



Facultad de Ingeniería
Escuela de Ingeniería Civil en Bioinformática

Identificación de elementos transponibles
como reguladores génicos durante el
proceso de regeneración de la aleta caudal
de *Danio rerio*.

Raúl Rojas Espinoza

Profesor Tutor: Braulio Valdebenito Maturana

Profesor Informante: Gonzalo Riadi Mahias

Memoria para optar al título de Ingeniero Civil en Bioinformática

Talca-Chile

Fecha: 30/11/2021

CONSTANCIA

La Dirección del Sistema de Bibliotecas a través de su unidad de procesos técnicos certifica que el autor del siguiente trabajo de titulación ha firmado su autorización para la reproducción en forma total o parcial e ilimitada del mismo.



Talca, 2021

Agradecimientos

Al llegar a este punto me gustaría agradecer a todas las personas que de una u otra forma fueron partícipes de este camino de desarrollo profesional y personal. Primero, a la gran mayoría de profesores que tuve en la universidad, que me permitieron llegar a esta etapa, son muchos, pero me gustaría destacar a los profesores Alejandro Valdés, Hassaine Mokhtar y Felipe Besoain por además de ser grandes docentes, estar conmigo en momentos personales complicados y darme siempre una mano para dichos problemas. Al profesor Mauricio Arenas por permitirme desarrollar mi lado innovador/emprendedor y poder ayudarme a desarrollar mis habilidades blandas para poder comunicarme de una mejor manera. Y principalmente a los profesores que me inspiraron a trabajar en el área genómica y que me acompañan en este proyecto, el profesor Gonzalo Riadi y al profesor Braulio Valdebenito. Al profesor Gonzalo Riadi le debo mi interés por esta área desde que ingresé a la carrera, captando mi atención y mi interés por siempre estar nutriéndome y cuestionándome los eventos científicos que son de mi interés. Y al profesor Braulio Valdebenito por su excelente calidad humana y profesional, por sus consejos y por permitirme aprender de él y dedicarme horas y horas de su tiempo para poder llevar a cabo este proyecto y otros en los que hemos trabajado.

A mi familia que siempre estuvo ahí para cualquier cosa que necesitara. A mis padres Raúl y Ruth que sé lo orgullosos que estarán una vez que finalice este proceso, por siempre apoyarme, escucharme, aconsejarme y confiar ciegamente en mis capacidades. A mi hermano Rodrigo que siempre me celebró y estuvo conmigo en cada uno de mis logros, a mi hermana Valeska que la gran mayoría de cosas las hago pensando en ella y en su bienestar, a mi sobrina Valentina y mis ahijados Renata y Rodrigo por permitirme desconectarme muchas veces de la rutina, a mis tíos y tías por siempre creer incondicionalmente en mí, apoyarme y desearme lo mejor en todos mis desafíos, a mis primos por escucharme y compartir con ellos partidos, asados e innumerables eventos, a mis suegros por apoyarme de distintas formas durante estos años, a todos mis compañeros con los que interactué de una u otra manera durante esta aventura y que me permitieron aprender de ellos y de las distintas circunstancias que nos deparó la universidad. Finalmente, agradezco a mi pareja Victoria por ser partícipe de toda esta larga travesía de principio a fin, siendo siempre un pilar importante en momentos difíciles en donde necesitaba ayuda y también me agradezco a mí por darme siempre nuevas oportunidades cuando me caí, por querer siempre ser una mejor versión de mi mismo y por continuar progresando a mi manera a pesar de todas las adversidades que me he enfrentado.

Tabla de contenido

ÍNDICE DE FIGURAS.....	6
ÍNDICE DE TABLAS	8
ÍNDICE DE FIGURAS SUPLEMENTARIAS.....	9
ÍNDICE DE TABLAS SUPLEMENTARIAS	10
RESUMEN	11
ABSTRACT	12
1. INTRODUCCIÓN.....	13
1.1. REGENERACIÓN Y <i>DANIO RERIO</i>	13
1.2. EXPRESIÓN GÉNICA.....	19
1.3. EXPRESIÓN DIFERENCIAL DE GENES	20
1.4. REGULACIÓN DE LA EXPRESIÓN GÉNICA	20
1.4.1. ACCESIBILIDAD A LA CROMATINA	22
1.4.1.1. ATAC-SEQ.....	24
1.4.2. TRANSCRIPCIÓN.....	25
1.4.2.1. RNA-SEQ	27
1.5. ELEMENTOS REGULADORES DE TRANSCRIPCIÓN.....	29
1.6. ELEMENTOS REPETIDOS.....	31
1.6.1. ELEMENTOS TRANSPONIBLES (TES)	32
1.6.1.1. ROL DE TES EN EVENTOS REGULATORIOS	32
1.6.1.2. IDENTIFICACIÓN DE TES EXPRESADOS.....	34
2. HIPÓTESIS.....	38
3. OBJETIVO GENERAL	38
4. OBJETIVOS ESPECÍFICOS	38
5. METODOLOGÍA.....	39
5.1. MATERIALES	39
5.1.1. SET DE DATOS	39
5.1.2. HARDWARE	41
5.1.3. SOFTWARE	41

5.2.	MÉTODOS	43
5.2.1.	MÉTODOS PARA PREPARACIÓN DE ARCHIVOS	44
5.2.2.	MÉTODOS PARA OBJETIVO ESPECÍFICO N°1	45
5.2.3.	MÉTODOS PARA OBJETIVO ESPECÍFICO N°2	48
5.2.4.	MÉTODOS PARA OBJETIVO ESPECÍFICO N°3	50
6.	RESULTADOS	53
6.1.	PRE-PROCESAMIENTO DE ARCHIVOS	53
6.1.1.	ANÁLISIS DE CALIDAD	53
6.1.2.	CLASIFICACIÓN DE TES ANOTADOS	56
6.2.	RESULTADOS OBJETIVO ESPECÍFICO N°1	57
6.2.1.	CANTIDAD DE TES TRANSCRITOS	57
6.2.2.	ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE TES Y GENES	59
6.3.	RESULTADOS OBJETIVO ESPECÍFICO N°2	67
6.4.	RESULTADOS OBJETIVO ESPECÍFICO N°3	72
6.4.1.	RELACIÓN DE TES SOBRE-EXPRESADOS CON GENES	72
6.4.2.	RELACIÓN DE TES CON TFBSS CON TFS ASOCIADOS A REGENERACIÓN	82
7.	DISCUSIÓN	89
8.	CONCLUSIONES	92
9.	REFERENCIAS	94
10.	ANEXO	101
10.1.	FIGURAS SUPLEMENTARIAS	101
10.2.	TABLAS SUPLEMENTARIAS	107

Índice de Figuras

FIGURA 1. FASES DE REGENERACIÓN EN TEJIDOS.....	15
FIGURA 2. PORCENTAJE DE ELEMENTOS REPETIDOS EN <i>DANIO RERIO</i>	17
FIGURA 3. REGENERACIÓN EN ALETA CAUDAL PRODUCIDA POR AMPUTACIÓN.	19
FIGURA 4. ETAPAS DE REGULACIÓN GÉNICA EN EL NÚCLEO CELULAR EUCARIONTE.	21
FIGURA 5. ESTADOS DE ACCESIBILIDAD A LA CROMATINA.....	23
FIGURA 6. REPRESENTACIÓN DEL MECANISMO DE ACCIÓN DE TRANSPOSASA TN5 EN ATAC- SEQ.	24
FIGURA 7. ACTIVIDAD DE ISOFORMAS DE LOS FACTORES DE TRANSCRIPCIÓN.	26
FIGURA 8. METODOLOGÍA RNA-SEQ.....	28
FIGURA 9. REPRESENTACIÓN DE ELEMENTOS REGULADORES DE TRANSCRIPCIÓN.	30
FIGURA 10. MECANISMOS DE REGULACIÓN GÉNICA MEDIADOS POR TES.....	33
FIGURA 11. REGULACIÓN GÉNICA MEDIADA POR TES.....	34
FIGURA 12. TIPOS DE CLASIFICACIÓN DE LA EXPRESIÓN DE UN TE.	35
FIGURA 13. ESQUEMA GENERAL DE LA METODOLOGÍA PROPUESTA.....	52
FIGURA 14. DIAGRAMA DE VENN CORRESPONDIENTE A IDENTIFICACIÓN DE TES TRANSCRITOS POR AMBOS SOFTWARE EN TODAS LAS CONDICIONES TEMPORALES. .	58
FIGURA 15. ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE <i>SELF-EXPRESSED</i> TES EN ALETA CAUDAL DE <i>DANIO RERIO</i> , COMPARANDO LAS CONDICIONES 0DPA V/S 1DPA.	60
FIGURA 16. ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE GENES EN ALETA CAUDAL DE <i>DANIO RERIO</i> , COMPARANDO LAS CONDICIONES 0DPA V/S 1DPA.....	62
FIGURA 17. ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE <i>SELF-EXPRESSED</i> TES EN ALETA CAUDAL DE <i>DANIO RERIO</i> , COMPARANDO LAS CONDICIONES 0DPA V/S 4DPA.	63
FIGURA 18. ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE GENES EN ALETA CAUDAL DE <i>DANIO RERIO</i> , COMPARANDO LAS CONDICIONES 0DPA V/S 4DPA.....	64
FIGURA 19. ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE <i>SELF-EXPRESSED</i> TES EN ALETA CAUDAL DE <i>DANIO RERIO</i> , COMPARANDO LAS CONDICIONES 1DPA V/S 4DPA	65
FIGURA 20. ANÁLISIS DE EXPRESIÓN DIFERENCIAL DE GENES EN ALETA CAUDAL DE <i>DANIO RERIO</i> , COMPARANDO LAS CONDICIONES 1DPA V/S 4DPA.....	66

FIGURA 21. REPRESENTACIÓN PROCESO DE IDENTIFICACIÓN DE PEAKS, FOOTPRINTS Y MOTIVOS RELACIONADOS CON TES.	69
FIGURA 22. REPRESENTACIÓN DE LA EXPRESIÓN DEL TE Y DEL GEN SOBRE-EXPRESADO EN LA CONDICIÓN 0DPA VS 1DPA.	77
FIGURA 23. RED BIOLÓGICA DEL PRODUCTO PROTEICO DEL GEN <i>ANXA2A</i>	79
FIGURA 24. REPRESENTACIÓN DE LA EXPRESIÓN DEL TE Y DEL GEN SOBRE-EXPRESADO EN LA CONDICIÓN 1DPA VS 4DPA.	80
FIGURA 25. RED BIOLÓGICA DEL PRODUCTO PROTEICO DEL GEN <i>HSPA12B</i>	81
FIGURA 26. RED BIOLÓGICA DEL TF <i>GATA3</i>	87
FIGURA 27. RED BIOLÓGICA DEL TF <i>LEF1</i>	88

Índice de Tablas

TABLA 1. ESTADÍSTICAS DEL ENSAMBLE DE <i>DANIO RERIO</i> VERSIÓN GRCz11.	17
TABLA 2. LIBRERÍAS GENERADAS MEDIANTE RNA-SEQ EN EL TRABAJO DE THOMPSON Y COLABORADORES (2020).	40
TABLA 3. LIBRERÍAS GENERADAS MEDIANTE ATAC-SEQ EN EL TRABAJO DE THOMPSON Y COLABORADORES (2020).	41
TABLA 4. ESPECIFICACIONES GENERALES DE <i>EXXACT</i>	41
TABLA 5. SOFTWARE UTILIZADO.	43
TABLA 6. PAQUETES DE <i>R</i> UTILIZADOS.	43
TABLA 7. ESTADÍSTICAS FINALES PARA LIBRERÍAS RNA-SEQ.	55
TABLA 8. ESTADÍSTICAS FINALES DE LIBRERÍAS ATAC-SEQ.	55
TABLA 9. ESTADÍSTICAS DE TES ANOTADOS.	56
TABLA 10. CANTIDAD DE SELF-EXPRESSED TES EXPRESADOS EN 0DPA VS 1DPA.	61
TABLA 11. CANTIDAD DE GENES EXPRESADOS EN 0DPA VS 1DPA.	62
TABLA 12. CANTIDAD DE SELF-EXPRESSED TES EXPRESADOS EN 0DPA VS 4DPA.	63
TABLA 13. CANTIDAD DE GENES EXPRESADOS EN 0DPA VS 4DPA.	64
TABLA 14. CANTIDAD DE SELF-EXPRESSED TES EXPRESADOS EN 1DPA VS 4DPA.	65
TABLA 15. CANTIDAD DE GENES EXPRESADOS EN 1DPA VS 4DPA.	66
TABLA 16. CANTIDAD DE REGIONES ENRIQUECIDAS (PEAKS) Y SU INTERSECCIÓN CON TES.	67
TABLA 17. CANTIDAD DE SITIOS <i>FOOTPRINTS</i> Y SU INTERSECCIÓN CON TES.	68
TABLA 18. TIPOS DE TES QUE PRESENTAN MOTIVOS DE TFBSs EN 0DPA.	70
TABLA 19. TIPOS DE TES QUE PRESENTAN MOTIVOS DE TFBSs EN 1DPA.	71
TABLA 20. TIPOS DE TES QUE PRESENTAN MOTIVOS DE TFBSs EN 4DPA.	71
TABLA 21. TES SOBRE-EXPRESADOS CERCANOS A GENES SOBRE-EXPRESADOS.	73
TABLA 22. FUNCIONES BIOLÓGICAS DE GENES SOBRE-EXPRESADOS, CERCANOS A TES SOBRE-EXPRESADOS.	75
TABLA 23. MOTIVOS ENCONTRADOS ASOCIADOS A REGENERACIÓN.	84
TABLA 24. FUNCIONES BIOLÓGICAS DE TFs ASOCIADOS A MOTIVOS ENCONTRADOS.	86

Índice de Figuras Suplementarias

FIGURA SUPLEMENTARIA 1. <i>MEAN QUALITY SCORES</i> PARA LAS 11 LIBRERÍAS DE RNA-SEQ (PRE-FILTRO).....	102
FIGURA SUPLEMENTARIA 2. <i>PER SEQUENCE QUALITY SCORES</i> PARA LAS 11 LIBRERÍAS DE RNA-SEQ (PRE-FILTRO).....	102
FIGURA SUPLEMENTARIA 3. <i>MEAN QUALITY SCORES</i> PARA LAS 12 LIBRERÍAS DE ATAC-SEQ (PRE-FILTRO).	103
FIGURA SUPLEMENTARIA 4. <i>PER SEQUENCE QUALITY SCORES</i> PARA LAS 12 LIBRERÍAS DE ATAC-SEQ (PRE-FILTRO).....	103
FIGURA SUPLEMENTARIA 5. <i>MEAN QUALITY SCORES</i> PARA LAS 11 LIBRERÍAS DE RNA-SEQ (POST-FILTRO).....	104
FIGURA SUPLEMENTARIA 6. <i>PER SEQUENCE QUALITY SCORES</i> PARA LAS 11 LIBRERÍAS DE RNA-SEQ (POST-FILTRO).....	104
FIGURA SUPLEMENTARIA 7. <i>MEAN QUALITY SCORES</i> PARA LAS 12 LIBRERÍAS DE ATAC-SEQ (POST-FILTRO).....	105
FIGURA SUPLEMENTARIA 8. <i>PER SEQUENCE QUALITY SCORES</i> PARA LAS 12 LIBRERÍAS DE ATAC-SEQ (POST-FILTRO).....	105
FIGURA SUPLEMENTARIA 9. DIAGRAMA DE VENN CORRESPONDIENTE A LA CANTIDAD TOTAL DE TES TRANSCRITOS EN TODAS LAS CONDICIONES TEMPORALES.	106

Índice de Tablas Suplementarias

TABLA SUPLEMENTARIA 1. TES CON PRESENCIA DE MÁS MOTIVOS DIFERENTES POR CADA CONDICIÓN.	108
TABLA SUPLEMENTARIA 2. TES QUE PRESENTAN SITIOS DE UNIÓN A TFS RELACIONADOS A REGENERACIÓN.....	112

La regeneración es un proceso biológico en el cual se reparan células, tejidos, órganos o extremidades completas de un organismo tras sufrir una lesión en ellos. Ocurre naturalmente en todos los organismos, sin embargo, en la mayoría de ellos decae a medida que alcanzan su madurez reproductiva. Uno de los organismos modelos para estudiar la regeneración es el pez cebra (*Danio rerio*). Este organismo presenta a lo largo de toda su vida capacidades regenerativas de muchos de sus órganos, siendo una de las más estudiadas la aleta caudal. Pero ¿Cómo sucede este proceso? ¿Qué genes y/o elementos participan durante el evento regenerativo?

En particular, esta investigación se centró en identificar los principales agentes participantes del evento de regeneración de la aleta caudal de *Danio rerio*, desde un punto de vista regulatorio, es decir, los genes relacionados a regeneración y los elementos reguladores que promueven o reprimen su transcripción. En la actualidad se conoce un diverso catálogo de genes y potenciadores que se activan durante la regeneración de la aleta caudal de *Danio rerio*.

Muchos de los elementos reguladores de un genoma provienen de secuencias no codificantes y en muchos casos estas son aportados por elementos repetidos, en concreto los elementos transponibles (TEs). Diversos estudios han implicado TEs como elementos reguladores en tejidos somáticos y enfermedad. Aproximadamente la mitad del genoma de *Danio rerio* corresponde a TEs, por lo que se puede especular que en este organismo también haya regulación de genes mediada por TEs. Así, caracterizar su rol en el contexto del evento regenerativo, puede permitir identificar nuevos mecanismos moleculares asociados a regeneración. Además, tanto este como otros análisis, pueden ser considerados como puerta de entrada a potenciales terapias y/o herramientas biotecnológicas para tratar daños en órganos y tejidos de otros organismos.

Considerando todo lo anterior, se planteó la pregunta ¿los TEs participan en la regulación de la expresión génica durante la regeneración de la aleta caudal? Y si es así, ¿qué tipos de TEs participan? ¿De qué manera están potencialmente regulando la expresión de los genes asociados a regeneración?

En este trabajo se buscó analizar y explorar el rol regulatorio que podrían presentar estos elementos durante el evento regenerativo, utilizando datos de metodologías RNA-Seq, para generar un perfil transcriptómico de un momento determinado del proceso de genes y TEs cercanos y datos de ATAC-Seq para identificar regiones abiertas de cromatina donde participen y se unan al DNA factores de transcripción que regulen el evento regenerativo. Se encontraron TEs de tipo DNA y LTR que se estaban sobre-expresando posterior al proceso de amputación, en regiones intrónicas de genes relacionados a regeneración, como lo son *anxa2a* y *hspa12b*. A su vez, se identificaron un total de 184 TEs que estaban actuando como sitios de unión a 6 factores de transcripción distintos relacionados a regeneración. Estos resultados apoyan la hipótesis del rol regulatorio de estos elementos en procesos biológicos como el regenerativo, sin embargo, deben ser tomados con precaución y confirmarse con nuevos experimentos/metodologías biológicas que confirmen su rol en el evento.

Abstract

Regeneration is a biological process in which cells, tissues, organs, or complete limbs of an organism are repaired after suffering an injury to them. It occurs naturally in all organisms, however, in most organisms it declines as they reach reproductive maturity. One of the model organisms to study regeneration is the zebrafish (*Danio rerio*). This organism presents throughout its life regenerative capacities of many of its organs, one of the most studied being the caudal fin. But how does this process take place? What genes and/or elements are involved during the regenerative event?

This research focused on identifying the main agents involved in the regeneration event of the *Danio rerio* caudal fin, from a regulatory point of view, i.e., the genes related to regeneration and the regulatory elements that promote or repress their transcription. A diverse catalog of genes and enhancers that are activated during *Danio rerio* caudal fin regeneration is currently known.

Many of the regulatory elements in a genome come from non-coding sequences and in many cases these are provided by repeated elements, in particular transposable elements (TEs). Several studies have implicated TEs as regulatory elements in somatic tissues and disease. Approximately half of the *Danio rerio* genome corresponds to TEs, so it can be speculated that TE-mediated gene regulation also occurs in this organism. Thus, characterizing their role in the context of the regenerative event may allow us to identify new molecular mechanisms associated with regeneration. Furthermore, both this and other analyses can be considered as a gateway to potential therapies and/or biotechnological tools to treat damage in organs and tissues of other organisms.

Considering all of the above, the question was raised: Are TEs involved in the regulation of gene expression during caudal fin regeneration? And if so, what types of TEs are involved? In what way are they potentially regulating the expression of regeneration-associated genes?

In this work we sought to analyze and explore the regulatory role that these elements could play during the regenerative event, using data from RNA-Seq methodologies to generate a transcriptomic profile of a specific moment of the process of nearby genes and TEs and ATAC-Seq data to identify open chromatin regions where transcription factors that regulate the regenerative event participate and bind to DNA. DNA and LTR TEs were found to be overexpressing after the amputation process, in intronic regions of genes related to regeneration, such as *anxa2a* and *hspa12b*. In turn, a total of 184 TEs were identified that were acting as binding sites for 6 different transcription factors related to regeneration. These results support the hypothesis of the regulatory role of these elements in biological processes such as regeneration, however, they should be taken with caution and confirmed with new biological experiments/methodologies to confirm their role in the event.

1. Introducción

Los eventos que se presentan posteriores a la ocurrencia de una lesión involucran varias vías tanto intracelulares que buscan reparar el tejido afectado. Mientras que algunos organismos son capaces de regenerar el tejido u órgano dañado de una manera completa, en la gran mayoría el proceso de reparación deriva en una masa de tejido fibrótico conocido como cicatriz (Thompson et al., 2020)(Gurtner et al., 2008). Para estos últimos organismos a medida que van alcanzando su madurez reproductiva, el proceso regenerativo decae en su función.

Son solo algunos pocos organismos los que presentan una alta capacidad de regeneración durante toda su vida y son motivo de estudio para poder comprender estos mecanismos y asociar mecanismos genéticos específicos al evento de regeneración. Dentro de estos organismos con una alta capacidad regenerativa se puede encontrar a la rana africana (*Xenopus laevis*), al gusano platelminto (*Planaria torva*), la salamandra ajolote (*Ambystoma mexicanum*), la salamandra gallipato (*Pleurodeles waltl*) y al pez cebra (*Danio rerio*). En particular, este trabajo estará enfocado en este último organismo, debido a que es uno de los organismos modelo con más estudios realizados a la fecha, en materia de regeneración (Marques et al., 2019).

1.1. Regeneración y *Danio rerio*

El proceso regenerativo de un organismo se puede definir como aquel en donde se reemplazan células, tejidos, órganos, o incluso partes completas de su cuerpo, luego de perderlas, ya sea por accidentes, o daños externos.

Este proceso ocurre naturalmente en todos los organismos, sobre todo en etapas de gestación y proliferación. Sin embargo, la mayoría de los organismos a medida que se desarrollan, tienden a perder o disminuir en su capacidad

regenerativa. Así también, la regeneración puede ocurrir en múltiples niveles de organización biológica, mientras algunos organismos son capaces de regenerar una extremidad completa, otros se limitan a regenerar algunos tipos de células específicas, como podría ser un axón cortado (Seifert & Muneoka, 2018).

Diversos estudios se han enfocado en la comprensión del proceso regenerativo en los instantes posteriores a una lesión, en donde una serie de eventos concatenados permiten en conjunto la restauración de la integridad y homeostasis del tejido perdido (Gurtner et al., 2008). Como se mencionó anteriormente, estos eventos involucran varias vías intracelulares e intercelulares y procesos en los que se destacan la formación de coágulos y reepitelización (regeneración epitelial), cicatrización de heridas y la formación de una masa de células no diferenciada en proliferación (Gilbert et al., 2015; Gurtner et al., 2008).

En términos generales, la regeneración involucra eventos de proliferación y diferenciación celular, con lo que se logra recuperar la actividad funcional del tejido. A su vez, dentro de los eventos inmediatos a la lesión se destaca la formación del blastema. El blastema provee a la zona de la lesión, una proliferación continua, que causará el crecimiento estructural, mediante la formación de nuevas células que posteriormente se diferenciarán para poder lograr reconstruir la zona del cuerpo perdida o dañada (Santos-Ruiz et al., 2002). En consecuencia, el tejido regenerado cuenta con las mismas propiedades fisiológicas y morfológicas del tejido perdido (Figura 1), (Marques et al., 2019).

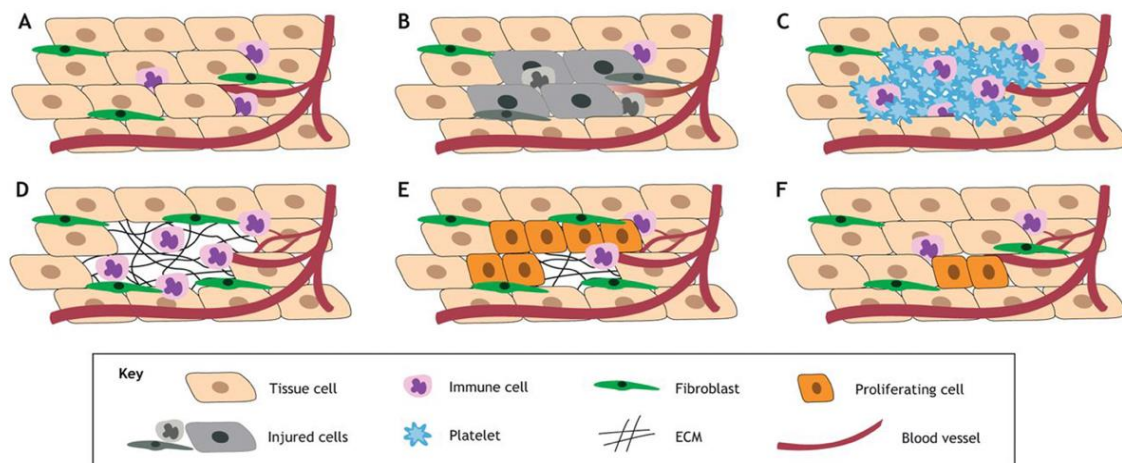


Figura 1. Fases de regeneración en tejidos. El recuadro “Key”, muestra la correspondencia de distintos elementos con un símbolo, para cada tipo de célula implicada en el proceso diferenciadas por colores. En (A) se presenta el tejido sin lesión. En (B) se visualiza una zona lesionada representada por células de color gris y además se visualiza el vaso sanguíneo más próximo a la zona dañada con un tono más tenue que el de los demás. En (C) se ve la respuesta inmune (color rosa), y el comienzo del cierre de la herida producto de la acción de plaquetas (color celeste). En (D) se visualiza la eliminación de desechos y comienzo de la formación de una matriz extracelular (líneas negras), mientras las células inmunes comienzan a limpiar el área. Además, los fibroblastos (en verde), proliferan y producen fibras. Por último, se comienzan a generar nuevos vasos sanguíneos. En (E) se puede ver la proliferación celular (células de color naranja), mientras la matriz extracelular comienza a desaparecer a medida que se logra poblar el área con las nuevas células. Finalmente, en (F) se visualiza el área que anteriormente estaba dañada, prosiguiendo su proceso de reparación, ya con la mayoría de las células regeneradas. Figura tomada de (Marques et al., 2019).

El pez cebra, *Danio rerio*, es considerado un organismo modelo en regeneración (Marques et al., 2019). Es un pez teleósteo¹ de entre 2 y 5 cm de largo, con una notable capacidad de regenerar varios órganos después de una lesión. La regeneración en este organismo se lleva a cabo en tiempos mucho más cortos que en otros con igual capacidad de regeneración (Poss et al., 2003), permitiendo estudiar de una mejor manera todos los procesos involucrados .

¹ Infracalse de peces que poseen un esqueleto completamente osificado

Además, *Danio rerio* cuenta con numerosas características que lo hacen ser un organismo modelo, no solo para la regeneración, sino que, para el estudio de otros fenómenos biológicos, como por ejemplo el desarrollo del sistema nervioso entérico (ENS), angiogénesis, cáncer de piel, entre otros. De acuerdo a Khan & Alhewairini (2019), se pueden destacar las siguientes características al respecto:

- Tasa de fecundidad; *Danio rerio* alcanza la madurez sexual aproximadamente a las 10 semanas y cuando se mantiene en condiciones óptimas puede poner alrededor de 200 huevos a la semana.
- Rapidez de desarrollo del embrión; La etapa embrionaria de *Danio rerio* comprende aproximadamente entre 48 y 72 horas. Posterior a esto, el embrión eclosiona del huevo. Dentro de la etapa embrionaria, alrededor de las primeras 18 horas pueden verse diversos órganos y tejidos como el cerebro, ojos u orejas y a las 24 horas se terminan de formar la mayoría de los órganos primarios.
- Transparencia de embriones; Durante esta etapa primaria, los embriones presentan una transparencia óptica que los convierte en organismos ideales para la obtención de imágenes microscópicas de distintos procesos celulares.
- Bajo costo de mantención; *Danio rerio* es un organismo fácil de criar y debido a sus proporciones es escogido por sobre otros para la realización de estudios.

Dentro de un contexto genómico, *Danio rerio* cuenta con un ensamble a nivel de cromosoma en su versión GRCz11. Las principales estadísticas de esta versión se detallan en la Tabla 1.

Largo del genoma (bp)	1.373.471.384
Nº de cromosomas	25
Nº de scaffolds	1.917
Nº de genes codificantes	25.592

Nº de genes no codificantes	6.599
Nº de pseudogenes	315
Nº de transcritos	59.876

Tabla 1. Estadísticas del ensamble de *Danio rerio* versión GRCz11. Ensamble de *Danio rerio*, obtenidas de Ensembl (Yates et al, 2016) (GRCz11).

Aproximadamente el 70% de los genes de *Danio rerio* presentan al menos un ortólogo en el humano, siendo utilizado con éxito para estudiar la actividad de estos genes en relación con enfermedades humanas (Howe et al., 2013). Por otro lado, **y altamente pertinente para este trabajo**, aproximadamente el 52.2% del genoma de *Danio rerio* corresponde a elementos repetidos, siendo este un valor inusualmente alto, ya que los demás peces teleósteos presentan un promedio de 30% de elementos repetidos. Considerando el genoma completo de *Danio rerio*, cerca del 50 % corresponde a elementos transponibles (*Transposable Elements*, referidos de aquí en adelante como TEs), teniendo 39% de TEs de tipo II y 11% de TEs tipo I (Figura 2)(Howe et al., 2013). La importancia de estos elementos y de su actividad transcripcional se describe en detalle en el punto 1.6.1.

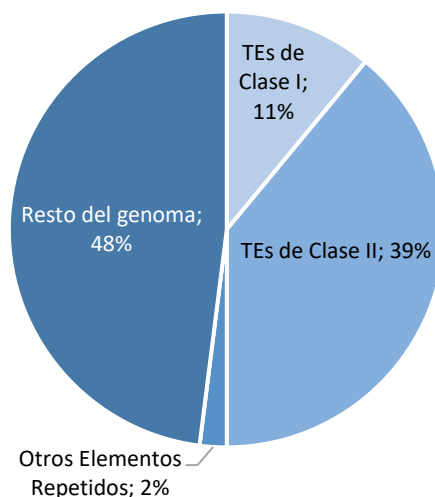


Figura 2. Porcentaje de elementos repetidos en *Danio rerio*. Gráfico representativo de elementos repetidos en el genoma de *Danio rerio*. TEs de clase I: 11%. TEs de clase II 39%. Otros elementos repetidos: 2%.

Entre algunos de los órganos y tejidos que más se han estudiado en este organismo, en materia de regeneración, se encuentra el cerebro, riñón, corazón, páncreas, retina y aleta (Marques et al., 2019).

Danio rerio cuenta con 5 tipos de aletas, todas con capacidad de regeneración. Estas aletas son dorsal, pectoral, pélvica, anal y caudal. En concreto la aleta caudal, es una de las más estudiadas debido a su accesibilidad y facilidad para realizar el seguimiento visual del proceso (Sehring & Weidinger, 2020). Esta aleta, es de fácil acceso para realizar una cirugía y provocar una lesión que permita evaluar los eventos regenerativos posteriores a la amputación. Estos eventos se pueden organizar de acuerdo a horas post-amputación (hpa), o a días post-amputación (dpa). Así también, la lesión en este órgano no compromete la supervivencia del organismo (Poss et al., 2003).

Una vez realizada la amputación a la aleta, una de las primeras respuestas a la lesión es la migración de células epiteliales que conforman un epitelio para cubrir la herida. Este epitelio dentro de las próximas horas a su formación, se convierte en una estructura multicapa denominada epidermis (Sehring & Weidinger, 2020).

Como se mencionó anteriormente, en la mayoría de los organismos vertebrados decae la capacidad regenerativa al alcanzar su madurez reproductiva, por lo que al sufrir una herida o amputación desarrollan una cicatrización posterior a la formación de la epidermis para terminar el proceso de reparación, lo que impide que los eventos regenerativos como la proliferación y diferenciación celular puedan actuar. Sin embargo, en *Danio rerio* esto no ocurre así (Figura 3). En el caso de la aleta caudal, luego de la amputación comienza inmediatamente el proceso regenerativo después del cierre de la herida. Así, dentro de los dos días posteriores a la amputación ocurre la formación de un blastema (Figura 3B). Posteriormente, ocurre la regeneración como tal, en donde hay una rápida proliferación celular. Esto resulta en el aumento de tamaño y restauración de la aleta en alrededor de 20 días (Figura 3C) (Pfefferli & Jazwińska, 2015).

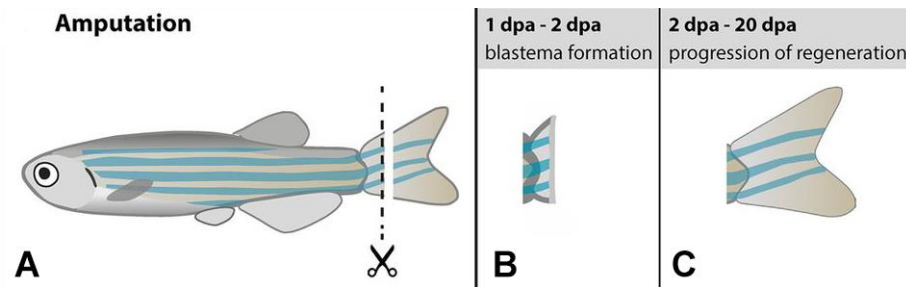


Figura 3. Regeneración en aleta caudal producida por amputación. En (A) se esquematiza la amputación de la aleta caudal de *Danio rerio*. En (B) surge la formación de epitelio y blastema dentro de los 2 días posteriores a la amputación (dpa) y posteriormente en (C), la regeneración completa de la aleta dentro de 20 dpa. Figura adaptada de (Chassot et al., 2016).

Para poder estudiar a un nivel molecular los cambios y eventos que ocurren en la regeneración, se han aplicado metodologías genómicas, dentro de las cuales una de las más utilizadas es el análisis de expresión génica.

1.2. Expresión Génica

En general, todas las células de un organismo eucarionte contienen más o menos el mismo material genético. No obstante, la mayoría de las células se encuentran diferenciadas, a fin de tener una morfología y función particular dentro de un organismo. Esto se logra gracias a diversas combinaciones particulares de genes que se están expresando y reprimiendo (Ralston & Shaw, 2008). El proceso mediante el cual parte del DNA dirige la síntesis de proteínas, o en algunos casos solo RNA, se conoce como expresión génica (Cain et al., 2016).

A medida que ocurren eventos como la cicatrización de heridas o la regeneración, son muchos los genes que cambian su expresión (Goldman & Poss, 2020). Este cambio de expresión génica, ya sea activación o represión de genes, es regulado por diversas señales tanto internas como externas de la célula (Hoopes, 2008).

1.3. Expresión Diferencial de Genes

En una célula promedio se tienen diversos genes, algunos de estos con características “típicas” y/o comunes a todas estas, es decir, que cumplen un rol esencial para el funcionamiento de cualquier tipo de célula. Sin embargo, las células también cuentan con un conjunto de genes exclusivos de cada tipo de célula en particular, permitiéndole que estas lleven a cabo una función específica. En base a esto, las diferencias que presentan los distintos tipos de células no se debe a la presencia de distintos tipos de genes, sino, a la expresión diferencial de estos (Cain et al., 2016). Comúnmente, las metodologías que se realizan para analizar la expresión diferencial de genes comprenden a un conjunto de células del mismo tipo, pero en distintas condiciones. Así, en base a estas condiciones, se puede identificar que genes están siendo sobre-expresados o reprimidos en un momento determinado.

1.4. Regulación de la Expresión Génica

Como se detalló anteriormente, la expresión génica es el proceso mediante el cual el DNA dirige la síntesis de proteínas y/o la transcripción. La expresión de un gen depende de varios factores, los cuales se pueden regular en varias etapas, de acuerdo con el nivel o al tipo de molécula que participe (DNA, RNA o proteína). Dentro de las principales etapas de regulación que ocurren dentro del núcleo celular eucarionte, esquematizadas en Figura 4, están:

- Accesibilidad a la cromatina
- Transcripción
- Procesamiento del transcrito primario de RNA
- Transporte del RNA al citoplasma

Este proyecto se centrará en el análisis de dos de estas etapas que ocurren en el núcleo celular, es decir, no se considerarán eventos de control post-transcripcional, debido a la disponibilidad del set de datos a utilizar. En concreto estos dos procesos son: la modificación o accesibilidad a la cromatina y la transcripción (Cain et al., 2016).

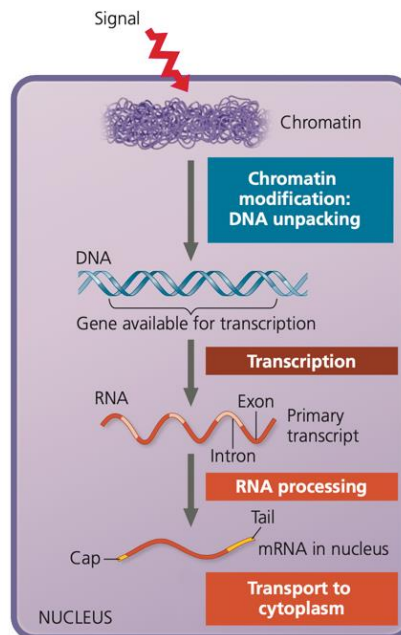


Figura 4. Etapas de regulación génica en el núcleo celular eucarionte. Diagrama que representa las etapas de regulación génica de una célula eucarionte dentro de su núcleo. Cada etapa está representada por un rectángulo de color. La primera etapa, representada por el rectángulo “Chromatin modification, DNA unpacking”, comienza con una señal (“Signal”), simbolizada por una flecha roja, que permite el desempaquetamiento de la cromatina en las zonas donde se producirá la transcripción. La segunda etapa, denominada “Transcription”, es donde, por medio de diversas proteínas que se fijan al DNA, se promueve la transcripción de un gen a partir del DNA disponible representado por la doble hélice de color celeste. En la tercera etapa, llamada “RNA processing”, se muestra que, a partir de un transcrito primario, representado por la hebra de RNA que cuenta tanto con exones (en rojo) e intrones (en amarillo), se puede obtener por medio de *splicing* alternativo, un transcrito maduro que lleva la adición de nucleótidos a los extremos del mRNA (representados por “Cap” y “Tail”). Finalmente, en la última etapa, denominada “Transport to cytoplasm”, se observa, como este transcrito maduro se exporta desde el núcleo hacia el citoplasma. Figura adaptada de (Cain et al., 2016).

1.4.1. Accesibilidad a la cromatina

El DNA, en organismos eucariontes, no se encuentra totalmente accesible en las células, sino que está empaquetado en una estructura compactada denominada cromatina (Orphanides & Reinberg, 2002). La cromatina está altamente compactada por una matriz de nucleosomas, los cuales son una estructura que consta de un núcleo de octámeros de proteínas histonas envuelto por aproximadamente 147 pares de bases de DNA y separados por DNA que las enlaza (Tsompana & Buck, 2014). La posición que tomen los nucleosomas dentro del genoma cumple una función reguladora directa, permitiendo o no el acceso al DNA, de distintas proteínas involucradas en el proceso de transcripción y de la expresión de genes. Esto, particularmente se debe a que la distribución de los nucleosomas no es permanente, y se modifica de manera dinámica, siendo un factor crítico para la organización y función de la cromatina (Klemm et al., 2019).

En base a la posición y comportamiento de los nucleosomas y al reclutamiento y participación de numerosos elementos reguladores y/o factores de transcripción, la cromatina se puede presentar en un estado abierto (Eucromatina) o cerrado (Heterocromatina). Una representación de los distintos estados de la cromatina mediado por las distintas posiciones que tomen los nucleosomas se ilustra en la Figura 5. Algunos elementos reguladores se ubican selectivamente en donde la cromatina es más accesible o donde está abierta, produciendo así eventos de transcripción génica. Por el contrario, en lugares donde la cromatina está cerrada o empaquetada, se presenta una situación desfavorable para los elementos reguladores. Al no poder unirse dichos elementos regulatorios, ocurre un silenciamiento genético en esas zonas del genoma (Sun et al., 2019).

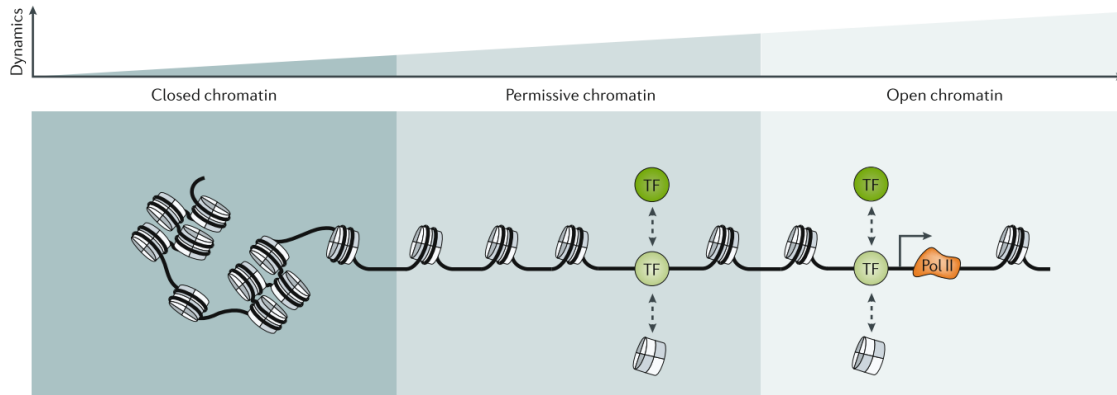


Figura 5. Estados de accesibilidad a la cromatina. De acuerdo con la interacción de diversos agentes reguladores, los estados determinan un estado cerrado, un estado permisivo o un estado abierto. En la parte superior de la figura se visualiza un gráfico de nivel, que simboliza el estado dinámico de la cromatina a lo largo del genoma de acuerdo con el posicionamiento de los nucleosomas. Se ve representada en 3 niveles: cromatina cerrada, cromatina permisiva y cromatina abierta. Esta distribución dinámica de la cromatina aumenta a medida que la cromatina se va desempaquetando. En la parte inferior de la imagen, a la izquierda (estado cerrado de la cromatina) se visualizan una serie de cilindros enrollados y bastante próximos entre sí. Estos son los nucleosomas y la línea que los enrolla es una secuencia de DNA. Los nucleosomas son elementos estructurales centrales de la cromatina, que permiten empaquetarla y regular la accesibilidad a ella según su composición y el lugar donde se encuentren. En el cuadrado central (estado permisivo de la cromatina) se pueden visualizar estos nucleosomas un poco más alejados, permitiendo que la cromatina esté más desempaquetada en algunas regiones, lo que a su vez permite la unión de factores de transcripción (TF, círculos verdes) al DNA. En el cuadrado que se visualiza más a la derecha (estado abierto de la cromatina), se visualiza como los nucleosomas están mucho más alejados y la unión de los factores de transcripción permite que otras proteínas interactúen con la cadena de DNA, como la enzima RNA polimerasa II (representada por la figura naranja Pol II), para que se inicie el proceso de transcripción. Figura tomada de (Klemm et al., 2019).

1.4.1.1. ATAC-Seq

Una metodología para poder evaluar la accesibilidad a la cromatina en el genoma es ATAC-Seq (*Assay for Transposase-Accessible Chromatin using sequencing*, en español ensayo de cromatina accesible por transposasa usando secuenciación), en donde se utiliza una versión hiperactiva de la transposasa Tn5. Precisamente, esta hiperactividad de Tn5 hace que se una y corte el DNA con mucha más frecuencia en la cromatina accesible (sin nucleosomas) que en la inaccesible (con nucleosomas). Posteriormente, se generan fragmentos de DNA a los cuales se les agregan adaptadores, para poder secuenciarlos (Buenrostro et al., 2016). Una representación de este proceso se muestra en la

Figura 6.

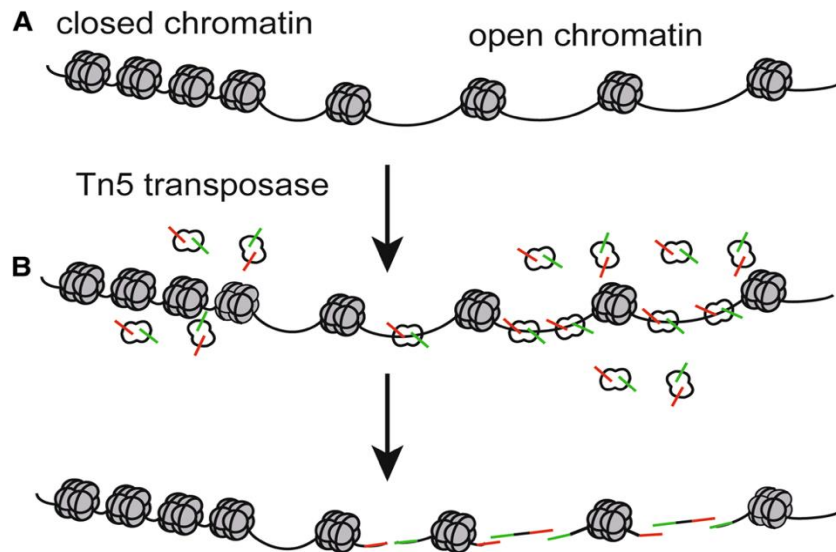


Figura 6. Representación del mecanismo de acción de transposasa Tn5 en ATAC-Seq. En (A), se pueden observar las figuras circulares en color gris (nucleosomas) empaquetando de tal forma el DNA generando dos estados, cromatina cerrada (a la izquierda) y cromatina abierta (a la derecha). En (B), se visualiza la acción de la transposasa Tn5, figura de color blanco con líneas verdes y rojas, que representan adaptadores para secuenciación. La Tn5 se une a la cromatina con mucha más frecuencia en el estado abierto de la cromatina que en el cerrado. Finalmente, bajo la segunda flecha en la parte final de la figura, se visualiza que el corte de las regiones se realiza con mayor frecuencia en donde la cromatina está abierta. Estas regiones contienen a sus extremos los adaptadores de

secuenciación (líneas verdes y rojas) para generar fragmentos de DNA a secuenciar. Figura tomada de (Sun et al., 2019).

1.4.2. Transcripción

El proceso de transcripción es el cual donde se sintetiza RNA a partir de DNA, mediante la acción de una enzima denominada RNA polimerasa. La secuencia producida se denomina como transcrito primario (Cain et al., 2016).

En esta etapa ocurren los principales eventos de regulación de muchos genes. Estos eventos, en su mayoría, están mediados por un grupo de proteínas llamados factores de transcripción. Los factores de transcripción reconocen y se unen al DNA en regiones específicas como pueden ser la región promotora, dominios de unión conservados, o también directamente sobre otro factor de transcripción. Una vez se une a la región determinada de la secuencia, que comúnmente se encuentra cerca del inicio de la transcripción de un gen, el factor de transcripción ejercerá la regulación sobre él. Esta regulación puede resultar en la activación o en el silenciamiento de los genes (Phillips, 2008). A su vez, también existen isoformas de los factores de transcripción. Estas isoformas provienen de un evento de *splicing* alternativo que, como se detalló anteriormente en la Figura 4, es una de las etapas de regulación génica celular. Estas nuevas configuraciones, en lugar de regular la expresión de un gen, pueden llegar a regular distintos eventos transcripcionales completamente distintos, como se visualiza en la Figura 7.

Diversos estudios han mostrado que la tasa de transcripción de un gen está regulada por el equilibrio relativo de factores de transcripción activadores e inhibidores (Latchman, 2001).

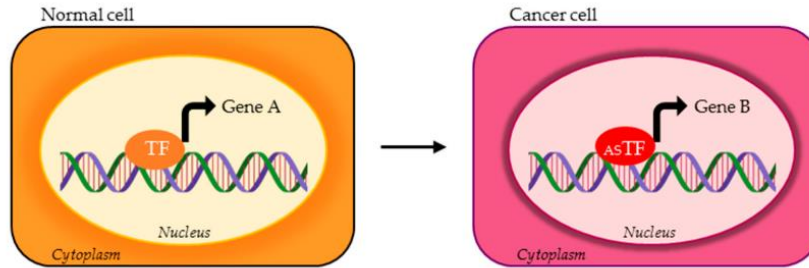


Figura 7. Actividad de isoformas de los factores de transcripción. Al lado izquierdo de la figura se ve la representación de una célula normal esquematizada por un rectángulo de color naranja. Dentro de éste, se visualiza un óvalo que representa el núcleo celular. En este núcleo se ve la doble cadena de DNA y sobre ella un óvalo de color naranja, el cual representa al factor de transcripción (TF). Este TF promueve la expresión del gen A. En contraste, al lado derecho de la figura, se visualiza una célula cancerígena, representada por color rosa, también con su núcleo en el interior de ella en forma de óvalo. Dentro de este núcleo se visualiza la doble cadena de DNA con la unión, esta vez, de una isoforma del primer factor de transcripción (asTF). La unión de esta isoforma resultará en la expresión de otro gen distinto, el gen B. Figura tomada de (Belluti et al., 2020).

La transcriptómica es el estudio del transcriptoma, es decir, el conjunto de transcritos que son producidos por el genoma en circunstancias específicas. Es clave a la hora de evaluar los elementos funcionales del genoma y comprender un determinado fenómeno biológico. Dentro de sus objetivos principales está clasificar los tipos de transcritos, determinar la estructura transcripcional de los genes, patrones de *splicing* y cuantificar los niveles de expresión bajo distintas condiciones (Wang et al., 2009).

1.4.2.1. RNA-Seq

La metodología más utilizada actualmente para estudiar el transcriptoma es RNA-Sequencing (de aquí en adelante abreviado como RNA-Seq), cuyo objetivo es proporcionar el conjunto completo de todos los transcritos expresados por un conjunto de células en una condición determinada. En concreto en la metodología RNA-Seq (

Figura 8), el conjunto de moléculas de RNA de un tejido es convertido a una biblioteca de fragmentos de cDNA² con adaptadores unidos a los extremos. Luego de esto, se realiza la secuenciación de estos pequeños fragmentos resultando en lecturas o *reads*. Los *reads* generados comúnmente se alinean contra el genoma de referencia del organismo estudiado. De no contar con genoma de referencia, se puede realizar un ensamble *de novo* de transcritos. Finalmente, se estiman los niveles de expresión de cada gen (Wang et al., 2009).

RNA-Seq además permite conocer los límites intrón-exón dentro del genoma con una notable precisión, a una resolución a nivel de base.

La integración de los datos obtenidos por RNA-Seq con los datos de ATAC-Seq pueden ser claves para lograr una asociación de los cambios de la expresión génica con los cambios de la accesibilidad a la cromatina. Se pueden integrar los datos de ambas metodologías para identificar genes con expresión diferencial y regiones que tengan una accesibilidad a la cromatina diferencial (Yan et al., 2020).

² Secuencia de DNA complementaria al RNA.

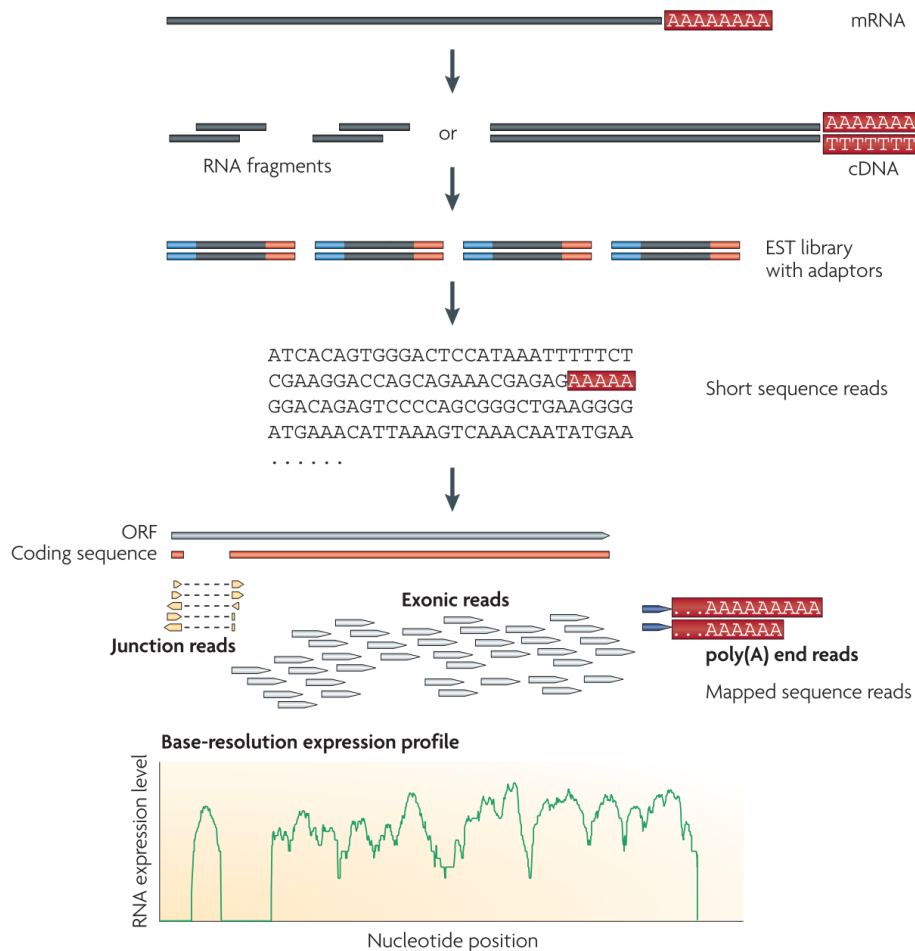


Figura 8. Metodología RNA-Seq. La metodología RNA-Seq comienza con las secuencias completas de un conjunto de RNAs, que en este caso está representado por una línea gris larga en la parte superior de la figura. Además, esta secuencia posee una cola poly A, representada por el rectángulo rojo con letras A. Bajo la primera flecha vertical se esquematiza la generación de la biblioteca, que puede ser en fragmentos de RNA (representados a la izquierda de la figura), o de DNA (derecha de la imagen). Bajo la segunda flecha vertical se visualizan las bibliotecas, las cuales ya cuentan con la unión de adaptadores a sus extremos (líneas azules y naranjas). Bajo la tercera flecha vertical se representan las secuencias cortas de cDNA (reads) que se obtienen mediante secuenciación. Bajo la cuarta flecha vertical se visualiza como estos reads se alinean a un genoma o transcriptoma de referencia, en este caso está representado por un ORF (línea gris) y su secuencia codificante (línea color naranja). Los reads que se obtienen pueden ser de 3 tipos, Junction reads, exonic reads o poly(A) end reads. Finalmente, en el final de la imagen se visualiza un gráfico de expresión de RNA de los reads generados. Figura tomada de (Wang et al., 2009).

1.5. Elementos Reguladores de transcripción

Como se detalló anteriormente, la regulación génica es importante para que diversas células realicen funciones específicas en distintos tejidos. Esta regulación es mediada por elementos reguladores y concretamente, uno de estos son los sitios de unión a los factores de transcripción, también conocidos como TFBSs por sus siglas en inglés (*Transcription Factor Binding Sites*), que son diversas secuencias no codificantes que determinan la conectividad de las redes moleculares y que se activan mediante la unión de factores de transcripción (Doane & Elemento, 2017). La identificación de estos elementos es clave para poder caracterizar de manera correcta los patrones precisos de expresión, ya que la unión de los factores de transcripción a estas secuencias cambia la posición del nucleosoma, generando a su vez un cambio en la cromatina a un estado más abierto (Ahituv, 2012).

Se pueden definir varios tipos de elementos reguladores de transcripción, que pueden interactuar con genes y/o entre sí (Figura 9). Entre los más caracterizados, según (Ahituv, 2012), se tienen a:

- Promotores: Es uno de los elementos reguladores mejor caracterizados. Se ubica al comienzo del gen y sirve como sitio donde se ensambla la maquinaria de transcripción y en donde comienza la transcripción del gen.
- Potenciadores: Los potenciadores se unen a los promotores en lugares específicos. Pueden regular la expresión génica en varios momentos. Pueden ejercer una regulación en *cis* o *trans*. La regulación en *cis* comprende la regulación de un gen dentro de la misma región cromosómica y la regulación en *trans* un gen ubicado en un cromosoma diferente. También son conocidos como *Enhancers*.
- Silenciadores: Los silenciadores desactivan la expresión génica en momentos y lugares específicos.

- Aisladores: Secuencias de DNA que generan límites reguladores *cis* que impiden que los elementos reguladores de un gen afecten a los genes vecinos. También son conocidos como *Insulators*.

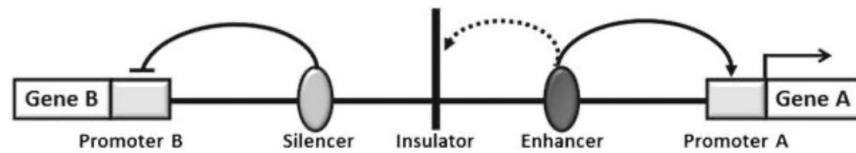


Figura 9. Representación de elementos reguladores de transcripción. Los promotores B y A (representados por cuadrados grises), se encuentran cada uno junto a un gen ("Gene B" y "Gene A", respectivamente). Un potenciador (*Enhancer*, representado por un óvalo gris oscuro), actúa sobre el promotor A (línea sólida levemente ovalada con punta de flecha), regulándolo activamente, favoreciendo la transcripción del gen A. Por otro lado, un aislador (representado por una línea vertical negra), evita que el potenciador logre ejercer acción sobre el promotor B (línea punteada). Por último, un silenciador (representado por un óvalo gris claro), regula negativamente al gen B (línea sólida sin punta de flecha). Figura tomada de (Ahituv, 2012).

Muchos de estos elementos reguladores provienen de los TEs, un tipo de elemento repetido (Feschotte, 2008). Es por esto, que su identificación puede ser clave para estudiar eventos de regulación génica que participen en el proceso regenerativo.

1.6. Elementos Repetidos

Los elementos repetidos son secuencias que están en más de una ubicación a lo largo de un genoma. Están presentes en distintas proporciones dentro de diversos genomas, por ejemplo, aproximadamente un 80% en *Zea mays*, un 3% en *Saccharomyces cerevisiae*, aproximadamente un 45% en *Homo sapiens* (Lerat, 2010) y como se mencionó anteriormente un 52% en *Danio rerio*. Estos elementos repetidos pueden proporcionar pistas claves sobre la dinámica de los cromosomas, las fuerzas evolutivas y los mecanismos para el intercambio de información genética entre organismos. (Izsvák et al., 1997).

Existen diversos tipos de elementos repetidos, dentro de los cuales se destacan 5 grupos principales (Billingsley et al., 2019). Estos grupos son:

- Tandem repeats y DNA Satélite: Secuencias de DNA que se repiten de tal manera que las repeticiones se encuentran una al lado de la otra.
- Simple repeats: Secuencias cortas de DNA de entre 2 y 6 pares de bases que se repiten.
- Pseudogenes: Secuencias de DNA que se asemejan a un gen, pero no poseen su funcionalidad, puesto que se han inactivado en el curso de la evolución de estas.
- Segmentos duplicados: Secuencias de DNA largas, comúnmente sobre 1 kbp³ de longitud que tienen secuencias casi totalmente idénticas entre sí.
- TEs: Secuencias de DNA que tienen la capacidad de moverse de manera autónoma y no autónoma (según el tipo) a distintas partes del genoma. Son el grupo más importante y abundante de elementos repetidos.

³ Kilo pares de bases. 1 Kbp = 1.000 pares de bases de DNA.

1.6.1. Elementos Transponibles (TEs)

Los TEs, como se mencionó anteriormente, son elementos con la capacidad de moverse dentro del genoma. Se pueden organizar en 2 clases, según el intermediario que utilicen para moverse (Lerat, 2010):

- Clase I: Utilizan un intermediario de RNA para transponerse utilizando un mecanismo de “copiar y pegar”. Estos a su vez se subdividen en el grupo LTR (*Long Terminal Repeat*, repeticiones terminales largas) o no LTR, dentro de los cuales se encuentran los LINE (*Long Interspersed Nuclear Elements*, elementos nucleares dispersos largos) y SINE (*Short Interspersed Nuclear Elements*, elementos nucleares dispersos cortos).
- Clase II: Utilizan un intermediario de DNA para transponerse utilizando un mecanismo de “cortar y pegar”.

1.6.1.1. Rol de TEs en eventos regulatorios

Los TEs presentan capacidades reguladoras génicas en todo el genoma incluyendo, por ejemplo, efectos *cis* y *trans* reguladores de genes (Slotkin, 2018). La actividad transcripcional de los TEs puede ser clave para eventos de expresión génica, ya que proveen al genoma de diversos tipos de estos elementos reguladores que pueden terminar influyendo directamente en la regulación de genes cercanos, tanto a un nivel transcripcional y post-transcripcional (Feschotte, 2008).

Dependiendo del sitio donde se encuentre un TE puede llegar a interrumpir la expresión génica, provocando cambios genéticos, epigenéticos o ambos (Elbarbary et al., 2016). Esto, convierte a los TEs en secuencias de interés para diversos estudios de regulación génica. Algunas de las principales formas que tienen los TEs para regular la expresión génica se visualizan en la

Figura 10 y Figura 11.

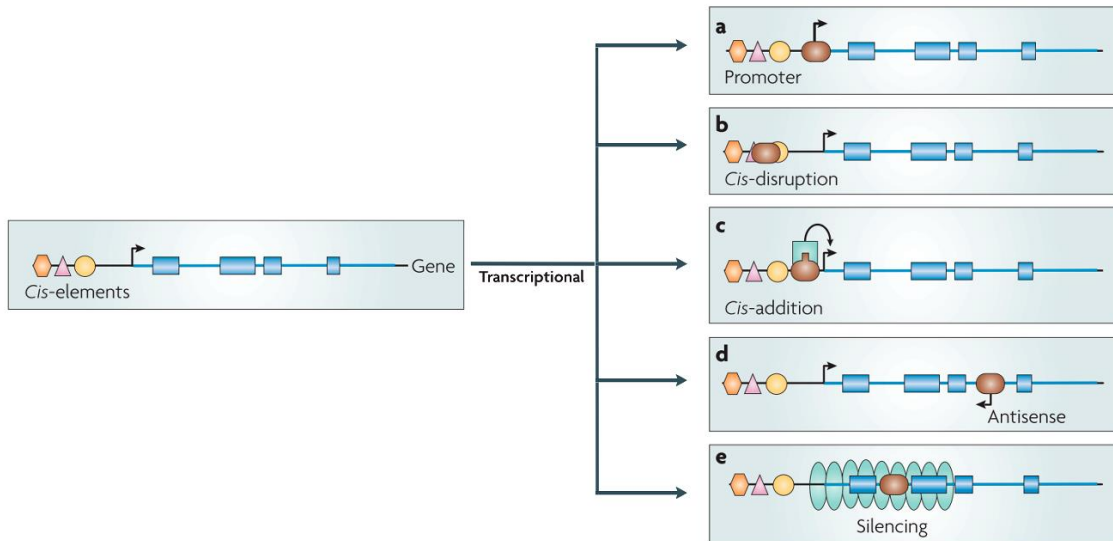


Figura 10. Mecanismos de regulación génica mediados por TEs. En el panel de la izquierda se visualiza una serie de elementos reguladores cis (*Cis-elements*), representados por un hexágono naranja, un triángulo violeta y un círculo amarillo. Luego, la flecha indica el inicio de la transcripción y el respectivo conjunto de exones (rectángulos azules) e intrones (líneas azules) del gen (*Gene*). A la derecha, se indican los mecanismos de regulación génica que puede tener un TE (ovalado de color café): **a.** actuando como promotor, generando un nuevo sitio de inicio de transcripción, **b.** TE que interrumpe la acción de los elementos reguladores cis, **c.** adición de un nuevo elemento regulador como podría ser un sitio de unión a un factor de transcripción, **d.** la inserción de un TE en un intrón, la cual promueve la transcripción antisentido que potencialmente interferirá con la transcripción, y **e.** donde la inserción un TE puede promover la formación de heterocromatina (representada por óvalos de color verde), silenciando a la expresión de genes a su alrededor. Figura adaptada de (Feschotte, 2008)

Se ha reportado que uno de los mejores ejemplos de que TEs adquieren una función reguladora, proviene de fragmentos de ellos que están conservados entre algunos organismos. Esta conservación y acumulación de TEs, genera una fuente de potenciales elementos reguladores para el organismo, a partir de los cuales, dichos elementos evolucionan de una manera *de novo* mediante mutaciones puntuales (Feschotte, 2008).

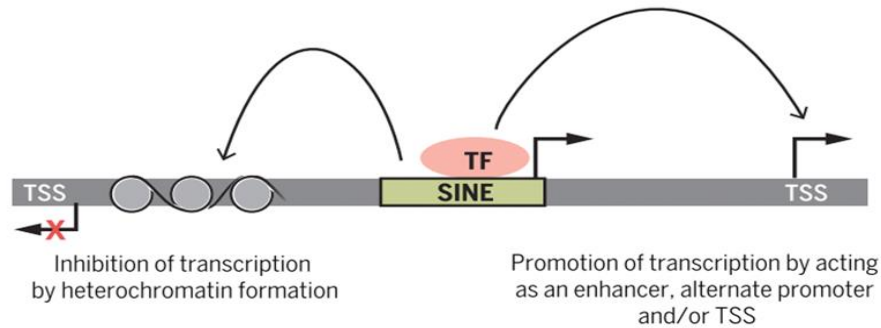


Figura 11. Regulación génica mediada por TEs. Los TEs (en este caso un elemento SINE representado por un rectángulo verde), pueden promover o inhibir la transcripción de genes cercanos, ya que funciona como sitio de unión para un factor de transcripción (óvalo rosa TF). Los TF son proteínas que se unen al DNA para regular el proceso de transcripción. Se visualiza al costado izquierdo de la figura en la que el TE actúa, inhibiendo la transcripción, formando una región cerrada de la cromatina por la representación de círculos grises cercanos al sitio de inicio de la transcripción (TSS). Al costado derecho de la figura se visualiza como el TE promueve la transcripción al actuar como potenciador en una zona cercada al TSS. Figura tomada de (Elbarbary et al., 2016).

1.6.1.2. Identificación de TEs expresados

Una de las principales limitaciones que tiene la caracterización de TEs dentro de cualquier organismo, está dada por su naturaleza repetitiva y el lugar donde estos se encuentren y estén ejerciendo acción. Esto, principalmente por que los TEs pueden estar transcribiéndose dentro de un gen (TEs dependientes de un gen), utilizando la maquinaria transcripcional de este, o en regiones intergénicas o regiones intrónicas de genes no transcritos (TEs autoexpresados), como se describe en la Figura 12.

Así también, el origen de expresión de los TEs es otro de los grandes desafíos, a la hora de caracterizar estos elementos. Debido a sus variados números en cantidad de repeticiones herramientas como *TEcandidates* (Valdebenito-Maturana & Riadi, 2018), son de gran utilidad para poder predecir el origen de expresión de los TEs.

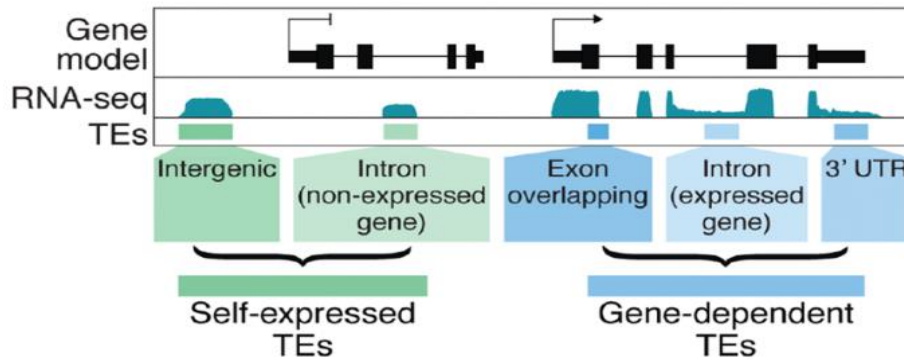


Figura 12. Tipos de clasificación de la expresión de un TE. En el primer sub-cuadro de la figura se visualiza la representación del modelo genético (“Gene model”), el cual representa, al costado izquierdo, la estructura de un gen no expresado, evidenciado por la flecha con punta de línea vertical, además de sus respectivas regiones UTR (rectángulos negros horizontales), exones (rectángulos negros verticales) e intrones (líneas delgadas negras). Al costado derecho del mismo cuadro se visualiza la estructura de un gen expresado, con los mismos componentes que el anterior, pero con la línea de expresión representada por una flecha con punta de lanza. En el segundo sub-cuadro se visualiza la actividad transcripcional de la metodología RNA-Seq, en donde, los distintos *peaks* demuestran regiones que se transcriben en un determinado momento. En el último sub-cuadro, se visualizan las regiones correspondientes a TEs dentro de la representación general. De acuerdo con la posición que se encuentre un determinado TE, este se clasificará según la dependencia o no de genes del organismo. En el lado izquierdo de la figura, se tiene a la clasificación de TEs autoexpresados “Self-expressed TEs”, representados por el esquema en color verde, en donde los TEs se encuentran en zonas intergénicas y zonas intrónicas de genes no expresados. Por el contrario, en el lado derecho de la figura, se tiene a la clasificación de TEs expresados dependientes de un gen “Gene-dependent TEs”, caracterizados por el esquema en color azul, en donde los TEs se encuentran sobrelapando regiones exónicas, regiones intrónicas y regiones UTR de un gen que se está expresando.

La comprensión y caracterización de los TEs que están ejerciendo una regulación génica en el proceso de regeneración en *Danio rerio*, puede **permitir tener una visión mucho más amplia de cómo se produce el evento regenerativo en sí** y de cómo estos elementos influyen para que el proceso se complete en su totalidad. Además, particularmente la regeneración como evento

biológico, tiene un gran interés y potencial dentro del mundo científico, ya que es un proceso que está en vías de aprendizaje y comprensión. Caracterizar y entender el rol de los TEs, aportará al entendimiento de esta área en particular y podría tener un gran impacto en cómo se trata la regeneración en un futuro.

Particularmente, los TEs no se distribuyen al azar dentro del genoma, ya que estos elementos, por motivos evolutivos, tienden a tener una preferencia por insertarse dentro de determinadas regiones del genoma del organismo, resguardando o facilitando su propagación y mitigando los efectos adversos de su actividad. Esto se suma a que se ha reportado actividad de TEs tanto en células somáticas como en células germinales (Bourque et al., 2018). Por lo tanto, la cantidad y diversidad de TEs que pueda tener un organismo, en este caso *Danio rerio*, y de cómo estos aporten o interactúen en diversos procesos biológicos como el regenerativo, provienen de las propiedades específicas de estos elementos, así como de las fuerzas evolutivas que actuaron en el organismo *Danio rerio* para que esté presente una gran capacidad regenerativa en la mayoría de sus órganos por sobre otras especies.

Es en este contexto, que diversos estudios han caracterizado TEs como importantes elementos reguladores de genes (Chuong et al., 2017). Sin embargo, **aún no se ha explorado su potencial rol en el evento de regeneración de la aleta de *Danio rerio***. Dado el alto porcentaje de TEs en el genoma de este organismo, surge la interrogante de si estos pueden estar involucrados en regulación génica durante el evento de regeneración.

En la actualidad, se conocen diversos genes implicados en el proceso de regeneración, específicamente asociados a eventos particulares, dado que muestran una expresión diferencial en el proceso. También se han realizado estudios a fin de determinar elementos reguladores asociados a la activación de estos genes. Uno de estos estudios es el realizado por Thompson y colaboradores (2020), en el que se utilizó tanto la metodología de RNA-Seq junto a la de ATAC-

Seq para identificar elementos reguladores que dirigen la expresión génica durante el proceso regenerativo de aleta caudal en *Danio rerio*. Sin embargo, tanto en este como en el resto de los estudios, no han analizado TEs. Esto llama la atención aún más considerando la alta proporción de TEs que presenta *Danio rerio*, como se mencionó anteriormente. Por otro lado, desde un punto de vista bioinformático, se pueden hacer varias críticas al trabajo de Thompson y colaboradores: (i) utilizaron herramientas obsoletas, como *TopHat*, para realizar el alineamiento de las secuencias de RNA-Seq al genoma de referencia, y (ii) utilizaron una versión antigua del genoma de *Danio rerio*. Estos son factores que pueden derivar en resultados erróneos.

Teniendo en cuenta la alta proporción de TEs en el genoma de *Danio rerio*, sumado a los numerosos genes que potencialmente participan en los eventos regenerativos y que tienen elementos reguladores que permiten su expresión o silenciamiento, surgen las interrogantes: ¿los TEs participan en la regulación de la expresión génica de los eventos que comprenden la regeneración? Y si es así, ¿Qué tipos de TEs participan? ¿De qué clase y familia son? ¿Regulan a genes cercanos (“en *cis*”) implicados en regeneración?, o ¿Están participando como sitios de unión a factores de transcripción para la expresión de estos genes?

En este trabajo se buscará responder estas preguntas, con la finalidad de evaluar el potencial biológico que representa la expresión y la actividad de los TEs en el evento regenerativo de aleta caudal de *Danio rerio*.

2. Hipótesis

Los elementos transponibles de *Danio rerio* están participando, mediante su actividad transcripcional y/o como sitio de unión a factores de transcripción, en la regulación de genes durante el proceso de regeneración de aleta caudal.

3. Objetivo General

Asociar elementos transponibles transcripcionalmente activos y aquellos que sean sitios de unión de factores de transcripción con genes involucrados en el proceso de regeneración de aleta caudal en *Danio rerio*.

4. Objetivos Específicos

- N°1: Identificar TEs transcripcionalmente activos durante el evento de regeneración de aleta caudal en *Danio rerio* mediante uso de datos RNA-Seq.
- N°2: Identificar TEs que actúen como sitios de unión de factores de transcripción durante el evento de regeneración de aleta caudal en *Danio rerio* mediante el uso de datos ATAC-Seq.
- N°3: Correlacionar datos de TEs obtenidos con genes involucrados en el proceso regenerativo.

5. Metodología

La metodología de este trabajo se subdividió en dos apartados, considerando tanto los recursos que se utilizaron, como el procedimiento que conllevó cada tarea con la finalidad de cumplir con los objetivos antes planteados. Así, en el primer apartado se indicarán los materiales que se utilizaron durante todo el trabajo, mientras que en el segundo apartado se detallará como se usaron estos materiales para cumplir cada objetivo específico.

5.1. Materiales

Los materiales de este trabajo corresponden a: el set de librerías de secuenciación (que se usaron como archivos de entrada para gran parte de los análisis), el conjunto de *software* para la realización de cada paso de la metodología propuesta y una breve descripción del *hardware* donde se ejecutaron la mayoría de estos procedimientos.

5.1.1. Set de datos

Como se detalló anteriormente, este trabajo utilizó el set de datos disponible públicamente del estudio de Thompson y colaboradores (2020). En dicho trabajo se analizaron grupos de muestras correspondientes a distintos puntos en el tiempo de la etapa regenerativa y utilizando distintas metodologías (RNA-Seq y ATAC-Seq). Se utilizó este último parámetro para separar las muestras en tablas (Tabla 2 y Tabla 3, respectivamente), donde cada columna significa:

- ID de muestras: identificador en la base de datos *Sequence Read Archive* (SRA), en la cual se encuentran alojados los archivos de secuenciación del trabajo de Thompson y colaboradores.

- Condición: Tiempo en días post-amputación (*days post amputation*, abreviado como *dpa* en cada una de las tablas). Las distintas condiciones son: recién amputados (0 dpa), 1 día después de la amputación (1 dpa) o 4 días después de la amputación (4 dpa).
- Largo de *reads*: largo promedio de las lecturas en cada muestra.

ID de muestra	Condición	Largo de <i>reads</i>
SRR11305406	0 dpa	51
SRR11305407	0 dpa	51
SRR11305410	0 dpa	51
SRR11305412	0 dpa	51
SRR11486778	0 dpa	51
SRR11305413	1 dpa	51
SRR11305414	1 dpa	51
SRR11305415	1 dpa	51
SRR11305408	4 dpa	51
SRR11305408	4 dpa	51

Tabla 2. Librerías generadas mediante RNA-Seq en el trabajo de Thompson y colaboradores (2020).

ID de muestra	Condición	Largo de <i>reads</i>
SRR11305426	0 dpa	51
SRR11305427	0 dpa	51
SRR11305428	0 dpa	51
SRR11305432	0 dpa	51
SRR11305433	0 dpa	51
SRR11305434	0 dpa	51
SRR11305435	1 dpa	51
SRR11305436	1 dpa	51
SRR11305437	1 dpa	51
SRR11305429	4 dpa	51

SRR11305430	4 dpa	51
SRR11305431	4 dpa	51

Tabla 3. Librerías generadas mediante ATAC-Seq en el trabajo de Thompson y colaboradores (2020).

5.1.2. Hardware

Las tareas asociadas al procesamiento de los archivos de secuenciación se llevaron a cabo en el computador de alto rendimiento *Exxact*, cuyo acceso fue generosamente otorgado por el Dr. Gonzalo Riadi. Un resumen de las especificaciones de este computador se detalla en la Tabla 4.

Modelo procesador	Intel(R) Xeon(R) CPU E7-8867 v3 @ 2.50GHz
Número de procesadores	128 (2x64)
Número de cores por procesador	16
Arquitectura del procesador	64 bits
Memoria RAM	252 GB
Capacidad disco duro	48 TB

Tabla 4. Especificaciones generales de *Exxact*.

5.1.3. Software

Todo el software que se utilizó durante el desarrollo de este trabajo se presenta en base a su nombre junto a su referencia, versión y utilidad en la Tabla 5.

Nombre (Referencia)	Versión	Utilidad
SRA Toolkit (Leinonen et al., 2011)	2.8.2	Colección de herramientas para manejar archivos de secuenciación provenientes de SRA.

FastQC (Andrews, 2010)	0.11.9	Herramienta para visualizar y comprobar la calidad de reads de un archivo <i>fastq</i> .
MultiQC (Ewels et al., 2016)	1.10.1	Herramienta que resume el análisis de calidad de lecturas generado con FastQC, de múltiples muestras, generando reportes <i>html</i> con las principales características.
Trimmomatic (Bolger et al., 2014)	0.36	Herramienta para pre-procesar reads en términos de <i>filtering</i> y <i>trimming</i> , según sus calidades.
TEcandidates (Valdebenito-Maturana & Riadi, 2018)	2.0.3	Herramienta para identificar el origen de expresión de TEs expresados.
SQulRE (Yang et al., 2019)	0.9.9.9a	Software para cuantificar la expresión de TEs a nivel de locus.
Bowtie2 (Langmead & Salzberg, 2012)	2.3.2	Alineador de secuencias ultrarrápido y eficiente en especial para alinear secuencias cortas.
SAMtools (Danecek et al., 2021)	1.12	Conjunto de herramientas para manipular archivos de alineamientos en formato SAM/BAM, dentro de las cuales se incluyen indexar, ordenar, entre otras.
MACS3 (Zhang et al., 2008)	2.2.7.1	Herramienta utilizada para procesar datos de ATAC-Seq.
HINT-ATAC (Li et al., 2019)	0.13.1	HINT-ATAC es parte de <i>Regulatory Genomic Toolbox</i> (RGT), una colección de herramientas para el análisis de elementos reguladores. En concreto, HINT-ATAC se utiliza para identificar sitios de unión de factores de transcripción activos.

BEDTools (Quinlan & Hall, 2010)	2.30	Conjunto de herramientas para comparar y analizar intervalos genómicos.
IGV (Robinson et al., 2011)	2.9.4	Explorador visual interactivo de datos genómicos.

Tabla 5. Software utilizado.

Adicionalmente, se utilizó el lenguaje de programación estadística *R* (R Core Team, 2020) en su versión 4.0.5. En él se usaron diversas colecciones de funciones denominadas paquetes. Los principales paquetes de *R* utilizados se detallan en la Tabla 6.

Nombre (Referencia)	paquete	Versión	Utilidad
ATACseqQC (Ou et al., 2018)		1.14.4	Herramienta para realizar análisis de calidad de datos ATAC-Seq.
DESeq2 (Love et al., 2014)		1.30.1	Herramienta para el análisis de expresión génica diferencial.
ggplot2 (Wickham, 2016)		3.3.3	Librería para la creación de gráficos.
ggvenn (Linlin Yan, 2020)		0.1.8	Librería para la creación de gráficos de tipo Venn.
Sushi (Phanstiel et al., 2014)		1.30.0	Herramienta para visualizar información genómica.

Tabla 6. Paquetes de *R* utilizados.

5.2. Métodos

Los procedimientos para completar los objetivos utilizando los materiales anteriormente mencionados, se describen a continuación. Se comienza con un proceso importante que es la preparación de los archivos utilizados y su respectiva

comprobación de estado antes de comenzar los distintos análisis. Posteriormente, se indicará una metodología para cada objetivo específico planteado que se realizó.

5.2.1. Métodos para preparación de archivos

De manera general, a partir de los IDs de SRA mostrados en la Tabla 2 y Tabla 3, se descargaron los respectivos archivos mediante el software *SRA Toolkit*. Una vez descargados, se comprobó la calidad de estos, mediante *FastQC* y *MultiQC*. Según la calidad de los reads y/o presencia de adaptadores, se realizó un proceso de *filtering* (filtrar los reads de acuerdo con una calidad promedio mínima establecida) y *trimming* (recortar los adaptadores de los reads si es que estos estuvieran aún presentes), a fin de obtener un set de reads de buena calidad. Por otro lado, se procedió a descargar los archivos de anotación, genoma y elementos repetidos del organismo *Danio rerio*. Un esquema general de esta etapa se muestra en la Figura 13 en el apartado de color azul.

En detalle los pasos que se realizaron en esta etapa son:

- **i. Descarga de archivos fastq, genoma y anotación de elementos repetidos**
 - Descarga de los archivos “*fastq*” según su *id* de muestra mediante la herramienta *fastq-dump* desde la base de datos SRA.
 - Descarga del archivo de genoma de referencia de *Danio rerio* versión GRCz11 en formato FASTA y sus respectivos archivos de anotación de genes y de elementos repetidos.
 - Se comprobó la integridad de los archivos descargados, basándose en la estructura de un archivo con extensión “*fastq*”. Se realizó un script en Bash, que comprobó la cantidad de líneas de los archivos dividida en 4 versus el número de reads indicado en SRA para la respectiva muestra. Si se descargaron todos los reads, los valores

deberán ser iguales, de manera contraria el script pudo notificar que librería estaba mal descargada.

- **ii. Análisis de calidad**
 - Una vez descargados y comprobados los archivos, se procedió a realizar el análisis de calidad de estos. Esto se hizo con las herramientas *FastQC* y *MultiQC*, con lo que se obtuvo un reporte en formato *html*, con las principales estadísticas y calidades de las lecturas. Además, se realizó la comprobación si los archivos traen adaptadores de secuenciación en sus extremos.
- **iii. Limpieza de adaptadores (*Trimming*)**
 - Si se da el caso de existencia de adaptadores de secuenciación en los extremos de las lecturas, estos serán eliminados con la herramienta *Trimmomatic*.
- **iv. Eliminación de reads de mala calidad (*Filtering*)**
 - Así también, se filtraron los reads según una calidad promedio mínima establecida. Dicha calidad se obtiene utilizando el nivel de calidad *Phred* y estará establecida, en un inicio, en 30.

5.2.2. Métodos para objetivo específico N°1

Una vez pre-procesados los archivos, se procedió a utilizarlos en la realización de los distintos objetivos según corresponda. Para el objetivo específico N°1 se buscó identificar TEs transcripcionalmente activos, por lo que para este objetivo se seleccionaron los archivos correspondientes a datos de RNA-Seq. Una vez seleccionados, se realizó la identificación de TEs transcripcionalmente activos. Esta identificación se realizó mediante la ejecución de dos software en paralelo. Por un lado, la ejecución de *SQUIRE*, que entregó una cuantificación de los TEs expresados y *TEcandidates* que reportó el origen de la expresión de estos elementos. Los resultados obtenidos se intersectaron con el software *BEDTools*,

con la finalidad de tener un archivo de TEs identificados por cada condición a evaluar. Finalmente, estos archivos fueron utilizados en el paquete de *R*, *DESeq2*, para la realización de un análisis de expresión diferencial. Estos resultados se visualizaron también mediante el paquete *ggplot2* de *R*. Un esquema general asociado a los métodos del objetivo específico N°1, se visualiza en la Figura 13 en color verde.

El protocolo que se realizó para este método es el siguiente:

- **1.1. Identificación de TEs transcripcionalmente activos**

- Se usaron los archivos de RNA-Seq como entrada para *SQUIRE* y *TEcandidates*.
- Para la identificación de TEs a un nivel cuantitativo, *SQUIRE* recibió como datos de entrada estos archivos. Este software cuenta con una serie de módulos o etapas. La primera utilizada fue *squire Fetch*, para preparar los archivos. La siguiente etapa que se empleó fue la de *squire Clean*, que se encargó de filtrar el archivo de elementos repetidos ingresado, con la finalidad de obtener un archivo con las coordenadas solo de los TEs. Posteriormente, se ejecutó la etapa de alineamiento de reads al genoma mediante *squire Map*. Finalmente, la última etapa realizada fue la de cuantificación, en donde el módulo *squire Count*, se encargó de cuantificar los TEs entregando un archivo de salida de recuento tanto de genes como de estos elementos.
- A su vez, el software *TEcandidates* permitió identificar el origen de la expresión de estos TEs. Para su funcionamiento se usaron los archivos correspondientes al genoma de referencia de *Danio rerio*, los archivos de reads de las muestras a evaluar, y el archivo de anotación de TEs que otorga *SQUIRE* en el paso anterior, a fin de tener resultados comparables. *TEcandidates* entregó un archivo de los TEs candidatos en formato *bed*.

- **1.2. Filtro de archivo de TEs según clasificación genómica**
 - Si bien el archivo de TEs entregado por la etapa de *squire Clean* contenía en su mayoría este tipo de elementos, también estaba la presencia de otro tipo de elementos repetidos, por lo que se realizó un nuevo filtro de estos. En este nuevo filtro, realizado en *bash*, solo se consideró conservar los TEs de tipo DNA, LTR, LINE y SINE.
 - El segundo filtro que se realizó fue el de TEs según su posición dentro del genoma de *Danio rerio*. Únicamente se conservaron los TEs que estuvieran intersectando con regiones intergénicas e intrónicas de genes, a modo de discriminar TEs que pudiesen estar en regiones exónicas y que se estuvieran transcribiendo gracias a la maquinaria de dichas regiones. Este archivo se obtuvo utilizando el archivo de TEs previamente filtrado según el punto anterior, intersectado con el genoma de *Danio rerio*, todo esto, utilizando el software *BEDTools*, mediante el módulo *bedtools intersect*.
- **1.3. Intersección de TEs encontrados**
 - Los archivos resultantes, tanto por *SQUIRE* y *TEcandidates*, fueron intersectados según la condición temporal particular a analizar, con el software *BEDTools*, mediante el módulo *bedtools intersect*. Asimismo, se realizaron gráficos de tipo Venn, para poder visualizar la cantidad de TEs que identificaron ambos software y la intersección entre ellos.
- **1.4. Análisis de expresión diferencial**
 - Una vez se obtuvo el archivo final intersectado, se efectuó el análisis de expresión diferencial. Este análisis se realizó en el entorno y lenguaje de programación *R*. Dentro de *R*, se utilizó el paquete *DESeq2*, en donde se procesaron los archivos de acuerdo con su identificador de muestra, condición y cantidad de TEs anotados. Una vez aquí, estos se agruparon según condición y se realizó al análisis utilizando las condiciones temporales a estudiar.
 - Finalmente, para visualizar todos los resultados, se utilizó *ggplot2*. En dicha herramienta se realizaron gráficos del tipo “*volcano plot*” para

ver TEs expresados diferencialmente identificados por *SQUIRE*, así como también los identificados por *SQUIRE* y *TEcandidates* (intersección), esto a modo de poder visualizar el contraste de ambos software. En este paso también se realizó el análisis de genes con algún cambio de expresión diferencial.

5.2.3. Métodos para objetivo específico N°2

Para la realización de este objetivo, el cual consiste en identificar sitios de unión a factores de transcripción, se utilizaron los archivos correspondientes a la metodología ATAC-Seq. Los archivos FASTQ de ATAC-Seq se alinearon contra el genoma de referencia de *Danio rerio* con el software *Bowtie2*. El archivo resultante de este alineamiento se indexó y ordenó con *Samtools*. Posteriormente, se analizaron las regiones accesibles a la cromatina con la utilización del software *MACS2*. El archivo resultante, que contenía las regiones enriquecidas de reads (peaks), fue intersectado con los TEs anotados. Finalmente, el resultado de la intersección de *MACS3* con los TEs, junto con el genoma de referencia fueron utilizados para la identificación de sitios de unión a factores de transcripción mediante *HINT-ATAC*. Un esquema general de los métodos de este objetivo se visualiza en la Figura 13 en color naranja.

El protocolo por seguir para este método es el siguiente:

- **2.1. Alineamiento de reads contra el genoma de referencia**
 - Con cada librería correspondiente a ATAC-Seq se realizó el alineamiento contra el genoma de referencia. Este proceso se ejecutó con el software *Bowtie 2*. Primero, se llevó a cabo la construcción del índice del genoma con *bowtie2-build*. Luego, se realizó el alineamiento de los reads al genoma con *bowtie2*. Este alineamiento dio como resultado un archivo en formato *SAM* (*Sequence Alignment Map*).

- **2.2. Indexación y orden del archivo de alineamiento**
 - El archivo *SAM* debió ser indexado, ordenado y convertido a *BAM* para posteriormente poder trabajar con él. Este proceso se realizó con el software *Samtools*, el cual por medio de *samtools sort* (para ordenar el archivo) entregó un archivo en formato *BAM* (Binary Alignment Map), el que se indexó con *samtools index*. Adicionalmente, en este paso se efectuó un análisis de distintas métricas post-alineamiento, con el paquete *ATACseqQC* de R.
- **2.3. Identificación de áreas enriquecidas**
 - Posteriormente se identificaron las zonas de enriquecimiento genómico por medio del software *MACS3*. Este software, en base a los archivos *BAM* anteriormente generados con el alineamiento, entregó un conjunto de archivos con las zonas más enriquecidas, las cuales se podrán ver a más detalle con el visualizador *IGV* o el paquete de R, *Sushi*.
- **2.4. Intersección de áreas enriquecidas con TEs identificados**
 - Luego de obtener los archivos con las zonas más enriquecidas, estos se interseccionaron con el archivo de TEs anotados utilizando el software *BEDTools*. Esto con la finalidad de obtener la identificación solo de estos elementos presentes en estas zonas específicas.
- **2.5. Identificación de sitios de unión a factores de transcripción**
 - Para la identificación de estos sitios se utilizó el software *HINT-ATAC*. Primeramente, con el algoritmo *hint-atac footprinting* se detectaron las regiones pertenecientes a los peaks en donde potencialmente se podían unir factores de transcripción. Esto se realizó utilizando el archivo obtenido en el punto anterior correspondiente al de *peaks* interseccionado con TEs y a los correspondientes archivos *BAM*, generados por el alineamiento de cada una de las librerías de reads. Este nuevo archivo de salida de regiones footprints, volvió a interseccionarse con el archivo de TEs filtrados anotados, para nuevamente considerar regiones pertenecientes a TEs.

- El archivo correspondiente a las regiones footprints identificadas, fue el input para ejecutar el segundo algoritmo de HINT-ATAC utilizado, el cual fue el de *hint-atac motifanalysis*. Con este pasó se obtuvo los motivos encontrados en estas regiones, asociados a sitios de unión a factores de transcripción. Este procedimiento entregó una tabla con cada uno de los motivos encontrados asociados a un TF en particular, en cada región anotada en el archivo *bed* de footprints.

5.2.4. Métodos para objetivo específico N°3

El método que se utilizó para el último objetivo específico está determinado por la asociación de los datos obtenidos anteriormente, buscando una relación entre estos y con genes implicados en el proceso regenerativo de *Danio rerio*. Para esto, se utilizó *bedtools* y la creación e implementación de diversos scripts en *Bash* para poder analizar la totalidad de datos. Finalmente, tanto para los TEs como los genes encontrados, se realizó una búsqueda bibliográfica a fin de comprender su relación o no con el proceso regenerativo en *Danio rerio* y exponer conclusiones de lo encontrado. Un esquema general de los métodos de este objetivo se visualiza en la Figura 13 en color morado.

El protocolo que se siguió para este método se subdividió según los resultados obtenidos por cada uno de los objetivos planteados. En resumen, los pasos fueron los siguientes:

- **3.1. Relacionar TEs con expresión diferencial con genes cercanos con expresión diferencial**
 - Una vez obtenidos todos los archivos y resultados del objetivo específico N°1, se buscó relacionar los TEs con los genes identificados, en el caso de que ambos estuviesen sufriendo un cambio de expresión diferencial. Esta relación se realizó en base a un

criterio de cercanía (< 5000 pb). Este procedimiento se efectuó utilizando el software *BEDTools* mediante el comando *bedtools closest*.

- **3.2. Identificar TEs que presenten TFBSs**

- Con respecto a los resultados del objetivo específico N°2, se asociaron los TEs que habían sido detectados con presencia de motivos de sitios de unión a factores de transcripción, al evento regenerativo. Dentro de este procedimiento, se obtuvo desde la base de datos ZFIN, una lista de 138 registros correspondientes a factores de transcripción relacionados a regeneración. Seguidamente, estos identificadores se buscaron en los archivos entregados por HINT-ATAC, a modo de filtrar este archivo según los motivos de interés.

- **3.3. Búsqueda en literatura y bases de datos biológicas**

- Finalmente, para ambos resultados se realizó una búsqueda exhaustiva en literatura y en distintas bases de datos biológicas, tanto de TEs como de genes, con la finalidad de determinar y analizar la participación de estos elementos en los eventos de regeneración de *Danio rerio*. La búsqueda bibliográfica se centro en plataformas web centradas en el organismo *Danio rerio* como la base de datos ZFIN.
- Además, se realizaron representaciones de las redes regulatorias de los productos proteicos de cada gen identificado, destacando el potencial rol del TE identificado en cualquiera de los casos. Dichas representaciones fueron realizadas utilizando la base de datos STRING, mediante su herramienta web.

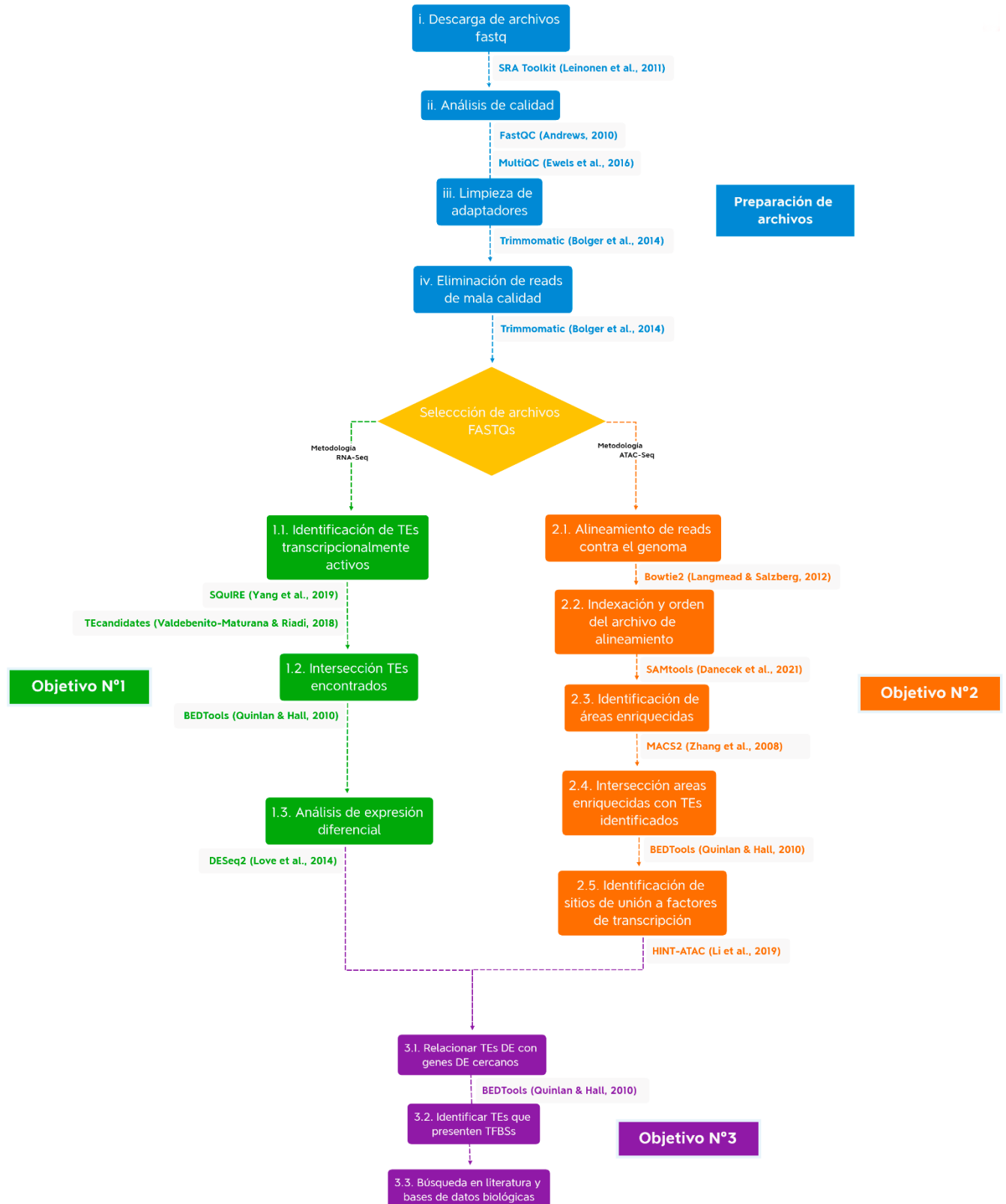


Figura 13. Esquema general de la metodología propuesta.

6. Resultados

Los resultados del presente trabajo se subdividirán según los objetivos específicos anteriormente planteados, esto con la finalidad de mantener un orden elaborado al planteamiento previo de la propuesta presentada.

6.1. Pre-procesamiento de archivos

Previo a ejecutar todos los procesos detallados en la metodología y concretamente en este caso para el objetivo específico N°1, se procedió a ejecutar un pre-procesamiento de los archivos a utilizar. Esto con la finalidad de poder trabajar con archivos sin problemas, con una calidad aceptable y debidamente filtrados y/o ordenados para poder reducir en una mayor medida los resultados artefactuales.

6.1.1. Análisis de calidad

Como paso principal para realizar todos los análisis de esta investigación, y luego de la descarga y comprobación de todas las librerías a utilizar, se realiza un análisis de calidad de estas con las herramientas *FastQC* (Andrews, 2010) y posteriormente *MultiQC* (Ewels et al., 2016).

Los resultados del análisis de calidad de cada una de las librerías entregados por *FastQC*, son tomados como *input* para un análisis general global de *MultiQC*. Este procedimiento se ejecutó para todas las librerías RNA-Seq y ATAC-Seq por separado.

En particular, el criterio para evaluar la calidad se centró principalmente en el puntaje de calidad media y en el puntaje de calidad por secuencia o *read*. Los

gráficos de calidad asociados a cada tipo de metodología se encuentran en el apartado Anexo, en la Figura Suplementaria 1 y Figura Suplementaria 2 para los datos de la metodología RNA-Seq y en la Figura Suplementaria 3 y Figura Suplementaria 4 para los datos de metodología ATAC-Seq.

Concretamente, los criterios que se emplearon para el proceso de filtering y trimming de las librerías fueron la calidad y el largo mínimo del *read*. Se eliminaron reads cuya calidad promedio fuera menor a 30 *Phred*, con esto se alcanza un nivel de precisión de que la base secuenciada sea la correcta en un 99.9% (1 entre 1000), lo que determinará que en promedio el read tenga esta precisión por base. A su vez se realizó un proceso de trimming, dejando reads con un largo mínimo de 35 pb, ya que según estudios como el de Valdebenito-Maturana y colaboradores (2021), este es largo mínimo de read para poder analizar TEs considerando su característica de multimapeo en el genoma. Este proceso se realizó con el software *Trimmomatic* (Bolger et al., 2014).

Las estadísticas finales de las librerías a utilizar se detallan en la Tabla 7 para las librerías de RNA-Seq y en la Tabla 8 para las librerías de ATAC-Seq.

ID librería	Condición	Largo promedio de reads pre-filtering	Largo promedio de reads post-filtering	Número reads pre-filtering	Número reads post-filtering
SRR11305406	0dpa	51 pb	48 pb	41.862.747	40.900.347
SRR11305407	0dpa	51 pb	46 pb	32.921.356	32.171.309
SRR11305410	0dpa	51 pb	48 pb	55.897.313	55.541.279
SRR11305411	0dpa	51 pb	47 pb	58.035.513	57.621.402
SRR11305412	0dpa	51 pb	48 pb	60.689.620	60.292.209
SRR11486778	0dpa	51 pb	45 pb	47.303.554	47.048.792

SRR11305413	1dpa	51 pb	49 pb	59.877.496	59.466.669
SRR11305414	1dpa	51 pb	49 pb	61.944.645	61.538.014
SRR11305415	1dpa	51 pb	48 pb	57.334.335	56.950.187
SRR11305408	4dpa	51 pb	47 pb	49.817.955	48.762.654
SRR11305409	4dpa	51 pb	47 pb	50.473.039	49.433.739

Tabla 7. Estadísticas finales para librerías RNA-Seq.

ID librería	Condición	Largo promedio de reads pre-filtering	Largo promedio de reads post-filtering	Número reads pre-filtering	Número reads post-filtering
SRR11305426	0dpa	51 pb	42 pb	33.377.690	32.924.370
SRR11305427	0dpa	51 pb	42 pb	31.684.061	31.248.330
SRR11305428	0dpa	51 pb	42 pb	39.692.002	39.126.185
SRR11305432	0dpa	51 pb	41 pb	50.567.154	49.941.970
SRR11305433	0dpa	51 pb	41 pb	42.494.276	41.935.091
SRR11305434	0dpa	51 pb	42 pb	43.675.713	43.106.246
SRR11305435	1dpa	51 pb	44 pb	39.999.034	39.428.051
SRR11305436	1dpa	51 pb	43 pb	91.937.674	90.680.214
SRR11305437	1dpa	51 pb	43 pb	62.627.105	61.815.126
SRR11305429	4dpa	51 pb	41 pb	35.805.937	35.323.050
SRR11305430	4dpa	51 pb	43 pb	147.632.283	145.743.770
SRR11305431	4dpa	51 pb	42 pb	35.838.887	35.377.379

Tabla 8. Estadísticas finales de librerías ATAC-Seq.

Posterior a la aplicación del *filtering* y *trimming* aplicado con *Trimmomatic* los gráficos de calidad promedio y por secuencia se visualizan en la Figura Suplementaria 5 y Figura Suplementaria 6 para las librerías RNA-Seq y en la Figura Suplementaria 7 y Figura Suplementaria 8 para librerías ATAC-Seq.

6.1.2. Clasificación de TEs anotados

Dado que los TEs se pueden encontrar en distintas regiones, ya sea regiones intergénicas, intrónicas o exónicas, se procedió a clasificarlos en esas categorías de acuerdo a su ubicación relativa a genes. Para esto en una primera etapa, se filtró el archivo de elementos repetidos de *Danio rerio*, con la finalidad de obtener solo los elementos transponibles de tipo DNA, LTR, SINE y LINE. Adicionalmente a esto, se tomó la decisión de filtrar este archivo de anotación de TEs disponibles, separando y trabajando solamente con aquellos TEs que no intersectaran con regiones exónicas. Esta decisión fue tomada para evitar confusiones en los casos en donde el gen asociado a un TE exónico se está expresando, causando erróneamente que el TE fuese reportado como expresado también. Esta nueva anotación, llamada para fines prácticos como “Self-expressed TEs”, contiene únicamente TEs que intersectan con regiones intergénicas e intrónicas del archivo de anotación de *Danio rerio*. Las estadísticas de esta nueva clasificación de TEs se muestran en la Tabla 9.

Tipo	Cantidad
Elementos repetidos anotados en <i>Danio rerio</i>	2.556.853
TEs anotados en <i>Danio rerio</i> (DNA, LRT, SINE, LINE)	2.440.919
TEs intrónicos anotados	1.378.522
TEs intergénicos anotados	1.032.059

Tabla 9. Estadísticas de TEs anotados.

6.2. Resultados objetivo específico N°1

Concretamente, para los resultados del objetivo específico N°1, la expresión de TEs se analizó utilizando los software *SQUIRE* y *TEcandidates*, según lo reportado en el apartado de Metodología. A estas estimaciones de TEs transcritos en cada una de las etapas temporales, se agregó la de genes para poder analizar si presentan algún cambio de expresión diferencial. En este apartado se reportan los análisis por cada una de las etapas, así como las correspondientes comparaciones temporales realizadas.

6.2.1. Cantidad de TEs transcritos

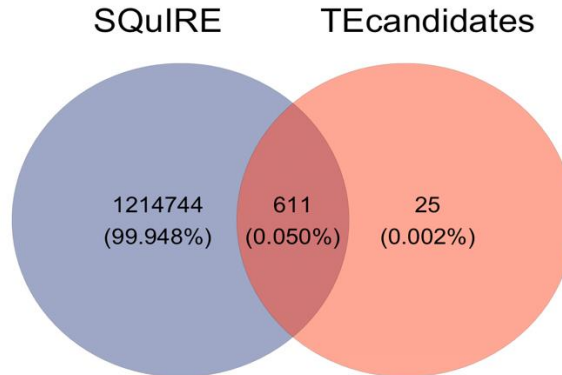
Una vez clasificados los TEs de acuerdo a su posición relativa a un gen, se realizó la identificación de TEs transcritos utilizando los softwares *SQUIRE* y *TEcandidates*. La cantidad y el traslape de los TEs identificados por cada software se muestran en la Figura 14, según las condiciones temporales de las librerías empleadas (0dpa, 1 dpa y 4dpa respectivamente).

Para fines prácticos, en la visualización de la expresión diferencial de estos elementos, se incluirán los TEs identificados tanto por *SQUIRE* como por *TEcandidates*, con tal de contrastar la diferencia de sensibilidad en ambos software. Se ha reportado que el traslape de estos resultados puede ayudar a identificar con mayor precisión a los TEs expresados a un nivel de locus (Valdebenito-Maturana et al., 2021).

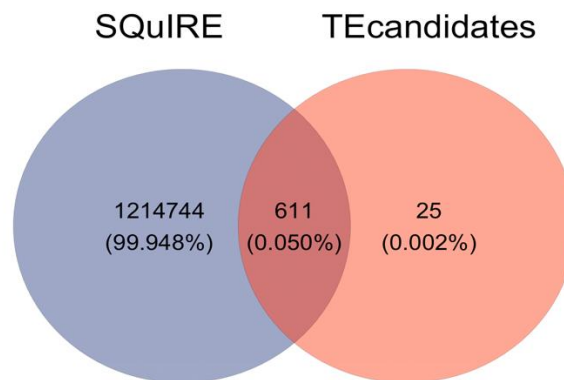
Una representación de todos los elementos a utilizar y como se distribuyen a lo largo de las condiciones temporales, se visualiza en la Figura Suplementaria 9.

TEs transcritos según condición

A) 0dpa



B) 1dpa



C) 4dpa

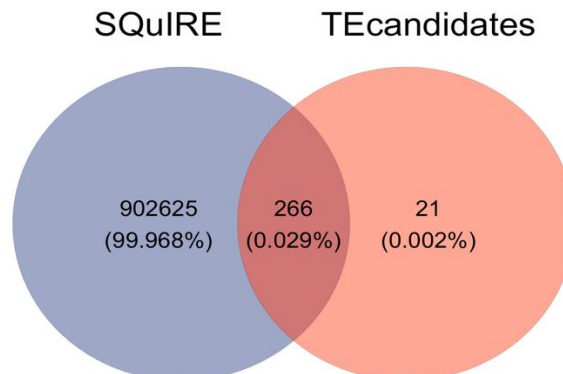


Figura 14. Diagrama de Venn correspondiente a identificación de TEs transcritos por ambos software en todas las condiciones temporales. En A, se visualizan los TEs transcritos para la condición 0dpa, en B para los TEs transcritos en 1dpa y en C para los 4dpa. En todas estas,

el círculo azul representa la cantidad y proporción total de TEs identificados por el software *SQUIRE*, en naranja los TEs identificados por *TEcandidates* y al medio de estos el traslape de TEs identificados entre ambos software.

6.2.2. Análisis de expresión diferencial de TEs y genes

Ya realizado el proceso de conteo de reads de RNA-Seq por TEs y genes en cada condición temporal, se procedió a realizar el análisis de expresión diferencial, tanto para TEs (identificados por ambos software y solo por *SQUIRE*) y para genes (identificados por *SQUIRE*). Las comparaciones realizadas fueron 0dpa vs 1dpa, 0dpa vs 4dpa y 1dpa vs 4dpa.

Los análisis de expresión diferencial se realizaron con la librería de R, *DESeq2*, y los resultados se muestran en gráficos del tipo *Volcano plot* utilizando líneas punteadas para los valores de corte de *p-value* ajustado y de $\log_2(\text{Fold Change})$. Para el *p-value* ajustado, se considera un punto de corte en el gráfico de aproximadamente 1.3, es decir un *p-value* ajustado de 0.05. En otras palabras, se controla que el porcentaje de falsos positivos sea de 5%. Para el $\log_2(\text{Fold Change})$ de 2: se considera un gen o TE como sobre-expresado si su $\log_2(\text{Fold Change})$ es mayor o igual que 2, y como sub-expresado si su $\log_2(\text{Fold Change})$ es menor o igual que -2. Esto corresponde a cambios de expresión de al menos 4 veces.

Los resultados de los análisis de expresión diferencial de TEs se muestran en la Figura 15 para la comparación 0dpa vs 1dpa, en la Figura 17 para 0dpa vs 4dpa y en la Figura 19 para 1dpa vs 4dpa. En dichas figuras se representan mediante círculos de color rojo tenue, los TEs sobre-expresados identificados por *SQUIRE* y en un color rojo fuerte los TEs sobre-expresados identificados por ambos software (*SQUIRE* y *TEcandidates*). Asimismo, los círculos con un color azul tenue se visualizan los TEs sub-expresados identificados por *SQUIRE* y en un color azul fuerte los TEs sub-expresados, identificados por ambos software. Finalmente, también se muestran los elementos que no presentan ningún cambio de expresión

diferencial. Estos se visualizan por medio de círculos con un color gris claro, para los TEs que no tienen un cambio de expresión diferencial identificados por *SQuIRE* y un color gris oscuro para los TEs que no tienen un cambio de expresión diferencial identificados por ambos software.

Por otro lado, los resultados relacionados a expresión diferencial de genes se representan en la Figura 16 para la comparación 0dpa vs 1dpa, en la Figura 18 para 0dpa vs 4dpa y en la Figura 20 para 1dpa vs 4dpa. En dichas figuras se representan mediante círculos rojos los genes sobre-expresados, en azul los genes sub-expresados y en gris los genes que no presentaron un cambio de expresión diferencial. Esta paleta de colores y su correspondiente representación se mantiene para la totalidad de gráficos de expresión diferencial tanto de TEs como de genes.

Análisis de expresión diferencial de self-expressed TEs en aleta caudal de *Danio rerio*. 0dpa vs 1dpa

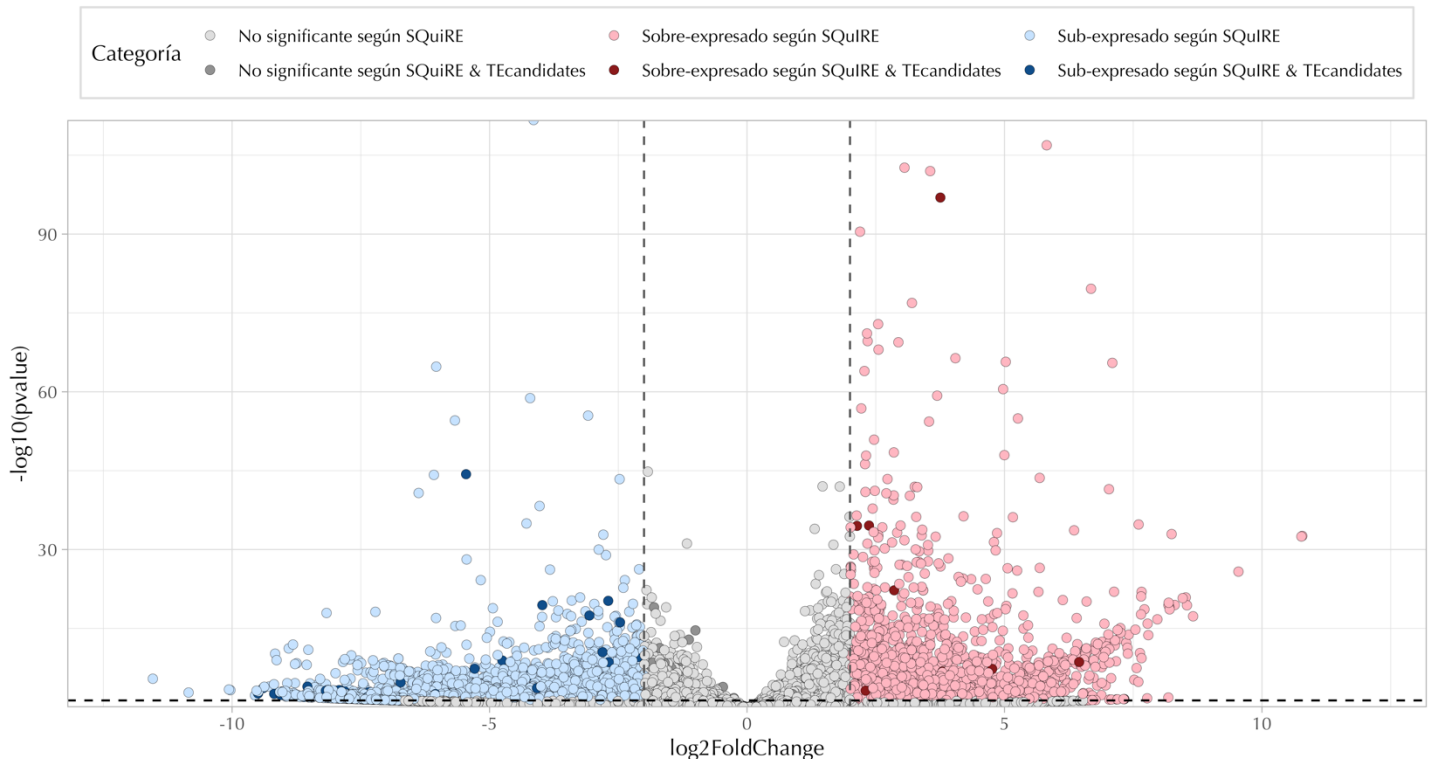


Figura 15. Análisis de expresión diferencial de *Self-expressed* TEs en aleta caudal de *Danio rerio*, comparando las condiciones 0dpa v/s 1dpa.

Las estadísticas de los TEs transcritos identificados por ambos software para la comparación temporal 0dpa vs 1dpa se visualizan en la Tabla 10.

TEs expresados para condición 0dpa vs 1dpa		
	<i>SQUIRE</i>	<i>SQUIRE & TEcandidates</i>
TEs sin cambio significativo	46.675	245
TEs sobre-expresados	7.018	15
TEs sub-expresados	2.632	109

Tabla 10. Cantidad de Self-expressed TEs expresados en 0dpa vs 1dpa.

De estos TEs, se seleccionaron los que tenían un cambio de expresión diferencial relacionado a una sobre-expresión detectada por ambos software, es decir 15 TEs.

De estos 15 TEs, 10 son de tipo DNA (66.7%), 3 de tipo LTR (20%) y 2 de tipo LINE (13.3%). Esta elección se realizó para poder asociar la sobre-expresión de estos TEs con genes con un cambio de expresión diferencial durante el proceso regenerativo, de acuerdo con la cercanía de ambos. Dichos resultados, tanto para esta condición como para las siguientes que se evaluaron, se encuentran en el apartado 6.4.

Análisis de expresión diferencial de genes en aleta caudal de *Danio rerio*. 0dpa vs 1dpa

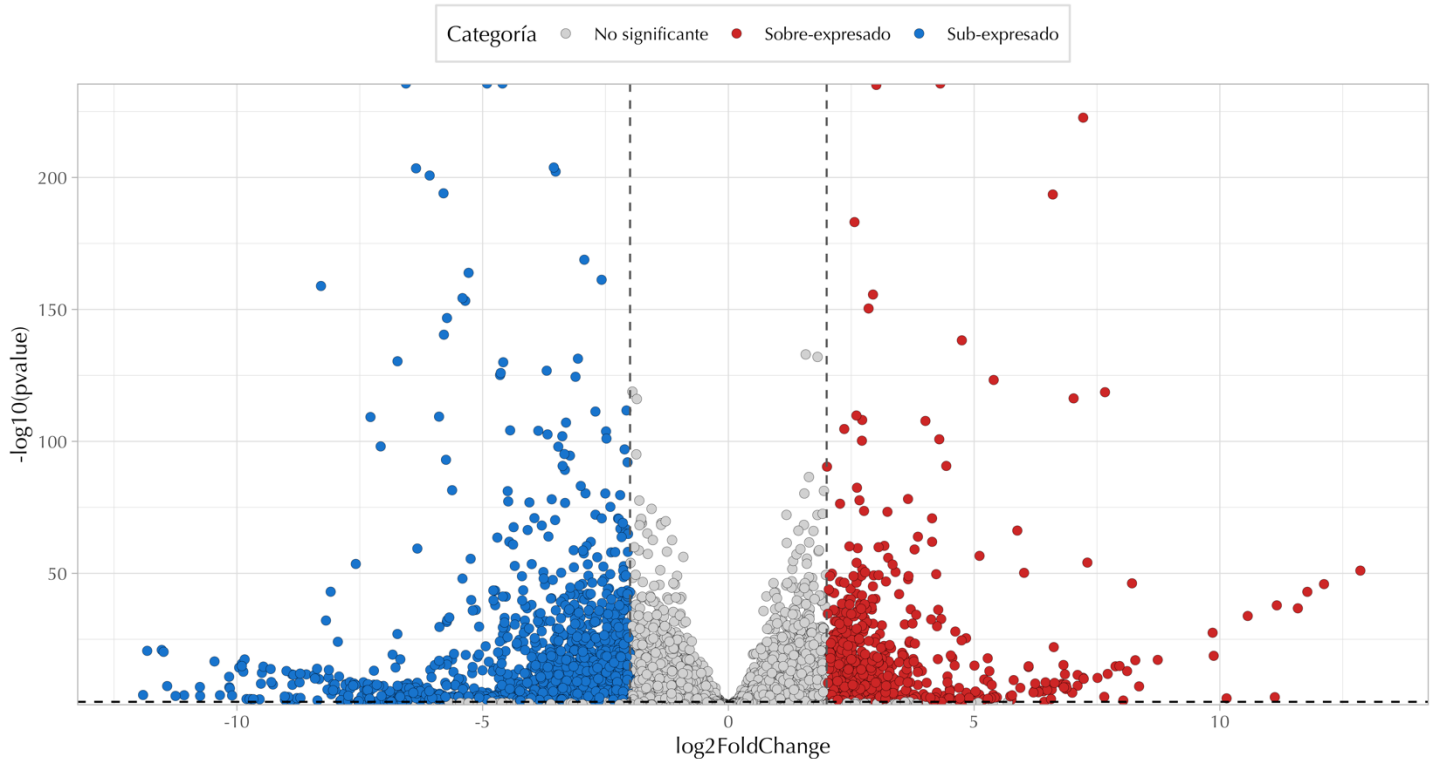


Figura 16. Análisis de expresión diferencial de genes en aleta caudal de *Danio rerio*, comparando las condiciones 0dpa v/s 1dpa.

Las estadísticas asociadas a la cantidad de genes identificados se pueden ver en la Tabla 11.

Genes expresados para condición 0dpa vs 1dpa		
Genes sin cambio significativo	Genes sobre-expresados	Genes sub-expresados
13.344	589	1506

Tabla 11. Cantidad de genes expresados en 0dpa vs 1dpa.

La siguiente comparación temporal por analizar fue la de 0dpa vs 4dpa. Los gráficos de expresión diferencial tanto de TEs como de genes se visualizan en las Figura 17 y Figura 18 respectivamente. Así también, las estadísticas asociadas a

las cantidades de TEs y genes con expresión diferencial obtenidos se pueden visualizar en la Tabla 12 para los TEs y en la Tabla 13 para los genes.

Análisis de expresión diferencial de self-expressed TEs en aleta caudal de *Danio rerio*. 0dpa vs 4dpa

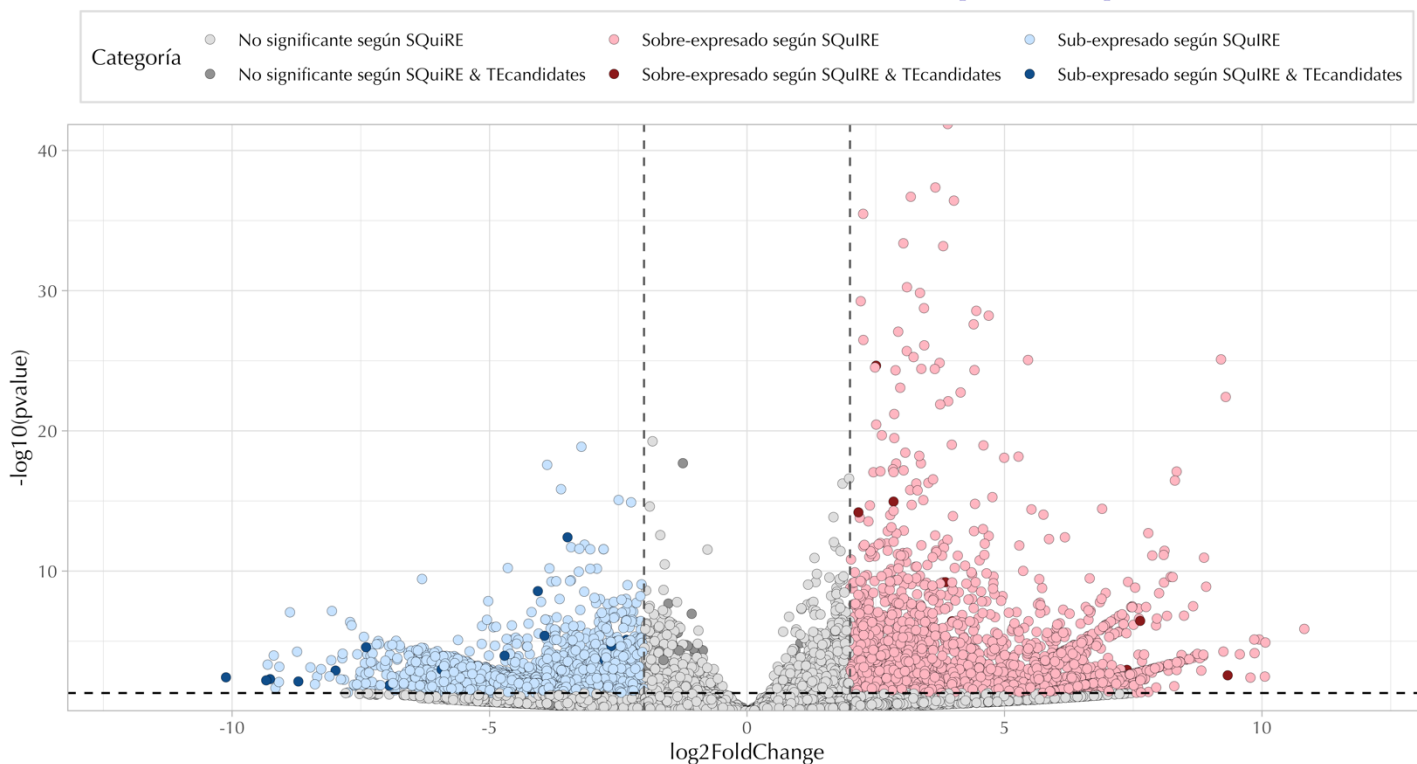


Figura 17. Análisis de expresión diferencial de Self-expressed TEs en aleta caudal de *Danio rerio*, comparando las condiciones 0dpa v/s 4dpa.

TEs expresados para condición 0dpa vs 4dpa		
	<i>SQuIRE</i>	<i>SQuIRE & TEcandidates</i>
TEs sin cambio significativo	51.676	342
TEs sobre-expresados	2.720	19
TEs sub-expresados	4.853	33

Tabla 12. Cantidad de Self-expressed TEs expresados en 0dpa vs 4dpa.

De estos 19 TEs, 11 son de tipo DNA (58%), 5 de tipo LTR (26%) y 3 de tipo LINE (16%). Del mismo modo que el análisis de expresión anterior, estos TEs se analizaron según su distancia a genes con algún tipo de cambio de expresión diferencial.

Análisis de expresión diferencial de genes en aleta caudal de *Danio rerio*. 0dpa vs 4dpa

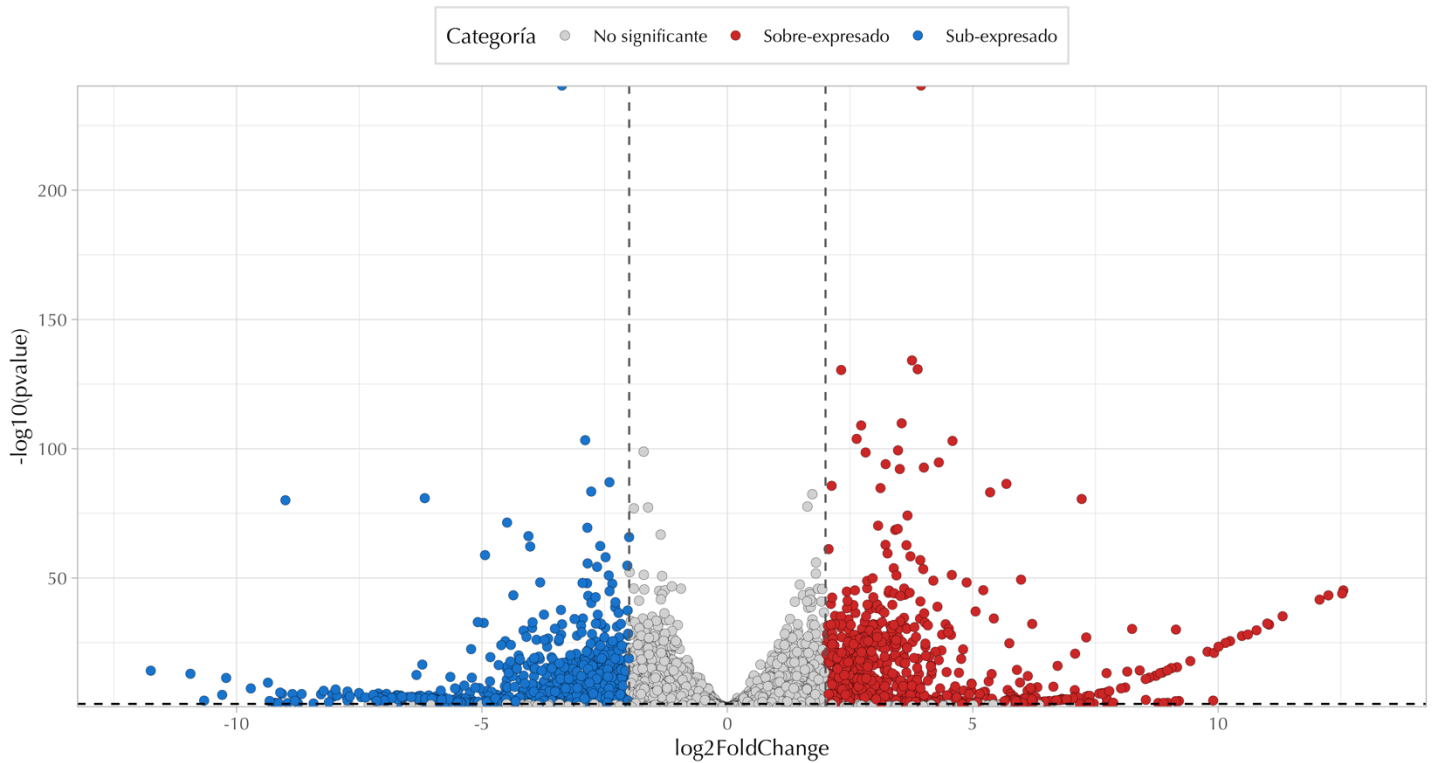


Figura 18. Análisis de expresión diferencial de genes en aleta caudal de *Danio rerio*, comparando las condiciones 0dpa v/s 4dpa.

Genes expresados para condición 0dpa vs 4dpa		
Genes sin cambio significativo	Genes sobre-expresados	Genes sub-expresados
13.646	811	982

Tabla 13. Cantidad de genes expresados en 0dpa vs 4dpa.

Finalmente, la última comparación temporal por analizar fue la de 1dpa vs 4dpa. Los gráficos de expresión diferencial tanto de TEs como de genes se visualizan en la Figura 19 y Figura 20 respectivamente. Del mismo modo, las estadísticas asociadas a las cantidades de TEs y genes con expresión diferencial obtenidos se pueden visualizar en la Tabla 14 para los TEs y en la Tabla 15 para los genes.

Análisis de expresión diferencial de self-expressed TEs en aleta caudal de *Danio rerio*. 1dpa vs 4dpa

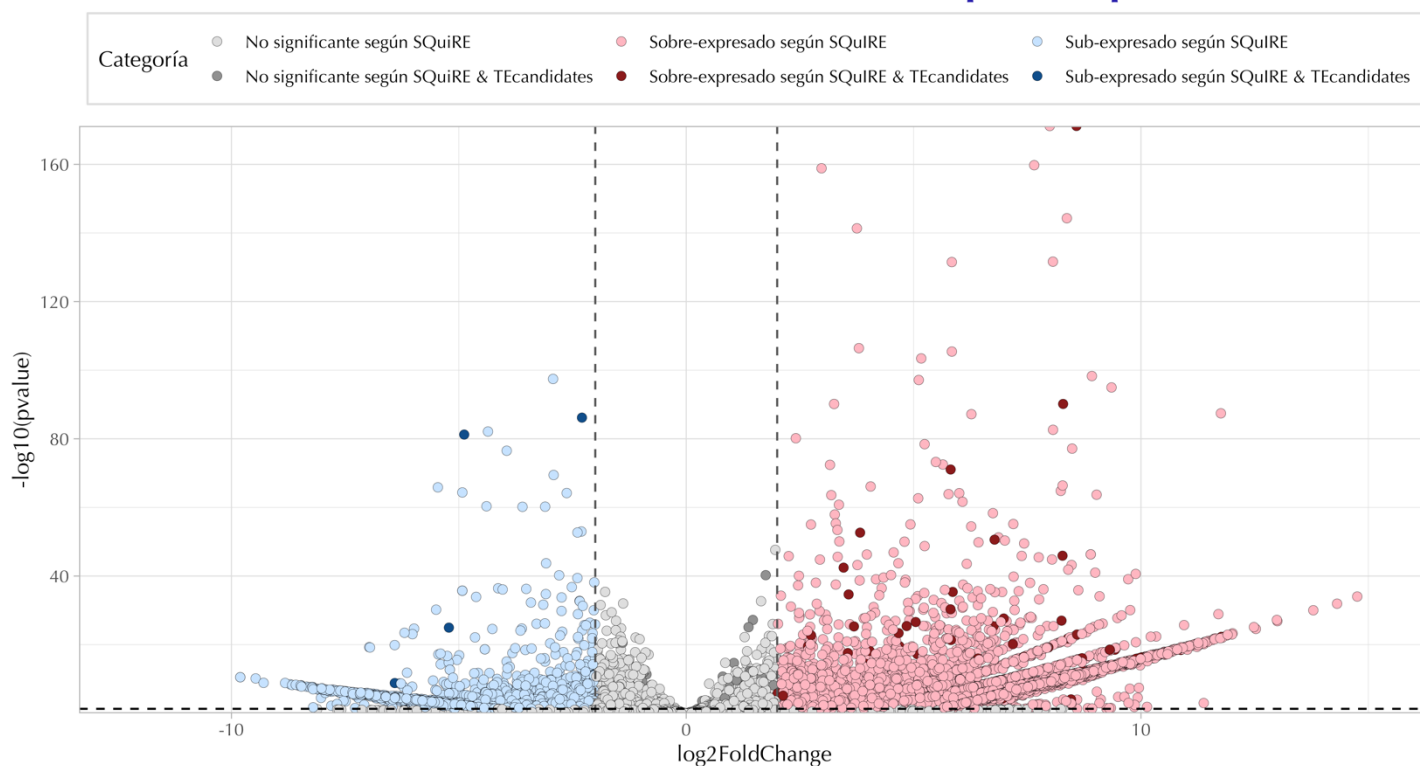


Figura 19. Análisis de expresión diferencial de Self-expressed TEs en aleta caudal de *Danio rerio*, comparando las condiciones 1dpa v/s 4dpa

TEs expresados para condición 1dpa vs 4dpa		
	SQUIRE	SQUIRE & TEcandidates
TEs sin cambio significativo	26.556	114
TEs sobre-expresados	16.658	91
TEs sub-expresados	4.639	18

Tabla 14. Cantidad de Self-expressed TEs expresados en 1dpa vs 4dpa.

De estos 91 TEs sobre-expresados, 62 son de tipo DNA (68%), 18 de tipo LTR (20%) y 11 de tipo LINE (12%). Del mismo modo que el análisis de expresión anterior, estos TEs se analizaron según su distancia a genes con algún tipo de cambio de expresión diferencial.

Análisis de expresión diferencial de genes en aleta caudal de *Danio rerio*. 1dpa vs 4dpa

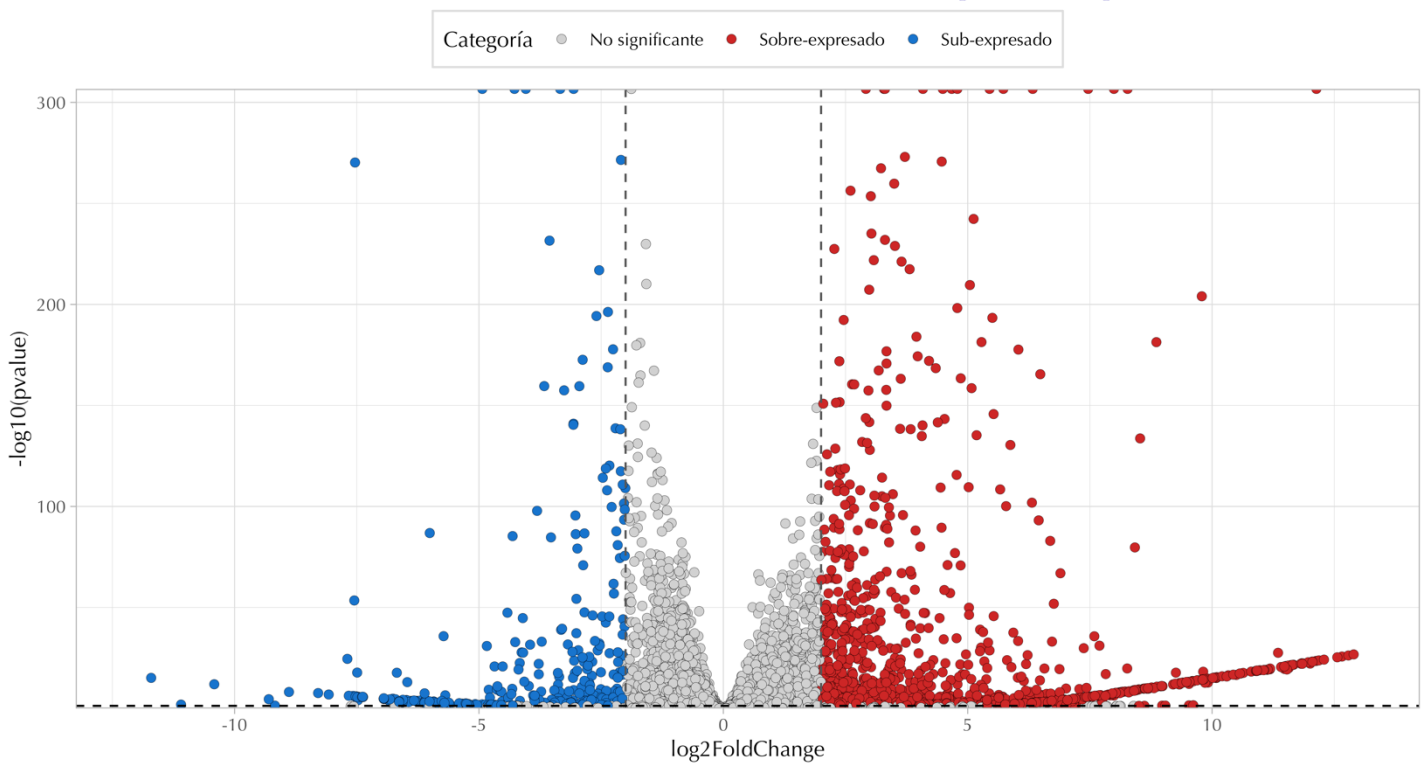


Figura 20. Análisis de expresión diferencial de genes en aleta caudal de *Danio rerio*, comparando las condiciones 1dpa v/s 4dpa.

Genes expresados para condición 0dpa vs 1dpa		
Genes sin cambio significativo	Genes sobre-expresados	Genes sub-expresados
13.954	1141	344

Tabla 15. Cantidad de genes expresados en 1dpa vs 4dpa.

En concreto, y como se mencionó anteriormente, solo se seleccionaron TEs sobre-expresados identificados por ambos software, en todas las condiciones evaluadas, ya que se buscó asociar su sobre-expresión con genes cercanos a su locus. El número final de TEs sobre-expresados durante el evento de regeneración de *Danio rerio* es de 125, teniendo 15 en la comparación 0dpa vs 1dpa, 19 en 0dpa vs 4dpa y 91 en 1dpa vs 4dpa.

6.3. Resultados objetivo específico N°2

Los datos de secuenciación generados por la metodología ATAC-Seq se alinearon al genoma de *Danio rerio* y luego de eso se realizó el *peak calling*. Esto se hizo para identificar regiones con enriquecimiento de *reads*, denominadas *peaks*. Estos *peaks* se agruparon según cada condición temporal a evaluar (0dpa, 1dpa y 4dpa). Esta agrupación se obtuvo intersectando la cantidad total de *peaks* que se obtuvo de cada réplica por condición.

Luego, se evaluó la intersección entre *peaks* con la anotación de TEs, con la finalidad de obtener regiones enriquecidas con traslape de TEs. El resultado de este procedimiento se muestra en la Tabla 16.

Condición temporal	Peak totales	Peaks intersectados con TEs
0dpa	6.541	438 (7%)
1dpa	34.374	3.751 (11%)
4dpa	59.238	7.451 (13%)

Tabla 16. Cantidad de regiones enriquecidas (peaks) y su intersección con TEs.

Dada la naturaleza de la metodología ATAC-Seq, se detectan regiones abiertas de la cromatina (eucromatina). En ellas se pueden detectar regiones libre de nucleosomas y regiones donde pueden unirse TFs. Por sí solo, el *peak calling* no permite discriminar la diferencia entre ambas. Es por esto que a estos *peaks* que tenían algún traslape con TEs, se les procedió a identificar regiones o sitios *footprints*, los cuales representan a las regiones de estos *peaks* donde la cobertura de *reads* tiende a bajar por la presencia de un factor de transcripción que se une al DNA. Del mismo modo que el paso anterior, estos nuevos sitios *footprints*, se intersecaron con TEs. Las estadísticas para estas nuevas regiones *footprints*, así como las que intersectaban con TEs se muestran en la Tabla 17.

Condición temporal	<i>Footprints</i> totales	<i>Footprints</i> intersectados con TEs
0dpa	10.726	8.182 (76%)
1dpa	26.586	20.936 (78%)
4dpa	108.731	86.694 (80%)

Tabla 17. Cantidad de sitios *footprints* y su intersección con TEs.

El siguiente paso posterior a la obtención de sitios *footprints* es la identificación de motivos de secuencia, los cuales están vinculados a TFs. Es por esto por lo que a estos *footprints* identificados con alguna intersección con TEs, se les procedió a buscar patrones específicos de secuencia, de tal forma de asociarlos con sitios de unión a factores de transcripción. Estos patrones son conocidos como motivos. Esta búsqueda entregó un total de 405 motivos para la condición 0dpa, 5895 motivos para la condición 1dpa y 9618 motivos para la condición 4dpa.

Dichos motivos identificados, según su posición en el genoma, se asociaron a TEs, es decir, motivos que pertenecieran a TEs anotados, de tal modo de ver

cuántos TEs presentaban estos motivos y qué tipo de elementos eran. Esta agrupación de motivos en TEs dio como resultado un total de 75 TEs que contenían los 405 motivos de 0dpa, 834 TEs que contenían los 5895 motivos de 1dpa y 1356 TEs que contenían los 9618 motivos de 4dpa. En la Tabla Suplementaria 1, se pueden visualizar los TEs con presencia de más motivos diferentes por cada condición.

En la

Figura 21 se puede observar una representación gráfica de los pasos anteriormente detallados en este apartado, en donde desde una región enriquecida de *reads* (*peaks*), se obtuvieron las sub-regiones de sitios *footprints*, desde las cuales se identificaron los motivos recientemente mencionados.

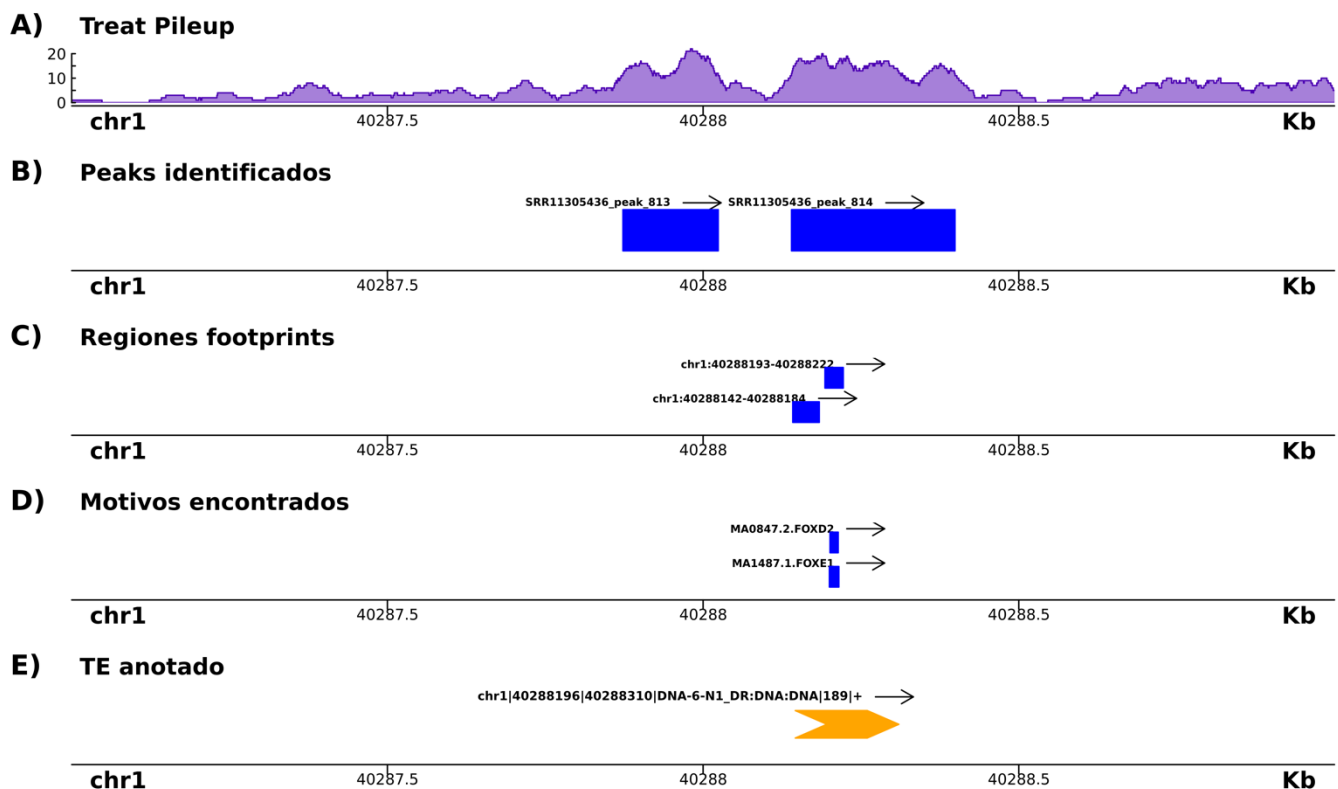


Figura 21. Representación proceso de identificación de peaks, footprints y motivos relacionados con TEs. En el apartado A se visualiza la representación de peaks en toda la región

genómica de interés. En B se visualiza concretamente el peak por medio de un rectángulo azul, con su correspondiente identificador. En C se tiene la identificación de regiones footprints representadas por rectángulos azules, también con su correspondiente identificador de región específica de la región. En D se visualizan los motivos encontrados en las regiones footprints anteriormente mencionadas. Estos se representan con un rectángulo azul acompañado del identificador del motivo, seguido del identificador del TF asociado. Finalmente, en E se representa el TE anotado que sobrelapa con todos los elementos antes mencionados. Este se visualiza mediante una línea amarilla con punta de flecha acompañado de su identificador. En todos los apartados, en el eje X se indica la región genómica donde se encuentran los elementos.

Para cada una de las condiciones temporales se detalla el tipo de TE que presenta motivos de sitios de unión a factores de transcripción. Para la condición 0dpa se encuentra en la Tabla 18, para la condición 1dpa en la Tabla 19 y finalmente para la condición 4dpa en la Tabla 20. En dichas tablas, la primera columna contiene el tipo de TE que presenta el o los motivos, la segunda columna la cantidad de TEs y finalmente la tercera columna el porcentaje que representa esta cantidad de elementos.

75 TEs presentan motivos de TFBSs en condición 0dpa		
Tipo TE	Cantidad	Porcentaje
DNA	56	75%
LTR	11	15%
LINE	7	9%
SINE	1	1%

Tabla 18. Tipos de TEs que presentan motivos de TFBSs en 0dpa.

834 TEs presentan motivos de TFBSs en condición 1dpa		
Tipo TE	Cantidad	Porcentaje
DNA	568	68%
LTR	162	20%
LINE	92	11%
SINE	12	1%

Tabla 19. Tipos de TEs que presentan motivos de TFBSs en 1dpa.

1356 TEs presentan motivos de TFBSs en condición 4dpa		
Tipo TE	Cantidad	Porcentaje
DNA	909	67%
LTR	260	19%
LINE	153	11%
SINE	34	3%

Tabla 20. Tipos de TEs que presentan motivos de TFBSs en 4dpa.

6.4. Resultados objetivo específico N°3

Para realizar el objetivo específico N°3, se parte desde los resultados de los objetivos específicos N°1 y N°2 para poder evaluar el rol que podrían tener los TEs en las distintas etapas de la regeneración de aleta caudal de *Danio rerio*.

6.4.1. Relación de TEs sobre-expresados con genes

Para la relación de TEs con genes asociados a regeneración se tomaron 4 criterios: 1. TE sobre-expresado, 2. gen con expresión diferencial, 3. que la expresión ocurriera en la misma hebra tanto para el TE como para el gen, y 4. finalmente que la distancia entre ambos (TE y gen) fuera entre 0pb (presencia del TE en el intrón del gen), hasta como máximo 5000pb (presencia del TE en regiones intergénicas cercanas al gen).

Bajo estos criterios, y considerando las comparaciones temporales que se detallaron en los resultados del objetivo específico N°1, se encontraron 2 TEs en regiones intrónicas para la comparación 0dpa vs 1dpa y 4 TEs para la comparación 1dpa vs 4dpa. Es importante destacar que para la condición 0dpa vs 4dpa no se obtuvieron resultados favorables en cuanto a distancia, llegando a obtener un gen más cercano a un TE sobre-expresado a una distancia de 14.677pb. Además, todas las relaciones encontradas se detectaron para genes sobre-expresados, al igual que el caso anterior, también en términos de distancia. Los identificadores, tanto de los TEs como los genes encontrados según los criterios antes mencionados se pueden encontrar en la Tabla 21.

Comparación temporal	Categoría TE	ID TE	Nombre Gen
0dpa vs 1dpa	Intrónico	chr20 3315189 3316177 DNA-N4_DR:DNA:DNA 22 -	<i>marcksa</i>
0dpa vs 1dpa	Intrónico	chr25 34020704 34021917 Tc1-5_DR:TcMar-Tc1:DNA 148 +	<i>anxa2a</i>
1dpa vs 4dpa	Intrónico	chr10 2822800 2823765 piggyBac-N1_DR:PiggyBac:DNA 34 +	<i>aebp1</i>
1dpa vs 4dpa	Intrónico	chr5 7721139 7722527 DNA-3-1_DR:DNA:DNA 146 -	<i>pdlim5a</i>
1dpa vs 4dpa	Intrónico	chr7 60118867 60120528 Gypsy-17-LTR_DR:Gypsy:LTR 12 +	<i>hsa12b</i>
1dpa vs 4dpa	Intergénico (1105 pb)	chr5 46204590 46205913 DNA-2-4_DR:DNA:DNA 119 -	<i>si:ch211-130m23.3</i>

Tabla 21. TEs sobre-expresados cercanos a genes sobre-expresados. En la columna 1 se detalla la comparación temporal realizada. En la columna 2 se describe la categoría del TE, de acuerdo con su ubicación dentro del genoma y además la distancia en pb al gen más cercano es indicada entre paréntesis. En el caso de TEs intrónicos dicha distancia se omite al ser 0 pb. En la columna 3 se visualiza el identificador del TE. Finalmente, en la columna 4 se representa el nombre del gen más cercano al TE identificado.

Seguidamente, se realizó una búsqueda de la función biológica de cada gen encontrado en estas relaciones, con tal de filtrar por aquellos que tenían algún grado de incidencia en el evento regenerativo. En la Tabla 22 se visualiza cada uno de los genes identificados con un resumen de su respectiva función.

Nombre gen	Función
<i>marcksa</i>	Permite la unión de filamentos de actina. Actúa en el desarrollo de órganos y desarrollo del sistema respiratorio. Se expresa en el cerebro, el ojo y el músculo esquelético.
<i>anxa2a</i>	Permite la unión de iones de calcio. Actúa en la membrana y en la región extracelular. Se expresa en varias estructuras, incluyendo las EVL (células envolventes propias del blastema), branquias, peridermo, entre otros.
<i>aebp1</i>	TF que reprime la transcripción, la actividad de unión específica al DNA de la RNA polimerasa II y la región reguladora de la transcripción de la misma.
<i>pdlim5a</i>	Participa en la organización del citoesqueleto, el desarrollo del corazón y el desarrollo de la estructura muscular. Forma parte de la actina filamentosa. Se expresa en la musculatura cefálica, el sistema digestivo, la aorta dorsal, el sistema nervioso, entre otros.

<i>hspa12b</i>	Participa en el proceso de angiogénesis germinal (formación de vasos sanguíneos nuevos a partir de vasos preexistentes). Se expresa en el cerebro, el sistema cardiovascular, cráneo, mesodermo ventral, entre otros.
<i>si:ch211-130m23.3</i>	Permite la actividad de unión del DNA monocatenario. Participa en la regulación positiva de la transcripción por la RNA polimerasa II.

Tabla 22. Funciones biológicas de genes sobre-expresados, cercanos a TEs sobre-expresados. En la columna 1 se visualiza el nombre del gen sobre-expresado que cuenta con un TE sobre-expresado cercano. En la columna 2 se describe un resumen de la función biológica del gen obtenida de la base de datos ZFIN (Ruzicka et al., 2019).

Se estudió la participación y relevancia en el proceso de regeneración de aleta caudal de *Danio rerio* de los genes mostrados anteriormente. Para la gran mayoría (4 de 6, es decir un 66.7%) se ha reportado algún tipo de relación con el proceso regenerativo. Para el presente trabajo, se enfocaron los resultados en base a la importancia del gen en el evento regenerativo, en los cuales estos genes podrían estar interactuando en una red regulatoria génica relacionada con dicho evento. Los genes escogidos fueron *anxa2a* que se está sobre-expresando en la comparación 0dpa vs 1 dpa y *hspa12b* que lo está haciendo en la comparación 1dpa vs 4dpa. Esta elección será abordada con más profundidad en la discusión del presente trabajo. De igual manera, no se descarta el impacto que puedan tener los otros genes no seleccionados en este apartado en el evento regenerativo, que probablemente a día de hoy, aún no estén bien estudiados o bien definidas sus redes regulatorias para su relación con dicho evento.

El gen *anxa2a* posee un rol trascendental en el proceso regenerativo, y se ha demostrado que su edición genética, es decir quitarlo del genoma tiene consecuencias negativas, desacelerando el evento (Quoseena et al., 2020). Se ha reportado que durante las primeras horas del evento regenerativo, tanto el gen *anxa2b* como el gen *anxa2a* sufren una regulación positiva llegando a su máxima sobre-expresión al día posterior a la amputación respecto al mismo tejido antes del proceso regenerativo (Saxena et al., 2016), información que se pudo comprobar en el presente trabajo. Ambos genes (*anxa2a* y *anxa2b*) son parte de la misma familia de *Anexinas* que codifican una proteína llamada *Anexina A2* (Saxena et al., 2016).

Particularmente, el gen *anxa2a* tiene una gran relevancia al participar y estar implicado en el proceso de regeneración tisular, es decir, regeneración de tejidos. Además, sumado a la regulación positiva que presenta al día posterior a la amputación, se encuentra en una de sus regiones intrónicas un TE de tipo DNA sobre-expresándose, lo que podría ser un indicio de que este TE podría estar ejerciendo alguna regulación sobre este gen en ese determinado momento. Una representación de la sobre-expresión del TE y del gen en esta comparación en particular se visualiza en la Figura 22.

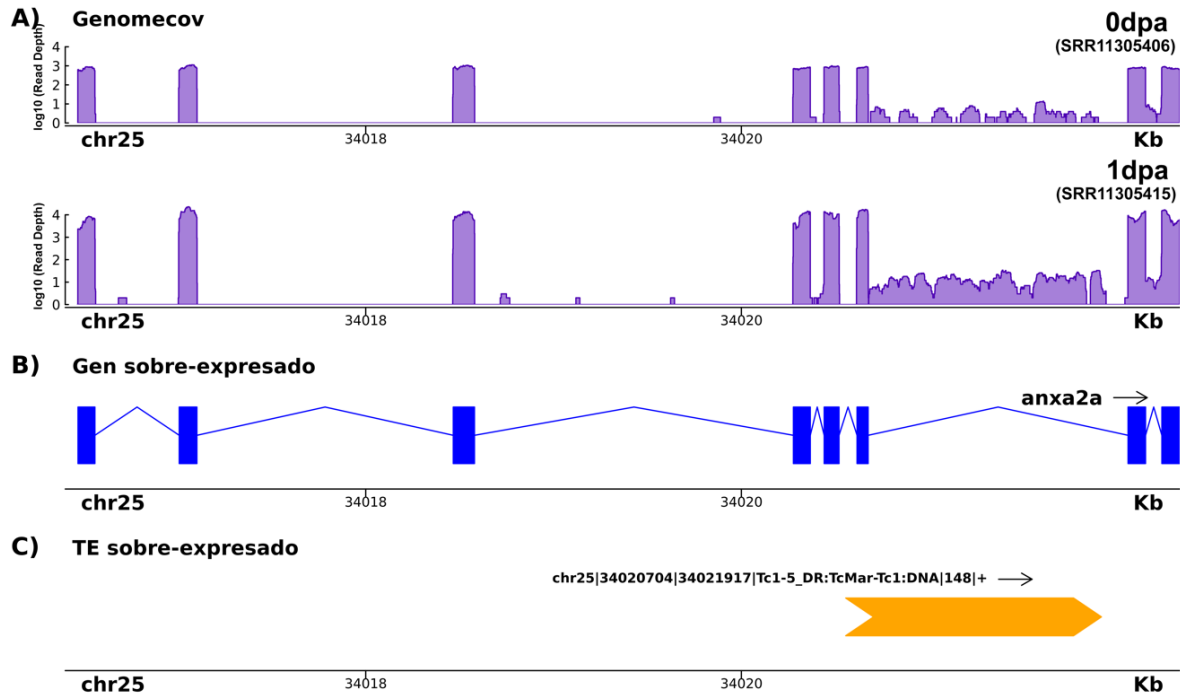


Figura 22. Representación de la expresión del TE y del gen sobre-expresado en la condición 0dpa vs 1dpa. En el inciso A se representa la cobertura genómica de *reads* en la región específica del cromosoma 25 de *Danio rerio*. Esta cobertura se representa en una escala logarítmica y se visualiza en forma de peaks de color morado. En el primer diagrama de cobertura de A, se visualiza una réplica correspondiente a la condición 0dpa (SRR11305406), mientras que en el segundo diagrama de A se tiene una réplica correspondiente a la condición 4dpa (SRR11305415). En el inciso B, se representa la arquitectura exón-intrón del gen sobre-expresado “*anxa2a*”. A la derecha del nombre del gen se representa la direccionalidad del gen en forma de flecha. Los rectángulos verticales azules representan las regiones exónicas mientras que las líneas representan regiones intrónicas. Finalmente, en C se visualiza el TE sobre-expresado en forma de rectángulo con punta de flecha amarillo. También en este caso se visualiza el nombre del TE, así como su direccionalidad. Como eje X de cada apartado se puede ver la posición específica en kilobases (Kb) del genoma en donde se encuentran los elementos representados.

Para comprender mejor el impacto del gen *anxa2a* en una escala biológica mayor, se realizó un análisis de red o interacciones proteína-proteína conocidas utilizando la base de datos STRING.

En particular, la red biológica a la que pertenece el producto proteico del gen *anxa2a* se visualiza en la

Figura 23. En ella se pueden apreciar las principales proteínas con las cuales presenta una relación funcional determinada (functional partners). En concreto, esta red biológica se centra en proteínas funcionales relacionados a la unión de calcio, como lo son la Proteína S100 (*s100a10a* y *s100a10b*) y Anexina (*anxa2b*, *zgc:101785* y la propia *anxa2a*). A su vez se distinguen 2 reguladores de splicing como lo son *LOC566639* y *ahnak*. Y finalmente 3 proteínas con estructura conocida (*tdr1*, *rps6ka1* y *ctsba*). Como dato a destacar se puede decir que *tdrd1* ejerce un rol represivo en los TEs para evitar su movilización, sin embargo, esto durante el evento de espermatogénesis, cuidando la integridad de la línea germinal. Igualmente, se evaluó la presencia de *tdr1* en esta y todas las condiciones temporales de este trabajo y en ninguna de estas reportó un cambio de expresión diferencial.

En la representación de la red biológica de cada gráfico generado por STRING, se visualiza un círculo de color rojo se representa el eje central de la red. Este posee relaciones con genes determinados como interacciones conocidas. Estas interacciones se representan en líneas de color rosa para interacciones determinadas experimentalmente y en líneas de color celeste para interacciones reportadas por bases de datos seleccionadas. Los genes que codifican proteínas con estructura conocida son representados por círculos que poseen una representación proteica en su interior. Finalmente, al gráfico reportado por *STRING*, se añade el cuadro TE en color amarillo, simbolizando así la relación que tiene el TE sobre-expresado identificado en la región intrónica del gen que codifica la proteína representada.

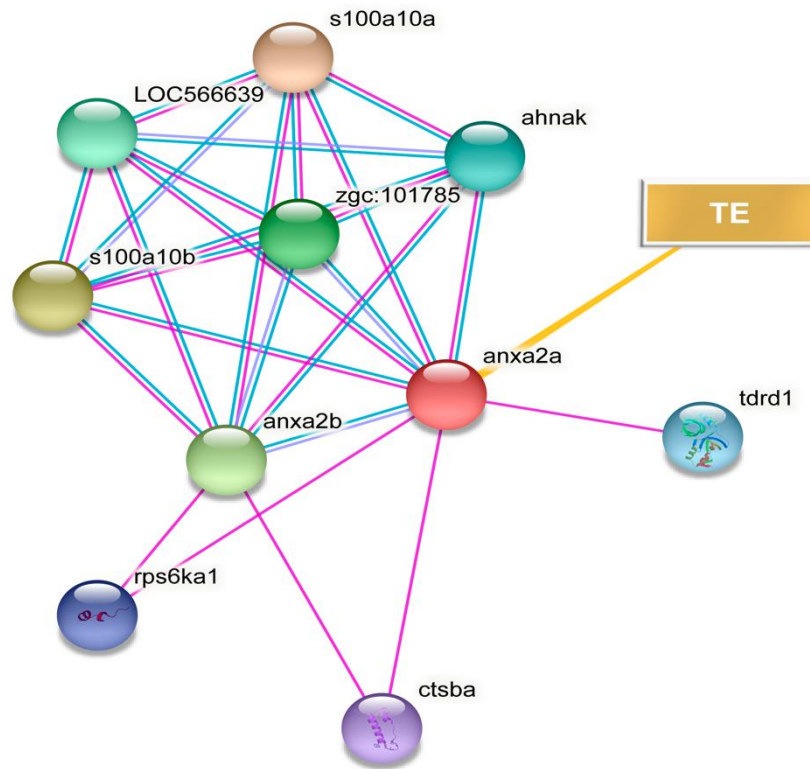


Figura 23. Red biológica del producto proteico del gen *anxa2a*. Representación de la red biológica proteica que comparte el producto proteico del gen *anxa2a*, donde además se considera el potencial rol del TE identificado.

Del mismo modo, durante el transcurso del evento regenerativo se sobre-expresan un mayor número de genes en el transcurso de 1dpa hasta 4dpa, que en las primeras etapas de dicho evento. Uno de ellos, sobre-expresado a los 4 días post-amputación, es el gen *hspa12b*. Dicho gen, cumple un rol de vital importancia para la formación de nuevos vasos sanguíneos a partir de vasos existentes. Como se vio anteriormente, una de las etapas implicadas en la regeneración de un tejido era precisamente la formación y reparación de los vasos sanguíneos. Existen estudios, que demuestran que *hspa12b* presenta una regulación positiva durante el proceso de angiogénesis que es un evento clave en la proliferación y migración de células para el desarrollo vascular de *Danio rerio* (Zhao et al., 2018).

En términos del objetivo de este trabajo, se encontraron 3 TEs intrónicos que están teniendo una regulación positiva en esta etapa. Como dato a destacar, los 3 TEs son de tipo LTR y se encuentran de manera contigua entre sí. La relación que podrían tener con la sobre-expresión del gen *hspa12b* podría ser mediada no solo por 1 TE, sino por un conjunto de estos, algo que se ve comúnmente en los elementos regulatorios, en donde no solo participa un único elemento, sino que es un conjunto de elementos que regulan la expresión génica (Rodríguez & Kang, 2020). Un esquema de la expresión del gen y los TEs se visualiza en la Figura 24.

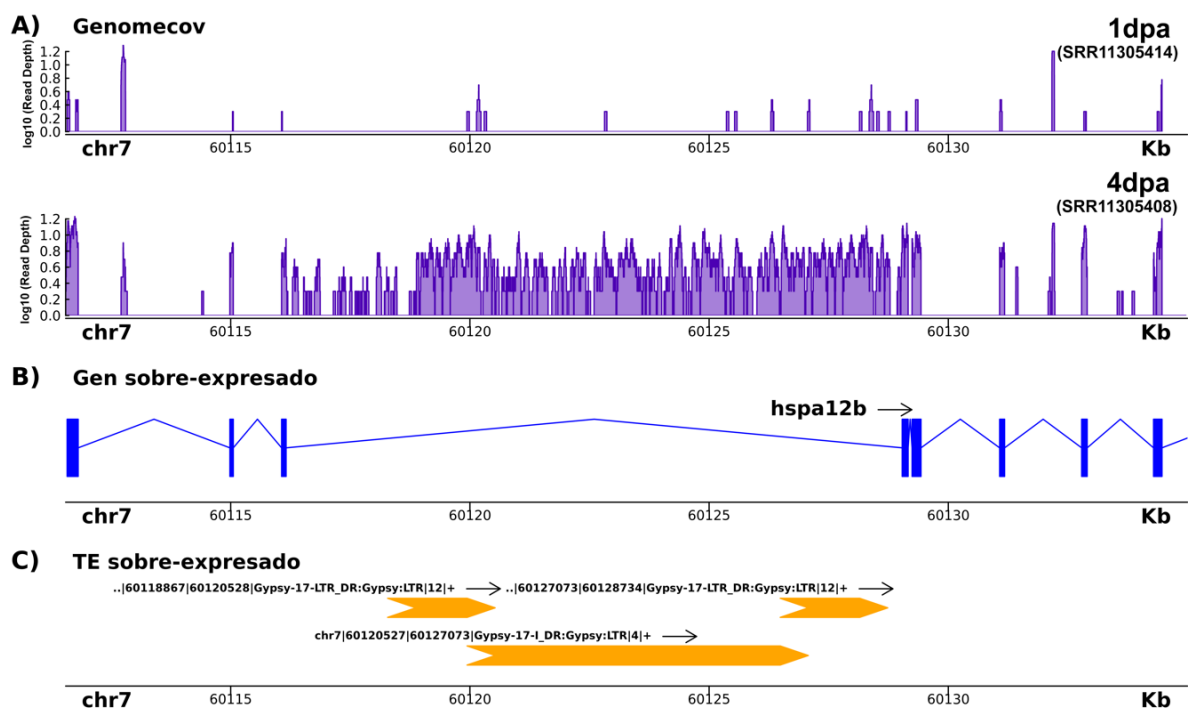


Figura 24. Representación de la expresión del TE y del gen sobre-expresado en la condición 1dpa vs 4dpa. En el inciso A se representa la cobertura genómica de reads en la región específica del cromosoma 7 de *Danio rerio*. Al igual que la figura anterior, esta cobertura se representa en una escala logarítmica y se visualiza en forma de peaks de color morado. En el primer diagrama de cobertura de A, se visualiza una réplica correspondiente a la condición 1dpa (SRR11305414), mientras que en el segundo diagrama de A se tiene una réplica correspondiente a la condición 4dpa (SRR11305408). En el inciso B, se representa al gen sobre-expresado “hspa12b”. A la derecha del nombre del gen se representa la direccionalidad del gen en forma de flecha. Los rectángulos verticales azules representan las regiones exónicas mientras que las líneas azules representan regiones intrónicas. Finalmente, en C se visualizan los TEs sobre-expresados en forma

de rectángulo con punta de flecha amarillo. También en este caso se visualizan los nombres de los TEs, así como su correspondiente direccionalidad. Como eje X de cada apartado se puede ver la posición específica en kilobases (Kb) del genoma en donde se encuentran los elementos representados.

Asimismo, al igual que en el caso anterior se presenta la red biológica del producto proteico del gen sobre-expresado, la cual se visualiza en la Figura 25. Los principales *functional partners* de *hspa12b* son una serie de proteínas de choque térmico con características de resistencia a la apoptosis como lo son *bag1*, *bag2*, *bag3*, *bag4* y *bag5*. Por otro lado, la relación con proteínas de la familia *Hsp40* como lo son *dnajc2*, *dnajc7*, *dnajb6a*, *dnajb1a*, todas ellas también relacionadas con la respuesta al estrés. Finalmente, *LOC557824* que también presenta funciones relacionadas a choque térmico.

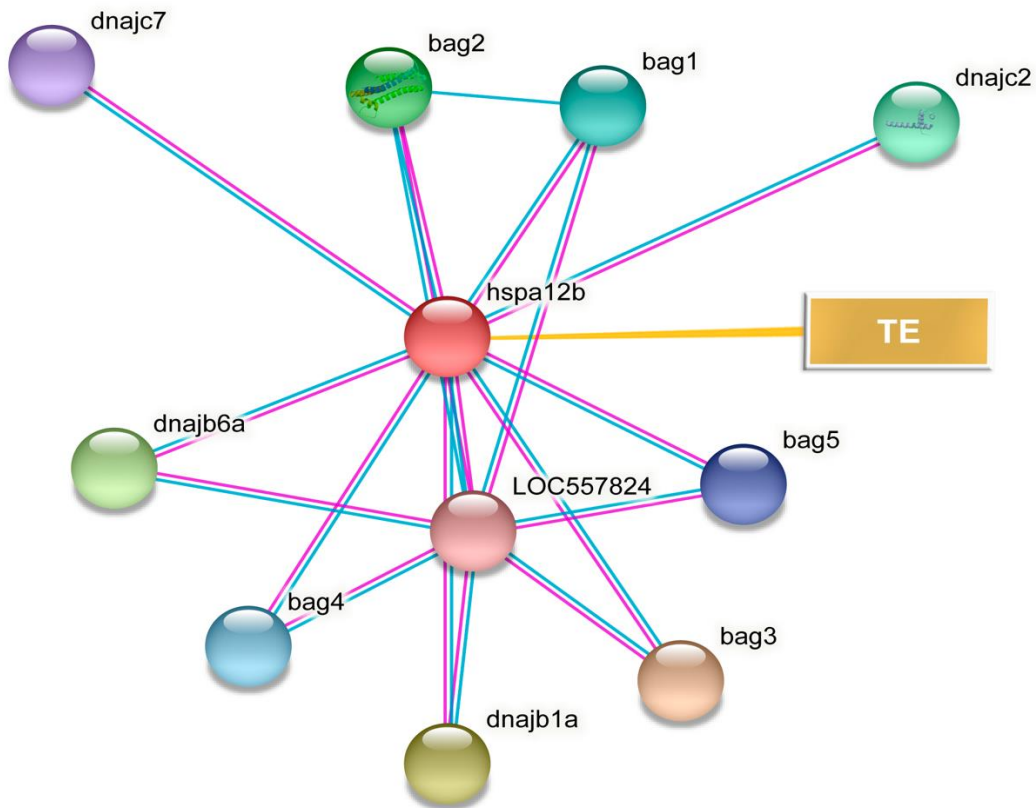


Figura 25. Red biológica del producto proteico del gen *hspa12b*. Representación de la red biológica proteica que comparte el producto proteico del gen *hspa12b*, donde además se considera el potencial rol del TE identificado.

6.4.2. Relación de TEs con TFBSs con TFs asociados a regeneración

De acuerdo con los resultados obtenidos del objetivo específico N°2, se evaluó qué TFBSs asociados a TEs están asociados a regeneración. De este modo, se obtuvieron los TEs a los cuales potencialmente se les unen factores de transcripción, y así vincularlos con en el proceso de regeneración de aleta caudal.

Este nuevo filtro se realizó, mediante una búsqueda exhaustiva en bases de datos asociadas al organismo *Danio rerio*. En concreto se obtuvo TFs asociados a regeneración desde la base de datos ZFIN (Ruzicka et al., 2019), según la metodología propuesta. En dicha base de datos se recopilieron 138 registros que se utilizaron para compararlos con los obtenidos desde los motivos de DNA asociados a TFBSs previamente obtenidos en el apartado 6.3. Finalmente, en términos de motivos, se obtuvo un total de 1 motivo asociado a un TFs implicado en regeneración en la etapa 0dpa, 6 motivos en la etapa 1dpa asociado a 6 TFs y 6 motivos asociados a 6 TFs en la etapa 4dpa.

De igual manera, se analizó la presencia de estos motivos en los TEs anotados del organismo. En particular para la condición 0dpa, el motivo identificado se encontró en 3 TEs distintos, los 6 motivos localizados en 1dpa se encontraron en 45 TEs distintos y los 6 motivos detectados en 4dpa se encontraron en 56 TEs distintos. Los identificadores de cada uno de estos TEs asociados a cada TF según condición se pueden encontrar en la Tabla Suplementaria 2.

Dichos motivos se analizaron de acuerdo con la presencia de la secuencia correspondiente al sitio de unión del factor de transcripción. Esto se realizó mediante los gráficos de secuencia tipo LOGOS asociados a la base de datos de perfiles de unión de TFs, JASPAR (Sandelin et al., 2004). Estos resultados se visualizan en la Tabla 23.

ID Motivo	Condición temporal donde se presenta	LOGOS
Pou5f1.Sox2	0dpa 1dpa 4dpa	
MA0037.3.GATA3	1dpa 4dpa	
MA0768.1.LEF1	1dpa 4dpa	

Como se mencionó anteriormente, cada uno de estos motivos encontrados presenta patrones claves de secuencia, los cuales están relacionados a sitios de unión a factores de transcripción específicos. Precisamente, estos factores de transcripción que se unen a estas regiones son detallados en términos de su función en la Tabla 24 siguiendo el mismo orden de la tabla anterior.

TF	Función
<i>sox2</i>	Actúa en el desarrollo de órganos y la regeneración de aletas. Forma parte de un complejo regulador de la transcripción.
<i>gata3</i>	Participa en varios procesos como la vía de señalización del receptor del factor de crecimiento de fibroblastos, la regulación positiva neurogénesis, regeneración del tejido muscular cardíaco, la morfogénesis de órganos embrionarios y la regeneración de aletas.
<i>lef1</i>	Actúa durante el desarrollo de órganos, particularmente el desarrollo de aletas y la regeneración de estas.
<i>tp63</i>	Actúa durante el proceso de morfogénesis de la epidermis, la vía de señalización apoptótica intrínseca en respuesta al estrés del retículo endoplasmático.

<i>nr3c1</i>	Actúa durante el transporte de cationes, la homeostasis de los iones metálicos y la respuesta de glucocorticoides. Se expresa en varias estructuras como el sistema digestivo, las branquias y la región pleuroperitoneal.
<i>sox10</i>	Actúa durante varios procesos, entre los que se destacan el desarrollo del sistema nervioso, migración de células de la cresta neural y el desarrollo de órganos sensoriales.

Tabla 24. Funciones biológicas de TFs asociados a motivos encontrados. En la columna 1 se visualiza el nombre del TF. En la columna 2 se describe un resumen de la función biológica del TF obtenida de la base de datos ZFIN.

En particular, los TFs con una mayor relación en regeneración concretamente de la aleta caudal son *gata3* y *lef1*. Los sitios de unión a estos factores de transcripción están siendo aportados por TEs, lo que sugiere que los TEs podrían estar involucrados en el evento de regeneración.

Concretamente, *gata3* participa en el proceso de regeneración, expresándose solamente después de una lesión en algún órgano. Es decir, actúa como un factor pro-regenerativo específico, inducido por una determinada lesión (Kizil et al., 2012). En base a la relevancia de los TEs que están actuando como sitio de unión a *gata3*, se considera un esquema de la red regulatoria biológica de dicho TF.

Según la base de datos STRING, las interacciones conocidas reportadas por bases de datos seleccionadas y validadas experimentalmente fueron 10, de las cuales se destaca la presencia de 2 proteínas quinasas, cofactores de transcripción, entre otras proteínas reguladoras. Además, se representa el TE que está actuando

como TFBSs. La representación de la red biológica del TF *gata3* se encuentra en la Figura 26.

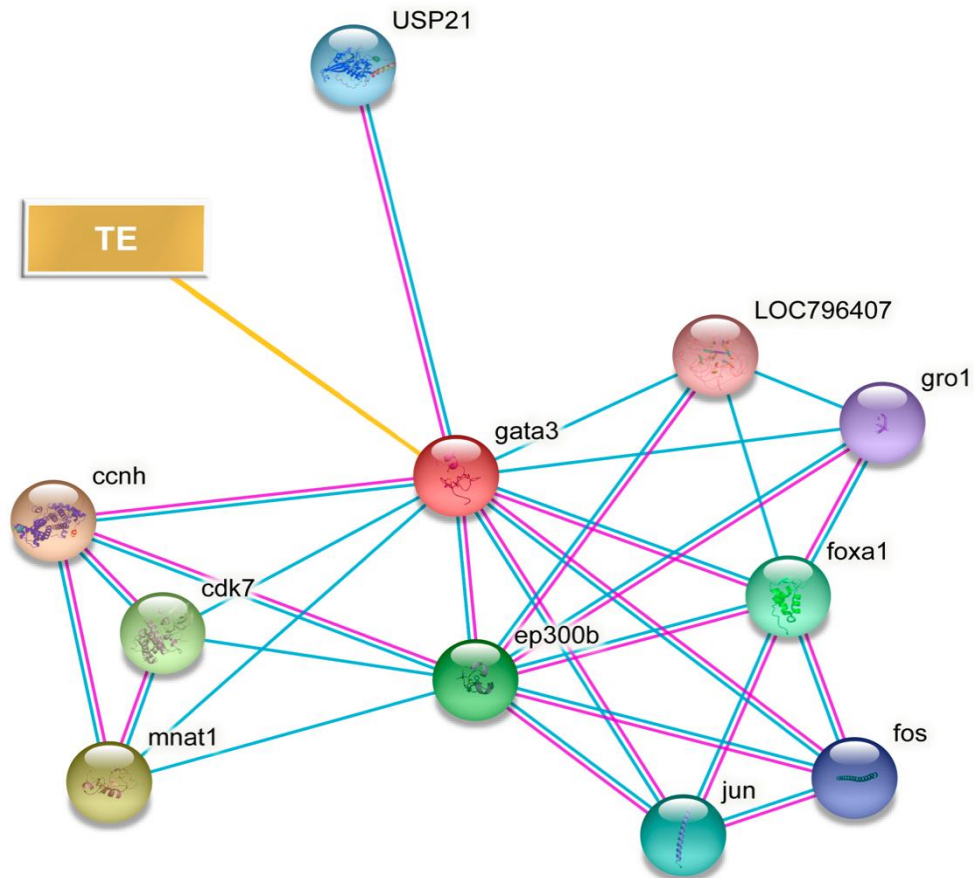


Figura 26. Red biológica del TF *gata3*. Representación de la red biológica proteica que comparte el TF *gata3*, donde además se considera el potencial rol del TE identificado como TFBSs.

Del mismo modo, el TF *lef1* participa en los eventos de regeneración, concretamente en la epidermis de la herida recién formada en la aleta caudal, manteniéndose en la capa epidérmica basal durante la formación del blastema, presentando funciones específicas en la regeneración de tejidos (Poss et al., 2000).

En este caso, el TF *lef1*, según la base de datos STRING, se obtuvo un total de 10 interacciones conocidas. Dentro de estos resultados, se destaca que la mayoría (6 de estos; *h2afx*, *hist2h2l*, *h2afz*, *zgc:101846*, *si:dkey-261m9.19* y *zgc:173585*)

son variantes de la histona H2A, mientras que las proteínas restantes corresponden a co-activadores de factores de transcripción. Además, considerando el rol de los TEs que presentan el motivo de unión a *lef1*, se visualiza la presencia del TE en la red biológica de *lef1*. La representación de dicha red del TF *lef1* se encuentra en la Figura 27.

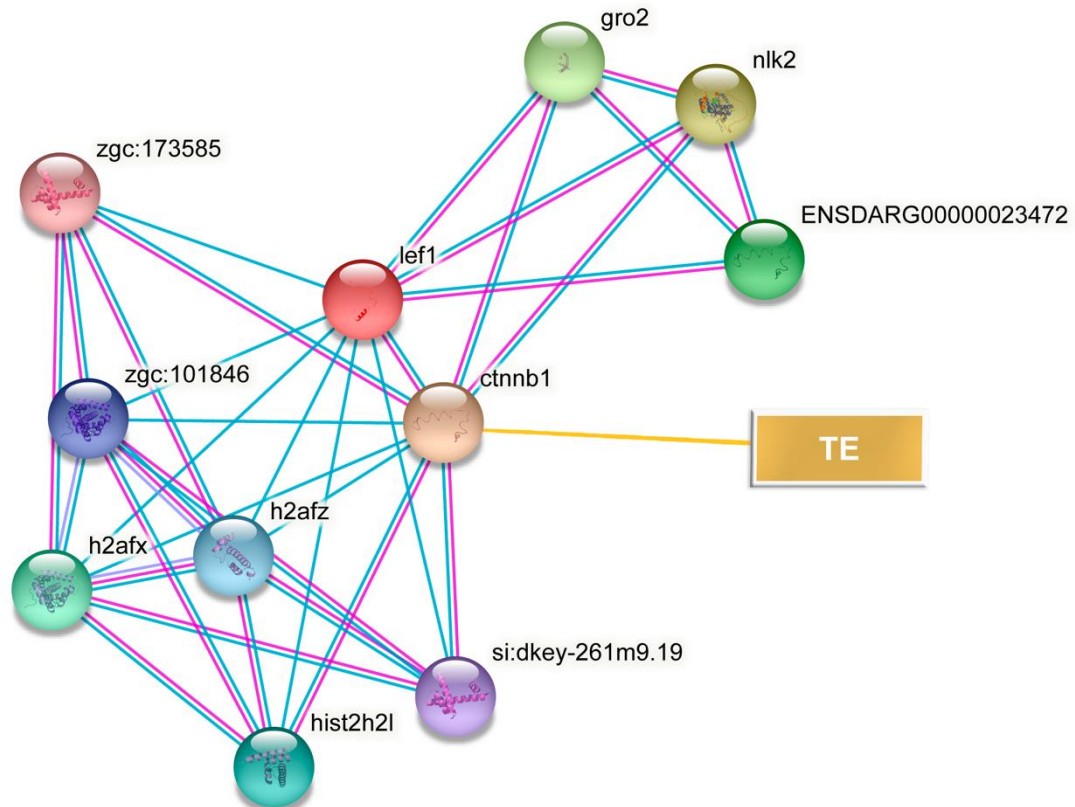


Figura 27. Red biológica del TF *lef1*.

Representación de la red biológica proteica que comparte el TF *lef1*, donde además se considera el potencial rol del TE identificado como TFBSs.

7. Discusión

Hasta hace un par de décadas, el rol de los TEs en cualquier evento biológico era totalmente impensado. Esto se debe a que originalmente se pensaba que eran “*junk DNA*”. Por otro lado, su análisis computacional es complicado por lo que suelen ser omitidos en la mayoría de trabajos. A pesar de que esta tarea aún sigue siendo complicada, por su naturaleza repetitiva y móvil dentro del genoma, con el desarrollo de nuevas metodologías se ha podido estudiar la potencial relación que los TEs pueden llegar a tener en distintos eventos celulares como la regulación génica.

En particular en el evento regenerativo son varios los genes que tienen algún grado de implicancia en las distintas etapas, pero como se detalló en este y otros trabajos, existe una compleja maquinaria que regula la expresión de todos estos genes, y muchas veces estos elementos no son precisamente genes si no, secuencias de DNA que pueden o no transcribirse para ejercer una regulación en determinados genes.

Como se demostró en el presente trabajo, el rol de los TEs puede tener una fuerte relación con el evento regenerativo ya sea con su regulación positiva cercana a genes o dentro de regiones intrónicas de genes sobre-expresados o actuando como sitios de unión a factores de transcripción implicados en el proceso de regeneración.

En primer lugar, la utilización de la última versión del genoma de referencia de *Danio rerio*, sin duda ayuda a que los resultados obtenidos tengan un mayor sustento sobre la base de estar trabajando con las anotaciones más actuales y las que presentan un mayor número de información relevante para todos los análisis.

Mediante el análisis de los transcritos presentes en cada condición se pudo evidenciar la gran cantidad de TEs que se transcribían en las distintas temporales,

particularmente detectados por el software *SQUIRE*. Sin embargo, para esta investigación se decidió ser estricto, al usar también el software *TEcandidates*, y solo trabajar con los resultados que entregaron ambos software. Esta decisión, sin duda dejó muchos TEs transcritos fuera del análisis, pero se considera un paso de gran importancia en pos de obtener resultados con una mayor sensibilidad.

Análisis como el de ATAC-Seq, ayudan a generar un contexto mucho más amplio de lo que sucede a nivel celular aparte de la transcripción en un determinado momento, ya que permite la identificación de los elementos que interactúan en la regulación transcripcional. En este trabajo se logró encontrar que regiones enriquecidas con reads de ATAC-Seq están en TEs. Esto es un dato de gran importancia a resaltar, ya que aproximadamente en el 20% de estas regiones se encontraban TEs, apoyando aún más la hipótesis sobre la participación de estos elementos en etapas de regulación génica.

La elección de genes, entre todos los que se encontraron cercanos a TEs sobre-expresados, se realizó basándose en la investigación previa de cada una de las características de las distintas etapas del evento regenerativo de la aleta caudal en *Danio rerio*. Por ejemplo, se ha reportado que el gen *anxa2a* tiene una función determinada en el proceso de regeneración tisular y obtiene su máxima regulación positiva a las 24 posteriores a la amputación (Saxena et al., 2016), dicho patrón de expresión fue comprobado en este trabajo como se pudo comprobar en el apartado de resultados. Por otro lado, el gen *hsps12b* tiene una función biológica implicada en la generación de nuevos vasos sanguíneos durante el evento regenerativo. Según la evidencia biológica, este proceso es clave mientras se regenera una determinada estructura funcional, en este caso la aleta caudal.

La gran mayoría de los genes identificados en este trabajo con una regulación positiva, son confirmados en base a los resultados de investigaciones anteriores, en donde diversos análisis demuestran el rol de estos genes en las distintas etapas del evento regenerativo. Distinto es el caso de los TEs, ya que, al no poseer estudios

previos que analizaran su rol en el evento regenerativo de la aleta caudal del pez cebra, no existe un punto de comparación o de confirmación para tales elementos. Sin embargo, considerando la naturaleza compleja del evento de regeneración, en donde participan un gran número de genes, es de esperar que también participen un gran número de elementos reguladores, entre los que pueden estar los TEs involucrados.

Si bien estos resultados resultan bastante prometedores y apoyan la hipótesis de que los TEs presentan algún grado de relación en el evento regenerativo desde un punto de vista regulatorio, es importante conocer las limitaciones de este trabajo y cómo pueden solventarse o complementarse a él por medio de otras áreas.

Lo primero es tener claro la naturaleza de este proyecto. Al ser un proyecto exploratorio, el análisis realizado se presenta como potenciales respuestas a las preguntas planteadas, es decir, estos resultados e interpretaciones se enmarcan en un contexto de evaluación en vías de ser confirmado experimentalmente. Precisamente, el paso experimental podría ser considerado como la complementariedad a esta investigación, comprobando los resultados acá planteados. Por ejemplo, realizar estudios de edición génica de los genes y/o TEs identificados en esta investigación con alguna relevancia en el contexto regenerativo, para verificar empíricamente la relación de estos en el evento sería el motivo final para corroborar los resultados acá entregados.

8. Conclusiones

En la presente investigación se buscó analizar e identificar el rol que presentan los TEs en el proceso biológico de regeneración de aleta caudal de *Danio rerio*. Mediante el procesamiento de datos desde tecnologías RNA-Seq y ATAC-Seq, se obtuvieron los siguientes resultados:

- Se identificaron 2 potenciales genes relacionados al proceso regenerativo que presentan un TE sobre-expresado en una de sus regiones intrónicas.
 - Uno de estos genes (*anxa2a*) se encontraba sobre-expresado al día posterior a la amputación comparado con la etapa recientemente amputada (0dpa). A su vez, este gen presentó un TE de tipo DNA, que se está sobre-expresando en una de sus regiones intrónicas. La expresión de este TE intrónico podría tener influencias en cómo se regula *anxa2a*.
 - Por otro lado, el gen *hspa12b* se encontraban sobre-expresado a los 4 días posteriores a la amputación comparado a la etapa del día posterior a la amputación. *hspa12b* presentó 3 TEs de tipo LTR que se estaban sobre-expresando en una de sus regiones intrónicas.
- Se identificaron en total 104 TEs que están actuando como sitios de unión a TFs que participan en alguna etapa del evento regenerativo.
 - 3 de estos TEs están aportando sitios de unión a 1 TF (*sox2*) en la condición temporal 0dpa. Estos 3 TEs son de tipo DNA.
 - 45 TEs están aportando sitios de unión a 6 TFs (*sox2*, *gata3*, *lef1*, *tp63*, *nr3c1*, *sox10*) en la condición temporal 1dpa. De estos 45 TEs, 30 son de tipo DNA, 12 tipo LINE y 3 de tipo LTR.
 - 56 TEs están aportando sitios de unión a 6 TFs (*sox2*, *gata3*, *lef1*, *tp63*, *nr3c1*, *sox10*) en la condición temporal 4dpa. De estos 45 TEs, 30 son de tipo DNA, 15 tipo LINE, 10 tipo LTR y 1 de tipo SINE.

Los datos resultantes apoyan la hipótesis planteada que buscaba comprender si los TEs tenían algún grado de relación regulativa en el proceso de regeneración de aleta caudal de *Danio rerio*. Sin embargo, tal como se mencionó en la discusión del presente trabajo, estos resultados deben ser comprobados empíricamente en laboratorio.

Estudios y análisis como estos podrán ayudar a dar, en un futuro cercano, mayores pruebas de la real función de los elementos transponibles que poseen los genomas, y así poder comprender, asociar y aplicar sus funciones regulatorias para futuras terapias y/o investigaciones de los procesos biológicos de mayor interés para la comunidad.

9. Referencias

1. Ahituv, N. (2012). Gene Regulatory Elements. In *Gene Regulatory Elements* (Vol. 9781461416, Issue September 2012, pp. 2–6). Springer. <https://doi.org/10.1007/978-1-4614-1683-8>
2. Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
3. Belluti, S., Rigillo, G., & Imbriano, C. (2020). Transcription Factors in Cancer: When Alternative Splicing Determines Opposite Cell Fates. *Cells*, 9(3), 760. <https://doi.org/10.3390/cells9030760>
4. Billingsley, K. J., Lättekivi, F., Planken, A., Reimann, E., Kurvits, L., Kadastik-Eerme, L., Kasterpalu, K. M., Bubb, V. J., Quinn, J. P., Kõks, S., & Taba, P. (2019). Analysis of repetitive element expression in the blood and skin of patients with Parkinson's disease identifies differential expression of satellite elements. *Scientific Reports*, 9(1), 1–9. <https://doi.org/10.1038/s41598-019-40869-z>
5. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
6. Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements 06 Biological Sciences 0604 Genetics. *Genome Biology*, 19(1), 1–12. <https://doi.org/10.1186/s13059-018-1577-z>
7. Buenrostro, J., Wu, B., Chang, H., & Greenleaf, W. (2016). ATAC-seq method. *Current Protocols in Molecular Biology*, 2015, 1–10. <https://doi.org/10.1002/0471142727.mb2129s109.ATAC-seq>
8. Cain, M. L., Buskirk, R. V., Minorsky, P. V., Reece, J. B., Urry, L. A., Gillen, C. M., & Wasserman, S. A. (2016). *Campbell Biology* (11 th). Pearson.
9. Chassot, B., Pury, D., & Jaźwińska, A. (2016). Zebrafish fin regeneration after cryoinjury-induced tissue damage. *Biology Open*, 5(6), 819–828.

<https://doi.org/10.1242/bio.016865>

10. Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. *Nature Reviews Genetics*, *18*(2), 71–86. <https://doi.org/10.1038/nrg.2016.139>
11. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), 1–4. <https://doi.org/10.1093/gigascience/giab008>
12. Doane, A. S., & Elemento, O. (2017). Regulatory elements in molecular networks. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *9*(3), 1–10. <https://doi.org/10.1002/wsbm.1374>
13. Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science*, *351*(6274). <https://doi.org/10.1126/science.aac7247>
14. Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
15. Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews. Genetics*, *9*(May), 397–405.
16. Gilbert, E. A. B., Delorme, S. L., & Vickaryous, M. K. (2015). The regeneration blastema of lizards: an amniote model for the study of appendage replacement. *Regeneration*, *2*(2), 45–53. <https://doi.org/10.1002/reg2.31>
17. Goldman, J. A., & Poss, K. D. (2020). Gene regulatory programmes of tissue regeneration. *Nature Reviews Genetics*, *21*(9), 511–525. <https://doi.org/10.1038/s41576-020-0239-7>
18. Gurtner, G. C., Werner, S., Barrandon, Y., & Longaker, M. T. (2008). Wound repair and regeneration. *Nature*, *453*(7193), 314–321. <https://doi.org/10.1038/nature07039>
19. Hoopes, L. (2008). Introduction to the gene expression and regulation topic room. *Nature Education*, *1*(1), 160.

20. Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., Collins, J. E., Humphray, S., McLaren, K., Matthews, L., McLaren, S., Sealy, I., Caccamo, M., Churcher, C., Scott, C., Barrett, J. C., Koch, R., Rauch, G. J., White, S., ... Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, *496*(7446), 498–503. <https://doi.org/10.1038/nature12111>
21. Izsvák, Z., Ivics, Z., & Hackett, P. B. (1997). Repetitive elements and their genetic applications in zebrafish. *Biochemistry and Cell Biology*, *75*(5), 507–523. <https://doi.org/10.1139/o97-045>
22. Khan, F. R., & Alhewairini, S. S. (2019). Zebrafish (*Danio rerio*) as a model organism. In *Current Trends in Cancer Management*. <https://doi.org/10.5772/intechopen.81517>
23. Kizil, C., Kyritsis, N., Dudczig, S., Kroehne, V., Freudenreich, D., Kaslin, J., & Brand, M. (2012). Short Article Regenerative Neurogenesis from Neural Progenitor Cells Requires Injury-Induced Expression of Gata3. *Developmental Cell*, 1–8. <https://doi.org/10.1016/j.devcel.2012.10.014>
24. Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, *20*(4), 207–220. <https://doi.org/10.1038/s41576-018-0089-8>
25. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
26. Latchman, D. S. (2001). Transcription factors: Bound to activate or repress. *Trends in Biochemical Sciences*, *26*(4), 211–213. [https://doi.org/10.1016/S0968-0004\(01\)01812-6](https://doi.org/10.1016/S0968-0004(01)01812-6)
27. Leinonen, R., Sugawara, H., & Shumway, M. (2011). The sequence read archive. *Nucleic Acids Research*, *39*(SUPPL. 1), 2010–2012. <https://doi.org/10.1093/nar/gkq1019>
28. Lerat, E. (2010). Identifying repeats and transposable elements in sequenced genomes: How to find your way through the dense forest of programs. *Heredity*, *104*(6), 520–533. <https://doi.org/10.1038/hdy.2009.165>

29. Li, Z., Schulz, M. H., Look, T., Begemann, M., Zenke, M., & Costa, I. G. (2019). Identification of transcription factor binding sites using ATAC-seq. *Genome Biology*, *20*(1), 45. <https://doi.org/10.1186/s13059-019-1642-2>
30. Linlin Yan. (2020). *Venn Diagram by ggplot2, with really easy-to-use API* (v0.1.8). <https://github.com/yanlinlin82/ggvenn>
31. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. <https://doi.org/10.1186/s13059-014-0550-8>
32. Marques, I. J., Lupi, E., & Mercader, N. (2019). Model systems for regeneration: Zebrafish. *Development (Cambridge)*, *146*(18). <https://doi.org/10.1242/dev.167692>
33. Orphanides, G., & Reinberg, D. (2002). A unified theory of gene expression. *Cell*, *108*(4), 439–451. [https://doi.org/10.1016/S0092-8674\(02\)00655-4](https://doi.org/10.1016/S0092-8674(02)00655-4)
34. Ou, J., Liu, H., Yu, J., Kelliher, M. A., Castilla, L. H., Lawson, N. D., & Zhu, L. J. (2018). ATCSeqQC. *BMC Genomics*, *19*(1), 169. <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-018-4559-3>
35. Pfefferli, C., & Jaźwińska, A. (2015). The art of fin regeneration in zebrafish. *Regeneration*, *2*(2), 72–83. <https://doi.org/10.1002/reg2.33>
36. Phanstiel, D. H., Boyle, A. P., Araya, C. L., & Snyder, M. P. (2014). Sushi.R: Flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics*, *30*(19), 2808–2810. <https://doi.org/10.1093/bioinformatics/btu379>
37. Phillips, T. (2008). Regulation of Transcription and Gene Expression in Eukaryotes. *Nature Education*, *1*(1), 199.
38. Poss, K. D., Keating, M. T., & Nechiporuk, A. (2003). Tales of regeneration in zebrafish. *Developmental Dynamics*, *226*(2), 202–210. <https://doi.org/10.1002/dvdy.10220>
39. Poss, K. D., Shen, J., & Keating, M. T. (2000). Induction of *lef1* during zebrafish fin regeneration. *Developmental Dynamics*, *219*(2), 282–286. [https://doi.org/10.1002/1097-0177\(2000\)9999:9999<::AID-](https://doi.org/10.1002/1097-0177(2000)9999:9999<::AID-)

DVDY1045>3.3.CO;2-3

40. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
41. Quoseena, M., Vuppaladadium, S., Hussain, S., Banu, S., Bharathi, S., & Idris, M. M. (2020). Functional role of annexins in zebrafish caudal fin regeneration – A gene knockdown approach in regenerating tissue. *Biochimie*, 175, 125–131. <https://doi.org/10.1016/j.biochi.2020.05.014>
42. Ralston, A. & Shaw, K. (2008). Gene Expression Regulates Cell Differentiation. *Nature Education*, 1(1), 127.
43. Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genome Viewer. *Nature Biotechnology*, 29(1), 24–26. <https://doi.org/10.1038/nbt.1754>. Integrative
44. Rodriguez, A. M., & Kang, J. (2020). Regeneration enhancers: Starting a journey to unravel regulatory events in tissue regeneration. *Seminars in Cell and Developmental Biology*, 97(March), 47–54. <https://doi.org/10.1016/j.semcdb.2019.04.003>
45. Ruzicka, L., Howe, D. G., Ramachandran, S., Toro, S., Van Slyke, C. E., Bradford, Y. M., Eagle, A., Fashena, D., Frazer, K., Kalita, P., Mani, P., Martin, R., Moxon, S. T., Paddock, H., Pich, C., Schaper, K., Shao, X., Singer, A., & Westerfield, M. (2019). The Zebrafish Information Network: New support for non-coding genes, richer Gene Ontology annotations and the Alliance of Genome Resources. *Nucleic Acids Research*, 47(D1), D867–D873. <https://doi.org/10.1093/nar/gky1090>
46. Sandelin, A., Alkema, W., Engström, P., Wasserman, W. W., & Lenhard, B. (2004). JASPAR: An open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Research*, 32(DATABASE ISS.). <https://doi.org/10.1093/nar/gkh012>
47. Santos-Ruiz, L., Santamaría, J. A., Ruiz-Sánchez, J., & Becerra, J. (2002). Cell proliferation during blastema formation in the regenerating teleost fin. *Developmental Dynamics*, 223(2), 262–272.

<https://doi.org/10.1002/dvdy.10055>

48. Saxena, S., Purushothaman, S., Meghah, V., Bhatti, B., Poruri, A., Meena Lakshmi, M. G., Sarath Babu, N., Narasimha Murthy, C. L., Mandal, K. K., Kumar, A., & Idris, M. M. (2016). Role of annexin gene and its regulation during zebrafish caudal fin regeneration. *Wound Repair and Regeneration*, 24(3), 551–559. <https://doi.org/10.1111/wrr.12429>
49. Sehring, I. M., & Weidinger, G. (2020). Recent advancements in understanding fin regeneration in zebrafish. *Wiley Interdisciplinary Reviews: Developmental Biology*, 9(1), 1–16. <https://doi.org/10.1002/wdev.367>
50. Seifert, A. W., & Muneoka, K. (2018). The blastema and epimorphic regeneration in mammals. *Developmental Biology*, 433(2), 190–199. <https://doi.org/10.1016/j.ydbio.2017.08.007>
51. Slotkin, R. K. (2018). The case for not masking away repetitive DNA. *Mobile DNA*, 9:15. <https://doi.org/10.1186/s13100-018-0120-9>
52. Sun, Y., Miao, N., & Sun, T. (2019). Detect accessible chromatin using ATAC-sequencing, from principle to applications. *Hereditas*, 156, 29. <https://doi.org/10.1186/s41065-019-0105-9>
53. Team, R. C. (2020). *R: A Language and Environment for Statistical Computing*. <https://www.r-project.org/>
54. Thompson, J. D., Ou, J., Lee, N., Shin, K., Cigliola, V., Song, L., Crawford, G. E., Kang, J., & Poss, K. D. (2020). Identification and requirements of enhancers that direct gene expression during zebrafish fin regeneration. *Development (Cambridge)*, 147(14). <https://doi.org/10.1242/DEV.191262>
55. Tsompana, M., & Buck, M. J. (2014). Chromatin accessibility: A window into the genome. *Epigenetics and Chromatin*, 7(1), 1–16. <https://doi.org/10.1186/1756-8935-7-33>
56. Valdebenito-Maturana, B., Arancibia, E., Riadi, G., Tapia, J. C., & Carrasco, M. (2021). Locus-specific analysis of Transposable Elements during the progression of ALS in the SOD1G93A mouse model. *PLoS ONE*, 16(10 October), 1–14. <https://doi.org/10.1371/journal.pone.0258291>
57. Valdebenito-Maturana, B., & Riadi, G. (2018). TEcandidates: Prediction of

- genomic origin of expressed transposable elements using RNA-seq data. *Bioinformatics*, 34(22), 3915–3916. <https://doi.org/10.1093/bioinformatics/bty423>
58. Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics in Western Equatoria State. *Nature Reviews Genetics*, 10(1), 57.
59. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
60. Yan, F., Powell, D. R., Curtis, D. J., & Wong, N. C. (2020). From reads to insight: A hitchhiker’s guide to ATAC-seq data analysis. *Genome Biology*, 21(1), 1–16. <https://doi.org/10.1186/s13059-020-1929-3>
61. Yang, W. R., Ardeljan, D., Pacyna, C. N., Payer, L. M., & Burns, K. H. (2019). SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Research*, 47(5), 1–16. <https://doi.org/10.1093/nar/gky1301>
62. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of CHIP-Seq (MACS). *Genome Biology*, 9(9). <https://doi.org/10.1186/gb-2008-9-9-r137>
63. Zhao, Y., Liu, C., Liu, J., Kong, Q., Mao, Y., Cheng, H., Li, N., Zhang, X., Li, C., Li, Y., Liu, L., & Ding, Z. (2018). HSPA12B promotes functional recovery after ischaemic stroke through an eNOS-dependent mechanism. *Journal of Cellular and Molecular Medicine*, 22(4), 2252–2262. <https://doi.org/10.1111/jcmm.13507>

10. Anexo

Apartado de Tablas y Figuras realizadas, relacionadas con la investigación principal.

10.1. Figuras suplementarias

Para la totalidad de gráficos de calidad obtenidos por MultiQC se dividen en 3 zonas, separando por calidad cada una de estas de acuerdo con el color presente. El fragmento donde el color es verde quiere decir que la calidad de las librerías es muy buena, cuando el color es amarillo la calidad es aceptable y cuando el color es rojo la calidad es mala.

En el caso de los gráficos de puntajes de calidad promedio (*Mean Quality Scores*) la línea verde representa cada una de las librerías con su correspondiente calidad promedio en un puntaje de Phred en una determinada posición del *read*.

Por otro lado, para los gráficos de puntajes de calidad por secuencia (*Per Sequence Quality Scores*) las líneas verdes representan los valores de calidad de un subconjunto de secuencias.

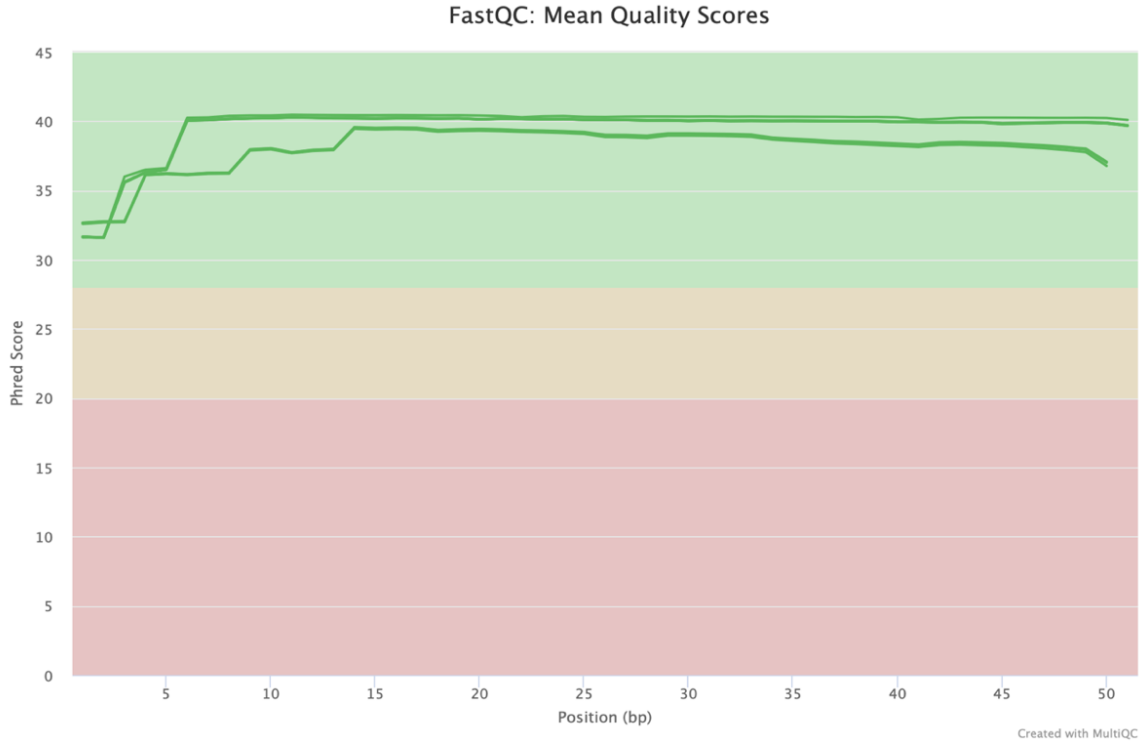


Figura Suplementaria 1. Mean Quality Scores para las 11 librerías de RNA-Seq (Pre-filtro).



Figura Suplementaria 2. Per Sequence Quality Scores para las 11 librerías de RNA-Seq (Pre-filtro).



Figura Suplementaria 3. Mean Quality Scores para las 12 librerías de ATAC-Seq (Pre-filtro).

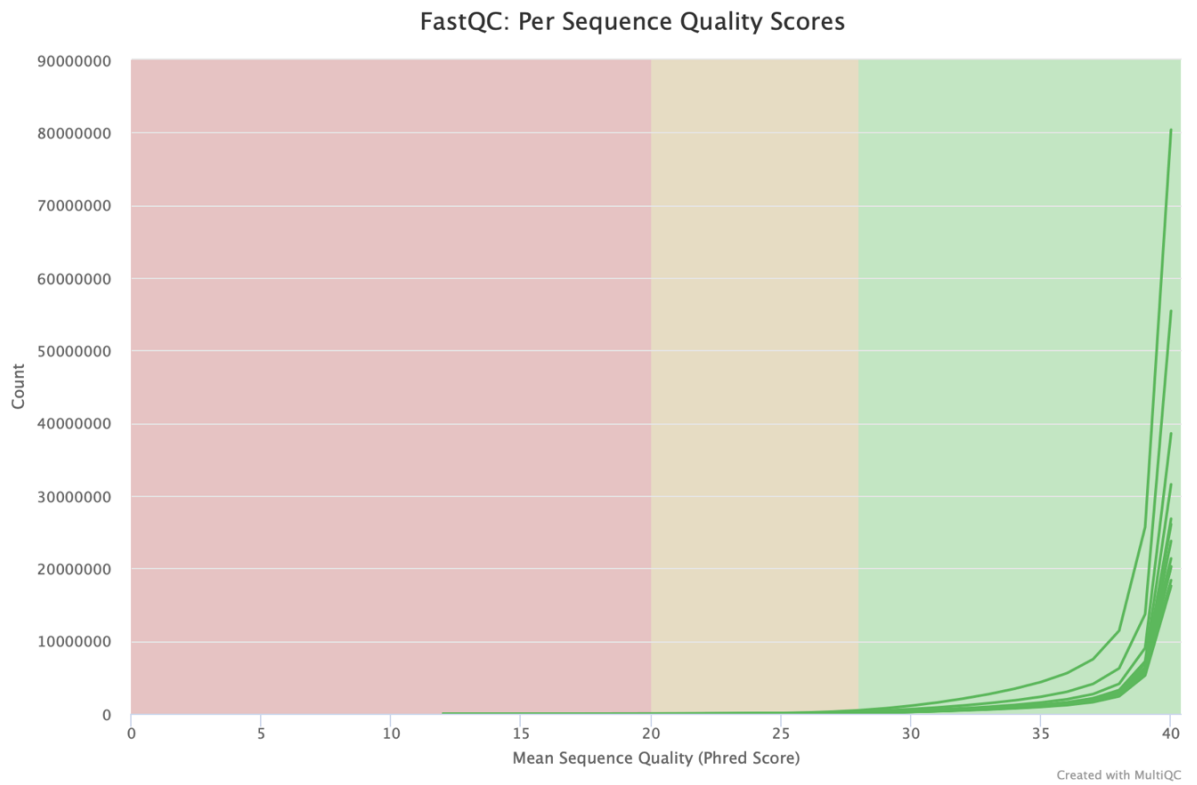


Figura Suplementaria 4. Per Sequence Quality Scores para las 12 librerías de ATAC-Seq (Pre-filtro).

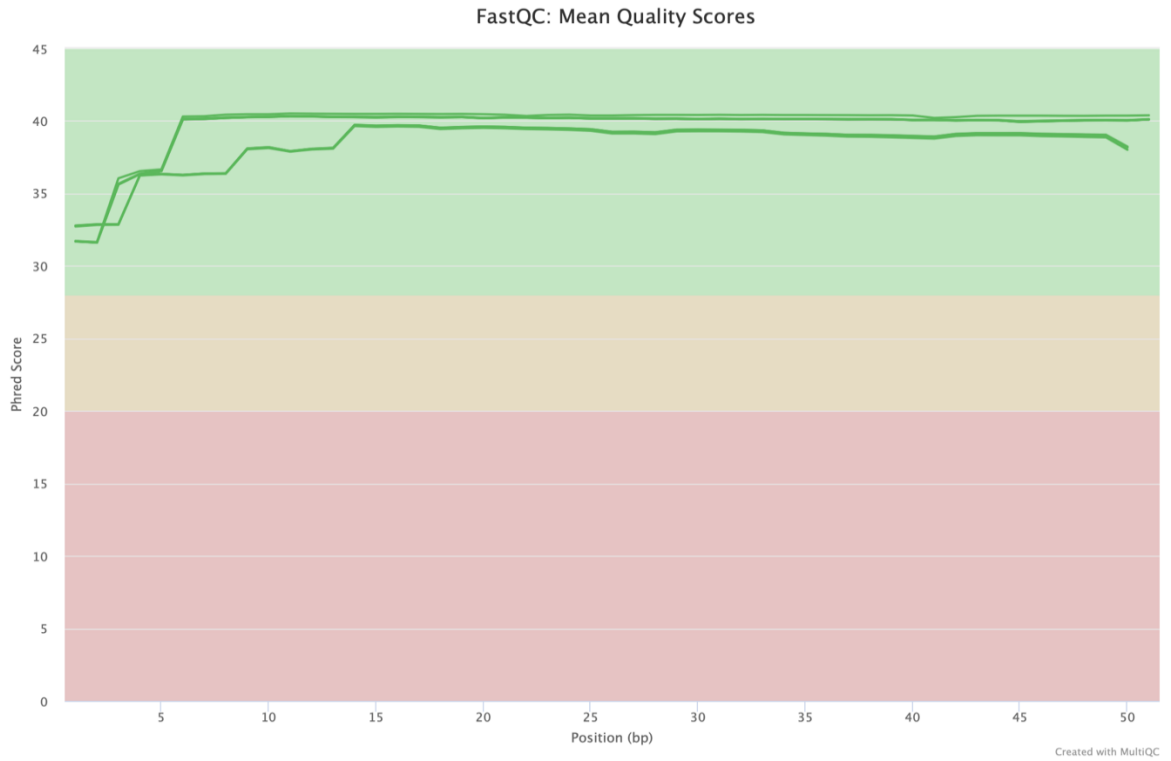


Figura Suplementaria 5. Mean Quality Scores para las 11 librerías de RNA-Seq (Post-filtro).



Figura Suplementaria 6. Per Sequence Quality Scores para las 11 librerías de RNA-Seq (Post-filtro).

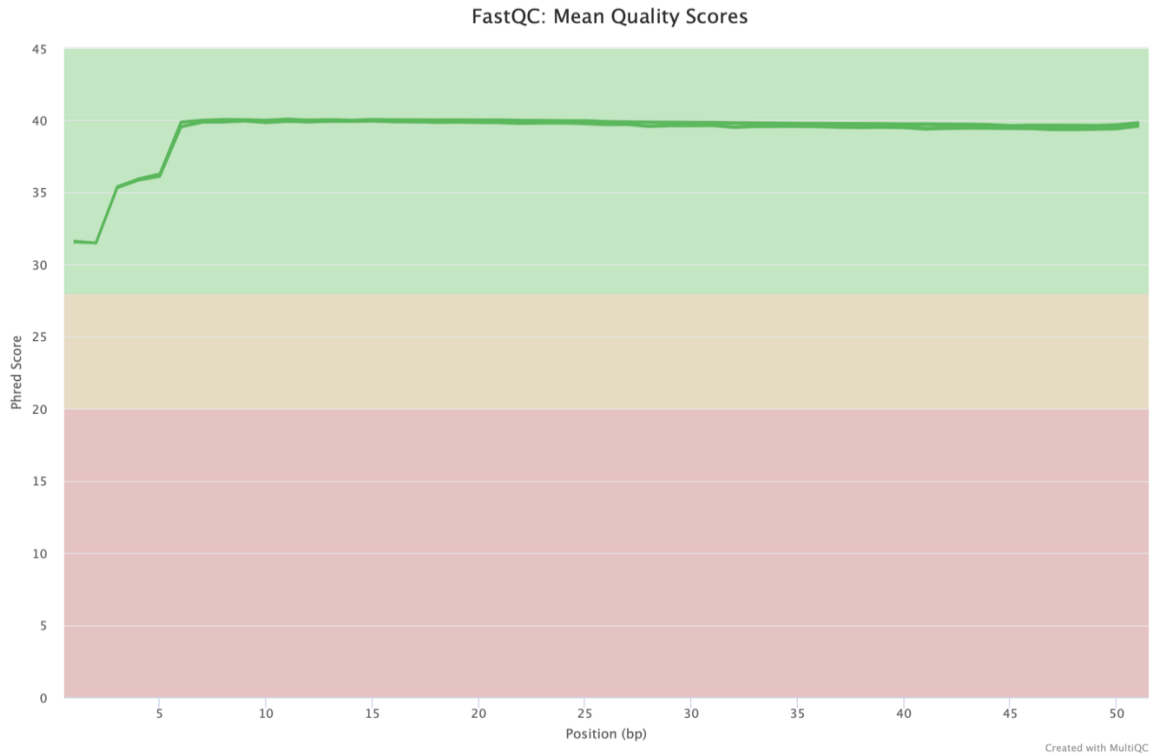


Figura Suplementaria 7. Mean Quality Scores para las 12 librerías de ATAC-Seq (Post-filtro).

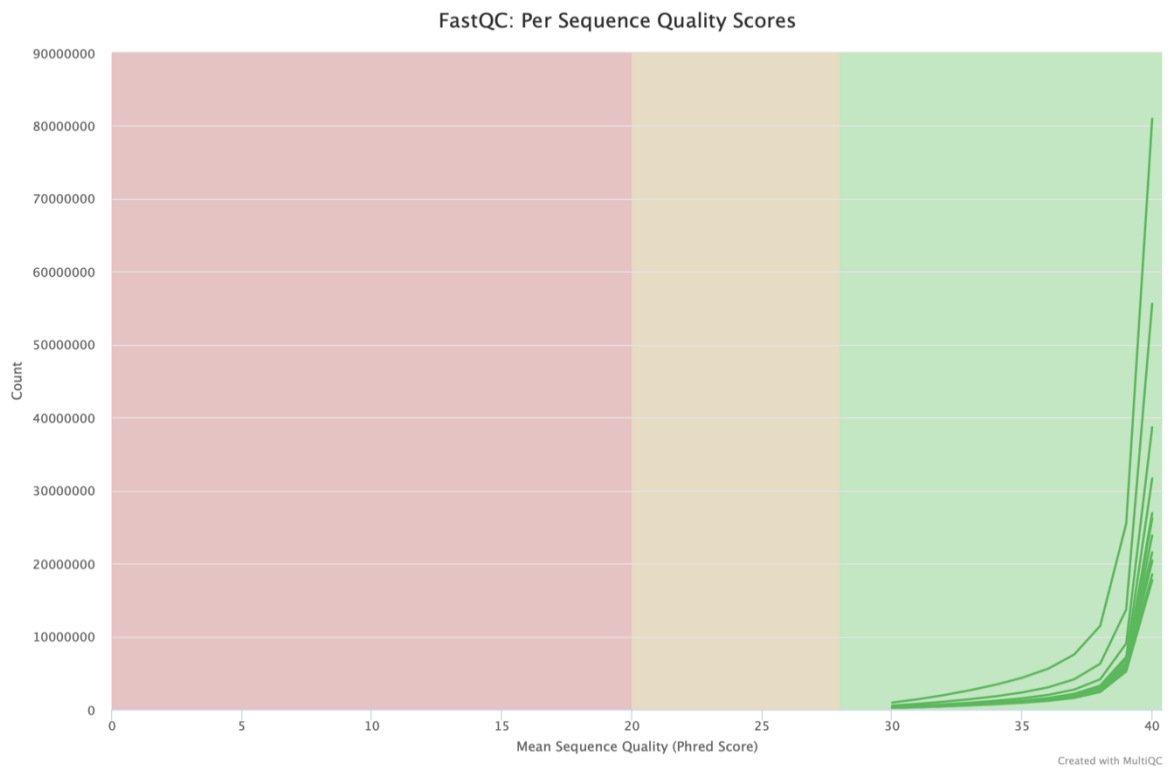


Figura Suplementaria 8. Per Sequence Quality Scores para las 12 librerías de ATAC-Seq (Post-filtro).

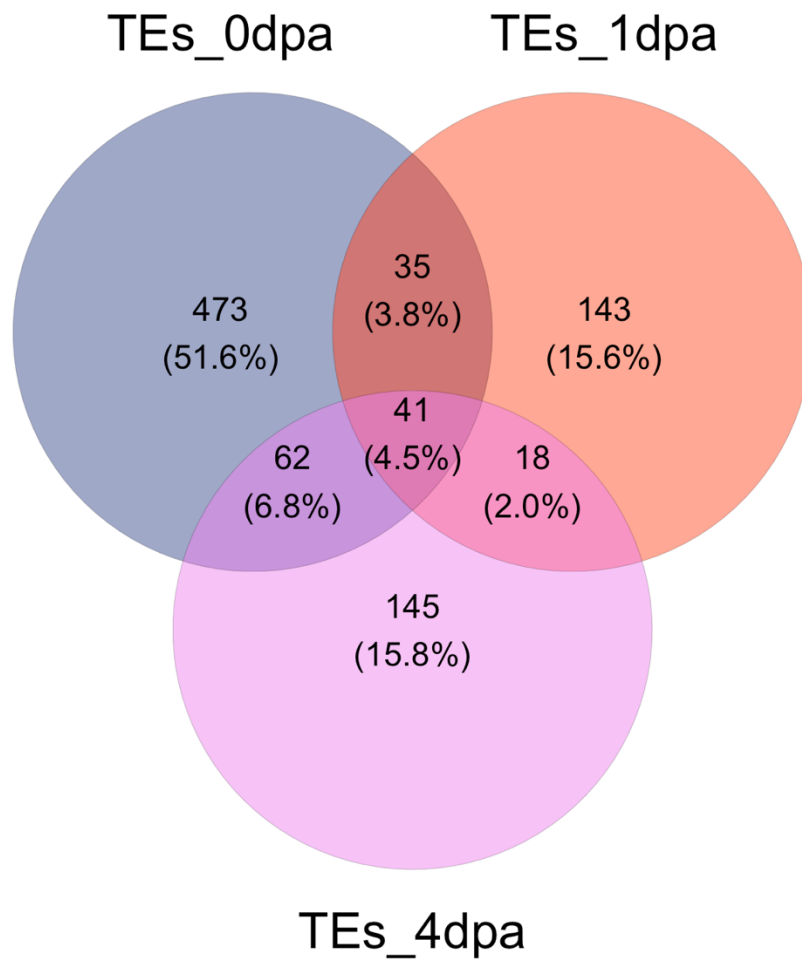


Figura Suplementaria 9. Diagrama de Venn correspondiente a la cantidad total de TEs transcritos en todas las condiciones temporales. En el círculo azul se representan la cantidad y proporción total de TEs identificados en la condición 0dpa (TEs_0dpa), en naranja los TEs identificados en la condición 1dpa (TEs_1dpa) y rosa los TEs identificados en la condición 4dpa (TEs_4dpa). Al medio de todas estas el traslape de TEs identificados entre condiciones con sus respectivos porcentajes del total de TEs transcritos.

10.2. Tablas suplementarias

Condición	Identificador TE	Cantidad de motivos	Motivos
0dpa	chr23 27827220 27828149 Gypsy40-LTR_DR:Gypsy:LTR 7 -	40	MA0004.1.Arnt:2,MA0093.3.USF1:2,MA0464.2.BHLHE40:2,MA0526.3.USF2:2,MA0603.1.Arntl:2,MA0608.1.Creb3l2:2,MA0620.3.MITF:2,MA0636.1.BHLHE41:2,MA0663.1.MLX:2,MA0664.1.MLXIPL:2,MA0692.1.TFEB:2,MA0822.1.HES7:2,MA0828.1.SREBF2(var.2):2,MA0829.2.SREBF1(var.2):2,MA0831.2.TFE3:2,MA0871.2.TFEC:2,MA1464.1.ARNT2:2,MA1571.1.TGIF2LX:2,MA1572.1.TGIF2LY:2,MA0072.1.RORA(var.2):1,MA1151.1.RORC:1
1dpa	chr17 44588944 44590096 Rex1-40_DRe:Rex-Babar:LINE 62 +	56	MA0605.2.ATF3:2,MA0639.1.DBP:2,MA0656.1.JDP2(var.2):2,MA0840.1.Creb5:2,MA1129.1.FOSL1::JUN(var.2):2,MA1136.1.FOSB::JUNB(var.2):2,MA1139.1.FOSL2::JUNB(var.2):2,MA0018.4.CREB1:1,MA0025.2.NFIL3:1,MA0062.3.GABPA:1,MA0076.2.ELK4:1,MA0098.3.ETS1:1,MA0103.3.ZEB1:1,MA0156.2.FEV:1,MA0473.3.ELF1:1,MA0474.2.ERG:1,MA0475.2.FLI1:1,MA0488.1.JUN:1,MA0492.1.JUND(var.2):1,MA0498.2.MEIS1:1,MA0518.1.Stat4:1,MA0598.3.EHF:1,MA0640.2.ELF3:1,MA0731.1.BCL6B:1,MA0750.2.ZBTB7A:1,MA0760.1.ERF:1,MA0761.2.ETV1:1,MA0762.1.ETV2:1,MA0763.1.ETV3:1,MA0764.2.ETV4:1,MA0765.2.ETV5:1,MA0774.1.MEIS2:1,MA0775.1.MEIS3:1,MA0820.1.FIGLA:1,MA0830.2.TCF4:1,MA0833.2.ATF4:1,MA0834.1.ATF7:1,MA0843.1.TEF:1,MA1114.1.PBX3:1,MA1127.1.FOSB::JUN:1,MA1131.1.FOSL2::JUN(var.2):1,MA1140.2.JUNB(var.2):1,MA1143.1.FOSL1::JUND(var.2):1,MA1145.1.FOSL2::JUND(var.2):1,MA1484.1.ETS2:1,M

			A1508.1.IKZF1:1,MA1631.1.ASCL1(var.2):1,MA1632.1.ATF2:1,MA1636.1.CEBPG(var.2):1
4dpa	chr4 58229519 58229728 DNA-2-18_DR:CMC-EnSpm:DNA 77 -	73	MA0068.2.PAX4:2,MA0628.1.POU6F1:2,MA0658.1.LHX6:2,MA0680.1.PAX7:2,MA0699.1.LBX2:2,MA0702.2.LMX1A:2,MA0704.1.Lhx4:2,MA0705.1.Lhx8:2,MA0710.1.NOTO:2,MA0721.1.UNCX:2,MA0725.1.VSX1:2,MA0726.1.VSX2:2,MA0780.1.PAX3:2,MA0886.1.EMX2:2,MA1481.1.DRGX:2,MA0027.2.EN1:1,MA0075.3.PRRX2:1,MA0100.3.MYB:1,MA0132.2.PDX1:1,MA0135.1.Lhx3:1,MA0463.2.BCL6:1,MA0601.1.Arid3b:1,MA0612.2.EMX1:1,MA0618.1.LBX1:1,MA0621.1.mix-a:1,MA0630.1.SHOX:1,MA0634.1.ALX3:1,MA0644.1.ESX1:1,MA0654.1.ISX:1,MA0661.1.MEOX1:1,MA0662.1.MIXL1:1,MA0674.1.NKX6-1:1,MA0675.1.NKX6-2:1,MA0681.2.PHOX2B:1,MA0701.2.LHX9:1,MA0703.2.LMX1B:1,MA0713.1.PHOX2A:1,MA0715.1.PROP1:1,MA0716.1.PRRX1:1,MA0717.1.RAX2:1,MA0718.1.RAX:1,MA0720.1.Shox2:1,MA0722.1.VAX1:1,MA0723.1.VAX2:1,MA0807.1.TBX5:1,MA0854.1.Alx1:1,MA0876.1.BSX:1,MA0879.1.Dlx1:1,MA0882.1.DLX6:1,MA0889.1.GBX1:1,MA0892.1.GSX1:1,MA0893.2.GSX2:1,MA1463.1.ARGFX:1,MA1519.1.LHX5:1,MA1558.1.SNAI1:1,MA1577.1.TLX2:1,MA1620.1.Ptf1a(var.3):1,MA1624.1.Stat5a:1

Tabla Suplementaria 1. TEs con presencia de más motivos diferentes por cada condición.

En la columna 1 se visualiza la condición temporal en la que se presentan los motivos. En la columna 2 el identificador del TE donde están los motivos. En la columna 3 se visualiza la cantidad de motivos que tiene ese TE en particular. Finalmente, en la columna 4 están los motivos asociados encontrados en ese TE en la condición en particular,

Condición	Identificador TE	TF asociado a Motivo
0dpa	chr25 31658736 31658862 ENSPM-6_DR:CMC-EnSpm:DNA 288 - chr9 45454182 45454242 hAT-N52_DR:hAT-Ac:DNA 100 + chr9 45454219 45454279 hAT-N52_DR:hAT-Ac:DNA 117 +	sox2
1dpa	chr1 30877682 30878001 DNA-8-13_DR:hAT?:DNA 149 + chr1 40443214 40443286 DNA-8-5_DR:hAT:DNA 194 + chr3 4765149 4765327 Gypsy56-I_DR:Gypsy:LTR 331 + chr3 10584477 10584871 DNA-8-13_DR:hAT?:DNA 106 + chr3 36420791 36421532 DNA-8-13_DR:hAT?:DNA 105 - chr4 64445725 64445913 Kolobok-N4_DR:Kolobok-T2:DNA 218 + chr5 13526274 13526348 TDR24:DNA:DNA 81 + chr6 9500000 9511293 DNA-X-12_DR:DNA:DNA 157 - chr6 14160459 14160627 ERV1-N3-I_DR:LTR:LTR 261 - chr6 14160539 14160654 HATN15_DR:hAT:DNA 219 + chr6 35157375 35157747 DNA-X-9_DR:DNA:DNA 207 - chr6 43021844 43022103 Looper-N7_DR:PiggyBac:DNA 120 + chr7 63360375 63361520 Harbinger3N_DR:PIF-Harbinger:DNA 51 + chr8 34556426 34556546 L2-33_DRe:L2:LINE 210 + chr8 34702069 34702189 L2-33_DRe:L2:LINE 210 + chr8 39581559 39582416 L1-17_DR:L1:LINE 314 + chr9 14023306 14023359 ANGEL:hAT:DNA 170 - chr9 32839863 32839991 ENSPM-6_DR:CMC-EnSpm:DNA 154 - chr9 45454073 45454131 hAT-N52_DR:hAT-Ac:DNA 138 + chr9 45454108 45454168 hAT-N52_DR:hAT-Ac:DNA 150 + chr9 45454145 45454205 hAT-N52_DR:hAT-Ac:DNA 133 + chr9 45454182 45454242 hAT-N52_DR:hAT-Ac:DNA 100 + chr9 45454219 45454279 hAT-N52_DR:hAT-Ac:DNA 117 +	<i>gata3</i> <i>lef1</i> <i>tp63</i> <i>sox2</i> <i>nr3c1</i> <i>sox10</i>

	<p>chr12 35945540 35945892 DNA-8-13_DR:hAT?:DNA 112 - chr13 15230674 15231665 hAT-N71_DR:hAT:DNA 12 - chr14 20352733 20352950 hAT-N54_DR:hAT-Ac:DNA 162 - chr15 13982009 13982525 DNA25TWA1_DR:DNA:DNA 201 + chr15 17387511 17387829 DNA-8-13_DR:hAT?:DNA 142 + chr18 30026072 30026192 L1-2_DR:L1-Tx1:LINE 212 - chr18 30034231 30034351 L1-2_DR:L1-Tx1:LINE 212 - chr18 30037258 30037378 L1-2_DR:L1-Tx1:LINE 212 - chr18 30039227 30039347 L1-2_DR:L1-Tx1:LINE 212 - chr18 30048649 30048769 L1-2_DR:L1-Tx1:LINE 212 - chr18 30054540 30054660 L1-2_DR:L1-Tx1:LINE 212 - chr18 30065012 30065132 L1-2_DR:L1-Tx1:LINE 212 - chr18 30066972 30067092 L1-2_DR:L1-Tx1:LINE 212 - chr18 30071122 30071242 L1-2_DR:L1-Tx1:LINE 212 - chr18 39813095 39814075 TC1DR3:TcMar-Tc1:DNA 87 - chr19 5580016 5580494 Copia-7-LTR_DR:Gypsy?:LTR 203 - chr19 47879461 47879515 LOOPERN3_DR:PiggyBac:DNA 186 + chr20 5186545 5187100 DNA25TWA1_DR:DNA:DNA 29 - chr21 26743310 26743575 EnSpm-N6_DR:CMC-EnSpm?:DNA 165 - chr24 19348531 19349695 hAT-N60_DR:hAT-Ac:DNA 103 - chr25 20647239 20647952 LOOPERN4_DR:PiggyBac:DNA 214 - chr25 31658736 31658862 ENSPM-6_DR:CMC-EnSpm:DNA 288 -</p>	
	<p>chr1 51898839 51898894 Polinton-1_DR:Maverick:DNA 185 - chr1 54783797 54783891 DNA-2-31_DR:TcMar-ISRm11:DNA 202 - chr2 7838950 7839639 TC1DR3:TcMar-Tc1:DNA 78 - chr2 13740132 13740580 LOOPERN3_DR:PiggyBac:DNA 126 + chr2 41856385 41856480 DNA-5-1_DR:DNA:DNA 213 - chr3 3610051 3610132 Nimb-2_DR:l:LINE 192 + chr3 36420791 36421532 DNA-8-13_DR:hAT?:DNA 105 - chr4 2469860 2470250 Gypsy40-LTR_DR:Gypsy:LTR 120 +</p>	

4dpa	chr4 51988246 51992102 BEL-37-I_DR:Pao:LTR 27 + chr4 62086054 62088405 Dada-tA_DR:Dada:DNA 37 - chr4 75242042 75242901 hAT-N26_DR:hAT?:DNA 65 + chr5 24872762 24873087 HE2_DR:tRNA-V:SINE 157 + chr6 936546 936749 Gypsy-31-LTR_DR:Gypsy:LTR 174 + chr6 14202019 14202166 DIRS1_DR:DIRS:LTR 230 - chr6 57477172 57477504 DNA-8-13_DR:hAT?:DNA 104 + chr7 44613031 44614085 EnSpm-6N1_DR:CMC-EnSpm:DNA 209 + chr7 53092917 53093800 Tc1-8B_DR:TcMar-Tc1:DNA 330 - chr7 66566460 66567889 ERV1-N3-LTR_DR:LTR:LTR 121 + chr8 34556426 34556546 L2-33_DRe:L2:LINE 210 + chr8 34727364 34727484 L2-33_DRe:L2:LINE 210 + chr8 49312839 49313547 hAT-N46_DR:hAT-Tip100:DNA 96 - chr9 45454367 45454427 hAT-N52_DR:hAT-Ac:DNA 100 + chr9 45454404 45454464 hAT-N52_DR:hAT-Ac:DNA 150 + chr9 45454441 45454501 hAT-N52_DR:hAT-Ac:DNA 150 + chr9 56122243 56122389 Polinton-1_DR:Maverick:DNA 253 - chr9 56122274 56122474 ERV1-N3-I_DR:LTR:LTR 221 - chr9 56122332 56122491 LOOPERN2_DR:Kolobok-T2:DNA 275 + chr13 2035209 2035813 TC1DR3:TcMar-Tc1:DNA 67 - chr14 49873844 49874707 LOOPERN3_DR:PiggyBac:DNA 117 + chr15 4745282 4746317 DNA8-3_DR:DNA:DNA 183 - chr15 17383847 17384168 DNA-8-13_DR:hAT?:DNA 108 + chr15 17387511 17387829 DNA-8-13_DR:hAT?:DNA 142 + chr15 35223825 35224118 Kolobok-N4_DR:Kolobok-T2:DNA 188 + chr15 43639267 43639466 ERV1-N3-I_DR:LTR:LTR 209 + chr16 16128535 16128990 Copia-6-I_DR:Gypsy?:LTR 380 + chr16 21974074 21986734 DNA-X-12_DR:DNA:DNA 169 - chr17 27966931 27967205 TC1DR3:TcMar-Tc1:DNA 192 + chr18 30026072 30026192 L1-2_DR:L1-Tx1:LINE 212 - chr18 30032265 30032385 L1-2_DR:L1-Tx1:LINE 212 -	<i>gata3</i> <i>lef1</i> <i>tp63</i> <i>sox2</i> <i>nr3c1</i> <i>sox10</i>
------	---	---

chr18 30034231 30034351 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 30037258 30037378 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 30039227 30039347 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 30048649 30048769 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 30050611 30050731 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 30052581 30052701 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 30054540 30054660 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 30065012 30065132 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 30071122 30071242 L1-2_DR:L1-Tx1:LINE 212 -	
chr18 39747477 39747670 LTR-11_DR-LTR:LTR:LTR 156 -	
chr19 9147290 9149946 DNA-8-13_DR:hAT?:DNA 26 -	
chr21 1297200 1298381 TC1DR3:TcMar-Tc1:DNA 57 -	
chr22 6064273 6064427 L2-33_DRe:L2:LINE 307 -	
chr23 45567877 45568291 Tc1-4_DR:TcMar-Tc1:DNA 111 +	
chr25 338880 339262 Tc1N1_DR:CMC-EnSpm?:DNA 157 +	
chr25 31656555 31656732 ERV1-N3-I_DR:LTR:LTR 335 -	
chr25 31658736 31658862 ENSPM-6_DR:CMC-EnSpm:DNA 288 -	

Tabla Suplementaria 2. TEs que presentan sitios de unión a TFs relacionados a regeneración. En la columna 1 se visualiza la condición temporal en la que se presentan los TFBSs. En la columna 2 el identificador del TE donde se encuentran. Finalmente, en la columna 3 están los identificadores de los TFs encontrados relacionados a regeneración.