimento y la referencia, la razón entre los valores es 1 y cae sobre la línea de identidad dibujada; si su abundancia aumenta respecto a la referencia, la razón es mayor que 1 y el punto se ubica sobre la línea de identidad.

#### 4.2.3. Obtención de perfiles de expresión temporal.

El diseño de las matrices consistió en sembrar un clon en duplicado en una membrana y luego sembrar una replica completa de esta membrana. Ambas membranas fueron hibridadas con la misma sonda, exponiéndolas así a iguales condiciones de hibridación. Este diseño permitió obtener cuatro señales de hibridación para un mismo clon. Los datos fueron filtrados descartando las señales de hibridación menores o iguales a cero, utilizando el coeficiente de variación (Dudoit y col., 2000) y el índice de calidad  $q_{com}$  (Wang y col., 2001) según lo descrito en materiales y métodos, recuperando un 48 % de los puntos sembrados. Con el objetivo de determinar si los cambios de expresión de cada gen entre algún estado y el estado de referencia (estado 2-3) son significativos, los datos filtrados y normalizados respecto a la expresión de *rp49* en cada membrana fueron sometidos a la prueba estadística  $t$  obteniendo así 149 genes cuya expresión cambia significativamente ( $p < 0.05$ ), que corresponden al 56 % de los genes únicos. De esta manera, podemos determinar cambios en la expresión y para cada gen obtener su perfil de expresión. Los perfiles individuales se muestran en la Tabla 5. En ella los valores de expresión de cada gen fueron normalizados respecto al valor máximo alcanzado en alguno de los estados analizados, obteniendo valores entre 0 y 1, e indicando con símbolos su nivel de expresión. De esta manera, si un gen en alguno de los estados no se expresa se le asigna un (-). Si se expresa, y el valor relativo de expresión varía entre 0.01 a 0.25 se le asigna un (+), entre 0.26 a 0.5 un (++) , entre un 0.51 y un 0.75 un (+++) y entre un 0.76 y 1 se le asigna un (++++).

**Tabla 5. Perfiles de expresión de genes que codifican proteínas.** Las abundancias relativas de cada gen que presentan diferencias significativas ( $p < 0.05$ ) con respecto al estado de referencia fueron normalizadas respecto al máximo valor alcanzado en alguno de los estados medidos. De esta manera la expresión de cada gen varía entre 0 y 1, valores que fueron divididos en intervalos: 0, (-), 0.01- 0.25, (+); 0.26-0.50, (++) ; 0.51-0.75, (+++) y 0.76-1, (++++).

Gen	Símbolo	E 2-3	E 5	E 6-7	E 8-9	E 10-12
CG12919	<i>egr</i>	+	++++	-	-	+++
CG6988	<i>Pdi</i>	++	++++	-	-	++++
CG14438	CG14438	+	++++	-	+	-
CG4440	CG4440	+	++++	-	+	+++
CG1624	<i>dpld</i>	+	++++	-	++	+++
CG31163	CG31163	+	++++	-	+++	++
CG6759	<i>cdc16</i>	++	++++	-	+++	+++
CG11276	<i>RpS4</i>	+	++++	-	++++	++
CG12026	CG12026	+	++++	-	++++	-
CG4887	CG4887	++	++++	-	++++	-
CG11307	CG11307	+	++++	+	-	+
CG17045	<i>yellow-e3</i>	+	++++	+	-	-
CG17765	CG17765	++	++++	+	-	++
CG14112	CG14112	+	++++	+	+	+
CG14317	CG14317	+	++++	+	+	+
CG17957	<i>Sry-alpha</i>	+	++++	+	+	+
CG14111	CG14111	+	++++	+	+++	-
CG4027	<i>Act5C</i>	++	++++	+	+++	-
CG5170	<i>Dp1</i>	+	++++	+	+++	+
CG8280	<i>Ef1alpha48D</i>	++	++++	+	+++	+++
CG15628	CG15628	+	++++	+	++++	++
CG3195	<i>RpL12</i>	+	++++	+	++++	+++
CG3851	<i>odd</i>	+	++++	+	++++	-
CG3921	CG3921	+	++++	+	++++	++
CG8472	<i>Cam</i>	+	++++	+	++++	+
CG1515	CG1515	+	++++	++	-	++
CG10417	CG10417	+	++++	++	++	+
CG11280	<i>trn</i>	+	++++	++	+++	++
CG13849	<i>Nop56</i>	+	++++	++	+++	++
CG17216	<i>KP78b</i>	+	++++	++	+++	++
CG33200	CG33200	+	++++	++	+++	++
CG5363	<i>cdc2</i>	++	++++	++	+++	+
CG10851	<i>B52</i>	+	++++	++	++++	++
CG11825	CG11825	+	++++	++	++++	+
CG33051	CG33051	+	++++	++	++++	+
CG4759	CG4759	+	++++	++	++++	++
CG5002	CG5002	+	++++	++	++++	++
CG8654	CG8654	+	++++	++	++++	+++
CG9677	<i>Int6</i>	+	++++	++	++++	+

CG11614	<i>nkd</i>	++	++++	+++	-	-
CG32850	CG32850	+	++++	+++	-	+++
CG8049	<i>Btk29A</i>	++	++++	+++	-	++++
CG8415	CG8415	++	++++	+++	-	-
CG2916	<i>septin 05</i>	+	++++	+++	+++	++
CG9063	CG9063	+	++++	+++	++++	++++
CG1956	<i>R</i>	++	++++	++++	++++	++
CG12276	<i>Aos1</i>	++	+++	-	++++	++++
CG3249	CG3249	+	+++	-	++++	+++
CG5814	<i>CycB3</i>	+	+++	+	++++	-
CG7970	CG7970	+	+++	+	++++	+
CG8092	CG8092	++	+++	+	++++	++
CG4111	CG4111	+	+++	++	-	++++
CG9553	<i>chic</i>	+	+++	++	-	++++
CG15138	<i>beat-IIIc</i>	+	+++	++	++++	++
CG15480	CG15480	+	+++	++	++++	+++
CG1884	CG1884	+	+++	++	++++	++
CG31800	CG31800	+	+++	++	++++	+
CG32982	CG32982	+	+++	++	++++	-
CG33070	CG33070	+	+++	++	++++	++
CG3730	<i>csul</i>	+	+++	++	++++	++
CG4547	CG4547	+	+++	++	++++	+
CG5820	<i>Gp150</i>	+	+++	++	++++	++
CG6686	CG6686	+	+++	++	++++	++
CG11605	<i>jbug</i>	+	+++	+++	-	++++
CG15319	<i>nej</i>	++	+++	++++	-	-
CG13427	CG13427	+	+++	++++	+	+
CG17921	<i>HmgZ</i>	+	++	-	++++	+++
CG2915	CG2915	+	++	-	++++	-
CG6327	CG6327	+	++	-	++++	++
CG9160	<i>mtacp1</i>	+	++	-	++++	+
CG11077	CG11077	+	++	+	++++	+
CG13389	<i>RpS13</i>	+	++	+	++++	++
CG14939	CG14939	+	++	+	++++	++
CG17117	<i>hth</i>	+	++	+	++++	+
CG17489	<i>yip6</i>	+	++	+	++++	++
CG18497	<i>spen</i>	+	++	+	++++	++++
CG18640	<i>Bj1</i>	+	++	+	++++	+
CG31253	CG31253	+	++	+	++++	+
CG31607	CG31607	+	++	+	++++	+
CG5277	<i>lp259</i>	+	++	+	++++	++++
CG5481	<i>lea</i>	+	++	+	++++	+
CG6234	CG6234	+	++	+	++++	+
CG6572	<i>Rpb7</i>	+	++	+	++++	+
CG6684	<i>RpS25</i>	+	++	+	++++	-
CG7005	<i>Esp</i>	+	++	+	++++	+++

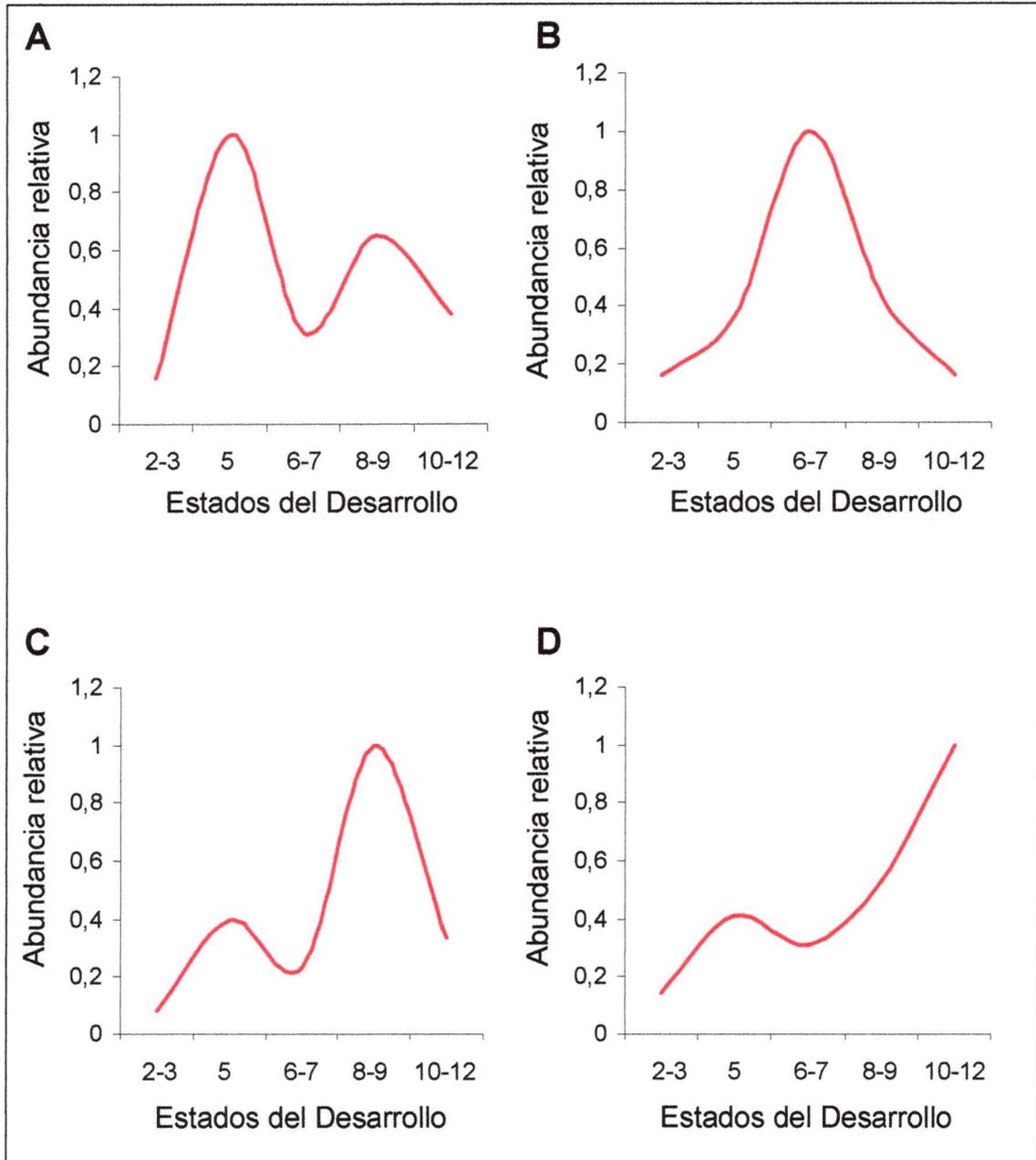
CG7590	<i>scylla</i>	+	++	+	++++	++
CG7892	<i>nmo</i>	+	++	+	++++	+
CG8050	<i>Cys</i>	+	++	+	++++	++
CG9771	<i>Dip2</i>	+	++	+	++++	+
CG11901	<i>Ef1gamma</i>	++	++	++	-	++++
CG1448	<i>inx3</i>	+	++	++	++++	++
CG3241	<i>msh-2</i>	+	++	++	++++	+
CG5059	CG5059	+	++	++	++++	++
CG7956	CG7956	+	++	++	++++	-
l(3)S118602a	l(3)S118602a	+	++	++	++++	++++
CG10307	CG10307	+	++	+++	++++	+
CG1913	<i>alphaTub84B</i>	++	++	+++	++++	+
CG10374	<i>Lsd-1</i>	++	++	++++	+	+
CG11212	<i>Ptr</i>	+	++	++++	++	+
CG4148	l(2)35Ea	+	++	++++	++	+
CG5448	<i>Syx1A</i>	++	++	++++	+++	-
CG13333	CG13333	+	++	++++	++++	++
CG1347	CG1347	+	+	-	++++	+
CG1677	CG1677	+	+	-	++++	-
CG16971	CG16971	+	+	-	++++	-
CG4494	<i>smt3</i>	++	+	-	++++	+
CG5138	<i>Ocho</i>	+	+	-	++++	+
CG1028	<i>Antp</i>	+	+	+	-	++++
CG18623	<i>Rtnl1</i>	+	+	+	++	++++
CG8776	CG8776	+	+	+	++	++++
CG1114	<i>Hph</i>	+	+	+	++++	++
CG12487	CG12487	+	+	+	++++	+
CG12582	CG12582	+	+	+	++++	-
CG15009	<i>ImpL2</i>	+	+	+	++++	++
CG17521	<i>Qm</i>	+	+	+	++++	+
CG18485	<i>stumps</i>	+	+	+	++++	++
CG2198	<i>Ama</i>	+	+	+	++++	+
CG2207	<i>Df31</i>	+	+	+	++++	+
CG2835	<i>G-salpa60A</i>	+	+	+	++++	++
CG32043	CG32043	+	+	+	++++	+
CG4463	<i>Hsp23</i>	+	+	+	++++	+
CG5185	<i>Tom</i>	+	+	+	++++	++
CG5486	<i>Ubp64E</i>	+	+	+	++++	+
CG7769	<i>DDB1</i>	+	+	+	++++	++
CG9696	<i>dom</i>	+	+	+	++++	+
CG9704	<i>Nrt</i>	+	+	+	++++	+++
CG9983	<i>Hrb98DE</i>	+	+	+	++++	+
CG40138	CG40138	+	+	++	++++	+++
CG9629	CG9629	+	+	+++	++++	+++
CG31131	CG31131	+	+	++++	+	+
CG2125	<i>ci</i>	+	+	++++	++	+

CG8846	<i>Thor</i>	+	+	++++	++	+
CG8960	CG8960	+	+	++++	++	+
CG12817	CG12817	+	-	-	++	++++
CG10218	<i>Mhcl</i>	+	-	-	++++	++
CG11092	CG11092	+	-	-	++++	+++
CG31739	CG31739	+	-	-	++++	-
CG16738	<i>slp1</i>	+	-	+	++++	+
CG2252	<i>fs(1)h</i>	+	-	+	++++	++
CG32386	<i>corn</i>	+	-	+	++++	++
CG6510	<i>RpL18A</i>	+	-	+	++++	+++
CG7071	CG7071	+	-	+	++++	++
CG11956	<i>SP1029</i>	+	-	++	++++	+
CG1225	<i>RhoGEF3</i>	+	-	++++	-	-
CG6395	<i>Csp</i>	++	-	++++	+	++
CG12369	<i>Lac</i>	+	-	++++	++	+
CG15095	<i>I(2)08717</i>	+	-	++++	++	+
CG5684	CG5684	++	-	++++	++	+
CG1358	CG1358	+	-	++++	++++	++

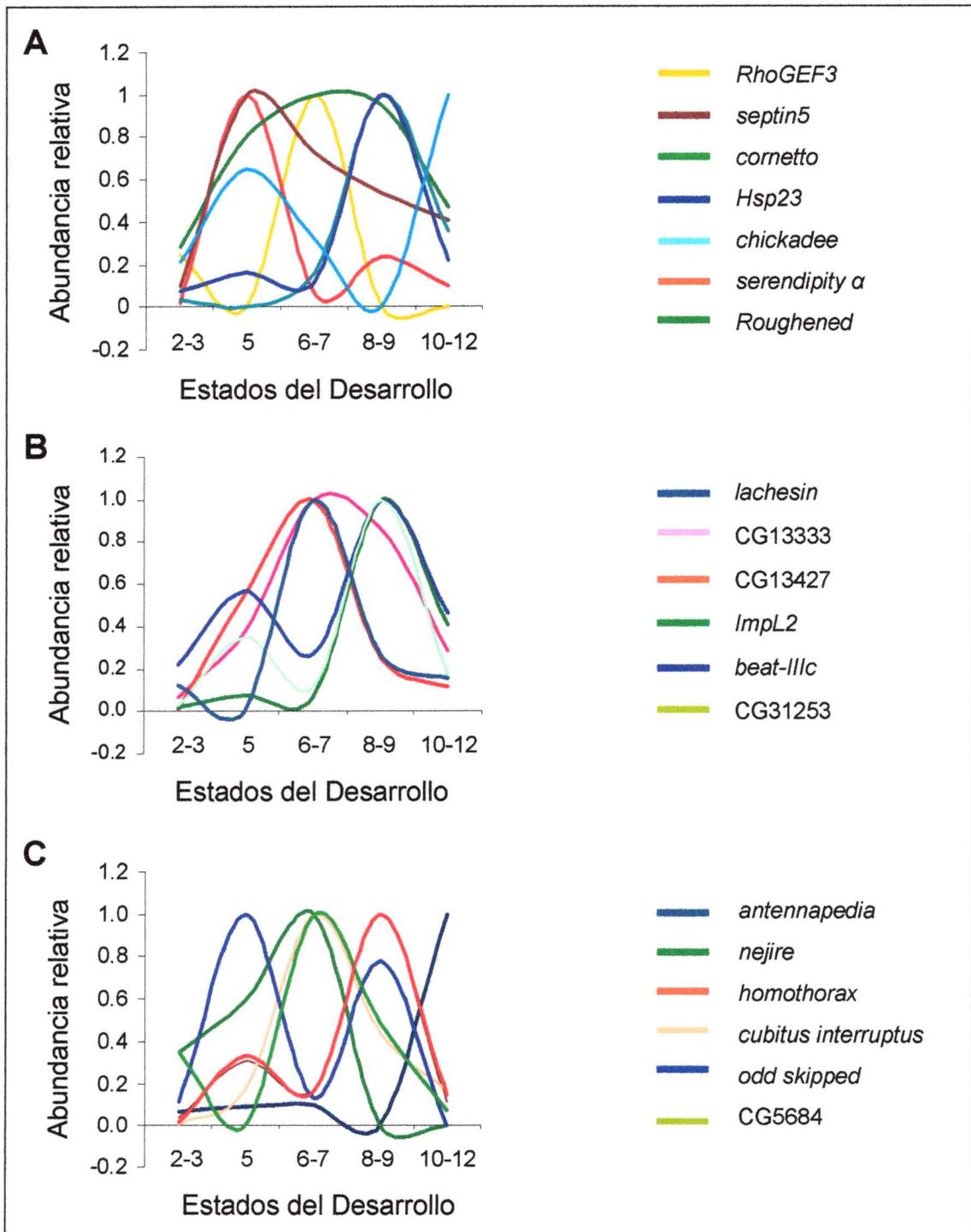
El comportamiento de este grupo de genes puede ser clasificado de acuerdo con el estado del desarrollo en que presentan su máximo nivel de abundancia relativa. La figura 18 muestra los perfiles de expresión promedio para grupos de genes, que presentan un máximo de expresión durante el estado 5 (figura 18 A), durante los estados 6-7 (figura 18 B), durante los estados 8-9 (figura 18 C) y durante los estados 10-12 (figura 18 D). La figura 19 muestra algunos de los perfiles individuales obtenidos para genes agrupados según sus características funcionales. Estos tres grupos corresponden a genes involucrados en la regulación del citoesqueleto (figura 19 A), genes que codifican proteínas secretadas e involucradas en la señalización extracelular (figura 19 B), y genes que codifican proteínas involucradas en la regulación de la transcripción, factores de transcripción y remodeladores de la cromatina (figura 19 C).

Estos ensayos confirman la expresión diferencial de estos genes en el estado de gastrulación respecto al estado de blastodermo sincicial. Además, sus perfiles de

expresión nos indican que algunos comparten comportamientos característicos durante los primeros estados del desarrollo sugiriendo que los distintos grupos obedecen a sistemas de regulación particulares y distintivos.



**Figura 18. Perfiles de expresión promedio.** Promedio de los perfiles de expresión de genes que presentan un máximo de expresión en estado 5 (A), estados 6-7 (B), estados 8-9 (C) y estados 10-12 (D).



**Figura 19. Perfiles de expresión de grupos funcionales.** Perfiles de expresión individuales de genes que comparten las funciones de regulación del citoesqueleto (A), señalización extracelular (B), y regulación de la transcripción (C).

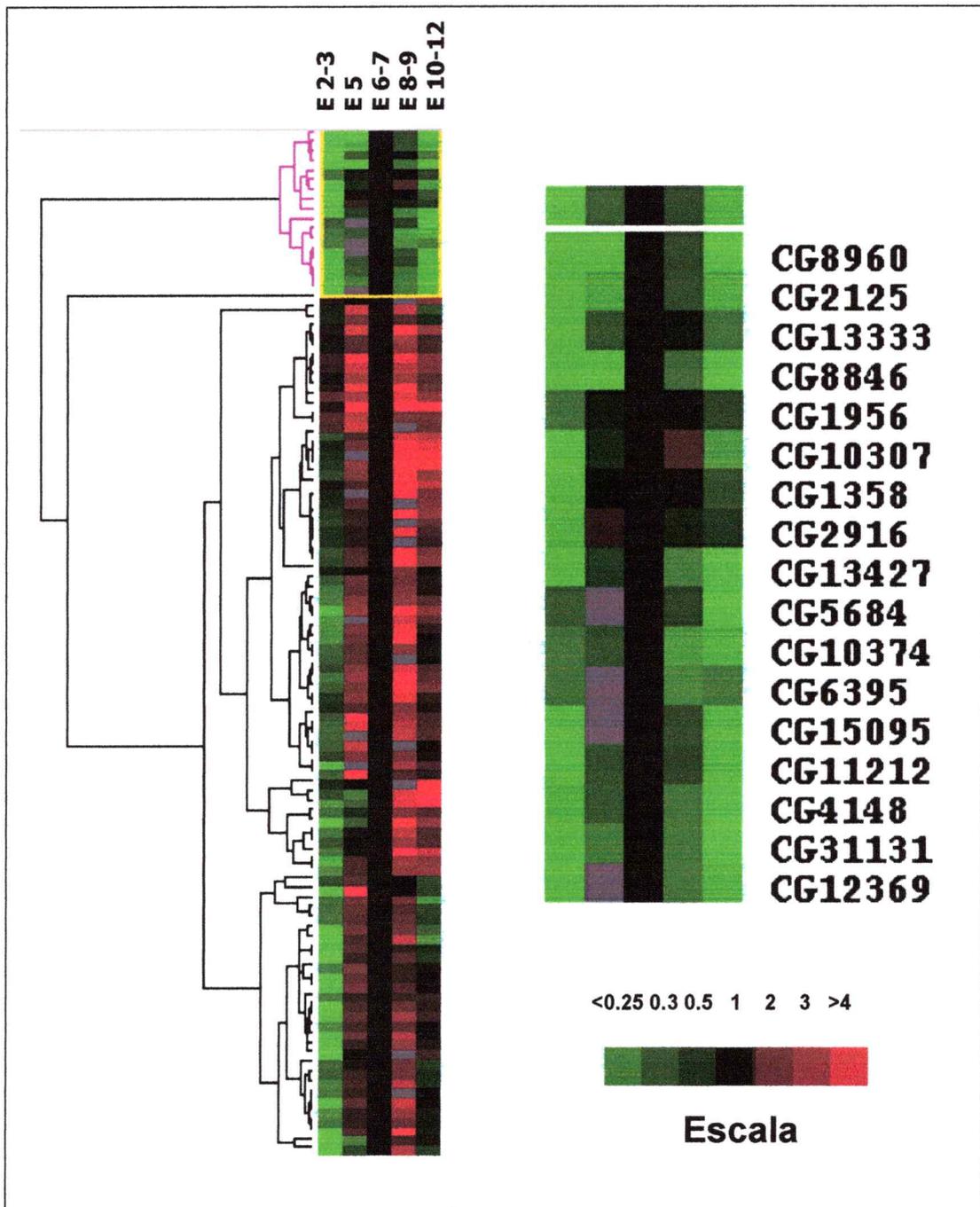
Además de los genes que codifican para proteínas, otros clones muestran cambios significativos de expresión. La Tabla 6 muestra los perfiles de expresión de 34 clones correspondientes al 54% de aquellos que alinearon con secuencias intergénicas e intrones, además de 4 clones que corresponden a RNAs no traducidos. Estos resultados confirman la existencia de genes no anotados dentro del genoma y además indica que estos genes presentan un comportamiento diferencial durante los estados del desarrollo analizados sugiriendo que ellos son regulados transcripcionalmente.

**Tabla 6. Perfiles de expresión de clones no anotados y RNAs no traducidos.** Las abundancias relativas de cada clon que presentan diferencias significativas ( $p < 0.05$ ) con respecto al estado de referencia fueron normalizadas respecto al máximo valor alcanzado en alguno de los estados medidos. De esta manera la expresión de cada clon varía entre 0 y 1, valores que fueron divididos en intervalos: 0, (-), 0.01- 0.25, (+); 0.26-0.50, (++); 0.51-0.75, (+++) y 0.76-1, (++++).

Característica	E 2-3	E 5	E 6-7	E8-9	E10-12
Intrón	+	++++	+++	+++	+
Región intergénica	+++	++++	+	++	+
Región intergénica	+	++++	++	+++	++
Región intergénica	+	++	++++	++	+++
Región intergénica	+	+	++++	++	++
Región intergénica	+	+	++++	+	+
Región Intergénica	+	+	++++	+	+
Región intergénica	+	+	++++	+	+
Región intergénica	+	+	++++	+	+
Intrón	++	+	++++	+	+
Región intergénica	+	+++	++	++++	++
Región intergénica	++	++++	++	++++	++++
Intrón	+	+	++	++++	+
Intrón	+	++++	++	++++	++++
Región intergénica	+	++	+	++++	++++
Región intergénica	+	++	+	++++	++
Región intergénica	+	+++	+	++++	++
Intrón	+	++	+	++++	++++
Intrón	+	++	+	++++	++
Región intergénica	+	++	+	++++	+++
Región intergénica	+	++	+	++++	++
Región intergénica	+	+	+	++++	++
Intrón	+	+	+	++++	++
Intrón	+	+	+	++++	+

Región intergénica	+	+	+	++++	++
Región intergénica	+	++++	+	++++	++
Región intergénica	+	+++	+	++++	+
Región intergénica	++	+++	+	++++	+
Intrón	+	+	+	++++	++
Región intergénica	+	++	+	++++	++++
Región intergénica	+	+	+	+++	++++
Región intergénica	+	++++	+	+++	++++
Intrón	+	+++	++	+	++++
Región intergénica	+	+	++	+	++++
RNA no traducido	+	++	+	++++	++
RNA no traducido	+	+	++	++++	+++
RNA no traducido	+	+	++	+++	++++
RNA no traducido	+	++	++	++++	++++

Con el objetivo de inferir relaciones funcionales entre los genes que cambian significativamente su nivel de abundancia relativa, se realizó un análisis de agrupamiento (Cluster jerárquico) el que se basa en el supuesto de que genes que comparten perfiles de expresión estarían co-regulados y serían requeridos en conjunto para cumplir alguna tarea específica dentro de la célula. Este análisis permite agrupar aquellos genes cuyos perfiles de expresión a lo largo de los estados comparados se comportan de manera similar. Los resultados se muestran en la figura 20, en donde vemos un dendrograma que resulta del cálculo de la razón entre el nivel de expresión de cada gen en cada estado y el nivel de expresión en el estado de referencia y en el que cada fila corresponde a un gen y cada columna a un estado. Los colores representan los niveles de expresión relativos al estado de referencia (estado 6-7). En la figura se destaca un grupo de genes que corresponde a los 17 genes que cambian significativamente su abundancia relativa y que alcanzan su máxima expresión durante los estados 6-7.



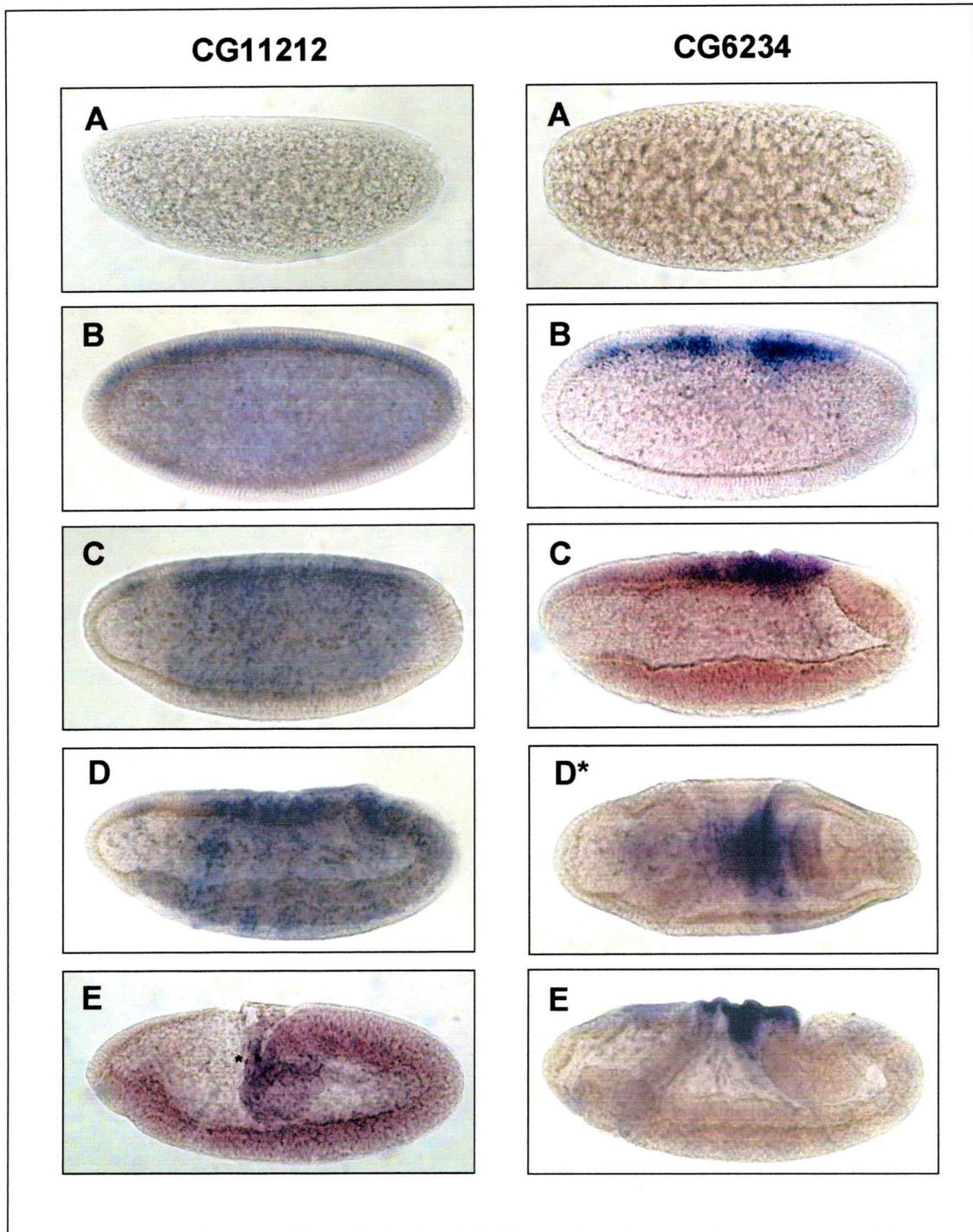
**Figura 20. Agrupamiento de genes de acuerdo a sus perfiles de expresión.** Se realizó un análisis de cluster jerárquico con aquellos clones que cambian significativamente de expresión en los estados analizados. Las distintas ramas del dendrograma muestran las distancias entre los genes (filas) y la razón de la expresión de cada gen en cada uno de los estados analizados (columnas) respecto al estado de referencia esta representada por la escala de colores que se indica.

#### 4.4. Expresión espacio temporal.

Los patrones de expresión espacial de los genes *Ptr* (CG11212) y CG6234 fueron estudiados mediante hibridaciones *in situ* en embriones tempranos. El máximo de expresión de estos genes ocurre en los estados 6-7 y en los estados 8-9, respectivamente.

El patrón de expresión del gen CG11212 nos confirma que no hay acumulación materna puesto que el transcrito no es detectado en estado de blastodermo sincicial (estados 2-3, figura 21, A). El transcrito es detectado por primera vez en el estado 5 (figura 21, B), en la región dorso-lateral del embrión. Durante los estados 6 y 7 (figura 21, C y D respectivamente) la expresión de CG11212 es detectada en las células que forman el surco cefálico y en el primordio del ectodermo neurogénico procefálico. En embriones de estado 7, CG11212 presenta un patrón de expresión segmentado a lo largo del eje antero-posterior del embrión y finalmente su expresión disminuye significativamente en embriones de estado 8 (figura 21 E).

El patrón de expresión del gen CG6234 también nos indica que no hay acumulación materna puesto que el transcrito no es detectado en el estado de blastodermo sincicial (figura 21, A). El transcrito es detectado por primera vez en el estado 5, durante la celularización, en las células dorsales del embrión (figura 21, B). Este transcrito se encuentra fuertemente expresado durante los estados 6 a 8 (figura 21 B-E) en células que componen la amnioserosa, una membrana extraembrionaria que surge desde el grupo de células epiteliales dorsales durante la celularización. Este patrón de expresión sugiere un posible vínculo de este gen en la diferenciación de la amnioserosa.



**Figura 21. Expresión espacio-temporal de los genes CG11212 y CG6234.** Vistas laterales de montajes completos de hibridaciones *in situ* en embriones. Expresión del gen CG11212 en embriones de estado 2 (A), estado 5 (B), estado 6 (C), estado 7 (D) y estado 9 (E), orientados de izquierda a derecha desde el polo anterior al posterior. Expresión del gen CG6234 en embriones de estado 2 (A), estado 5 (B), estado 6 (C), estado 8 (D\*, vista dorsal, E). Los embriones están orientados de izquierda a derecha desde el polo anterior al posterior.

## DISCUSIÓN

La formación y diferenciación del embrión de *Drosophila* es uno de los problemas mejor estudiados de la biología del desarrollo. La genética clásica, basada en la identificación de fenotipos mutantes, ha sido de gran ayuda en la búsqueda de genes cuya expresión está directamente relacionada con los acontecimientos que ocurren durante el desarrollo temprano. Estos genes codifican principalmente componentes de cascadas de transducción de señales, es decir, receptores, quinasas, fosfatasas y factores de transcripción que definen los distintos patrones de diferenciación celular (Wieschaus, 1996). En contraste, nuestro conocimiento de los mecanismos moleculares que son responsables de traducir estos patrones en tipos celulares y tejidos diferenciados es más limitado, principalmente porque muchos de los genes que están regulados por estas cascadas de señalización y que se expresan durante la gastrulación, han escapado al análisis genético.

Una forma alternativa de abordar este problema, no limitada por el fenotipo, es la aplicación de la Genómica Funcional, cuya metodología propone analizar un genoma, basándose en la información de los patrones temporales de expresión de los genes que lo constituyen, y cuyo objetivo es inferir las funciones biológicas de los genes y las interacciones producidas entre ellos. Dado que este tipo de estudios se apoya principalmente sobre la tecnología de hibridación en matrices y la utilización de la bioinformática, exigiendo una gran cantidad de recursos, he querido acotar el problema, aplicando una técnica de la biología molecular, la hibridación sustractiva por supresión (Diatchenko y col., 1996), que permite aislar genes diferencialmente expresados en distintas condiciones. Esto hizo posible obtener un grupo discreto de genes diferencialmente expresados durante la gastrulación, sobre el cual fueron

aplicadas las herramientas utilizadas por la Genómica Funcional, disminuyendo así la cantidad de recursos necesarios para efectuar este tipo de estudios. En esta tesis me he enfocado en la identificación de estos genes y su caracterización funcional basadas en criterios de similitud de secuencias, además he avanzado en su caracterización funcional mediante el estudio de la expresión de estos genes durante los primeros 12 estados del desarrollo. Los resultados de este trabajo permiten inferir algunas funciones que desempeñan estos genes durante la gastrulación. Además, aportan evidencias experimentales, tanto sobre genes que han sido anotados y aún no han sido caracterizados, así como de secuencias que no han sido anotadas como genes.

### **1. Características de la metodología utilizada.**

En este trabajo logré obtener una población de cDNA de *Drosophila melanogaster* enriquecido en secuencias diferencialmente expresadas durante la gastrulación, mediante la utilización de una hibridación sustractiva por supresión. Entre las ventajas de esta técnica podemos enumerar la recuperación de transcritos diferencialmente expresados de baja abundancia relativa, que suelen perderse por otras metodologías y la obtención de un cDNA que puede ser amplificado y usado directamente como sonda en experimentos de hibridación (Diatchenko y col, 1996). Los embriones utilizados en este trabajo fueron seleccionados manualmente según criterios morfológicos (Campos-Ortega y Hartenstein, 1985), lo que hizo difícil la obtención de la cantidad de RNA necesarios para el procedimiento de sustracción y para la generación de sondas representativas de cada estado. El uso de la metodología SMART permitió la síntesis y la amplificación de cDNA de doble hebra, enriquecido en transcritos completos. La principal ventaja de este método radica en que el RNA, total o mensajero, es amplificado sin alterar las abundancias relativas de

los transcritos presentes en la muestra original (Chenchik y col., 1998). La capacidad del método SMART de mantener la representatividad de cada transcrito en la muestra original ha sido verificada, en algunos casos, mediante la comparación de perfiles de expresión en experimentos de *macroarrays* que utilizaron sondas de cDNA amplificado por SMART y sondas obtenidas mediante de una reacción estándar de transcripción reversa (Zhumbabayeva y col., 2001) y en otros midiendo las abundancias relativas de genes por PCR cuantitativo en tiempo real usando como sustrato el cDNA SMART y cDNA sintetizado en una reacción estándar (Seth y col., 2003). Los resultados de estos experimentos indican que la representatividad de cada transcrito se mantiene luego de emplear la metodología SMART y demuestran la aplicabilidad de esta técnica en estudios donde se compara la expresión de múltiples genes. En el presente trabajo, este punto fue verificado a través de PCR cuantitativo en tiempo real, midiendo las abundancias de genes que, de acuerdo con la literatura, estarían presentes en mayor abundancia en embriones de cuatro diferentes estados de desarrollo (figura 15). Estos experimentos permiten inferir que el cDNA mantiene las abundancias relativas presentes en la muestra original y verificar la precisión en nuestra selección de embriones estado-específicos. Estrategias de hibridación sustractiva por supresión combinadas con síntesis de cDNA de doble hebra mediante el sistema SMART han sido utilizadas con éxito en diversos estudios, tanto de investigación básica como aplicada, con el fin de aislar y clonar genes diferencialmente expresados (Liu y col., 2003; Neuhaus y col., 2003; Wang y col., 2002; Squillace y col., 2002; González-Agüero y col., 2004).

La población de cDNAs obtenida luego de la sustracción fue clonada generando una genoteca estado-específica. Los resultados de caracterización de la genoteca, que incluyen verificación mediante amplificación de genes constitutivos utilizando cDNA

sustraído y sin sustraer como sustrato, hibridación en matrices e hibridación en Northern blots virtuales de una fracción de los clones, permitieron concluir que el procedimiento de sustracción fue realizado de manera adecuada, indicando que un 64 % de los clones aumentan su abundancia relativa al menos dos veces en embriones en estado 6-7 respecto a embriones en estado 2-3. La eficiencia esperada para esta técnica es de un 60% de genes diferencialmente expresados (Diatchenko y col., 1996). El aumento de la abundancia relativa en dos veces es el valor mínimo reportado para la identificación de genes diferencialmente expresados provenientes de una sustracción (Hadjiargyrou y col., 2002; King y col., 2003). Luego de secuenciar 588 clones (Tabla 1), nuestros análisis indican que estos corresponden a 268 genes únicos, de los cuales, solo el 33% está representado por más de un clon. De ellos, solo el 5,59% está representado por más de 5 clones. Esta baja ocurrencia de clones muy repetidos es esperable ya que el efecto de supresión por PCR posterior a la hibridación implica una ecualización de las abundancias relativas de todos los genes rescatados (Diatchenko y col., 1996).

Finalmente, entre las ventajas de la metodología utilizada en esta tesis podemos destacar que permite rescatar transcritos poco abundantes que se pierden en otros ensayos, como por ejemplo la hibridación sustractiva tradicional, y la ventaja de aislar genes que se expresan tempranamente sin sesgar por el dominio de expresión espacial.

## 2. Composición de la genoteca.

La tabla 1 resume los resultados del alineamiento mediante BLASTN contra la base de datos de los transcritos anotados por el proyecto genoma de *Drosophila* (BDGP, Berkeley Drosophila Genome Project). De los 588 clones secuenciados, solo 490 clones alinean significativamente ( $E > 10^{-11}$ ) con 283 genes anotados de los cuales 268 codifican para proteínas. De estos, solo 107 se encuentran bien caracterizados, principalmente a través del análisis fenotípico de embriones mutantes y en algunos casos mediante ensayos bioquímicos de su actividad. 23 genes están sólo parcialmente caracterizados, ellos poseen funciones asignadas sobre la base de su similitud de secuencia con genes previamente descritos o bien presentan dominios proteicos conservados que pueden ser relacionados con funciones particulares, sin embargo carecen de estudios funcionales *in vivo* o *in vitro*. 138 genes aislados aún no han sido caracterizados, ellos han sido anotados mediante algoritmos computacionales que incorporan criterios tales como: similitud de secuencias tanto con genes descritos como con genes putativos presentes en otras especies, alineamiento con ESTs (Expressed Sequence Tag) o presencia de atributos estructurales como: marcos de lectura abiertos de longitud determinada, señales de remoción de intrones, o distancias a regiones promotoras. Finalmente, es importante resaltar que para los 138 genes que no han sido caracterizados, el ser aislados en este análisis contribuye a su anotación, confirmando que se trata de genes transcripcionalmente activos y además, los convierte en candidatos para estudios posteriores, ya que podrían representar algún papel dentro de la dinámica del desarrollo temprano, especialmente considerando que una fracción de estos genes poseen ortólogos en otras especies como *H. sapiens* y *C. elegans*, (Tabla 2).

Los 98 clones que no presentan similitud significativa con genes que codifican para proteínas corresponden a intrones de genes, regiones intergénicas, transposones y genes que codifican para RNAs no traducidos. La anotación del genoma de *Drosophila* busca cumplir dos tareas principales: Predicción de secuencias de transcritos y proteínas y la predicción de funciones para cada una de estas proteínas (Adams y col., 2000). Esta anotación ha sido corregida en varias oportunidades desde la primera publicación de la secuencia eucromática en el año 2000 (Adams y col., 2000). La anotación se llevó a cabo mediante la utilización de distintos softwares como Genie (Reese y col., 2000) y Genscan (Burge y Karlin, 1997), pero los algoritmos en los que estos se basan, que incluyen criterios tales como: secuencias límites de intrones y exones, porcentaje de CG y uso de los codones, además de la información disponible respecto a homología de secuencias y alineamiento con ESTs, fallan principalmente en encontrar los extremos 3' y 5' de los transcritos. Otra fuente importante a la hora de anotar un gen es la comparación con otros genes conocidos, en particular con genes de especies evolutivamente cercanas como *D. pseudoscura* o *Anopheles gambiae*. Hild y col. (2003), utilizando parámetros menos estrictos mediante el software Fgenesh (Salamov y Solovyev, 2000) anotaron 2000 genes nuevos que fueron evaluados mediante ensayos de hibridación en membranas, RT-PCR e hibridaciones *in situ*, lo que nos indica que toda anotación de un gen debe concluir con la demostración de la presencia de su mensajero. Si bien la comparación de secuencias y la anotación computacional son herramientas útiles, es necesario contar con datos experimentales que demuestren la existencia de un gen, su organización y estructura. En este sentido, este trabajo aporta evidencias para 34 de los 63 clones anotados como regiones intergénicas e intrones, ya que estos se encuentran sobre-expresados en gastrulación y, además, muestran cambios significativos de expresión

en algún estado del desarrollo respecto al estado de referencia (Tabla 6). Una de las proyecciones de este trabajo es profundizar en el análisis de la estructura de estas secuencias a través de métodos alternativos de anotación, el análisis experimental de posibles isoformas generadas por procesamiento alternativo y la identificación de las secuencias completas de los transcritos de estos genes.

En relación con las secuencias que codifican para RNAs no traducidos, en nuestra genoteca aislamos transcritos que codifican para Heat shock RNA  $\omega$  (Hsp  $\omega$ ) expresado en el embrión desde el estado de blastodermo celular y para roX1, el cual se expresa desde las 2 horas de desarrollo embrionario (Meller, 2003) y está involucrado en el fenómeno de transcripción exacerbada (dosage compensation) del cromosoma X en los machos (Akhtar, 2003). Por otra parte, familias de retrotransposones han sido descritas en *Drosophila* y su expresión espacio-temporal durante la embriogénesis ha sido analizada (Ding y Lipshitz, 1994). Estos estudios indican que aunque los sitios de inserción cromosómica de estas familias varían entre distintas cepas silvestres, los patrones de expresión son conservados dentro de estas cepas. Esto se debería a que elementos regulatorios en Cis presentes en los mismos retrotransposones dirigen su propia expresión interactuando con factores de transcripción presentes en el embrión exhibiendo una expresión diferencial. Es debido a este fenómeno que encontramos elementos de transposición diferencialmente expresados durante el desarrollo. La contribución de estos elementos ha sido de carácter evolutivo, ya que un evento de escisión del elemento transponible puede dejar a un gen bajo el control transcripcional de los elementos en Cis del transposón (Ding y Lipshitz, 1994).

### 3. Caracterización de los genes aislados.

Gran parte de la información individual de cada gen fue obtenida mediante el uso de la base de datos de Gene Ontology, la que proporciona un vocabulario controlado, preciso y ordenado de manera jerárquica que permite describir los productos génicos sobre la base de su función molecular, proceso biológico en el que está involucrado y/o su localización celular (Gene Ontology Consortium, 2001). Sin embargo, debido a la estructura y al diseño de esta base de datos, la clasificación de un producto génico por Gene Ontology es engorrosa. Esto es consecuencia de que, en general, un producto génico posee más de una función molecular asignada debido a que la mayoría de las proteínas poseen varios dominios funcionales, y además el mismo producto génico puede participar en varios procesos biológicos. Esto resulta en la asignación de más de un término GO en cada ontología. Un ejemplo de esto es el producto del gen *fs(1)h* (CG2252) que en la ontología de función molecular tiene asignados los términos unión a DNA (GO:0003677) y actividad de proteína quinasa (GO:0004672), o el gen CG3921 que en la ontología de procesos biológicos tiene asignados los términos de metabolismo de aminoácidos (GO:0006520) y metabolismo de proteínas (GO:0019538) además de comunicación celular (GO:0007154) y señalización célula-célula (GO:0007267) entre otros. Otra característica inherente al diseño y la manera en que se ingresa la información de la base de datos es que muchos términos GO específicos pueden estar conectados a más de una jerarquía o rama de clasificación. Por esto al trabajar con los estratos superiores de clasificación, he preferido reflejar las frecuencias de aparición de las funciones moleculares o los procesos biológicos presentes en la genoteca y así ver cuantos genes comparten un término GO, en lugar de clasificar y encasillar cada gen de acuerdo a uno solo de los términos GO que tiene asignados, el cual tendría que ser escogido de manera

arbitraria, agregando un sesgo al análisis. Gene Ontology se nutre de los nuevos descubrimientos, por lo que está en constante crecimiento lo que implica que aunque es una gran ayuda a la hora de caracterizar productos génicos, aún es una base de datos que entrega una información parcial.

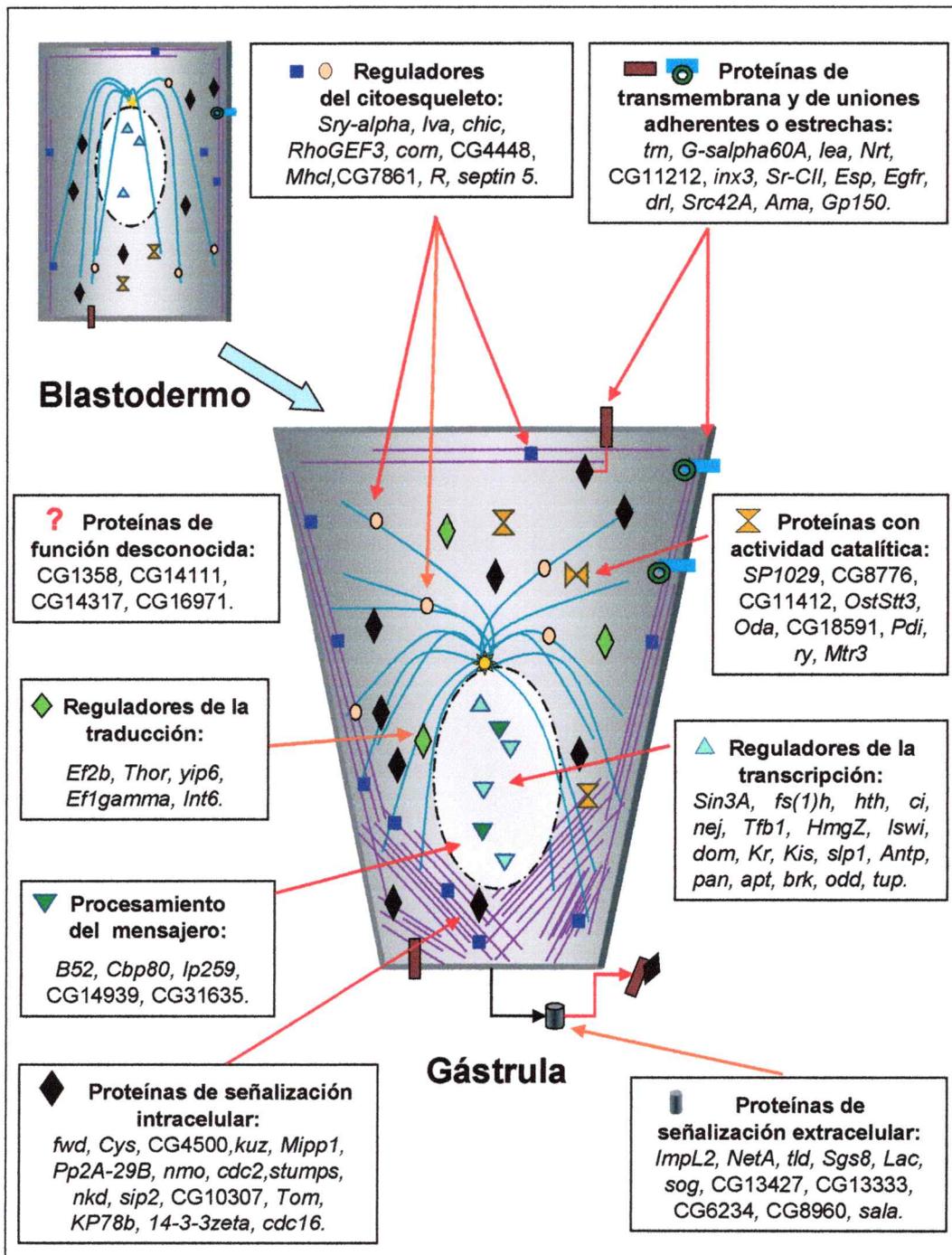
Teniendo esto en consideración, los análisis presentados aquí nos sirven para dar una mirada general a la constitución funcional de la genoteca vista desde la perspectiva de cada ontología. La comparación de la representatividad de cada término GO entre la genoteca y el genoma completo nos indica un enriquecimiento en la genoteca de proteínas con funciones moleculares involucradas en la transcripción y traducción, lo que es esperable, ya que en este periodo las proteínas y transcritos suplidos por la madre, están siendo degradados y las células deben asumir el control transcripcional y sintetizar sus proteínas de forma autónoma (Merrill y col., 1988; Wieschaus y Sweeton, 1988). Las funciones moleculares menos representadas respecto al genoma son aquellas involucradas en transducción de señales. Esto podría explicarse considerando el gran número de cascadas de señalización que se conocen operando durante el transcurso del ciclo de vida en *Drosophila*, desde esta perspectiva, este es un grupo de proteínas de mucho interés que ha sido ampliamente estudiado y por lo tanto estos términos GO pueden encontrarse sobre-representados en la porción de genes del genoma que tienen términos GO asignados. Sin embargo, considerando los procesos biológicos en los cuales participan los genes, la genoteca se muestra como un subconjunto de genes específicos que operan en morfogénesis y especificación de patrones además de migración y diferenciación celular en desmedro de procesos necesarios para mantener el estado estable de las células. En síntesis, tanto el análisis de las funciones moleculares como de los procesos biológicos sugieren que las células se encuentran en un estado dinámico en donde la mayoría de

los transcritos sintetizados son los encargados de llevar a cabo procesos morfogénicos.

#### 4. Genes de importancia en el desarrollo.

Entre los genes aislados en la genoteca encontramos algunos previamente descritos, entre ellos: 1) Reguladores transcripcionales, *odd skipped*, *sloppy paired*, *brinker* y *homothorax*, con funciones conocidas durante el desarrollo embrionario temprano (Coulter y col., 1990; Grossniklaus y col., 1992; Zhang y col., 2001; Inbal y col., 2001); 2) componentes de vías de transducción de señales, por ejemplo: miembros de la vía de señalización de Torso, responsable de la diferenciación de las estructuras posteriores del embrión. Estos genes, que se expresan en el ectodermo dorsal entre los estados 6-10 del desarrollo, incluyen *tolloid*, una endopeptidasa, *tailup*, un factor de transcripción, *short gastrulation*, *sog*, un factor de crecimiento; 3) reguladores del citoesqueleto *chickadee*, una profilina de *Drosophila* (Verheyen y Cooley, 1994), *lava lamp* una proteína que interacciona con microtúbulos y microfilamentos y que es requerida para la celularización del embrión (Sisson y col., 2000), *serendipity- $\alpha$*  una proteína de unión a microfilamentos, necesaria para la reorganización del citoesqueleto de actina durante la celularización (Ibnsouda y col., 1993) y *RhoGEF3*, un putativo intercambiador de nucleótidos de Rho que se expresa entre los estados 5-13 del desarrollo embrionario (Hicks y col., 2001).

Estos y otros genes que comparten características funcionales, han sido agrupados esquemáticamente en 9 categorías (figura 22) con el propósito de ilustrar los resultados obtenidos en la identificación y caracterización de los genes aislados a partir de los análisis presentados en esta tesis.



**Figura 22. Grupos funcionales encontrados en la genoteca.** La figura muestra el cambio morfológico que sufren las células al pasar del estado de blastodermo celular al de gástrula, Dentro de cada grupo funcional existen genes previamente caracterizados que cumplen funciones conocidas durante el desarrollo y otros parcialmente caracterizados sobre la base de sus secuencias y/o patrones de expresión espacio-temporal. Todos los genes fueron aislados en este trabajo.

Los 42 genes que componen la Tabla 3 corresponden tanto a genes caracterizados como no caracterizados, que fueron aislados en esta genoteca y en otros estudios utilizando, independientemente, técnicas de hibridación sustractiva e hibridaciones en matrices. En algunos casos se usaron, exitosamente, hibridación sustractiva de mutantes contra cepas silvestres, (Casal y Leptin, 1996), o ensayos de matrices de hibridación usando RNA de mutantes (Furlong y col., 2001) con el objetivo de encontrar genes específicos de mesodermo bajo el control transcripcional de *twist*. Entre los genes de función conocida que fueron aislados por estos autores se encuentran: *stumps* (CG18485), que codifica para un componente de la vía de señalización de FGF (Fibroblast Growth Factor) y esta relacionado con la diferenciación del mesodermo (Imam y col., 1999). El gen *tailup* (CG10619), codifica para un factor de transcripción con dominio homeobox que regula la transcripción de genes involucrados en la vía de señalización de Torso. Mutantes de *tailup* fallan en retraer la banda germinal luego de la gastrulación (Strecker y col., 1991). El gen *nejire* (CG15319), codifica un co-regulador de la transcripción involucrado en varias vías de señalización (Chen y col., 2000). Entre los genes que no han sido caracterizados se encuentran: CG4500, CG8776, CG32372, CG33200 (*ventrally-expressed-protein-D*) y CG40138. Los tres primeros codifican proteínas con dominios funcionales conservados y poseen ortólogos tanto en *H. sapiens*, como en *M. musculus*, y *C. elegans*. Por otra parte, las proteínas codificadas por los genes CG33200 y CG40138 no presentan dominios conocidos ni tienen ortólogos en otros organismos. En el caso de CG33200, el patrón de expresión de su transcrito es exclusivamente ventral durante los estados 5-10 del desarrollo (Casal y Leptin, 1996).

Recientemente, Stathopoulos y col. (2002), utilizaron hibridación en matrices contra sondas representativas del RNA de mutantes que carecen del eje dorso-ventral,

para identificar genes regulados transcripcionalmente por la proteína materna Dorsal, responsable de la formación de este eje en el embrión. Entre los genes aislados en nuestra genoteca y que aumentan su expresión en mutantes que desarrollan solo estructuras ventrales, debido a altas concentraciones de Dorsal en el núcleo, encontramos nuevamente a los genes CG4500, *stumps* (CG18485) y el gen *rosy* (CG7642), que codifica para una proteína involucrada en la pigmentación de los ojos (Reaume y col., 1991). Entre los genes expresados en mutantes dorsalizados que poseen bajas concentraciones de Dorsal en el núcleo encontramos los genes *tailup* (CG10619) y *eiger* (CG12919), que codifica para un receptor de factor de necrosis tumoral, involucrado en la apoptosis (Igaki y col., 2002). Estas evidencias indican que estos genes están directa o indirectamente regulados por Dorsal y probablemente cumplen una función durante la diferenciación de estructuras dorso-ventrales (Stathopoulos y col., 2002).

Los genes identificados en nuestra genoteca fueron también comparados con aquellos reportados en un análisis de la expresión de 534 genes de *Drosophila* en 66 intervalos de tiempo desde el inicio de la embriogénesis hasta 30 días de vida del adulto (Arbeitman y col., 2002). A pesar de las diferencias en nuestro procedimiento de colecta de embriones y en la técnica de hibridación utilizada, encontramos que 35 de los genes de la genoteca fueron también identificados por Arbeitman y col. (2002) como genes cigóticos de expresión temprana. En su análisis, estos autores detectaron altos niveles de expresión transitoria de genes cigóticos exclusivamente durante la celularización para un pequeño número de genes (21 de 534). Nuestro procedimiento sin embargo, basado en una selección más cuidadosa de embriones en estados precisos del desarrollo, nos permitió detectar genes con máximos de expresión en el estado de gástrula (Tabla 5). Nuestros resultados y los de Arbeitman y col., (2002)

confirman que los cambios de expresión génica son un requisito esencial en este proceso y plantean el desafío de dilucidar cual es el papel que desempeñan cada uno de los genes diferencialmente expresado durante la gastrulación.

## **5. Expresión temporal.**

La tecnología de hibridación en matrices es una de las técnicas con mayores proyecciones en todo el ámbito biotecnológico surgida en los últimos años. Su versatilidad y su bajo costo relativo a la cantidad de información que puede aportar son sus principales ventajas. En el desarrollo de esta tesis, ha sido una herramienta importante, tanto para verificar y caracterizar la genoteca como para la medición de la expresión de los genes en los estados del desarrollo estudiados. Si bien la literatura es abundante en cuanto a las aplicaciones de esta técnica, también son abundantes las distintas formas de extracción, tratamiento y análisis de los datos obtenidos. Los protocolos de extracción y tratamiento de datos utilizados en esta tesis fueron implementados paulatinamente de acuerdo con las problemáticas a resolver y las preguntas que surgieron.

La hibridación en matrices Tipo I fue apropiada para la caracterización de la genoteca. Con un costo mas bajo, su principal desventaja es el reducido número de clones que se pueden sembrar. Otra de las desventajas de esta técnica es la alta variabilidad respecto a cada evento de sembrado, que fue disminuida mediante el uso de un sembrador automático. El uso de este instrumento nos permitió diseñar nuevas matrices aumentando el numero de clones sembrados y el numero de replicas.

El mayor número de replicas nos permitió a su vez, aumentar la estrictez de los análisis con el objetivo de otorgarle una confianza estadística a los resultados obtenidos. Aunque el número de puntos de hibridación rescatados luego de descartar

aquellos puntos de mala calidad, es menor en experimentos de hibridación en membranas Tipo II respecto a las de Tipo I, estos datos son mas confiables y poseen una mayor resolución, ya que detectamos cambios de expresión menores a dos veces pero consistentes en 6 replicas ( $p < 0.01$ , tabla 4).

Esta mayor reproducibilidad del evento de sembrado nos permitió diseñar un experimento en el cual todos los clones secuenciados fueron hibridados con 5 sondas distintas representativas de 5 intervalos de tiempo de desarrollo. En este experimento, cada clon fue sembrado en duplicado y cada sonda fue hibridada con dos matrices de hibridación, con lo que se obtuvo un  $n = 4$  para cada clon sembrado, con lo que se realizo una prueba estadística t. En este análisis se rescatan los perfiles de expresión de cada gen individual, posibilitando el análisis de su expresión con una mayor confianza estadística. 149 genes cambian significativamente en al menos un estado respecto al estado de referencia, correspondiente a blastodermos sinciciales. He agrupado los genes de acuerdo al estado del desarrollo en que su expresión es máxima dentro de los estados probados, con el objetivo de tener una visión global de la expresión del conjunto de genes (tabla 5, figura 18). En general, observamos que un grupo de 38 genes alcanza su expresión máxima en el estado 5, durante la celularización, y un grupo de 17 genes cuyo máximo de expresión ocurre durante la gastrulación. Los restantes 94 genes tienen sus máximos de expresión en estados posteriores del desarrollo, es decir, estos genes se encienden al inicio de la gastrulación pero su expresión se incrementa en el tiempo analizado. El análisis de agrupamiento (cluster) nos indica que los 17 genes cuya expresión alcanza un máximo durante la gastrulación además comparten un comportamiento similar durante los estados probados sugiriendo que están sujetos a un sistema de regulación común, sin embargo la confirmación de este supuesto debe ser verificada a través de ensayos que

permitan identificar secuencias comunes de regulación en las regiones promotoras de los genes que constituyen este grupo y su actividad *in vivo*.

Una mirada más detallada de la expresión génica durante el desarrollo temprano se logró al agrupar genes que comparten funciones relacionadas. La figura 19 muestra los perfiles de expresión individual de los genes que componen algunos de los grupos esquematizados en la figura 22. Entre ellos destacamos: 1) el grupo de genes reguladores del citoesqueleto (figura 19 A), en el cual encontramos genes previamente mencionados, *serendipity α* y *chickadee*, además de *septin5*, que presentan un máximo de expresión en el estado 5. Todos ellos están involucrados en la reorganización del citoesqueleto de actina durante la celularización. Este grupo está compuesto además por los genes *Roughened (R)* y *RhoGEF3*, que codifican para una GTPasa y un intercambiador de nucleótidos de Rho (Asha y col., 1999; Hicks y col., 2001). Estos genes alcanzan su máxima expresión durante la gastrulación y ambos son potenciales reguladores de la dinámica del citoesqueleto de actina. Los genes *cornetto* y *Hsp23* se expresan durante la diferenciación de los neuroblastos (Bulgheresi y col., 2001; Haass y col., 1990) y coherente con ello, ambos genes presentan un aumento de su expresión en los estados 8-9, cuando comienza la diferenciación del sistema nervioso; 2) el grupo de genes relacionados con señalización extracelular (figura 19 B). Este grupo está formado principalmente por genes que codifican presuntas proteínas de secreción y se distingue por que carece de componentes que alcancen un máximo de expresión durante la celularización, sugiriendo que la función de estas proteínas sería necesaria luego de la formación de los tabiques celulares. Los genes *CG13427* y *CG13333* que alcanzan una expresión máxima durante la gastrulación presentan dominios proteicos de péptido señal y su distribución espacio temporal ha sido recientemente caracterizada (González-Agüero y col., 2004). Los

genes *beat-IIIc* e *ImpL2* que alcanzan un máximo de expresión durante los estados 8-9 codifican, respectivamente para una proteína de adhesión (Pipes y col., 2001) y para una proteína secretada con dominio de inmunoglobulina (Garbe y col., 1993), ambas necesarias para la diferenciación de tejido nervioso; 3) genes reguladores de la transcripción (figura 19 C), el perfil de expresión de estos genes coincide con la expresión descrita para ellos e ilustra su activación secuencial, en distintas vías de señalización, durante los estados examinados.

Este tipo de análisis nos ayuda a construir una imagen más detallada de los perfiles de expresión de estos genes que es de gran valor a la hora de hacer un análisis funcional de genes particulares. Por ejemplo, dentro del grupo de 17 genes cuyo máximo de expresión ocurre durante la gastrulación encontramos los siguientes:

1. *Roughened (R, CG1956)* codifica para rap 1, una GTPasa de la subfamilia Ras. Mutantes en este gen presentan alteraciones en la formación del surco ventral al inicio de la gastrulación, debido a anomalías en la organización del citoesqueleto (Asha y col., 1999).
2. *Syntaxin 1A (CG5448)*, codifica una proteína que participa en la regulación de la liberación de neurotransmisores (Schulze y col., 1995). Sin embargo, sus transcritos han sido identificados en etapas tempranas del desarrollo (3 horas post-fecundación, Burgess y col., 1997), antes de la diferenciación del sistema nervioso. Embriones mutantes para *Syntaxin 1A*, presentan alteraciones durante la celularización, sugiriendo que esta proteína juega un papel importante durante este proceso, y apoya la evidencia de que el tráfico de membranas es un requerimiento esencial durante la formación del blastodermo celular (Burgess y col., 1997).

3. *Csp* (cysteine string protein, CG6395) es un gen que codifica para una proteína con actividad chaperona involucrada en la secreción de neurotransmisores y es componente de las vesículas sinápticas (Umbach y col., 1994). *Csp* interacciona directamente con *Syntaxin* (Nie y col., 1999). Su expresión en estados tempranos del desarrollo no ha sido caracterizada, si bien en la base de datos BDGP Gene Expression (<http://www.fruitfly.org/cgi-bin/ex/insitu.pl>) se registra un aumento en la abundancia del transcrito de *Csp* desde el estado 5 del desarrollo.
4. *cubitus interruptus* (CG2125) codifica para un regulador transcripcional que activa o reprime a los genes blanco de la actividad de *hedgehog* (Motzny y Holmgren, 1995). Alcanza su máximo de expresión entre los estados 6-12 del desarrollo (<http://www.fruitfly.org/cgi-bin/ex/insitu.pl>). Su expresión es ubicua en el embrión hasta el estado de extensión de la banda germinal, ocurrida durante el estado 8 (Orenic y col., 1990)
5. *nejire* (CG15319) codifica para un co-activador de *cubitus interruptus* (Akimaru y col., 1997; Chen y col., 2000) induciendo la expresión de *patched*, *decapentaplejico* y *wingless*. *nejire* también sería necesario para la activación transcripcional de *twist* mediada por Dorsal (Akimaru y col., 1997b).
6. *Ptr* (*patched related*, CG11212) codifica para una proteína similar a *patched*. Al igual que *patched*, *Ptr* codifica para un putativo receptor con 11 dominios de transmembrana. Posee, además, un dominio SSD (Sterol Sensing Domain) que parece funcionar como un dominio de regulación involucrado en procesos como el tráfico de membranas, señalización celular y citoquinesis (Kuwabara y Labouesse, 2000). El patrón de expresión espacio-temporal de este gen se encuentra resumido en la figura 21.

7. *RhoGEF3* (CG1225) codifica para una presunta proteína con actividad de intercambiador de nucleótidos de Rho. Durante la gastrulación y la extensión de la banda germinal, *RhoGEF3* exhibe un patrón de expresión segmentado expresándose subsecuentemente en células ventrales, en grupos específicos de células subepidérmicas (Hicks y col., 2001).
8. *Lachesin* (CG12369) codifica para una proteína con dominio inmunoglobulina involucrada en la diferenciación de los neuroblastos precursores del sistema nervioso (Karlstrom y col., 1993). Alcanza su nivel máximo de expresión entre los estado 6-9 del desarrollo embrionario (<http://www.fruitfly.org/cgi-bin/ex/insitu.pl>).
9. *thor* (CG8846) codifica para un regulador traduccional que responde a la cascada de señalización activada por insulina, su sobre-expresión genera una reducción del tamaño de los discos imaginales y del tamaño celular (Miron y col., 2001).
10. *Lsd-1* (*Lipid storage droplet-1*, CG10374) codifica para una proteína asociada con la membrana de las vesículas de almacenamiento de lípidos en la célula (Miura y col., 2002), se expresa principalmente en larva y pupa, sin embargo en el lapso de tiempo estudiado, presenta un incremento expresión en los estado 6-7 (Arbeitman y col., 2002).
11. CG13333 codifica para una proteína no caracterizada que posee una secuencia consenso de péptido señal. CG13333 se expresa en estados tempranos del desarrollo en el ectodermo dorsal y mesoectodermo y en el primordio del ectodermo ventral (observaciones sin publicar). Entre los estados 10 y 13 es detectado en un grupo de células precursoras neurales (Brody y col., 2002). Tanto en esta tesis como en un análisis previo (González-agüero y col., 2004),

- CG13333 presentó un cambio en su abundancia relativa entre el estado gástrula y el de blastodermo sincicial mayor a 6 veces.
12. CG5684 codifica para una proteína no caracterizada que posee un dominio CAF1, característico del componente principal del complejo que degrada los extremos poly-A de los mensajeros. CG5684 se expresa a partir del estado 5 (Arbeitman y col., 2002).
  13. *l(2)35Ea* (CG4148) codifica para un factor de transcripción con dominio de dedos de Zinc relacionado con *snail*, y regularía la diferenciación de neuroblastos (Cai y col., 2001).
  14. *l(2)08717* (CG15095) codifica para una proteína cuyo homólogo en humano es un co-transportador Na/PO<sub>4</sub>. Su expresión en estados tempranos del desarrollo no ha sido caracterizada, si bien en la base de datos BDGP Gene Expression (<http://www.fruitfly.org/cgi-bin/ex/insitu.pl>) se registra un aumento en la abundancia del transcrito en el estado 5 del desarrollo.
  15. CG8960 es un gen cuya secuencia aminoacídica presenta una secuencia consenso de péptido señal, no contiene ningún otro dominio proteico conocido y no presenta ortólogos en otras especies. En la base de datos BDGP Gene Expression (<http://www.fruitfly.org/cgi-bin/ex/insitu.pl>) se registra un aumento en la abundancia del transcrito de CG8960 en los estados 6-7 del desarrollo. Este transcrito se expresa de manera ubicua en blastodermos celulares. Su expresión persiste hasta el estado 10, localizándose en los primordios del mesodermo y del endodermo anterior.
  16. CG13427 es un gen que codifica para una proteína cuyo extremo amino terminal presenta una secuencia consenso de péptido señal, no contiene ningún otro dominio proteico conocido y no presenta ortólogos en otras

especies. Tanto en esta tesis como en un análisis previo (González-agüero y col., 2004), CG13427 presentó el mayor nivel de cambio en su abundancia relativa entre el estado gástrula y el de blastodermo sincicial (>30 veces). Los transcritos de este gen se expresan principalmente en embriones de estados 5-7 en dominios discretos de células ectodérmicas (González-Agüero y col., 2004).

17. CG31131 es un gen cuya secuencia aminoacídica no contiene ningún dominio proteico conocido y no presenta ortólogos en otras especies.

En resumen, dentro del grupo de genes que aumentan su expresión durante la gastrulación encontramos genes de funciones conocidas involucrados en la vía de diferenciación de *hedgehog*, en la regulación de la dinámica del citoesqueleto y en el tráfico de membranas. Junto con ellos, identificamos genes para los que no existe información de su función. En algunos casos, ellos poseen atributos estructurales, que sumados a sus datos de expresión, permiten inferir algunas funciones. Por ejemplo, los datos obtenidos para *Ptr* permiten postular un papel para este gen dentro de una vía de señalización intracelular, posiblemente, y sobre la base de su similitud de secuencia con la proteína Patched de *Drosophila* y de otras especies, estamos en presencia de un componente aún no descrito de la vía activada por *hedgehog*. Esta idea es reforzada mediante el examen del patrón de expresión espacio-temporal descrito en la figura 21, en donde vemos que su transcrito se expresa en el ectodermo dorsal durante los primeros estados del desarrollo. Por otro lado, los genes CG13427, CG13333 y CG8960, no poseen dominios proteicos conservados ni presentan ortólogos en otras especies, pero al parecer codifican proteínas de secreción de pequeño tamaño (105-114 kDa.) El patrón de expresión temporal y espacial de CG13333 y CG13427 ha sido

analizado en González-Agüero y col. (2004). Estos genes se expresan en una ventana temporal estrecha en regiones particulares del embrión, esta información junto con las características de las proteínas codificadas por ellos los convierte en buenos candidatos para participar en la coordinación de los movimientos celulares durante la gastrulación a través de la comunicación intercelular. Cabe destacar que no fue posible identificar las contrapartidas de estos tres genes en especies cercanas a *Drosophila melanogaster*, *Drosophila pseudobscura* y *Anopheles gambiae*, para las cuales existe información de transcritos y/o secuencias genómica, sugiriendo que estos genes serían propios de *D. melanogaster*.

La información proporcionada en este trabajo, que combina el enfoque experimental y el enfoque bioinformático, permite inferir funciones de genes aún no caracterizados, uno de los problemas surgidos en la era post-genómica, caracterizada por mostrar una desigualdad entre el número de genes y las funciones asignadas a ellos. En el caso de *Drosophila*, solo el 48% de los genes anotados posee funciones moleculares asignadas en Gene Ontology mientras que solo el 28% han sido asignados a procesos biológicos particulares. Sin duda, será necesario avanzar en la caracterización funcional de los genes identificados en este trabajo para los cuales contamos ahora con evidencia experimental de su expresión temporal. Para ello es necesario implementar ensayos bioquímicos que proporcionen información directa de su actividad en la célula y analizar el efecto de alteraciones en su expresión durante el desarrollo embrionario mediante la utilización de las múltiples herramientas disponibles para este organismo modelo con el propósito de evaluar su participación en el proceso de la gastrulación.

## CONCLUSIONES

La aplicación de la hibridación sustractiva por supresión, combinada con la clonación de los genes aislados permitió construir una genoteca enriquecida en genes de expresión diferencial en gástrula, rescatando algunos genes bien caracterizados que están relacionados a los procesos celulares y moleculares que dan inicio a la morfogénesis de embriones de *Drosophila melanogaster* y que son responsables del fenotipo exhibido por el embrión durante esta etapa del desarrollo. Por lo tanto, muchos de los genes que aún no han sido caracterizados son buenos candidatos a desempeñar funciones similares. Debido a esto, podemos decir que los genes responsables de este fenotipo son posibles de identificar e individualizar mediante el examen cuantitativo de los niveles de abundancia de sus mensajeros.

La caracterización de la genoteca utilizando la base de datos de Gene Ontology, revela que ésta contiene una mayor proporción de genes que participan en la transcripción y traducción que en el genoma. De la misma manera existe una mayor proporción de proteínas que participan en procesos biológicos de morfogénesis.

Mediante el uso de hibridaciones en matrices se determinó que el 64% de los genes de la genoteca se sobre-expresan durante el estado de gastrulación.. Mediante esta metodología se obtuvieron los perfiles de expresión de 149 genes (55%) que presentan un cambio significativo ( $p < 0.05$ ) en al menos uno de los estados del desarrollo probados respecto al estado de blastodermo celular. Estos resultados indican que la expresión de los genes aislados presenta dos tipos de comportamientos expresión transitoria o sostenida a lo largo del desarrollo embrionario.

De los genes cuyo máximo de expresión tiene lugar durante la gastrulación encontramos principalmente genes de expresión en el ectodermo. Podemos agrupar

estos genes en tres categorías: un grupo que participa en vías de señalización; otro que participa en la regulación de la dinámica del citoesqueleto y un tercer grupo que participa en el tráfico de membranas. Existen además genes que no han sido caracterizados molecularmente, pero cuyos perfiles de expresión temporal y espacial sugieren su participación en los procesos morfogénicos que tienen lugar durante la gastrulación.

La metodología utilizada aportó una gran cantidad de información relacionada a la expresión de los genes aislados, muchos de ellos no caracterizados, que combinada con la información contenida en otras bases de datos, como Gene Ontology o InterPro, nos permite proyectar estos resultados hacia la formulación de posibles redes génicas funcionales. La información entregada es relevante en temas como genómica y biología molecular y aporta un conjunto de genes que son candidatos interesantes para estudios funcionales posteriores orientados a evaluar su participación en procesos particulares del desarrollo embrionario.

Finalmente, el desarrollo y la implementación de una tecnología tan versátil como la hibridación en matrices cobran relevancia por sus proyecciones en el ámbito del desarrollo de la biotecnología en nuestro país.

## REFERENCIAS BIBLIOGRÁFICAS

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., y col. 2000. The Genome Sequence of *Drosophila melanogaster*. *Science*. **287**:2185-2195.
- Akhtar, A. 2003. Dosage compensation: an intertwined world of RNA and chromatin remodelling. *Curr. Opin. Genet. Dev.* **13**:161-169.
- Akimaru, H., Chen, Y., Dai, P., Hou, D. X., Nonaka, M., Smolik, S. M., Armstrong, S., Goodman, R. H., Ishii, S. 1997. *Drosophila* CBP is a co-activator of cubitus interruptus in hedgehog signalling. *Nature*. **386**:735-738.
- Akimaru, H., Hou, D. X., Ishii, S. 1997b. *Drosophila* CBP is required for dorsal-dependent twist gene expression. *Nat. Genet.* **17**:211-214.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
- Arbeitman, M. N., Furlong, E. E., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis R. W., White K. P. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. *Science*. **297**:2270-2275.
- Asha, H., de Rooter, N. D., Wang, M. G., Hariharan, I. K. 1999. The Rap1 GTPase functions as a regulator of morphogenesis in vivo. *EMBO J.* **18**:605-615
- Ausubel, F. 1997. Short Protocols in Molecular Biology, Third Edition. John Wiley & Sons, Inc.
- Baugh, L. R., Hill, A. A., Slonim, D. K., Brown, E. L., Hunter, C. P. 2003. Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. *Development*. **130**:889-900.
- Bernard, K., Auphan, N., Granjeaud, S., Victorero, G., Schmitt-Verhulst, A. M., Jordan, B.R., Nguyen, C. 1996. Multiplex messenger assay: simultaneous, quantitative measurement of expression of many genes in the context of T cell activation. *Nucleic Acids. Res.* **24**:1435-1442.
- Boulay, J. L., Dennefeld, C., Alberga, A., 1987. The *Drosophila* developmental gene snail encodes a protein with nucleic acid binding fingers. *Nature*. **330**:395-398.
- Brody, T., Stivers, C., Nagle, J., Odenwald, W. F. 2002. Identification of novel *Drosophila* neural precursor genes using a differential embryonic head cDNA screen. *Mech. Dev.* **113**:41-59.
- Bulgheresi, S., Kleiner, E., Knoblich, J. A. 2001. Inscuteable-dependent apical localization of the microtubule-binding protein Cornetto suggests a role in asymmetric cell division. *J. Cell. Sci.* **114**:3655-3662.

- Burge, C., Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**:78-94.
- Burgess, R. W., Deitcher, D. L., Schwarz, T. L. 1997. The synaptic protein syntaxin1 is required for cellularization of *Drosophila* embryos. *J. Cell. Biol.* **138**:861-875
- Cai, Y., Chia, W., Yang, X. 2001. A family of snail-related zinc finger proteins regulates two distinct and parallel mechanisms that mediate *Drosophila* neuroblast asymmetric divisions. *EMBO J.* **20**:1704-1714
- Campos-Ortega & Hartenstein. 1985. The Embryonic Development of *Drosophila melanogaster*. Springer-Verlag.
- Casal, J., Leptin, M. 1996. Identification of novel genes in *Drosophila* reveals the complex regulation of early gene activity in the mesoderm. *Proc. Natl. Acad. Sci. USA.* **93**:10327-32.
- Casanova, J. 1990. Pattern formation under the control of the terminal system in the *Drosophila* embryo. *Development.* **110**:621-628.
- Celniker, S. E., Rubin, G. M. 2003. The *Drosophila melanogaster* genome. *Annu. Rev. Genomics. Hum. Genet.* **4**:89-117.
- Chen, Y., Goodman, R. H., Smolik, S. M. 2000. Cubitus interruptus requires *Drosophila* CREB-binding protein to activate wingless expression in the *Drosophila* embryo. *Mol. Cell. Biol.* **20**:1616-1625.
- Chenchik, A., Zhu, Y. Y., Diatchenko, L., Li, R., Hill, J., Siebert, P. D. 1998. Generation and use of high-quality cDNA from small amounts of total RNA by SMART PCR. En: Gene Cloning and Analysis by RT-PCR. pp. 305-319. (BioTechniques Books, MA).
- Coulter, D. E., Swaykus, E. A., Beran-Koehn, M. A., Goldberg, D., Wieschaus, E., Schedl, P. 1990. Molecular analysis of odd-skipped, a zinc finger encoding segmentation gene with a novel pair-rule expression pattern. *EMBO J.* **9**:3795-3804.
- Diatchenko, L., Lau, Y. F., Campbell, A. P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E. D., Siebert, P. D. 1996. Supresión Subtractive Hybridization: A metod for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA.* **93**:6025-6030.
- Ding, D., Lipshitz, H. D. 1994. Spatially regulated expression of retrovirus-like transposons during *Drosophila melanogaster* embryogenesis. *Genet. Res.* **64**:167-181.
- Dudoit, S., Yang, Y. H., Callow, M. J., Speed, T. P. 2000. Statistical methods for identifying diferentially expressed genes in replicated cDNA microarray experiments. Technical report # 578, <http://stat-www.berkeley.edu/users/terry/zarray/Html/matt.html>

- Eisen, M. B., Spellman, P. T., Brown, P. O., Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA.* **95**:14863-14868.
- Furlong, E. E., Andersen, E. C., Null, B., White, K. P., Scott, M. P. 2001. Patterns of gene expression during *Drosophila* mesoderm development. *Science.* **293**:1629-33.
- Garbe, J. C., Yang, E., Fristrom, J. W. 1993. IMP-L2: an essential secreted immunoglobulin family member implicated in neural and ectodermal development in *Drosophila*. *Development.* **119**:1237-1250
- Gene Ontology Consortium. 2001. Creating the Gene Ontology Resource: Design and Implementation. *Genome Res.* **11**:1425-1433
- Gonzalez-Agüero, M., Zúñiga, A., Pottstock, H. G., del Pozo, T., Gonzalez, M., Cambiazo, V. 2004. Identification of genes expressed during *Drosophila* gastrulation by using subtractive hybridization. Enviado a publicación.
- Grossniklaus, U., Pearson, R. K., Gehring, W. J. 1992. The *Drosophila* sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes. Dev.* **6**:1030-1051.
- Haass, C., Klein, U., Kloetzel, P. M. 1990. Developmental expression of *Drosophila melanogaster* small heat-shock proteins. *J. Cell. Sci.* **96**:413-418.
- Hadjjargyrou, M., Lombardo, F., Zhao, S., Ahrens, W., Joo, J., Ahn, H., Jurman, M., White, D. W., Rubin, C. T. 2002. Transcriptional profiling of bone regeneration. Insight into the molecular complexity of wound repair. *J. Biol. Chem.* **277**:30177-30182. Epub 2002 Jun 07.
- Hicks, M. S., O'Leary, V., Wilkin, M., Bee, S. E., Humphries, M. J., Baron, M. 2001. DrhoGEF3 encodes a new *Drosophila* DH domain protein that exhibits a highly dynamic embryonic expression pattern. *Dev. Genes Evol.* **211**:263-267.
- Hild, M., Beckmann, B., Haas, S. A., Koch, B., Solovyev, V., Busoid, C., Fellenberg, K., Boutros, M., Vingron, M., Sauer, F., Hoheisel, J. D., Paro, R. 2003. An integrated gene annotation and transcriptional profiling approach towards the full gene content of the *Drosophila* genome. *Genome Biol.* **5**:R3. Epub 2003 Dec 22.
- Ibnsouda, S., Schweisguth, F., de Billy, G., Vincent, A. 1993. Relationship between expression of serendipity alpha and cellularisation of the *Drosophila* embryo as revealed by interspecific transformation. *Development.* **119**:471-483. Erratum in: *Development* 119:971.
- Igaki, T., Kanda, H., Yamamoto-Goto, Y., Kanuka, H., Kuranaga, E., Aigaki, T., Miura, M. 2002. Eiger, a TNF superfamily ligand that triggers the *Drosophila* JNK pathway. *EMBO J.* **21**:3009-3018

- Imam, F., Sutherland, D., Huang, W., Krasnow, M. A. 1999. *stumps*, a *Drosophila* gene required for fibroblast growth factor (FGF)-directed migrations of tracheal and mesodermal cells. *Genetics*. **152**:307-318.
- Inbal, A., Halachmi, N., Dibner, C., Frank, D., Salzberg, A. 2001. Genetic evidence for the transcriptional-activating function of Homothorax during adult fly development. *Development*. **128**:3405-3413.
- Karlstrom, R. O., Wilder, L. P., Bastiani, M. J. 1993. Lachesin: an immunoglobulin superfamily protein whose expression correlates with neurogenesis in grasshopper embryos. *Development*. **118**:509-522.
- King, M. W., Nguyen, T., Calley, J., Harty, M. W., Muzinich, M. C., Mescher, A. L., Chalfant, C., N'Cho, M., McLeaster, K., McEntire, J., Stocum, D., Smith, R. C., Neff, A. W. 2003. Identification of genes expressed during *Xenopus laevis* limb regeneration by using subtractive hybridization. *Dev. Dyn.* **226**:398-409.
- Klingler, M. 1990. The organization of the antero-posterior axis. *Semin. Cell. Biol.* **1**:151-160.
- Kuwabara, P. E., Labouesse, M. 2002. The sterol-sensing domain: multiple families, a unique role?. *Trends. Genet.* **18**:193-201.
- Lehmann, R., Tautz, D. 1994. In situ hybridization to RNA. *Methods Cell. Biol.* **44**:575-598.
- Leptin, M. 1995. DROSOPHILA GASTRULATION: From Pattern Formation to Morphogenesis. *Annu. Rev. Cell. Dev. Biol.* **11**:189-212
- Leptin, M. 1999. Gastrulation in *Drosophila*: the logic and the cellular mechanisms. *EMBO J.* **18**:3187-3192.
- Levy, L. S., Manning, J. E. 1981. Messenger RNA sequence complexity and homology in developmental stages of *Drosophila*. *Dev. Biol.* **85**:141-149.
- Liu, Y. B., Wei, Z. X., Li, L., Li, H. S., Chen, H., Li, X. W. 2003. Construction and analysis of SSH cDNA library of human vascular endothelial cells related to gastroduodenal carcinoma. *World. J. Gastroenterol.* **9**:2419-2423.
- Meller, V.H. 2003. Initiation of dosage compensation in *Drosophila* embryos depends on expression of the roX RNAs. *Mech. Dev.* **120**:759-767.
- Merrill, P. T., Sweeton, D., Wieschaus, E. 1988. Requirements for autosomal gene activity during precellular stages of *Drosophila melanogaster*. *Development*. **104**:495-509.
- Miklos, G. L., Rubin, G. M. 1996. The role of the genome project in determining gene function: insights from model organisms. *Cell.* **86**:521-529.

- Miron, M., Verdu, J., Lachance, P. E., Birnbaum, M. J., Lasko, P. F., Sonenberg, N. 2001. The translational inhibitor 4E-BP is an effector of PI(3)K/Akt signalling and cell growth in *Drosophila*. *Nat. Cell. Biol.* **3**:596-601.
- Miura, S., Gan, J. W., Brzostowski, J., Parisi, M. J., Schultz, C. J., Londos, C., Oliver, B., Kimmel, A. R. 2002. Functional conservation for lipid storage droplet association among Perilipin, ADRP, and TIP47 (PAT)-related proteins in mammals, *Drosophila*, and *Dictyostelium*. *J. Biol. Chem.* **277**:32253-32257
- Motzny, C. K., Holmgren, R. 1995. The *Drosophila cubitus interruptus* protein and its role in the wingless and hedgehog signal transduction pathways. *Mech. Dev.* **52**:137-150.
- Neuhaus, T., Lutz, C., Stier, S., Totzke, G., Gruenewald, E., Fronhoffs, S., Sachinidis, A., Vetter, H., Ko, Y. D. 2003. The use of suppression subtractive hybridization for the study of SDF-1 $\alpha$  induced gene-expression in human endothelial cells. *Mol. Cell. Probes.* **17**:245-252.
- Nie, Z., Ranjan, R., Wenniger, J. J., Hong, S. N., Bronk, P., Zinsmaier, K. E. 1999. Overexpression of cysteine-string proteins in *Drosophila* reveals interactions with syntaxin. *J. Neurosci.* **19**:10270-10279
- Nüsslein-Volhard, C. 1996. The identification of genes controlling development in flies and fishes. En: Frängsmyr, T. (Eds.). Les Prix Nobel (The Nobel Prizes) 1995. pp. 235-260. Almquist & Wiksell International, Stockholm, Sweden.
- Nusslein-Volhard, C., Lohs-Schardin, M., Sander, K., Cremer, C. 1980. A dorso-ventral shift of embryonic primordia in a new maternal-effect mutant of *Drosophila*. *Nature.* **283**:474-476.
- Orenic, T. V., Slusarski, D. C., Kroll, K. L., Holmgren, R. A. 1990. Cloning and characterization of the segment polarity gene cubitus interruptus Dominant of *Drosophila*. *Genes Dev.* **4**:1053-1067.
- Parks, S., Wieschaus, E. 1991. The *Drosophila* gastrulation gene concertina encodes a G alpha-like protein. *Cell.* **64**:447-458.
- Pipes, G. C., Lin, Q., Riley, S. E., Goodman, C. S. 2001. The Beat generation: a multigene family encoding IgSF proteins related to the Beat axon guidance molecule in *Drosophila*. *Development.* **128**:4545-4552
- Postner, M. A., Wieschaus, E. F. 1994. The nullo protein is a component of the actin-myosin network that mediates cellularization in *Drosophila melanogaster* embryos. *J. Cell. Sci.* **107**:1863-73.
- Poulson, D. F. 1950. Histogenesis, organogenesis and differentiation in the embryo of *Drosophila melanogaster* (Meigen). En: Demerec, M. (ed) *Biology of Drosophila*. pp 168-274. Wiley, New York.

- Reaume, A. G., Knecht, D. A., Chovnick, A. 1991. The rosy locus in *Drosophila melanogaster*: xanthine dehydrogenase and eye pigments. *Genetics*. **129**:1099-1109.
- Reese, M. G., Kulp, D., Tammana, H., Haussler, D. 2000. Genie-gene finding in *Drosophila melanogaster*. *Genome Res.* **10**:529-538.
- Roth, S., Stein, D., Nusslein-Volhard, C. 1989. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the *Drosophila* embryo. *Cell*. **59**:1189-202.
- Rubin, G. M., Lewis, E. B. 2000. A Brief History of *Drosophila's* Contributions to Genome Research. *Science*. **287**: 2216-2218.
- Salamov, A. A., Solovyev, V. V. 2000. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**:516-522.
- Schulze, K. L., Broadie, K., Perin, M. S, Bellen, H. J. 1995. Genetic and electrophysiological studies of *Drosophila* syntaxin-1A demonstrate its role in nonneuronal secretion and neurotransmission. *Cell*. **80**:311-320
- Schweisguth, F., Lepesant, J. A., Vincent, A. 1990. The serendipity alpha gene encodes a membrane-associated protein required for the cellularization of the *Drosophila* embryo. *Genes Dev.* **4**:922-931.
- Seth, D., Gorrell, M. D., McGuinness, P. H., Leo, M. A., Lieber, C. S., McCaughan, G. W., Haber, P. S. 2003. SMART amplification maintains representation of relative gene expression: quantitative validation by real time PCR and application to studies of alcoholic liver disease in primates. *J. Biochem. Biophys. Methods*. **55**:53-66.
- Sisson, J. C., Field, C., Ventura, R., Royou, A., Sullivan, W. 2000. Lava lamp, a novel peripheral golgi protein, is required for *Drosophila melanogaster* cellularization. *J. Cell. Biol.* **151**:905-918.
- Squillace, R. M., Chenault, D. M., Wang, E. H. 2002. Inhibition of muscle differentiation by the novel muscleblind-related protein CHCR. *Dev. Biol.* **250**:218-230.
- Stathopoulos, A., Van Drenth, M., Erives, A., Markstein, M., Levine, M. 2002. Whole-genome analysis of dorsal-ventral patterning in the *Drosophila* embryo. *Cell*. **111**:687-701.
- Strecker, T. R., Yip, M. L., Lipshitz, H. D. 1991. Zygotic genes that mediate torso receptor tyrosine kinase functions in the *Drosophila melanogaster* embryo. *Proc. Natl. Acad. Sci. USA*. **88**:5824-5828.
- Sweeton, D., Parks, S., Costa, M., Wieschaus, E. 1991. Gastrulation in *Drosophila*: the formation of the ventral furrow and posterior midgut invaginations. *Development*. **112**:775-789.

- Thisse, B., Stoetzel, C., Gorostiza-Thisse, C., Perrin-Schmitt, F. 1988. Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. *EMBO J.* **7**:2175-2183.
- Umbach, J. A., Zinsmaier, K. E., Eberle, K. K., Buchner, E., Benzer, S., Gunderson, C. B. 1994. Presynaptic dysfunction in *Drosophila* csp mutants. *Neuron.* **13**:899-907
- Verheyen, E. M., Cooley, L. 1994. Profilin mutations disrupt multiple actin-dependent processes during *Drosophila* development. *Development.* **120**:717-728.
- Wang, X., Gosh, S., Guo, S-W. 2001. Quantitative Quality control in microarray image processing and data acquisition. *Nuc. Ac. Res.***15**:e75
- Wang, X., Wu, P., Xia, M., Wu, Z., Chen, Q., Liu, F. 2002. Identification of genes enriched in rice roots of the local nitrate treatment and their expression patterns in split-root treatment. *Gene.* **297**:93-102.
- Weigel, D., Jurgens, G., Klingler, M., Jackle, H. 1990. Two gap genes mediate maternal terminal pattern information in *Drosophila*. *Science.* **248**:495-498
- Wieschaus, E. 1996. From molecular patterns to morphogenesis: The lessons from *Drosophila*. En: Frängsmyr, T. (Eds.). Les Prix Nobel (The Nobel Prizes) 1995. pp. 235-260. Almquist & Wiksell International, Stockholm, Sweden.
- Wieschaus, E., Sweeton, D. 1988. Requirements for X-linked zygotic gene activity during cellularization of early *Drosophila* embryos. *Development.* **104**:483-493.
- Zhang, H., Levine, M., Ashe, H. L. 2001. Brinker is a sequence-specific transcriptional repressor in the *Drosophila* embryo. *Genes Dev.* **15**:261-266
- Zhu, X., Mahairas, G., Illies, M., Cameron, R. A., Davidson, E. H., Etensohn, C. A. 2001. A large-scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo. *Development.* **128**:2615-2627.
- Zhumabayeva, B., Diatchenko, L., Chenchik, A., Siebert, P. D. 2001. Use of SMART-generated cDNA for gene expression studies in multiple human tumors. *Biotechniques.* **30**:158-163.