

Escuela de Ingeniería Civil en Bioinformática.



Análisis de expresión de los Elementos Transponibles a nivel espacial en un modelo de Esclerosis Lateral Amiotrófica.

Memoria para optar al título de Ingeniero Civil en Bioinformática

Rocío Valdés Molina

Matrícula: 2016430023

Profesor tutor: Dr. Braulio Valdebenito Maturana

Profesor informante: Dr. Gonzalo Riadi

Memoria de título, Ingeniería civil en bioinformática
Semestre 2-2021

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, 2022

Resumen

La Esclerosis Lateral Amiotrófica (ELA) es una enfermedad neurodegenerativa mortal, y puede dividirse en dos tipos: ELA esporádica y la ELA familiar. Uno de los genes asociados a esta última es superóxido dismutasa 1 (SOD1), el cual influye en la enfermedad a través de una ganancia tóxica provocada por agregados de esta proteína mal plegada acumulada en el citoplasma. El modelo más utilizado y que recapitula la enfermedad es aquel basado en ratones transgénicos que llevan la mutación G93A de SOD1 (abreviada como SOD1^{G93A}).

Los elementos transponibles (*Transposable Elements*, TEs) son elementos genéticos que tienen la capacidad de moverse e integrarse en otra parte del genoma. A pesar de que la mayoría de ellos son inactivos, hay evidencia que los vincula con regulación génica tanto en condiciones normales como en enfermedad. Gran parte de los trabajos en donde han estudiado TEs en ELA ha sido mediante el método de RNA-Sequencing (RNA-Seq) estándar. Si bien este método permite caracterizar con alta resolución el transcriptoma en una determinada condición, pierde el contexto espacial de los genes expresados. Para resolver este problema, un nuevo avance es la técnica de Spatial Transcriptomics (ST) la cual permite conocer en qué secciones de un tejido se encuentran los transcritos. Recientemente se realizó un estudio ST en el modelo de ratón SOD1^{G93A}. Así, se logró caracterizar el repertorio genético en el cuerno ventral de la médula espinal, principal región anatómica en donde se encuentran las motoneuronas afectadas por ELA. No obstante, en dicho trabajo no estudiaron la distribución de los TEs a lo largo de la médula espinal. Dado el potencial rol regulatorio de TEs, y de que ya hay evidencia que muestra que estos son transcripcionalmente activos en ELA, en esta investigación se buscó determinar la expresión de TEs a nivel espacial en la médula espinal. La hipótesis de este trabajo es que los TEs están expresados preferencialmente en el cuerno ventral de la médula espinal en el modelo de ratón SOD1^{G93A}. Los resultados muestran que existen TEs expresados en el cuerno ventral, pero no se aprecia un enriquecimiento de expresión de estos en dicha zona, en comparación a otras regiones de la médula espinal, por lo que se propone estudios futuros en donde contemplen algoritmos estadísticos para medir la diferentes tipos de expresión.

Abstract

Amyotrophic Lateral Sclerosis (ALS) is a mortal neurodegenerative disease, and can be separated into two types: sporadic ALS and familial ALS. One of the genes associated with the latter is superoxide dismutase 1 (SOD1), which influences the disease through a toxic gain caused by aggregates of this misfolded protein accumulated in the cytoplasm. The most commonly used model that recapitulates the disease is based on transgenic mice carrying the G93A mutation of SOD1 (abbreviated as SOD1^{G93A}).

Transposable elements (TEs) are genetic elements that have the capacity to move and integrate into another part of the genome. Although most of them are inactive, there is evidence linking them to gene regulation both in normal conditions and in disease. The majority of the works in which TEs have been studied in ALS has been by the standard RNA-Sequencing (RNA-Seq) method. Although this method allows high-resolution characterization of the transcriptome in a given condition, it loses the spatial context of the expressed genes. To resolve this problem, a new advance is the Spatial Transcriptomics (ST) technique, which allows us to know in which sections of a tissue the transcripts are located. An ST study was recently carried out in the SOD1^{G93A} mouse model. Thus, it was possible to characterize the genetic repertoire in the ventral horn of the spinal cord, the main anatomical region where the motor neurons affected by ALS are located. However, in that work they did not study the distribution of TEs along the spinal cord. Given the potential regulatory role of TEs, and that there is already evidence showing that they are transcriptionally active in ALS, in this research we sought to determine the expression of TEs at the spatial level in the spinal cord. The hypothesis of this work is that TEs are preferentially expressed in the ventral horn of the spinal cord in the SOD1^{G93A} mouse model.

The results show that there are TEs expressed in the ventral horn, but there is no enrichment of their expression in this area, compared to other regions of the spinal cord, so future studies are proposed using statistical algorithms to measure if there is a difference in expression.

Índice

Resumen	1
Abstract	2
Índice	3
Índice de Figuras	5
Índice de Tablas	5
Índice de imágenes suplementarias.	6
Índice Tablas suplementaria	7
1. Introducción	7
1.1 Esclerosis Lateral Amiotrófica	8
1.2 Transcriptómica	12
1.3 Elementos Transponibles	17
Problemática	18
2. Hipótesis	19
3. Objetivos	19
Objetivo general.	19
Objetivos específicos.	19
4. Materiales y métodos.	20
4.1 Materiales.	20
4.1.1 Software y plataformas	20
4.1.2 Muestras de RNA-seq	21
4.1.3 Hardware	21
4.2 Metodología	22
4.2.1 Identificar TEs expresados en muestras de Spatial Transcriptomics provenientes del ratón transgénico SOD1G93A.	23
4.2.1.1 Descarga de archivos.	23
4.2.1.2 Construcción de índice.	23
4.2.1.3 Descarga de los datos de RNA-seq.	23
4.2.1.4 Obtención de archivo que contiene los códigos de barra y las coordenadas espaciales.	24
4.2.1.5 Análisis Spatial Transcriptomics.	24
4.2.1.6 Identificar TEs en los datos.	25
4.2.2 Analizar la distribución de los TEs expresados en las regiones anatómicas de la médula espinal.	25
4.2.2.1 Análizar la posición espacial de los TEs	25
6. Resultados	26
6.1 Análisis de calidad de los reads.	26
6.2 Identificar Elementos Transponibles expresados en muestras de Spatial Transcriptomics provenientes del ratón transgénico SOD1 G93A .	27

6.2.1 Realizar Spatial transcriptomics	27
6.2.2 Identificar TEs en los datos.	28
6.4 Analizar la distribución de los TEs expresados en las regiones anatómicas de la médula espinal.	29
6.4.1 Analizar distribución de TEs en los tejidos.	29
7. Conclusión.	32
Referencias	33
Material Suplementario	38

Índice de Figuras

Figura 1. Esquema de manifestaciones clínicas de ELA.	8
Figura 2. Distribución de genes por tipo de ELA.	10
Figura 3. Regiones anatómicas de la médula espinal.	12
Figura 4. Esquema general de RNA-seq.	13
Figura 5. Obtención de datos ST	14
Figura 6. Ejemplo de visualización de expresión génica mediante ST.	15
Figura 7. Expresión espacial de GFAP.	16
Figura 8. Clasificación de elementos transponibles.	17
Figura 9. Diagrama de flujo de metodología seguida.	22
Figura 10. Puntuaciones medias de calidad.	26
Figura 11. Puntaje de calidad secuencia.	27
Figura 12. Análisis de la muestra SRR7895713.	28
Figura 13. Análisis con SpatialTE de la muestra SRR7895713 del grupo Subfamily.	29
Figura 14. Procesamiento de imágenes.	30
Figura 15. Análisis con SpatialTE de la muestra SRR7895713 para los diferentes tipos de TEs.	31

Índice de Tablas

Tabla 1. Proteínas más expresadas en la corteza cerebral humana.	10
Tabla 2. Software y plataformas utilizadas.	20
Tabla 3. Tabla de datos analizados.	24

Índice de imágenes suplementarias.

Figura 1. Expresión de TEs en la médula espinal (muestra SRR7895536) usando output a nivel SubFamily de SpatialTE.	40
Figura 2. Expresión de TEs en la médula espinal (muestra SRR7895536) usando output a nivel Locus de SpatialTE.	41
Figura 3. Expresión de TEs en la médula espinal (muestra SRR7895537) usando output a nivel SubFamily de SpatialTE.	42
Figura 4. Expresión de TEs en la médula espinal (muestra SRR7895537) usando output a nivel Locus de SpatialTE.	43
Figura 5. Expresión de TEs en la médula espinal (muestra SRR7895538) usando output a nivel SubFamily de SpatialTE.	44
Figura 6. Expresión de TEs en la médula espinal (muestra SRR7895538) usando output a nivel Locus de SpatialTE.	45
Figura 7. Expresión de TEs en la médula espinal (muestra SRR7895712) usando output a nivel SubFamily de SpatialTE.	46
Figura 8. Expresión de TEs en la médula espinal (muestra SRR7895712) usando output a nivel Locus de SpatialTE.	47
Figura 9. Expresión de TEs en la médula espinal (muestra SRR7895713) usando output a nivel SubFamily de SpatialTE.	48
Figura 10. Expresión de TEs en la médula espinal (muestra SRR7895713) usando output a nivel Locus de SpatialTE.	49
Figura 11. Expresión de TEs en la médula espinal (muestra SRR7895714) usando output a nivel SubFamily de SpatialTE.	50
Figura 12. Expresión de TEs en la médula espinal (muestra SRR7895714) usando output a nivel Locus de SpatialTE.	51
Figura 13. Expresión de TEs en la médula espinal (muestra SRR7895715) usando output a nivel SubFamily de SpatialTE.	52
Figura 14. Expresión de TEs en la médula espinal (muestra SRR7895715) usando output a nivel Locus de SpatialTE.	53
Figura 15. Expresión de TEs en la médula espinal (muestra SRR7895716) usando output a nivel SubFamily de SpatialTE.	54
Figura 16. Expresión de TEs en la médula espinal (muestra SRR7895716) usando output a nivel Locus de SpatialTE.	55
Figura 17. Expresión de TEs en la médula espinal (muestra SRR7895717) usando output a nivel SubFamily de SpatialTE.	56

Figura 18. Expresión de TEs en la médula espinal (muestra SRR7895717) usando output a nivel Locus de SpatialTE.	57
Figura 19. Expresión de TEs en la médula espinal (muestra SRR7895718) usando output a nivel SubFamily de SpatialTE.	58
Figura 20. Expresión de TEs en la médula espinal (muestra SRR7895718) usando output a nivel Locus de SpatialTE.	69

Índice Tablas suplementaria

Tabla 1. Tabla de datos para estudio.	38
Tabla 2. Tabla de datos analizados.	39

1. Introducción

1.1 Esclerosis Lateral Amiotrófica

La Esclerosis Lateral Amiotrófica (ELA) es una enfermedad neurodegenerativa mortal cuya sintomatología no comienza al contraerse, lo que la hace difícil de estudiar. Es causada por la degeneración progresiva y selectiva de motoneuronas superiores e inferiores en el cerebro y en la médula espinal (Phatnani et al., 2013a). Su nombre fue acuñado por el neurólogo francés Jean-Martin, describiendo “esclerosis lateral” como cicatrización o endurecimiento de los tejidos en la médula espinal lateral y “amiotrófica” se refiere a atrofia muscular (Grad et al., 2017). Sus principales síntomas son debilidad muscular, disfagia, fasciculaciones y espasticidad (Figura 1). La enfermedad, por su naturaleza, es irreversible, y hasta el momento no tiene cura.

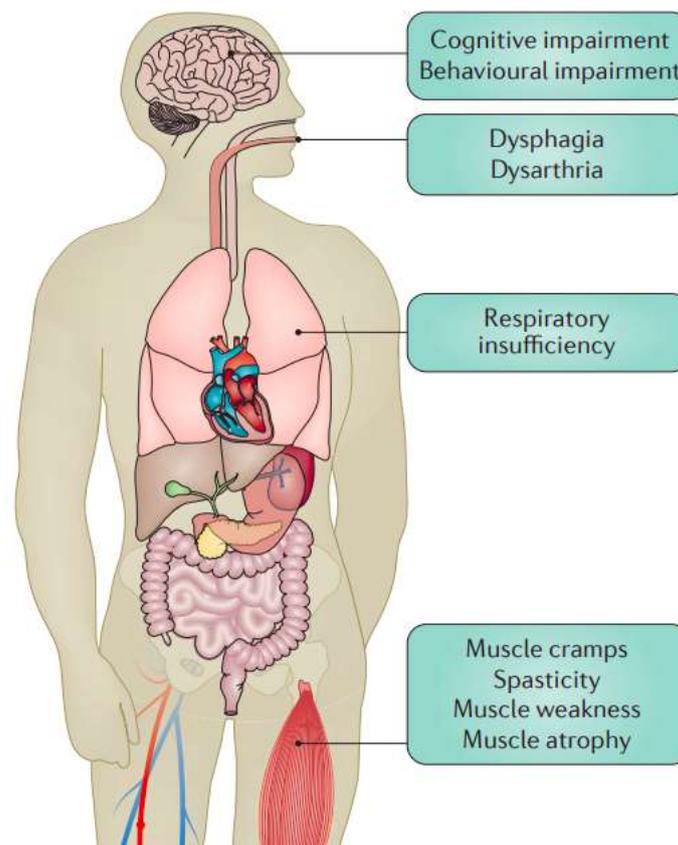


Figura 1. Esquema de manifestaciones clínicas de ELA. Se esquematiza el cuerpo humano y los principales órganos y tejidos afectados por la enfermedad. A nivel cerebral se asocia tanto deterioro cognitivo como del comportamiento, mientras que en los músculos en general se presentan calambres, espasticidad, debilidad y atrofia. Sistemas como el digestivo también se ven afectados, ya que esta enfermedad provoca disfagia. Otro sistema altamente afectado es el respiratorio a través de la insuficiencia respiratoria. Figura tomada de (Hardiman et al., 2017).

La ELA es una enfermedad relativamente rara, la cual posee un origen desconocido para la mayoría de sus pacientes. A pesar de esto es la enfermedad de motoneuronas más común en adultos, afectando a personas de todas las etnias, con una prevalencia aproximada de 2 casos por 100.000 habitantes, siendo más frecuente en hombres que en mujeres y además en personas blancas, las que presentan casi un doble de los casos registrados en personas de color (Mehta, 2018). La concentración etaria de esta enfermedad se encuentra en el término de la edad adulta, con una media de aparición a los 65 años, pero de igual manera, en alrededor del 10% de los casos se presenta antes de los 25 años. Estudios especulan que estos afectados de edad temprana deben tener una mayor carga genética de riesgo, asociándose a antecedentes de familiares positivos (Turner et al., 2012). La esperanza de vida de pacientes con ELA es baja, siendo el fallo respiratorio la causa más frecuente de hospitalización y muerte de estos. Esto ocurre mayoritariamente debido a una disfunción muscular respiratoria (Martínez-Llorens & Balañá Corberó, 2019). La mediana de supervivencia varía entre 20 a 48 meses posteriores al inicio de los síntomas, y sólo entre un 5 - 10% de los pacientes sobreviven más de 10 años.

Los casos de ELA pueden dividirse en dos grandes grupos: ELA esporádica, que no tiene un origen genético claro, comprende aproximadamente 90% de los casos, y ELA familiar, que corresponde a la proporción restante, representando a los pacientes que poseen familiares que han sufrido de la enfermedad. Hay algunos genes que se han asociado con ambos tipos de ELA (Figura 2), donde se destaca el gen superóxido dismutasa 1 (SOD1) (Riancho et al., 2019), el cual posee una longitud de 9.310 bp y se compone por 5 exones. Este gen codifica para una metaloenzima de 153 aminoácidos, tiene un peso de 16 kDa, su estructura cuaternaria está formada por hojas beta en forma de barril (Pansarasa et al., 2018). Su función principal es convertir los radicales libres superóxido intracelulares más un producto de desecho tóxico de la fosforilación oxidativa mitocondrial en un peróxido de hidrógeno que a su vez elimina mediante la acción de otras enzimas captadoras de radicales libres (Shaw, 2005). La SOD1 es abundante en el sistema nervioso central, representando ~1% de la proteína cerebral, y también se expresa en todos los demás tejidos. En particular en la corteza cerebral, es la proteína más abundante (Tabla 1, (Roudeau et al., 2021)).

Para SOD1 se han encontrado más de 100 mutaciones diferentes la mayoría corresponde a mutaciones sin sentido (*Mis-sense*), con un pequeño número de mutaciones por delección e inserción que dan como resultado polipéptidos SOD1 truncados (Shaw, 2005).

Gen	Abundancia relativa
SOD1	64.5 %
HBA1	7.4 %
HBB	4.6 %
ALB	2.4 %
PDRX6	2.1%

Tabla 1. Proteínas más expresadas en la corteza cerebral humana. Principales proteínas en la corteza cerebral humana detectadas en geles de enfoque isoelectrico y cromatología de exposición por tamaño. La abundancia relativa se redondea a 0.1% más cercano. Adaptado de (Roudeau et al., 2021)

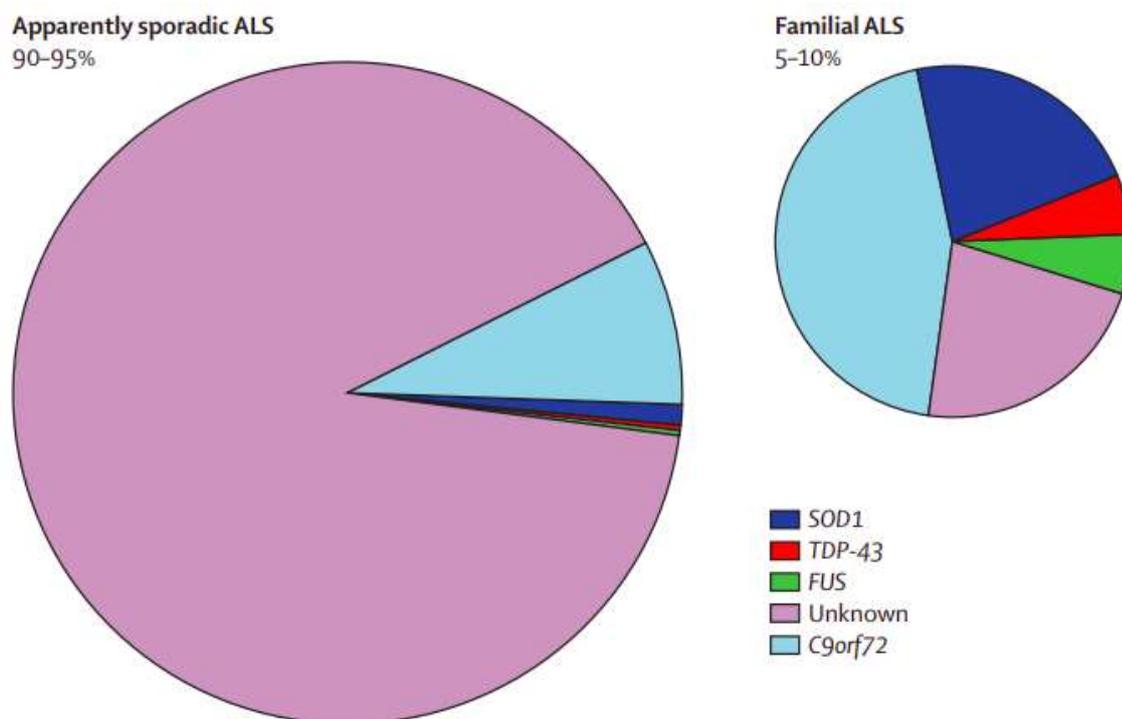


Figura 2. Distribución de genes por tipo de ELA. En la izquierda se aprecia la distribución de ELA esporádica, a la derecha ELA de origen familiar. La simbología correspondiente para ambos gráficos se encuentra en la parte inferior derecha. Figura tomada de (Turner et al., 2013).

La evidencia sugiere que las mutaciones de SOD1 causan ELA a través de una ganancia tóxica provocada por agregados de proteínas mal plegadas que se acumulan en citoplasma (Ghadge et al., 2020). Esto ocurre ya que la presencia de zinc en SOD1 ayuda a entregar estabilidad en su estructura (Shaw, 2005).

Para lograr un análisis de este gen se utilizan animales y células transgénicas (Julien & Kriz, 2006). Uno de los modelos más utilizados es el ratón transgénico que sobreexpresa SOD1 humana con una mutación de glicina a alanina en la posición 93 (SOD1^{G93A}). El desarrollo de la enfermedad en este modelo de ratón recapitula la progresión de la enfermedad observada en humanos y se estima que los conocimientos derivados de su estudio serán informativos para la patología de la ELA tanto esporádica como familiar (Phatnani et al., 2013b). Particularmente, el impacto de la mutación SOD1^{G93A} se refleja en una disminución de actividad enzimática, lo que conlleva a una menor eliminación de radicales libres dentro del núcleo. Como consecuencia de esto, hay mayor facilidad de que ocurran alteraciones y daños en el DNA, provocando alteraciones en el perfil de expresión de proteínas en las motoneuronas (Sau et al., 2007).

ELA afecta exclusiva y específicamente a las motoneuronas. Las motoneuronas a través de sus interacciones permiten regular y mantener los estados funcionales y la conectividad sináptica en la médula espinal (Cartarozzi et al., 2018), esto debido a que los axones que inervan las terminaciones primarias de los husos musculares en la periferia, los que se proyectan en el cuerno ventral y poseen conexiones mono sinápticas con las motoneuronas (Snider et al., 1992). A través de la técnica de microscopía de barrido láser de los fotones *in vivo* en ratones transgénicos, se logró visualizar que las motoneuronas se encuentran localizadas mayoritariamente en el cuerno ventral (Cartarozzi et al., 2018). Así, distintas regiones de la médula espinal se ven afectadas de distinta manera por la enfermedad. Se destaca este punto, ya que estas diferencias anatómicas son importantes en la problemática que se abordará en este trabajo. Las regiones anatómicas de la médula espinal se muestran en la Figura 3.

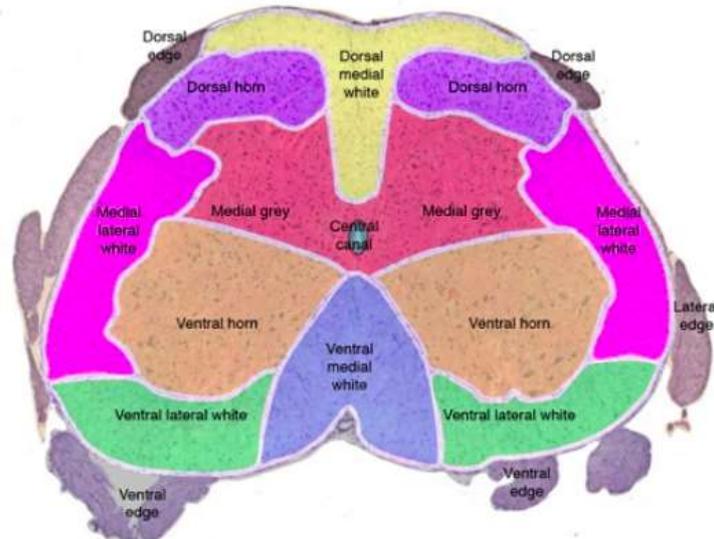


Figura 3. Regiones anatómicas de la médula espinal. Esquema que muestra 11 zonas marcadas con diferentes colores. De color naranja se aprecia el cuerno ventral donde se concentra la mayor proporción de motoneuronas de la médula espinal. Figura tomada de (Maniatis et al., 2019).

La actividad de genes a lo largo de las motoneuronas se ha estudiado a escala genómica recientemente por Maniatis et al. (2019). En particular, usaron la técnica de *Spatial Transcriptomics*. Antes de describir esta técnica, se describe a continuación la Transcriptómica.

1.2 Transcriptómica

El transcriptoma es el conjunto de todas las formas de RNA producidos por una célula o organismo en un momento determinado (Morozova et al., 2009). La comprensión y estudio del transcriptoma es esencial para la interpretación de elementos funcionales del genoma y revelar los constituyentes moleculares de células y tejidos, como el categorizar todas las especies de transcritos, incluidos RNAm y los RNA no codificantes, determinar la estructura transcripcional de los genes en términos de sitios de inicio, extremos 5' y 3', patrones de empalme y otras modificaciones postranscripcionales. Así también, el estudio del transcriptoma es ampliamente utilizado para comprender el desarrollo de diferentes enfermedades, cuantificando los niveles de expresión de cada transcrito durante el desarrollo de esta (Wang et al., 2009).

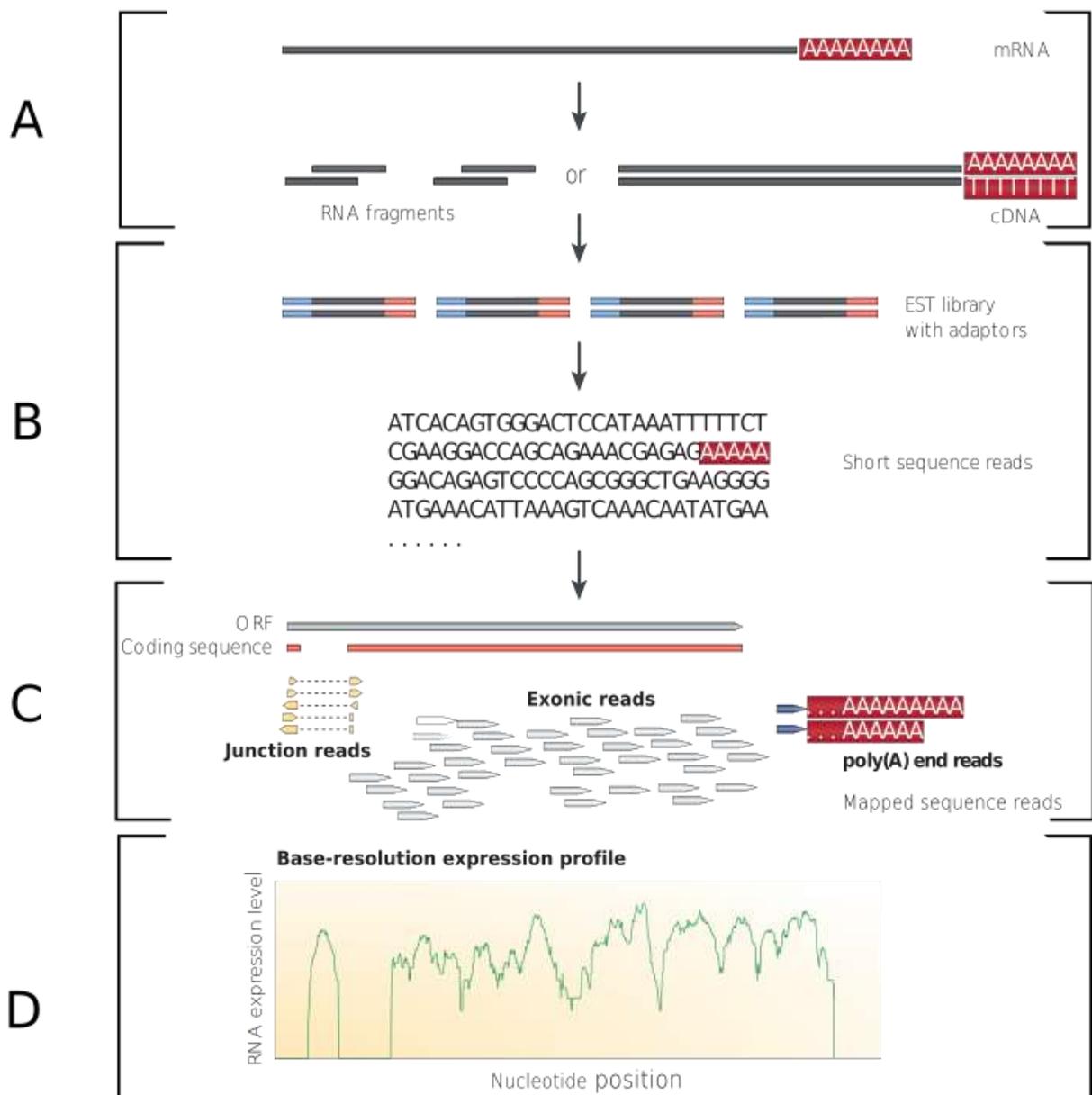


Figura 4. Esquema general de RNA-seq. **A.** Los RNA largos se convierten en una biblioteca de fragmentos de cDNA mediante la fragmentación de RNA. **B.** Se añaden los adaptadores de secuenciación, ilustrados de color azul, en cada fragmento de cDNA convirtiéndose en una secuencia corta de cada cDNA mediante tecnologías de secuenciación de alto rendimiento. **C.** Estas lecturas se alinean con genoma o transcriptoma de referencia y se clasifican en tres tipos: exonic reads, junction reads y poly(A) end-reads **D.** El alineamiento de los reads se usa para generar un perfil de expresión con resolución de base de cada gen. Figura adaptada de (Wang et al., 2009).

Actualmente, existe una gran cantidad de métodos para analizar el transcriptoma, siendo el más destacado la secuenciación de RNA (abreviado de aquí en adelante como RNA-Seq, Figura 4) (Lein et al., 2017). RNA-Seq, puede revelar la ubicación precisa de los elementos genéticos que se están transcribiendo, con resolución a nivel de base. RNA-Seq es muy preciso cuantificando niveles de expresión, además de permitir la detección de variaciones en transcritos (Wang et al., 2009).

Spatial Transcriptomics (ST) es una novedosa técnica basada en RNA-seq, la cual permite identificar cómo se expresan los transcritos a lo largo de una sección de un tejido (Navarro et al., 2017). El modo de acción de esta técnica se ejemplifica en la Figura 5, donde se ilustra el portaobjetos utilizado, el cual contiene cebadores oligo DT, que capturan mRNA mediante hibridización con su cola poli-A. Luego de esto, se hace una síntesis de cDNA mediante la transcripción inversa, mientras se mantiene la información espacial en las secciones de tejido histológico (Ståhl et al., 2016). La secuenciación de ST genera reads paired-end, donde la primera lectura (R1) contiene el código de barras espaciales y el identificador único molecular (nique Molecular Identifier, UMI). La segunda (R2) posee la información de la secuencia del transcrito (Yoosuf et al., 2020).

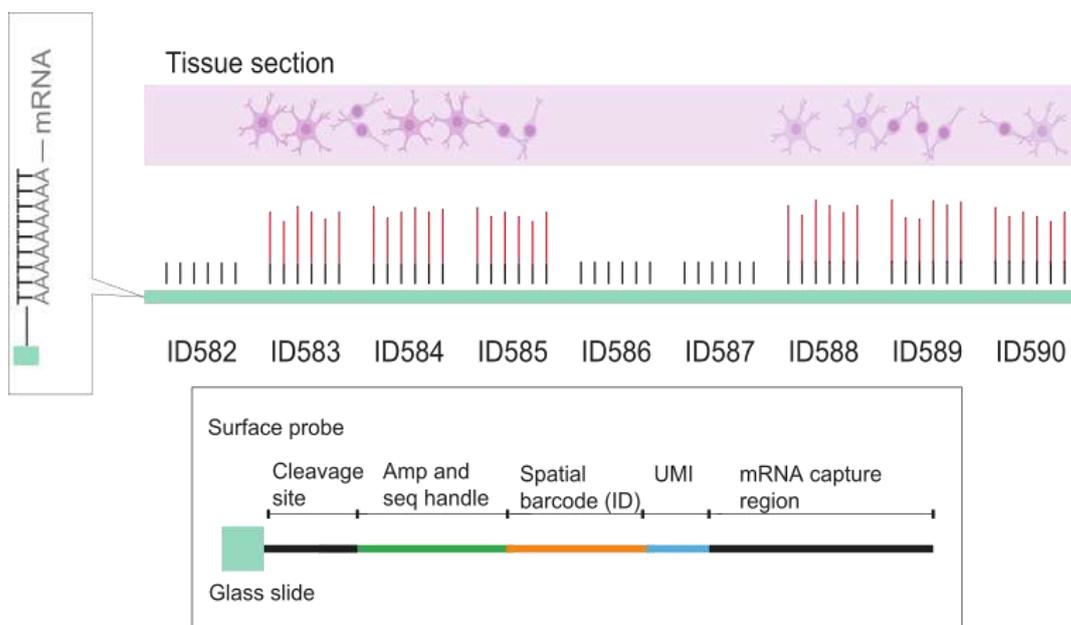


Figura 5. Obtención de datos de ST. Ilustración de cómo se capturan los RNAm en el arreglo que se encuentra en el portaobjetos sobre el cual se pone el tejido de interés. Cada casilla se encuentra ilustrada en la parte inferior, estas contienen sondas únicas con códigos de barras de DNA que incluye un sitio de unión de corte, un asa de amplificación y secuenciación, un código de barras espacial, un identificador molecular único (UMI) y una región de captura de oligo(dT). Figura adaptada de (Ståhl et al., 2016).

Para realizar los análisis computacionales de un proyecto ST se necesita la imagen del tejido en alta resolución y los datos de secuenciación generados con el protocolo descrito anteriormente. Esto permitirá proyectar la expresión génica en la imagen del tejido. Para ello, lo que se hace computacionalmente, es calcular los reads por gen y por posición espacial. Esto último se hace gracias al *barcode* explicado anteriormente. Así, se genera una matriz de conteos, donde los genes se representan como columnas y los códigos de barras espaciales están representados como filas. Por consiguiente, cada celda de la matriz representa el nivel de expresión de un gen dado en una coordenada específica de la imagen del tejido de alta resolución. En la Figura 6 se muestra un ejemplo de un análisis ST (Navarro et al., 2017).

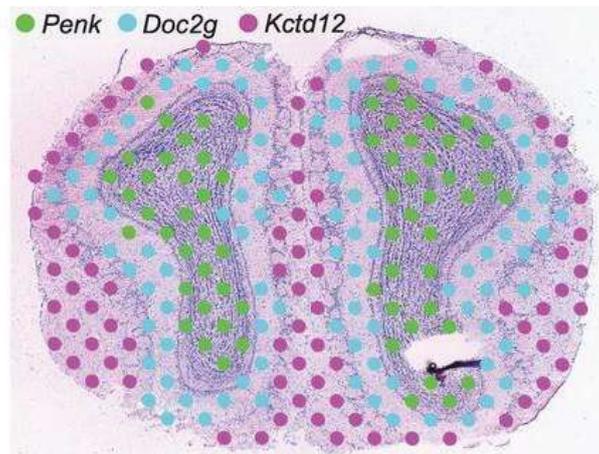


Figura 6. Ejemplo de visualización de expresión génica mediante ST. Imagen de bulbo olfatorio de ratón sometido a partir del cual se hizo Spatial Transcriptomics, donde es posible apreciar e identificar la resolución espacial de tres genes descritos en la parte superior de la imagen. Figura tomada de (Ståhl et al., 2016).

En el trabajo de (Maniatis et al., 2019) se realizó secuenciación tipo ST. Con ello, estudiaron la expresión de genes a lo largo de disposiciones espaciales, de diferentes eventos moleculares que impulsan interacciones entre las motoneuronas utilizando ST en ratones transgénicos SOD1^{G93A} y tejido postmortem de los pacientes con ELA. En este artículo se aprecia expresión de la proteína ácida fibrilar glial (GFAP), la cual se encuentra de forma elevada en células relacionadas con el sistema nervioso central. Además dicho gen es ampliamente propuesto como biomarcador en pacientes con enfermedades neurodegenerativas, debido a que estos en muestras de sangre presentan una notoria alza de esta proteína (Jeffrey et al., 2018). El objetivo de (Maniatis et al., 2019) con ilustrar de forma espacial GFAP, fue representar de esta forma subpoblaciones regionales de astrocitos que muestran distintas dinámicas espacio-temporal a lo largo de ELA. En la figura 7 se aprecia la diferencia de expresión espacial de GFAP en dos puntos temporales.

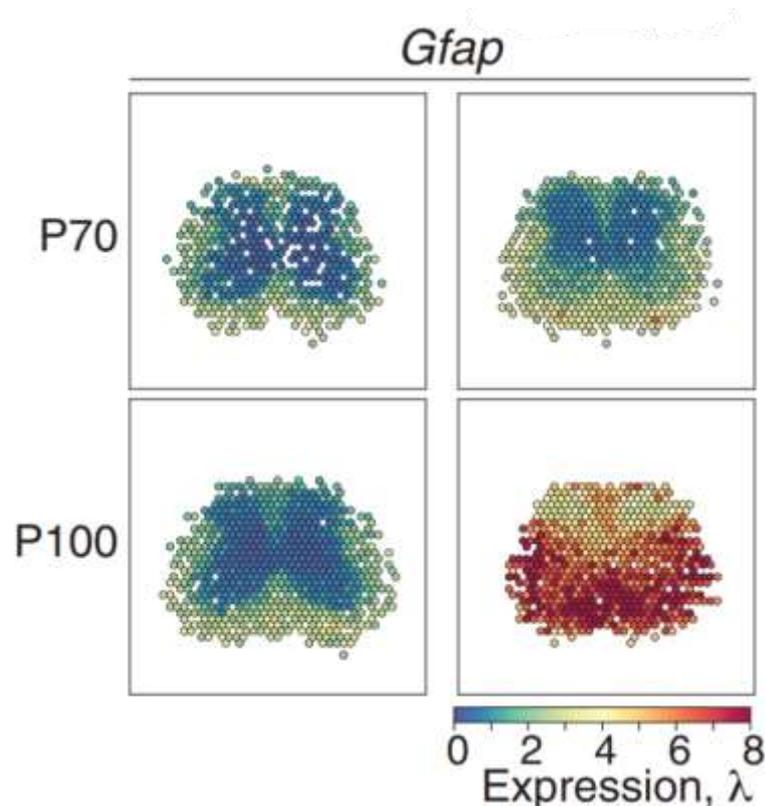


Figura 7. Expresión espacial de GFAP. ST realizado para visualizar las muestras de ratón $SOD1^{G93A}$ al día postnatal 70 (P70) y postnatal 100 (P100). Adaptado de (Maniatis et al., 2019).

Existe evidencia de que algunos retrotransposones están expresados en pacientes que sufren de ELA, y que esto parece ser independiente de si la ocurrencia de ELA es esporádica ó con un origen familiar (Tam et al., 2019). Adicionalmente, el grupo de mi tutor ya ha propuesto un modelo de expresión de Elementos Transponibles en el ratón transgénico $SOD1^{G93A}$ (Valdebenito-Maturana et al., 2021). Considerando que en el trabajo de Maniatis no se analizó la expresión de Elementos Transponibles, y los resultados mostrados en el trabajo recién citado, el próximo paso en esta investigación es analizar la expresión espacial de los Elementos Transponibles.

1.3 Elementos Transponibles

Los elementos transponibles (en inglés, *Transposable Elements*, abreviados de aquí en adelante como TEs) poseen la capacidad de moverse e integrarse a lo largo del genoma. Los TEs se encuentran en casi la totalidad de organismos eucariontes, y se clasifican en dos grupos (Figura 8): los transposones de DNA, que se transponen directamente (se escinden su posición original) , y los retrotransposones que utilizan un intermediario de RNA generando

una nueva copia mediante transcripción reversa (Elbarbary et al., 2016). Producto de esto último, se dice que el mecanismo de los retrotransposon es de “copiar y pegar”, mientras que el de los transposones de DNA se dice que es de “cortar y pegar” (Siomi et al., 2011). Los retrotransposones a su vez se dividen en dos grandes grupos basándose en su estructura: en primer lugar están los retrotransposones *Long Terminal Repeats* (LTRs), que poseen repeticiones largas en sus extremos, están formados por retrovirus endógenos, que comprenden aproximadamente un 8% del genoma humano, y por otro lado encontramos a los retrotransposones sin terminaciones largas o no LTR, que constituyen un 30% del genoma humano (Elbarbary et al., 2016). Los retrotransposones no LTR se subdividen según la capacidad de los elementos para movilizarse: los *Short Interspersed Nuclear Elements* (SINEs) los cuales no poseen autonomía de movimiento, y los *Long Interspersed Nuclear Elements* (LINEs), los cuales se desplazan de forma autosuficiente (Bodea et al., 2018).

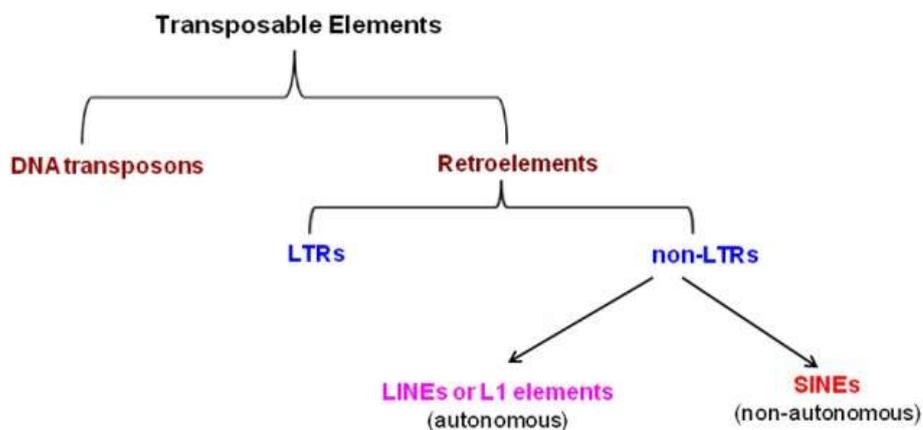


Figura 8. Clasificación de elementos transponibles. Los TEs se organizan en la clase DNA y retroelementos. Los retroelementos se subdividen según la presencia de *Long Terminal Repeats* (LTRs), y aquellos que no tienen LTRs (non-LTRs) se dividen en LINEs y SINEs. Figura tomada de (Golaconda Ramulu et al., 2012).

Hay variados estudios que sugieren que los TEs se expresan y están activos en el cerebro. Se ha especulado que estos podrían inducir diversidad genética y mutagénesis, influir en el desarrollo del cerebro humano, causar trastornos y enfermedades como Parkinson, Cáncer cerebral, Alzheimer, esquizofrenia, ELA, entre otras (Ahmadi et al., 2020). En análisis genómicos post mortem en personas con ELA se ha podido apreciar que los TEs se expresan, siendo lo opuesto a su comportamiento en un humano sano (Tam et al., 2019). Esto se puede explicar ya que en otros modelos analizados de ELA de origen familiar se ha visto mutaciones en el gen TARDBP que codifica la proteína TDP-43 (Saberri et al., 2015) y una de sus funciones primordiales es la regulación del silenciamiento de la actividad en

retrotransposones (Tam et al., 2019), siendo el principal mecanismo postranscripcional de control de retrotransposones en el tejido somático (Krug et al., 2017). Más recientemente, se ha mostrado que los TEs están expresados en el modelo de ratón SOD1^{G93A} y que podrían estar implicados en la progresión de la enfermedad (Valdebenito-Maturana et al., 2021).

A pesar de la evidencia que hay indicando un potencial rol de los TEs en ELA, no se ha estudiado su distribución espacial de estos en la médula espinal debido a la falta de herramientas para esto. Recientemente se diseñó el software SpatialTE (bvaldebenitom, 2021/2021) , el cual permite la inclusión de TEs en análisis ST. Con la integración de esta información se podría examinar si los TEs se encuentran mayoritariamente en el cuerno ventral de la médula espinal. De ser así, se podría proponer una relación de TEs con la degeneración y daño en motoneuronas.

Problemática

Durante los últimos años han habido varios avances en estudios de ELA mediante RNA-Seq, pero no ha sido hasta recientemente que se hizo un trabajo de ST enfocado en la enfermedad (Maniatis et al., 2019). En este trabajo vieron los cambios de expresión de genes y su distribución espacial en la médula espinal del ratón SOD1^{G93A}. Sin embargo, no estudiaron TEs, como se mencionó anteriormente. Este punto es importante, ya que diversos trabajos anteriores han especulado sobre la expresión de TEs en ELA, pero hasta la fecha, su distribución y expresión con resolución espacial no ha sido explorada. Entender la posición espacial de los TEs, nos permitiría proponer un rol causal de la enfermedad de ELA, ya que como se nombró anteriormente las motoneuronas se encuentran mayoritariamente en el cuerno ventral de la médula espinal y de esta manera si los TEs se encuentran en mayor medida en esa región, esto daría cuenta de que puede haber una influencia mayor de los TEs en ELA.

2. Hipótesis

En el modelo de ratón transgénico SOD1^{G93A} de la enfermedad de Esclerosis Lateral Amiotrófica, la expresión de Elementos Transponibles se encuentra preferencialmente en el cuerno ventral de la médula espinal.

3. Objetivos

Objetivo general.

Analizar la expresión de los Elementos Transponibles a nivel espacial en modelo de Esclerosis Lateral Amiotrófica.

Objetivos específicos.

1. Identificar TEs expresados en muestras de Spatial Transcriptomics provenientes del ratón transgénico SOD1^{G93A}.
2. Analizar la distribución de los TEs expresados en las regiones anatómicas de la médula espinal.

4. Materiales y métodos.

4.1 Materiales.

Los materiales usados se organizaron en tres categorías: 1. Software, 2. Muestras de RNA-seq y 3. Hardware

4.1.1 Software y plataformas

Nombre	Descripción
Ensembl	Base de datos, que proporciona un marco bioinformático para organizar la biología en torno a las secuencias de genomas grandes. Se encuentra un constante crecimiento para incluir una gama progresiva de datos, además para construir vistas comparativas de secuencias de genomas a medida que las secuencias de otros vertebrados son integradas (Hubbard et al., 2002).
STAR	Herramienta para indexar y realizar alineamientos de RNA-seq. Posee una alta velocidad de mapeo y además una alta sensibilidad (Dobin et al., 2013).
SRATools	Programa para la obtención de los datos de secuenciación disponibles en la base de datos SRA. Esta herramienta permite obtener los archivos en formato FASTQ (Leinonen et al., 2011).
ST Pipeline	Software que contiene las herramientas y scripts necesarios para realizar procesamiento de datos de Spatial Transcriptomics. Funciona realizando una matriz de conteos donde los genes son representados por columnas y los códigos de barras espaciales por filas. Cada celda de esta matriz generada representa el nivel de expresión de un gen dado en un lugar determinado del tejido (Navarro et al., 2017).
SpatialTE	Herramienta para incluir los TEs en análisis de Spatial Transcriptomics (Valdebenito-Maturana, 2021).
ST Analysis	Conjunto de herramientas para la visualización, procesamiento y análisis de datos de ST (Navarro, 2016/2021).
Inkscape	Programa de manipulación de imágenes (Developers).

Tabla 2. Software y plataformas utilizadas. cada una con su respectiva descripción.

4.1.2 Muestras de RNA-seq

Los datos utilizados fueron publicados junto al artículo “*Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis*” (Maniatis et al., 2019), y están públicamente disponibles en SRA código SRP162456. Estos datos son de RNA-Seq, generados con la metodología *Spatial Transcriptomics*. La secuenciación se generó desde la médula espinal del ratón transgénico SOD1^{G93A}, y los respectivos ratones control WT, a lo largo de 4 puntos temporales (p30, p70, p100 y p120). Para este proyecto, se utilizarán las muestras de ratón SOD1^{G93A} en la etapa p120, ya que esta etapa presenta una alta degeneración de motoneuronas.

4.1.3 Hardware

Este proyecto se llevó a cabo en el supercomputador de la doctora Wendy Gonzalez, el cual posee el nombre Clustertalca3, cuenta con 80 GPUs de tipo 8 x NVIDIA Corporation GP102 GeForce GTX 1080 Ti (rev a1), CPU Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10GHz y un disco duro total de 30 TB.

4.2 Metodología

El flujo de trabajo se encuentra de forma resumida en la Figura 9. Luego de la figura se describen las actividades punto por punto, según a qué objetivos específicos corresponden.

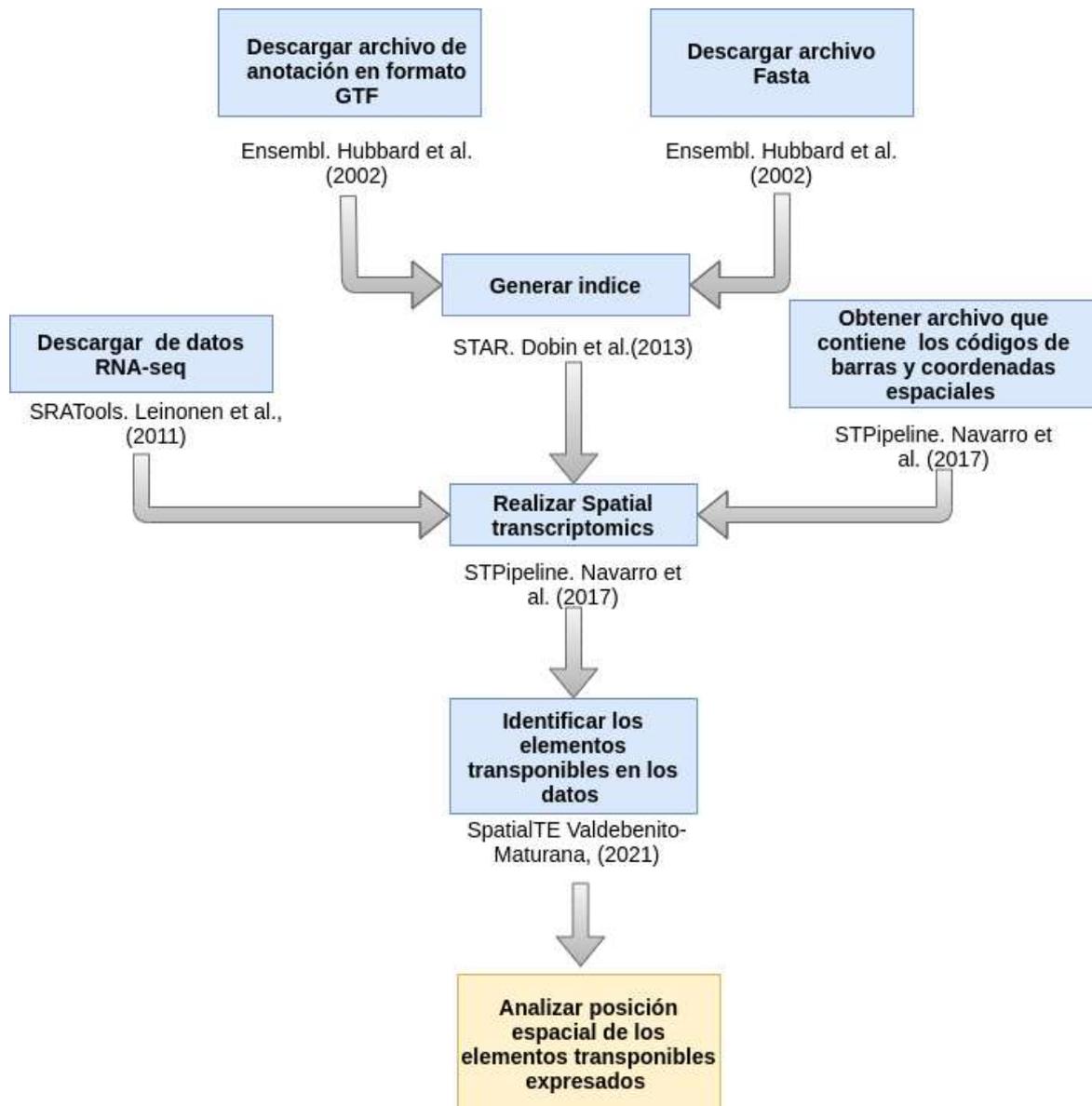


Figura 9. Diagrama de flujo de metodología seguida. De color azul se puede apreciar la estrategia de trabajo ejecutada en el objetivo específico 1, donde se descargó el archivo de anotación en formato GTF, y el archivo FASTA, ambos desde Ensembl. Con estos se generó un índice a través de STAR. Posteriormente, se obtuvieron los datos de RNA-seq a través de SRATools, y los códigos de barras y coordenadas espaciales desde el material suplementario del paper de STPipeline. Por otro lado de color amarillo están representados los procedimientos para aplicar el objetivo específico 2 donde se identificaron los TEs en los datos y se analizó la posición espacial de estos elementos en la médula espinal.

4.2.1 Identificar TEs expresados en muestras de Spatial Transcriptomics provenientes del ratón transgénico SOD1^{G93A}.

4.2.1.1 Descarga de archivos.

El primer paso de la ejecución fue la descarga de los archivos de anotación en formato GTF y archivo de genoma en formato FASTA. GTF, es un formato de archivo que contiene la información sobre la estructura de los genes, mientras que FASTA contiene las secuencias de ácidos nucleicos. Ambos archivos son descargados desde la base de datos Ensembl.

4.2.1.2 Construcción de índice.

Ya obtenidos los archivos, se procedió con la realización del índice del genoma con el software *Spliced Transcripts Alignment to a Reference* (STAR). Esta indexación permite lograr una rápida velocidad en los alineamientos de reads. Con esto se logra ahorrar tiempo y memoria del servidor (Li et al., 2009).

Se utilizará el comando base entregado por el manual de de STAR, para el cual es necesario el número de hebras del dispositivo a utilizar, el nombre del directorio en donde se guardarán los nuevos ficheros creados, los archivos FASTA y de anotación en formato GTF y por último el largo menos 1 de los reads.

4.2.1.3 Descarga de los datos de RNA-seq.

Se descargaron los archivos correspondientes a las muestras p120 de ratones SOD1^{G93A}, mediante SRATools. Esta herramienta es de utilización sencilla, pidiendo solo el código de acceso de los datos que se desea obtener. Estas son 37 muestras, que se pueden apreciar en la tabla 2 del material suplementario, de las cuales se realizó el estudio para los 10 de estas expuestas en la tabla 3, la cual posee para cada muestra el código Run, GEO, los largos de secuencias y el código de barra al cual corresponden. Esto debido al tiempo necesario para el procesamiento de los datos y el tiempo entregado para realizar el estudio

Run	GEO_Accession	Largo seq 1	largo seq2	Barcode
SRR7895536	GSM3399139	30	50	L6
SRR7895537	GSM3399140	30	50	L6
SRR7895538	GSM3399141	30	50	L6
SRR7895712	GSM3399315	30	55	L7
SRR7895713	GSM3399316	30	55	L7
SRR7895714	GSM3399317	30	55	L7
SRR7895715	GSM3399318	30	55	L7
SRR7895716	GSM3399319	30	55	L7
SRR7895717	GSM3399320	30	55	L7
SRR7895718	GSM3399321	30	55	L7

Tabla 3. Tabla de datos analizados. Datos seleccionados para ser analizados, de entre los 37 descritos en la Tabla 1 del material suplementario.

4.2.1.4 Obtención de archivo que contiene los códigos de barra y las coordenadas espaciales.

Este archivo contiene tres columnas: la primera corresponde al código de barras, la segunda las coordenadas en el eje X, por último la tercera columna posee las coordenadas del eje Y, permitiendo la identificación de las posiciones espaciales. Se obtiene desde el material suplementario del paper “*ST Pipeline: an automated pipeline for spatial mapping of unique transcripts*” (Navarro et al., 2017) en donde es posible encontrar este archivo para diferentes versiones de la técnica *Spatial Transcriptomics*.

4.2.1.5 Análisis Spatial Transcriptomics.

El análisis de Spatial Transcriptomics de las muestras de RNA-Seq se llevó a cabo con ST Pipeline. ST Pipeline toma como entrada los archivos FASTQ, el archivo de códigos de barra, el archivo de anotación de genes en formato GTF, y un índice genómico generado a través de STAR. ST Pipeline realiza el alineamiento de los reads, y la posterior asignación de reads a posiciones espaciales según su código de barras. ST Pipeline entrega como resultado una matriz que tiene nombres de genes como columnas y coordenadas espaciales como filas,

donde cada celda de la matriz representa el nivel de expresión de un gen dado en un lugar determinado.

4.2.1.6 Identificar TEs en los datos.

La identificación de TEs se realizó mediante SpatialTE, el cual usa los resultados ST Pipeline y un archivo BED de anotación de TEs. Con esto, el programa calcula los reads por TEs.. Según el MAPQ promedio de los TEs, estos se dividen en específicos de locus (MAPQ = 255, solo reads únicos) y a nivel de subfamilia (MAPQ < 255, mezcla de reads únicos y multi-mapeos). Los resultados son expuestos como una matriz que contiene los identificadores de TE como columnas y las coordenadas espaciales en las filas, similar al resultado final de ST Pipeline para genes (Valdebenito-Maturana, 2021). Para este trabajo, solo se utilizaron los resultados a nivel de subfamilia, ya que se optó por no enfocarse en los locus de expresión de TEs, si no en cuanto estaban expresados. Se ha reportado que las estimaciones a nivel de subfamilia son más confiables, al no tener la complicación de tratar de identificar el origen de expresión (Jin et al., 2015).

4.2.2 Analizar la distribución de los TEs expresados en las regiones anatómicas de la médula espinal.

4.2.2.1 Análizar la posición espacial de los TEs

Este paso es el más costoso en términos de tiempo, ya que se realizó de forma manual a través de un programa de manipulación de imágenes. En este trabajo se utilizó Inkscape.

En particular, la matriz de expresión de TEs entregada por SpatialTE se transforma en una imagen vectorial, la cual se debe superponer con la imagen de alta resolución del tejido. Una vez realizado esto, para obtener una mejor calidad de imagen se procedió a la limpieza de las figuras, eliminando todos los puntos de expresión que se encontraban fuera del tejido en estudio. Esto ocurre en ST debido a un fenómeno denominado difusión de RNA, el que provoca que los RNA del tejido se difunda a lo largo de la placa.

Posteriormente según la anotación de regiones anatómicas del trabajo de Maniatis y colaboradores (2019), se evaluó si hay preferencias espaciales de expresión de TEs en el cuerno ventral del tejido.

6. Resultados

6.1 Análisis de calidad de los reads.

Para medir la calidad se evaluaron dos gráficos obtenidos a través de FastQC. En primer lugar la puntuación media de calidad en las secuencias, el cual indica el promedio por cada posición del reads y se encuentra ilustrado en el gráfico 1, en donde se puede contemplar la distribución promedio de calidad se encuentra en la parte superior, indicando que su calidad es relativamente buena. Por otro lado se tiene el puntaje de calidad por secuencia, que denota el número de reads junto a los puntajes promedio de calidad, además este tipo de gráfico permite percibir la existencia de subset de reads que poseen una calidad baja, en el gráfico 5 se percibe que alza comienza pasando el 20 en la escala de puntajes de calidad promedio por secuencia y se acentúa pasando los 30 de esta misma escala, mostrando todos los peaks de los datos. Estos dos gráficos nos permiten tener una visión amplia de la calidad de los reads en las muestras, no se realizó un análisis profundo ya que los Software utilizados requieren los reads sin pre-procesar. De todas formas, estos resultados indican que la secuenciación se realizó con una calidad aceptable.



Figura 10. Puntuaciones medias de calidad. En el eje Y establece los puntajes de calidad promedio por base, mientras el eje X indica la posición de cada base a lo largo del read, donde cada línea corresponde a un archivo reads.

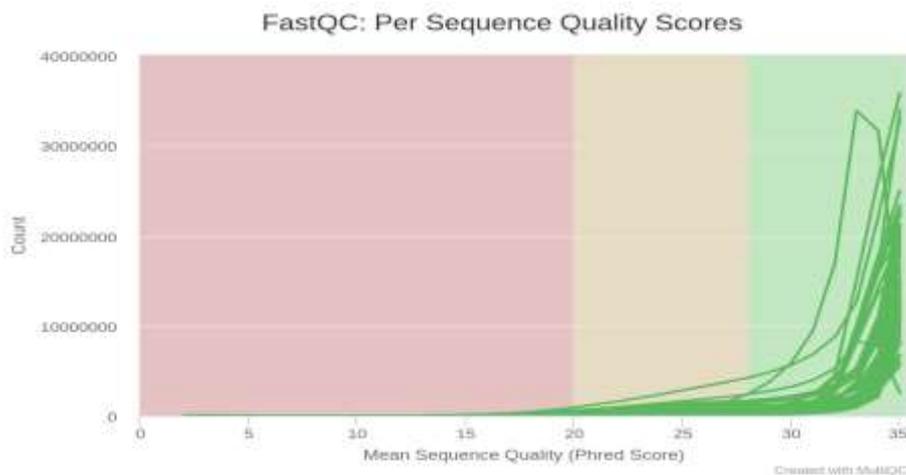


Figura 11. Puntaje de calidad por secuencia. En el eje Y se muestra la cantidad total de secuencias analizadas, y el eje X la calidad promedio por secuencia en una escala de puntajes Phred33.

6.2 Identificar Elementos Transponibles expresados en muestras de Spatial Transcriptomics provenientes del ratón transgénico SOD1^{G93A}.

6.2.1 Realizar Spatial transcriptomics

El análisis computacional de los datos ST del ratón transgénico SOD1^{G93A} del tejido médula espinal en la etapa p120 fue realizado con ST Pipeline con los parámetros básicos. Con esto, como resultado se obtiene un archivo en formato TSV, el cual es una matriz que posee los nombres de los genes como columnas y las coordenadas espaciales en las filas, en donde cada celda tiene la información del nivel de expresión de un gen dado en un lugar dado. El archivo TSV se procesa con el script *st_data_plotter.py* del paquete ST Analysis, publicado por el mismo grupo desarrollador de ST Pipeline. *st_data_plotter.py* se ejecutó la opción `--show-gene`, la que permite indicar el gen que se desea analizar, además se añade la imagen asociada a la muestra, como la que se muestra en la figura con la opción `--image-file`.

Como control positivo de la ejecución correcta de esta etapa, se hizo de forma preliminar el análisis de la expresión espacial de SOD1 en la muestra SRR7895713 que contiene 4 secciones de médula espinal.

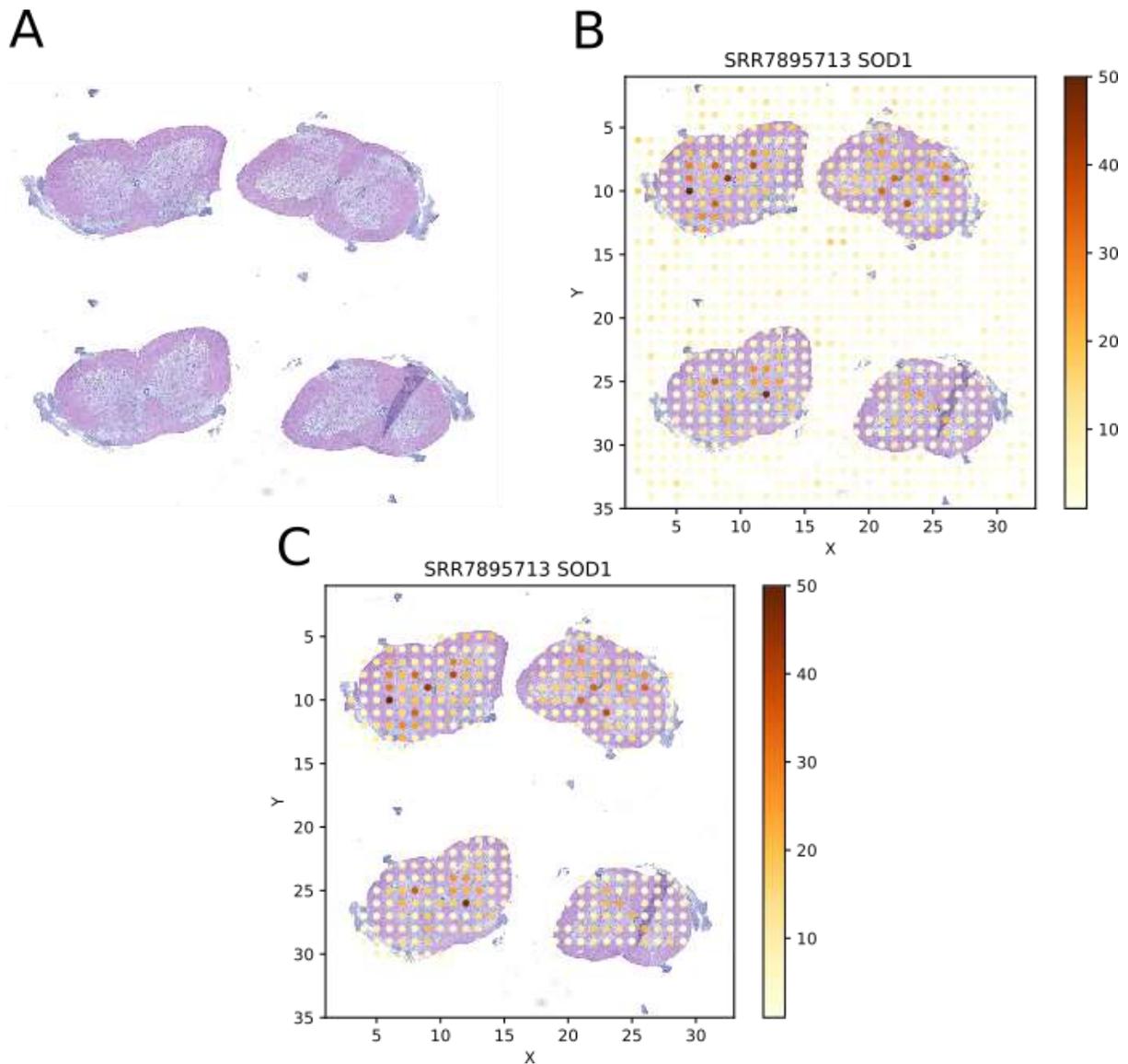


Figura 12. Análisis ST de la muestra SRR7895713. A. Imagen del tejido, **B.** Expresión de Sod1 a lo largo del tejido, la escala al costado indica el número de reads (spots con menos expresión están más claros, mientras que spots con más expresión están más oscuros). **C.** Imagen de la expresión de SOD1 con los puntos de expresión fuera del tejido eliminado.

6.2.2 Identificar TEs en los datos.

Luego de ejecutar ST Pipeline para todas las muestras seleccionadas, se ejecutó SpatialTE. Para el análisis en las imágenes, se utilizó el output que entrega a nivel de subfamilia. Las matrices obtenidas fueron procesadas con ST Analysis, pero, en el caso de la expuesta en la Figura 10 no se utilizó la opción --show-gene, por lo tanto por defecto este resultado nos muestra la expresión total de los TEs en la muestra.

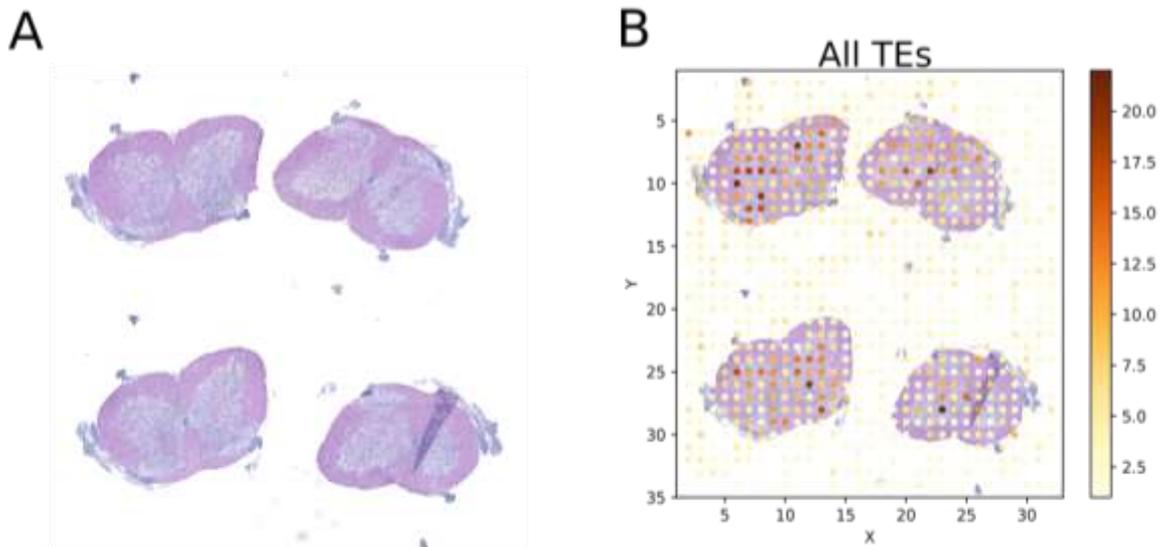


Figura 13. Análisis con SpatialTE de la muestra SRR7895713 del grupo subfamily . A. Imágenes del tejido. **B.** Expresión de todos los TEs a lo largo de diferentes imágenes del tejido, la escala al costado indica el número de reads (spots con menos expresión están más claros, mientras que spots con más expresión están más oscuros).

6.4 Analizar la distribución de los TEs expresados en las regiones anatómicas de la médula espinal.

6.4.1 Analizar distribución de TEs en los tejidos.

Previo al análisis se debe realizar una limpieza de las figuras, la cual permite visualizar de forma más óptima el tejido con su debida expresión de TEs. Para lograr esto se eliminaron todos los puntos de expresión que no se encuentran sobre el tejido analizado, este trabajo fue realizado con Inkscape con su herramienta de desacoplo de imágenes que permite seleccionar todos los objetos de la imagen por separado, posibilitando la eliminación seleccionada y prolija de estos puntos no deseados. En la Figura 14 se muestra el resultado de esto para la misma muestra exhibida en la Figura 13.

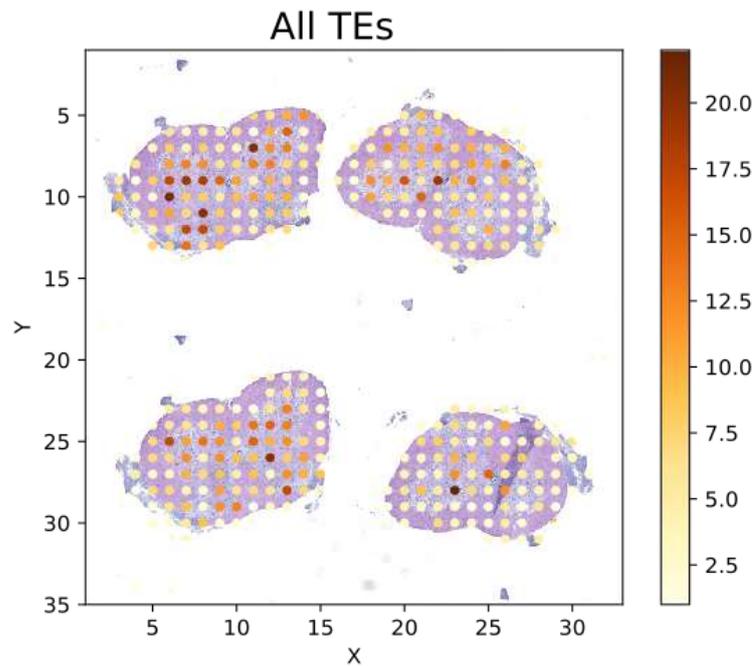


Figura 14. Procesamiento de imágenes. Imagen con los puntos de expresión fuera del tejido eliminados.

Con la imagen limpia se procede a seleccionar la mejor resolución del tejido dentro de esta misma muestra, puesto que como se muestra en la figura 14 hay algunas partes del tejido que presentan una baja resolución, se escogió la segunda médula espinal de la primera fila y se procedió a marcar el cuerno ventral dentro del tejido.

El procedimiento desde la Identificación TEs en los datos se repitió para un total de 10 muestras. Debido a que los resultados son consistentes entre ellos (se observan patrones similares de expresión de TEs), seleccionó la muestra SRR7895713 como representativa para la siguiente parte de este trabajo, mientras que el resto de resultados se encuentran en Figuras Suplementarias 1 a 20. En la Figura 15 se aprecia la diferencia en la expresión en los distintos tipos de TEs, se puede distinguir que el TE de tipo DNA es que posee menor expresión, mientras el de tipo LTR el que contiene una mayor expresión. En esta figura se realizó un análisis considerando la expresión de TEs en el cuerno ventral comparado a las otras partes de la médula espinal, llegando al resultado que no hay un aparente enriquecimiento en la zona deseada, en ninguno de los tipos de TEs.

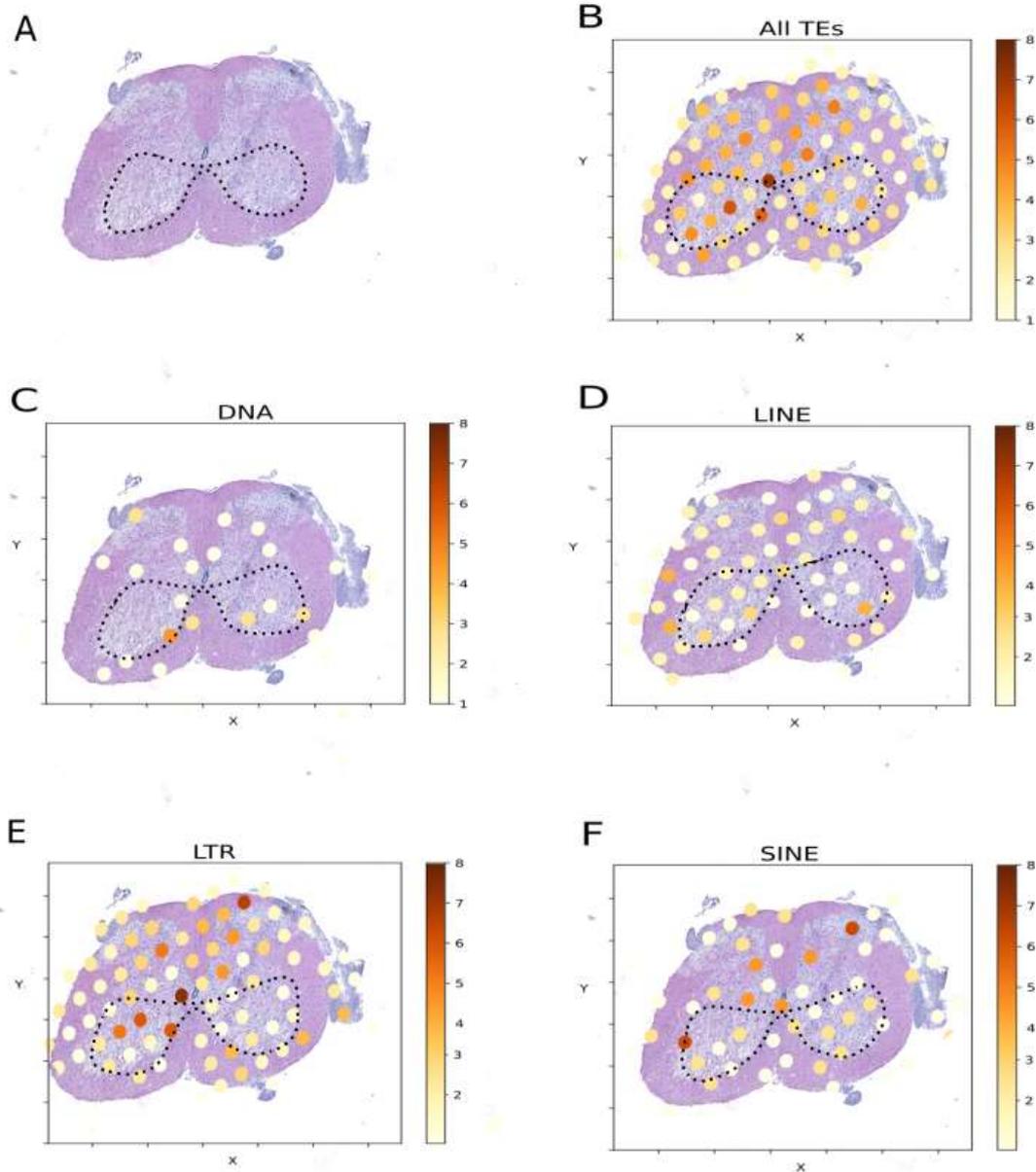


Figura 15. Análisis con SpatialTE de la muestra SRR7895713 para los diferentes tipos de TEs.
A. Imagen del tejido seleccionado. **B.** Expresión de todos los TEs en el tejido **C.** Expresión de TEs de tipo DNA en el tejido. **D.** Expresión de TEs de tipo LINE en el tejido. **E.** Expresión de TEs tipo LTR en el tejido. **F.** Expresión de TEs de tipo SINE en el tejido.

7. Conclusión.

La relevancia de este estudio radica en que este trabajo es el primer acercamiento en donde se contemplan TEs con resolución espacial a lo largo de la médula espinal del ratón SOD1^{G93A}. El primer resultado que se destaca y está en línea con la evidencia que hay disponible a la fecha es que se pueden detectar TEs expresados en este modelo de ELA. Sin embargo, y contrario a la hipótesis planteada, de que los TEs se encuentran expresados en el cuerno ventral, en este trabajo se observó que hay un gran nivel de dispersión de los diferentes tipos de TEs a lo largo de la médula espinal.

Dentro de las principales limitaciones de este trabajo están la cantidad de muestras utilizadas (n=10). En el trabajo de Maniatis y colaboradores (2019) al ser un grupo más amplio de investigadores pudieron analizar de forma eficiente todas las muestras (n=37). Así, se propone seguir con futuros trabajos, ya que si bien no existe una intensificación de TEs en la zona de estudio, en comparación a las otras regiones de la médula espinal, sí existe expresión de estos, por lo que aunque en menor cantidad podría ser relevante para el desarrollo de ELA. Esto a futuro se podría validar o testear de forma experimental.

Por otro lado, a diferencia del RNA-Seq estándar, en Spatial Transcriptomics no hay algoritmos estadísticos bien establecidos. Con avances metodológicos a futuro, se espera poder superar esto a fin de contar con una validación estadística de los resultados observados (por ejemplo, si existe una diferencia real entre zonas de expresión de TEs).

Referencias

- Ahmadi, A., De Toma, I., Vilor-Tejedor, N., Eftekhariyan Ghamsari, M. R., & Sadeghi, I. (2020). Transposable elements in brain health and disease. *Ageing Research Reviews*, *64*, 101153. <https://doi.org/10.1016/j.arr.2020.101153>
- Bodea, G. O., McKelvey, E. G. Z., & Faulkner, G. J. (2018). Retrotransposon-induced mosaicism in the neural genome. *Open Biology*, *8*(7), 180074. <https://doi.org/10.1098/rsob.180074>
- bvaldebenitom. (2021). *SpatialTE* [Shell]. <https://github.com/bvaldebenitom/SpatialTE> (Original work published 2021)
- Cartarozzi, L. P., Rieder, P., Bai, X., Scheller, A., Oliveira, A. L. R. de, & Kirchhoff, F. (2018). In vivo two-photon imaging of motoneurons and adjacent glia in the ventral spinal cord. *Journal of Neuroscience Methods*, *299*, 8-15. <https://doi.org/10.1016/j.jneumeth.2018.01.005>
- Developers, I. W. (s. f.). *Draw Freely | Inkscape*. Recuperado 27 de noviembre de 2021, de <https://inkscape.org/es/>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15-21. <https://doi.org/10.1093/bioinformatics/bts635>
- Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science*, *351*(6274), aac7247. <https://doi.org/10.1126/science.aac7247>
- Ghadge, G. D., Sonobe, Y., Camarena, A., Drigotas, C., Rigo, F., Ling, K. K., & Roos, R. P. (2020). Knockdown of GADD34 in neonatal mutant SOD1 mice ameliorates ALS. *Neurobiology of Disease*, *136*, 104702. <https://doi.org/10.1016/j.nbd.2019.104702>
- Golaconda Ramulu, H., Pontarotti, P., & Raouf, D. (2012). The rhizome of life: What about metazoa? *Frontiers in Cellular and Infection Microbiology*, *2*, 50. <https://doi.org/10.3389/fcimb.2012.00050>
- Grad, L. I., Rouleau, G. A., Ravits, J., & Cashman, N. R. (2017). Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS). *Cold Spring Harbor Perspectives in Medicine*, *7*(8), a024117. <https://doi.org/10.1101/cshperspect.a024117>
- Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E. M., Logroscino, G., Robberecht, W., Shaw, P. J., Simmons, Z., & van den Berg, L. H. (2017). Amyotrophic lateral sclerosis. *Nature Reviews. Disease Primers*, *3*, 17071. <https://doi.org/10.1038/nrdp.2017.71>
- Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki,

- L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., ... Clamp, M. (2002). The Ensembl genome database project. *Nucleic Acids Research*, 30(1), 38. <https://doi.org/10.1093/nar/30.1.38>
- Jeffrey, J., D’Cunha, H., & Suzuki, M. (2018). Blood Level of Glial Fibrillary Acidic Protein (GFAP) Does not Correlate With Disease Progression in a Rat Model of Familial ALS (SOD1G93A Transgenic). *Frontiers in Neurology*, 9, 954. <https://doi.org/10.3389/fneur.2018.00954>
- Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TETranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, 31(22), 3593-3599. <https://doi.org/10.1093/bioinformatics/btv422>
- Julien, J.-P., & Kriz, J. (2006). Transgenic mouse models of amyotrophic lateral sclerosis. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1762(11), 1013-1024. <https://doi.org/10.1016/j.bbadis.2006.03.006>
- Krug, L., Chatterjee, N., Borges-Monroy, R., Hearn, S., Liao, W.-W., Morrill, K., Prazak, L., Rozhkov, N., Theodorou, D., Hammell, M., & Dubnau, J. (2017). Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS. *PLOS Genetics*, 13(3), e1006635. <https://doi.org/10.1371/journal.pgen.1006635>
- Lein, E., Borm, L. E., & Linnarsson, S. (2017). The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science*, 358(6359), 64-69. <https://doi.org/10.1126/science.aan6827>
- Leinonen, R., Sugawara, H., & Shumway, M. (2011). The Sequence Read Archive. *Nucleic Acids Research*, 39(Database issue), D19. <https://doi.org/10.1093/nar/gkq1019>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078. <https://doi.org/10.1093/bioinformatics/btp352>
- Maniatis, S., Äijö, T., Vickovic, S., Braine, C., Kang, K., Mollbrink, A., Fagegaltier, D., Andrusivová, Ž., Saarenpää, S., Saiz-Castro, G., Cuevas, M., Watters, A., Lundeberg, J., Bonneau, R., & Phatnani, H. (2019). Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis. *Science*, 364(6435), 89-93. <https://doi.org/10.1126/science.aav9776>
- Martínez-Llorens, J., & Balañá Corberó, A. (2019). Esclerosis lateral amiotrófica: Valoración del fracaso muscular inspiratorio. *Archivos de Bronconeumología*, 55(7), 345-346. <https://doi.org/10.1016/j.arbres.2019.01.013>
- Mehta, P. (2018). Prevalence of Amyotrophic Lateral Sclerosis—United States, 2015. *MMWR. Morbidity and Mortality Weekly Report*, 67.

<https://doi.org/10.15585/mmwr.mm6746a1>

- Morozova, O., Hirst, M., & Marra, M. A. (2009). Applications of New Sequencing Technologies for Transcriptome Analysis. *Annual Review of Genomics and Human Genetics*, 10(1), 135-151. <https://doi.org/10.1146/annurev-genom-082908-145957>
- Navarro, J. F. (2021). *Spatial Transcriptomics Analysis Tools* [Python]. https://github.com/jfnavarro/st_analysis (Original work published 2016)
- Navarro, J. F., Sjöstrand, J., Salmén, F., Lundeberg, J., & Ståhl, P. L. (2017). ST Pipeline: An automated pipeline for spatial mapping of unique transcripts. *Bioinformatics*, 33(16), 2591-2593. <https://doi.org/10.1093/bioinformatics/btx211>
- Pansarasa, O., Bordoni, M., Diamanti, L., Sproviero, D., Gagliardi, S., & Cereda, C. (2018). SOD1 in Amyotrophic Lateral Sclerosis: “Ambivalent” Behavior Connected to the Disease. *International Journal of Molecular Sciences*, 19(5), 1345. <https://doi.org/10.3390/ijms19051345>
- Phatnani, H. P., Guarnieri, P., Friedman, B. A., Carrasco, M. A., Muratet, M., O’Keeffe, S., Nwakeze, C., Pauli-Behn, F., Newberry, K. M., Meadows, S. K., Tapia, J. C., Myers, R. M., & Maniatis, T. (2013a). Intricate interplay between astrocytes and motor neurons in ALS. *Proceedings of the National Academy of Sciences*, 110(8), E756-E765. <https://doi.org/10.1073/pnas.1222361110>
- Phatnani, H. P., Guarnieri, P., Friedman, B. A., Carrasco, M. A., Muratet, M., O’Keeffe, S., Nwakeze, C., Pauli-Behn, F., Newberry, K. M., Meadows, S. K., Tapia, J. C., Myers, R. M., & Maniatis, T. (2013b). Intricate interplay between astrocytes and motor neurons in ALS. *Proceedings of the National Academy of Sciences*, 110(8), E756-E765. <https://doi.org/10.1073/pnas.1222361110>
- Riancho, J., Gonzalo, I., Ruiz-Soto, M., & Berciano, J. (2019). ¿Por qué degeneran las motoneuronas? Actualización en la patogenia de la esclerosis lateral amiotrófica. *Neurología*, 34(1), 27-37. <https://doi.org/10.1016/j.nrl.2015.12.001>
- Roudeau, S., Trist, B. G., Carmona, A., Davies, K. M., Halliday, G. M., Rufin, Y., Claverol, S., Van Malderen, S. J. M., Falkenberg, G., Double, K. L., & Ortega, R. (2021). Native Separation and Metallation Analysis of SOD1 Protein from the Human Central Nervous System: A Methodological Workflow. *Analytical Chemistry*, 93(32), 11108-11115. <https://doi.org/10.1021/acs.analchem.1c01128>
- Saberi, S., Stauffer, J. E., Schulte, D. J., & Ravits, J. (2015). Neuropathology of Amyotrophic Lateral Sclerosis and Its Variants. *Neurologic Clinics*, 33(4), 855-876. <https://doi.org/10.1016/j.ncl.2015.07.012>
- Sau, D., De Biasi, S., Vitellaro-Zuccarello, L., Riso, P., Guarnieri, S., Porrini, M., Simeoni, S., Crippa, V., Onesto, E., Palazzolo, I., Rusmini, P., Bolzoni, E., Bendotti, C., & Poletti, A. (2007). Mutation of SOD1 in ALS: A gain of a loss of function. *Human Molecular*

- Genetics*, 16(13), 1604-1618. <https://doi.org/10.1093/hmg/ddm110>
- Shaw, P. J. (2005). Molecular and cellular pathways of neurodegeneration in motor neurone disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 76(8), 1046-1057. <https://doi.org/10.1136/jnnp.2004.048652>
- Siomi, M. C., Sato, K., Pezic, D., & Aravin, A. A. (2011). PIWI-interacting small RNAs: The vanguard of genome defence. *Nature Reviews Molecular Cell Biology*, 12(4), 246-258. <https://doi.org/10.1038/nrm3089>
- Snider, W. D., Zhang, L., Yusoof, S., Gorukanti, N., & Tsering, C. (1992). Interactions between dorsal root axons and their target motor neurons in developing mammalian spinal cord. *Journal of Neuroscience*, 12(9), 3494-3508. <https://doi.org/10.1523/JNEUROSCI.12-09-03494.1992>
- Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P. I., Sahlén, P., Mulder, J., Bergmann, O., ... Frisén, J. (2016a). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294), 78-82. <https://doi.org/10.1126/science.aaf2403>
- Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J. O., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P. I., Sahlén, P., Mulder, J., Bergmann, O., ... Frisén, J. (2016b). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294), 78-82. <https://doi.org/10.1126/science.aaf2403>
- Tam, O. H., Rozhkov, N. V., Shaw, R., Kim, D., Hubbard, I., Fennessey, S., Propp, N., Phatnani, H., Kwan, J., Sareen, D., Broach, J. R., Simmons, Z., Arcila-Londono, X., Lee, E. B., Deerlin, V. M. V., Shneider, N. A., Fraenkel, E., Ostrow, L. W., Baas, F., ... Hammell, M. G. (2019). Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. *Cell Reports*, 29(5), 1164-1177.e5. <https://doi.org/10.1016/j.celrep.2019.09.066>
- Turner, M. R., Barnwell, J., Al-Chalabi, A., & Eisen, A. (2012). Young-onset amyotrophic lateral sclerosis: Historical and other observations. *Brain: A Journal of Neurology*, 135(Pt 9), 2883-2891. <https://doi.org/10.1093/brain/aws144>
- Turner, M. R., Hardiman, O., Benatar, M., Brooks, B. R., Chio, A., Carvalho, M. de, Ince, P. G., Lin, C., Miller, R. G., Mitsumoto, H., Nicholson, G., Ravits, J., Shaw, P. J., Swash, M., Talbot, K., Traynor, B. J., Berg, L. H. V. den, Veldink, J. H., Vucic, S., & Kiernan, M. C. (2013). Controversies and priorities in amyotrophic lateral sclerosis. *The Lancet Neurology*, 12(3), 310-322. [https://doi.org/10.1016/S1474-4422\(13\)70036-X](https://doi.org/10.1016/S1474-4422(13)70036-X)

- 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), e0258291. <https://doi.org/10.1371/journal.pone.0258291>
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1), 57-63. <https://doi.org/10.1038/nrg2484>
- Yoosuf, N., Navarro, J. F., Salmén, F., Ståhl, P. L., & Daub, C. O. (2020). Identification and transfer of spatial transcriptomics signatures for cancer diagnosis. *Breast Cancer Research : BCR*, 22, 6. <https://doi.org/10.1186/s13058-019-1242-9>

Material Suplementario

Run	GEO_Accession	Largo seq 1	largo seq2	Barcode
SRR7895536	GSM3399139	30	50	L6
SRR7895537	GSM3399140	30	50	L6
SRR7895538	GSM3399141	30	50	L6
SRR7895541	GSM3399144	30	50	L6
SRR7895542	GSM3399145	30	50	L6
SRR7895571	GSM3399174	30	55	L6
SRR7895572	GSM3399175	30	55	L6
SRR7895577	GSM3399180	30	55	L7
SRR7895578	GSM3399181	30	55	L7
SRR7895579	GSM3399182	30	55	L7
SRR7895580	GSM3399183	30	55	L7
SRR7895581	GSM3399184	30	55	L7
SRR7895582	GSM3399185	30	55	L7
SRR7895583	GSM3399186	30	55	L7
SRR7895584	GSM3399187	30	55	L7
SRR7895585	GSM3399188	30	55	L7
SRR7895634	GSM3399237	30	55	L7
SRR7895635	GSM3399238	30	55	L7
SRR7895636	GSM3399239	30	55	L7
SRR7895637	GSM3399240	30	55	L7
SRR7895638	GSM3399241	30	55	L7
SRR7895639	GSM3399242	30	55	L7

SRR7895640	GSM3399243	30	55	L7
SRR7895641	GSM3399244	30	55	L7
SRR7895642	GSM3399245	30	55	L7
SRR7895643	GSM3399246	30	55	L7
SRR7895710	GSM3399313	30	55	L7
SRR7895711	GSM3399314	30	55	L7
SRR7895712	GSM3399315	30	55	L7
SRR7895713	GSM3399316	30	55	L7
SRR7895714	GSM3399317	30	55	L7
SRR7895715	GSM3399318	30	55	L7
SRR7895716	GSM3399319	30	55	L7
SRR7895717	GSM3399320	30	55	L7
SRR7895718	GSM3399321	30	55	L7
SRR7895719	GSM3399322	30	55	L7
SRR7895720	GSM3399323	30	55	L7

Tabla 1. Tabla de datos para estudio. Datos correspondientes al set de datos SRP162456, del genotipo transgénico SOD1G93A, cada uno con código run y GEO, sus largo de secuencias de los read y su correspondiente código de barcode.

Run	GEO_Accession	Largo seq 1	largo seq2	Barcode
SRR7895536	GSM3399139	30	50	L6
SRR7895537	GSM3399140	30	50	L6
SRR7895538	GSM3399141	30	50	L6
SRR7895712	GSM3399315	30	55	L7
SRR7895713	GSM3399316	30	55	L7
SRR7895714	GSM3399317	30	55	L7

SRR7895715	GSM3399318	30	55	L7
SRR7895716	GSM3399319	30	55	L7
SRR7895717	GSM3399320	30	55	L7
SRR7895718	GSM3399321	30	55	L7

Tabla 2. Tabla de datos analizados. Datos seleccionados para ser analizados, de entre los 37 de la tabla 1.

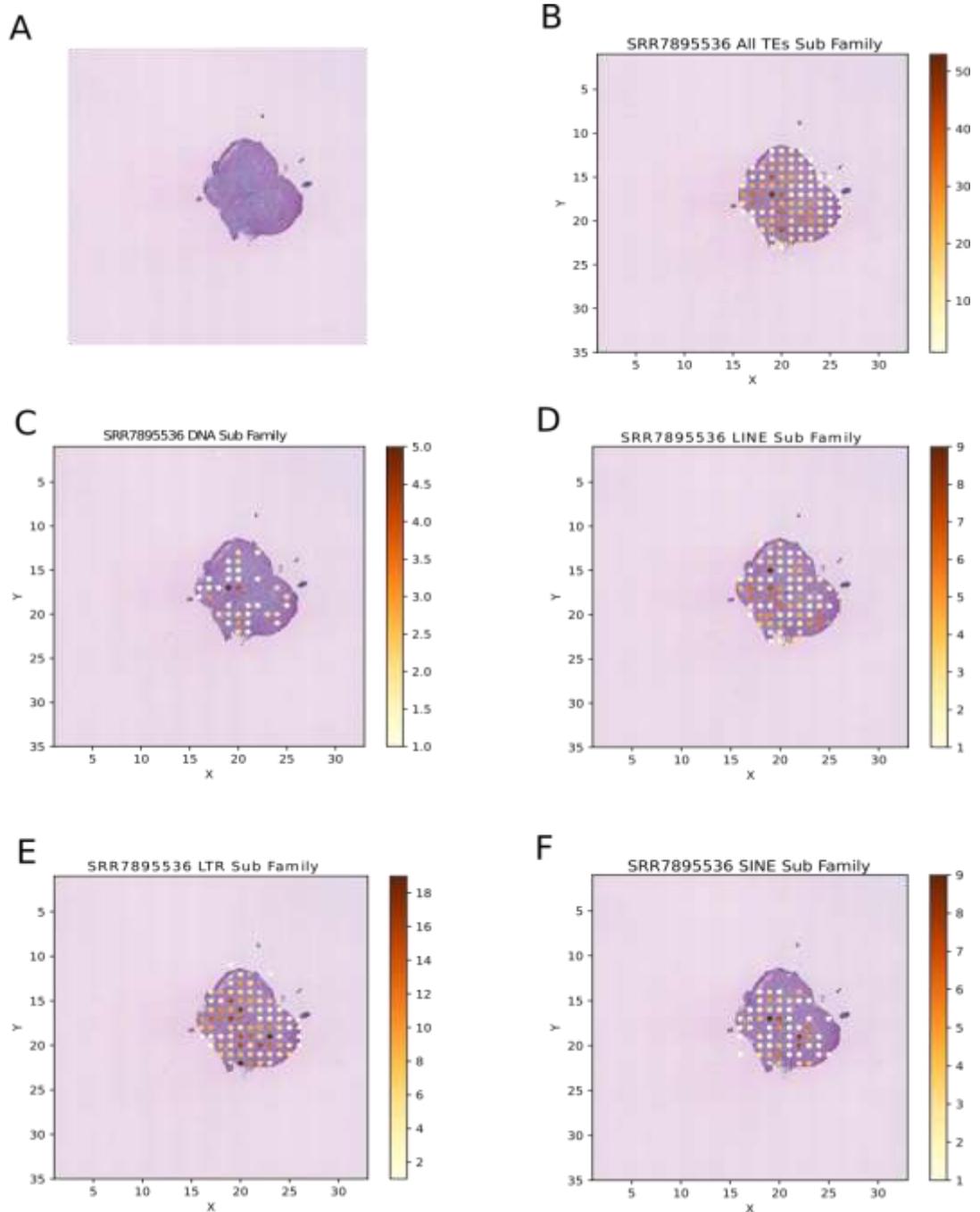


Figura 1. Expresión de TEs en la médula espinal (muestra SRR7895536) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

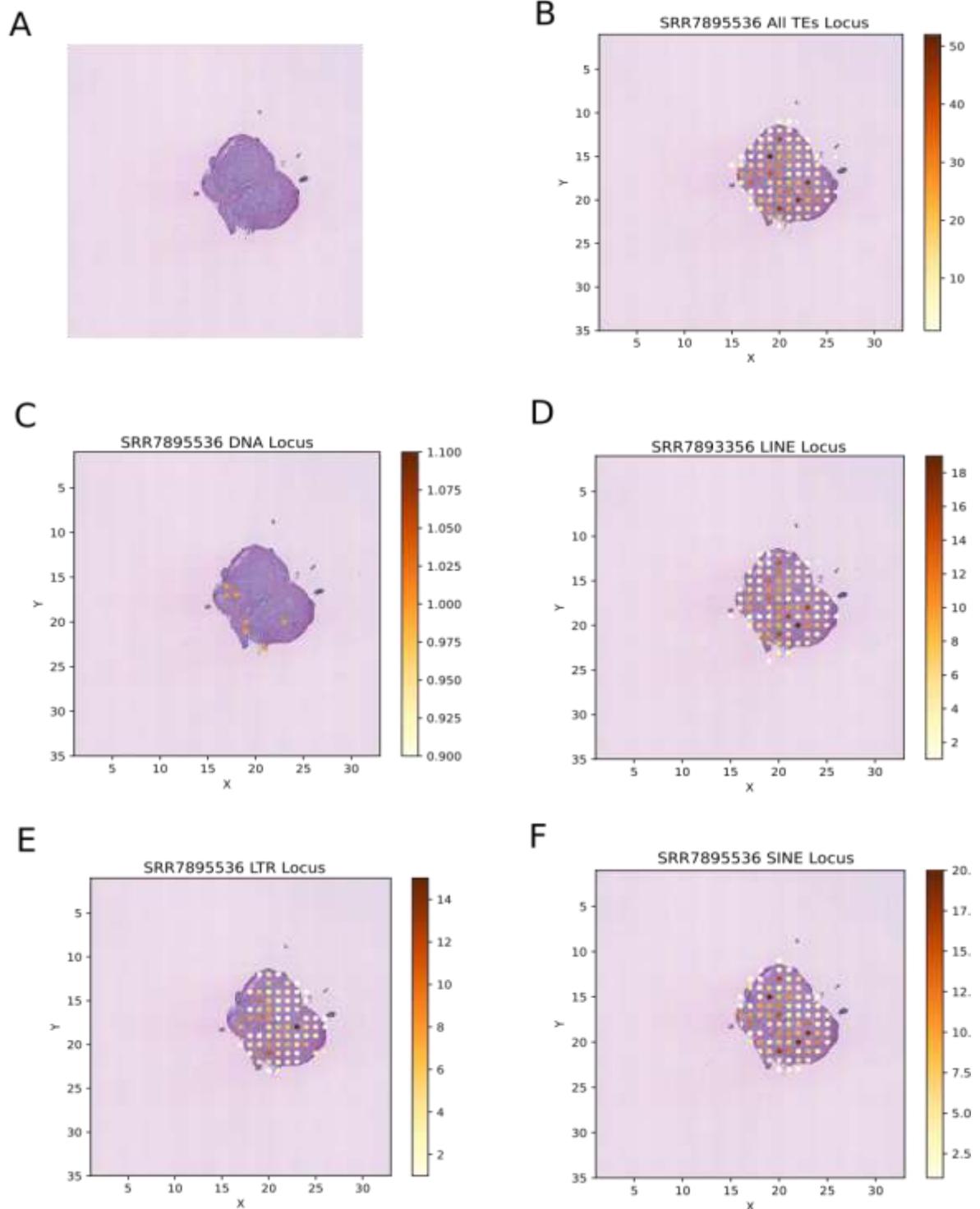


Figura 2. Expresión de TEs en la médula espinal (muestra SRR7895536) usando output a nivel de Locus. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

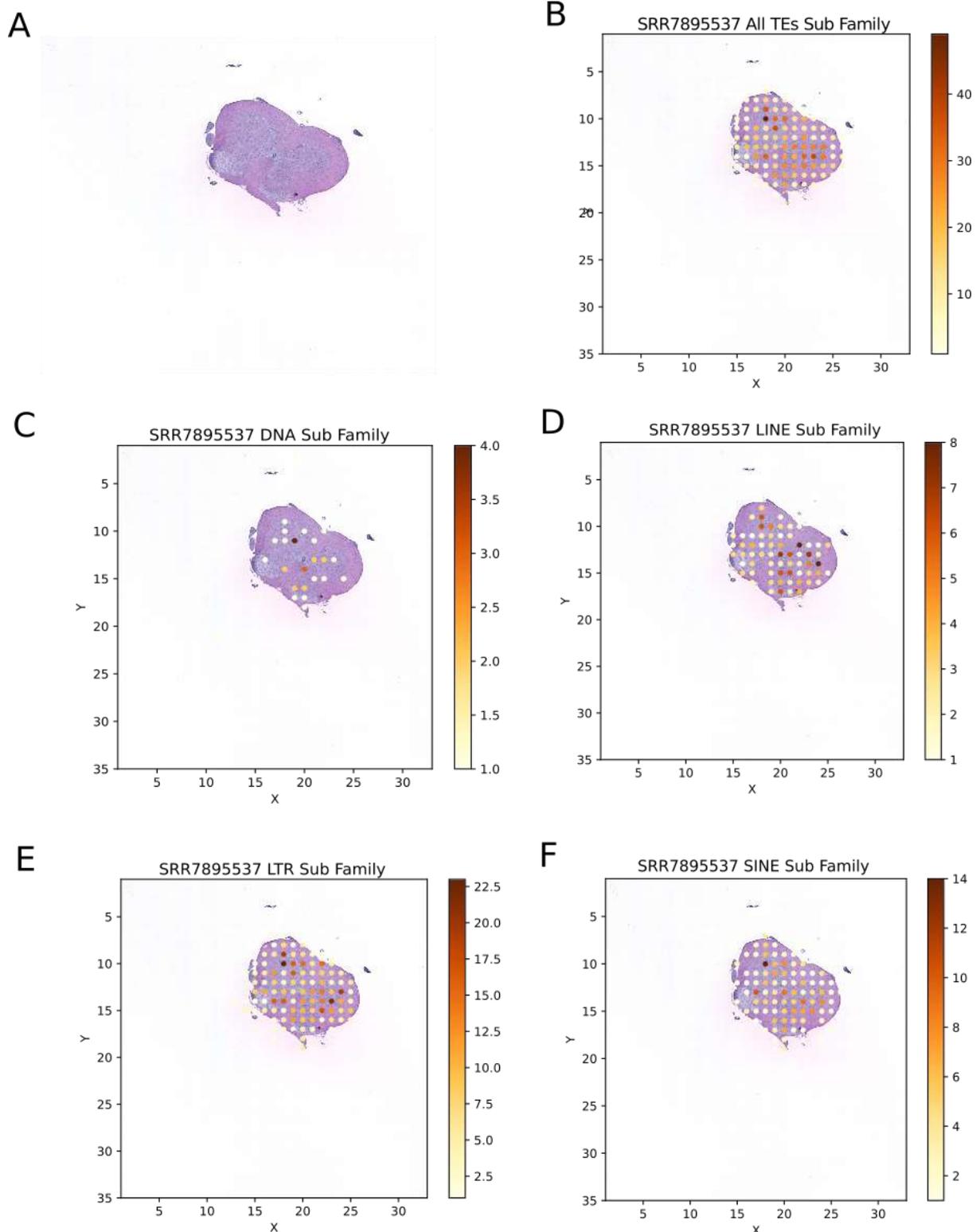


Figura 3. Expresión de TEs en la médula espinal (muestra SRR7895537) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

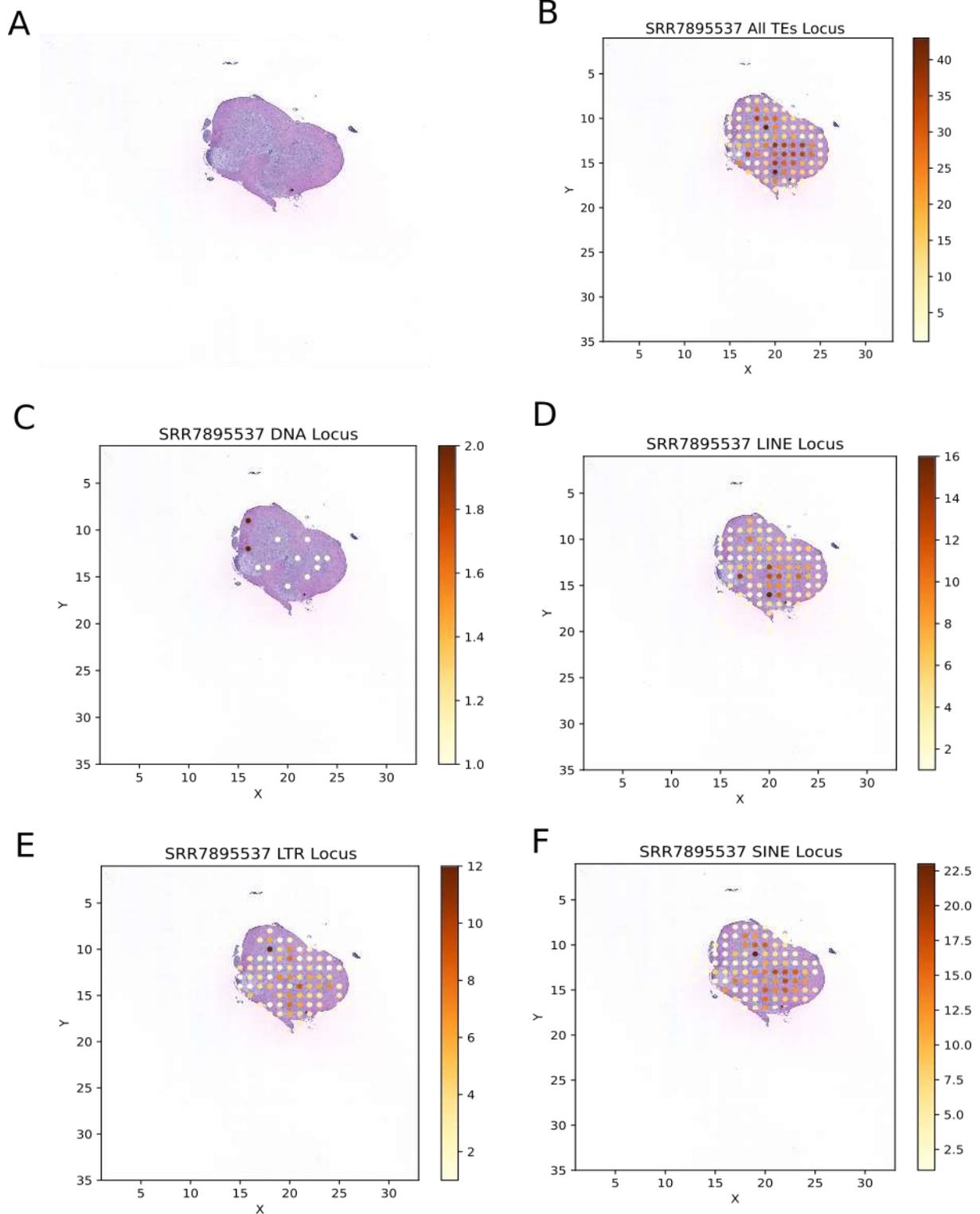


Figura 4. Expresión de TEs en la médula espinal (muestra SRR7895537) usando output a nivel de Locus. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

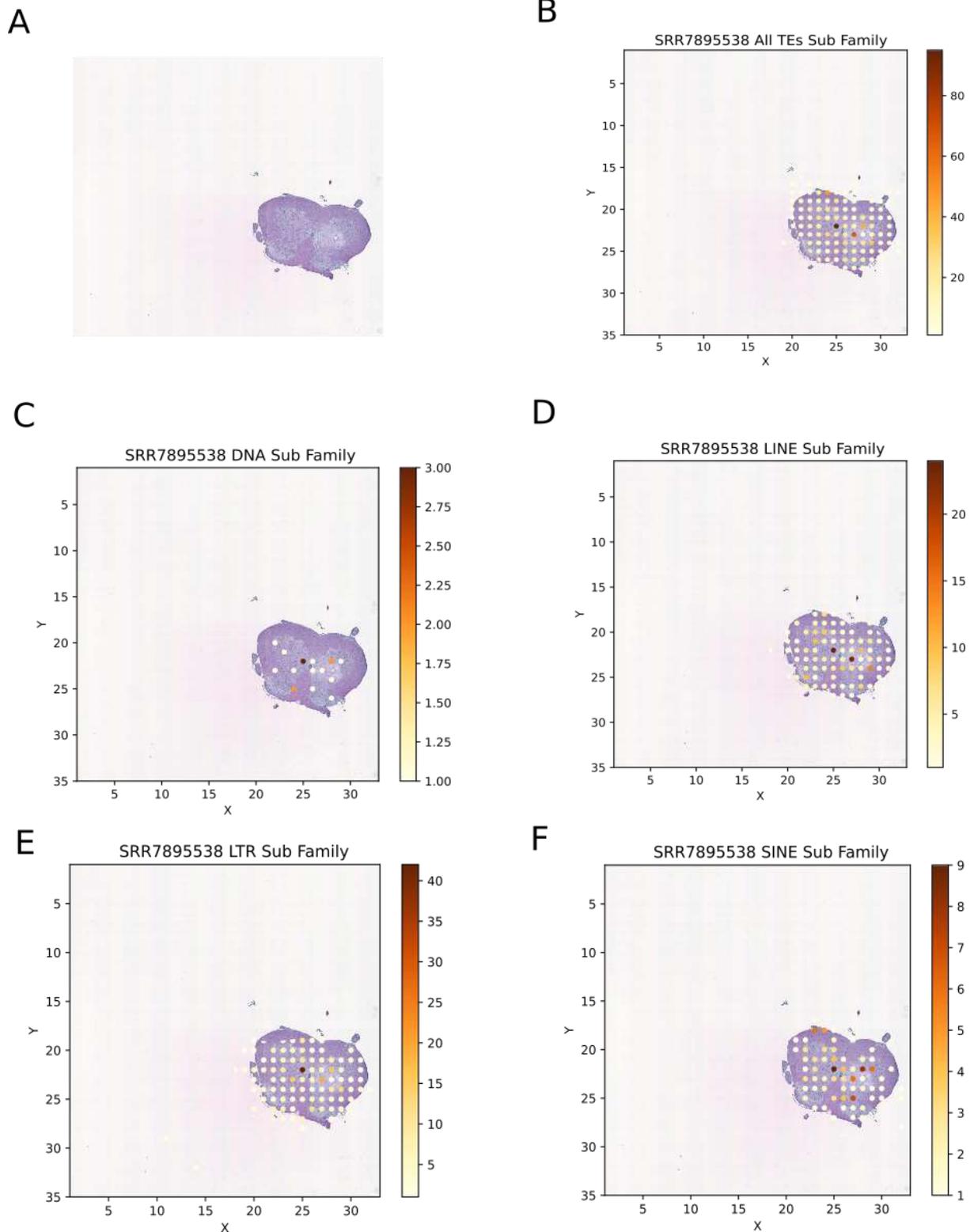


Figura 5. Expresión de TEs en la médula espinal (muestra SRR7895538) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

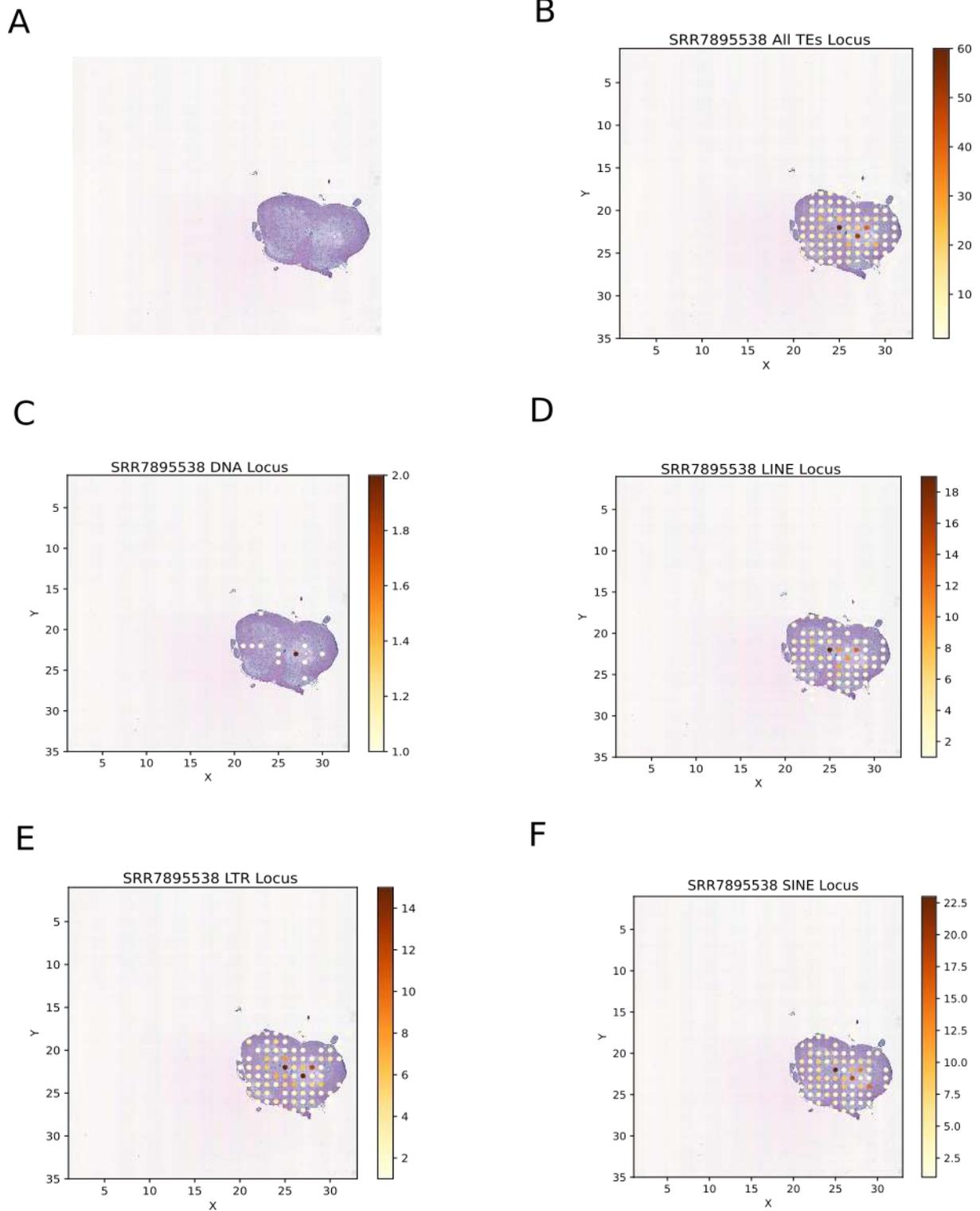


Figura 6. Expresión de TEs en la médula espinal (muestra SRR7895538) usando output a nivel Locus de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

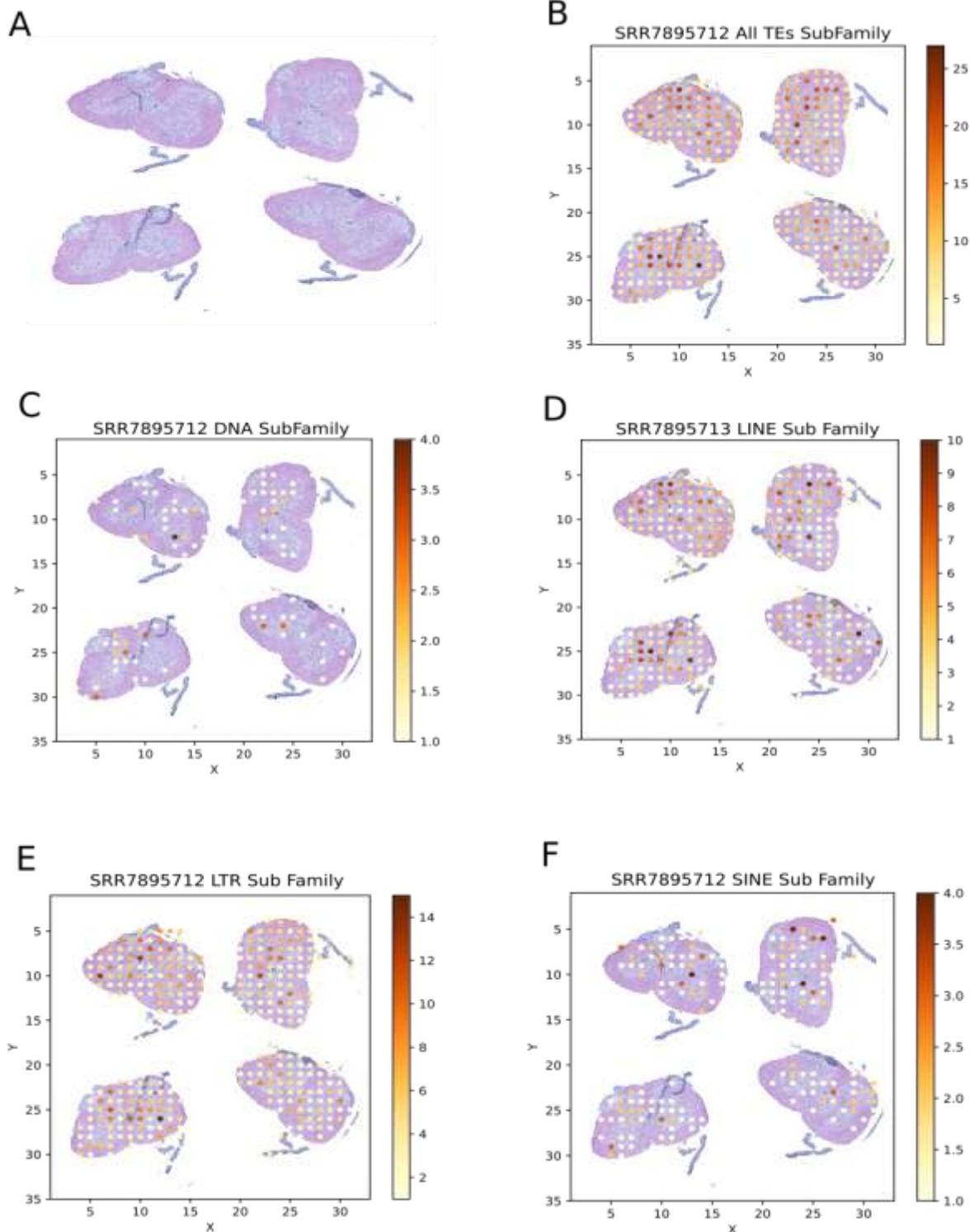


Figura 7. Expresión de TEs en la médula espinal (muestra SRR7895712) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

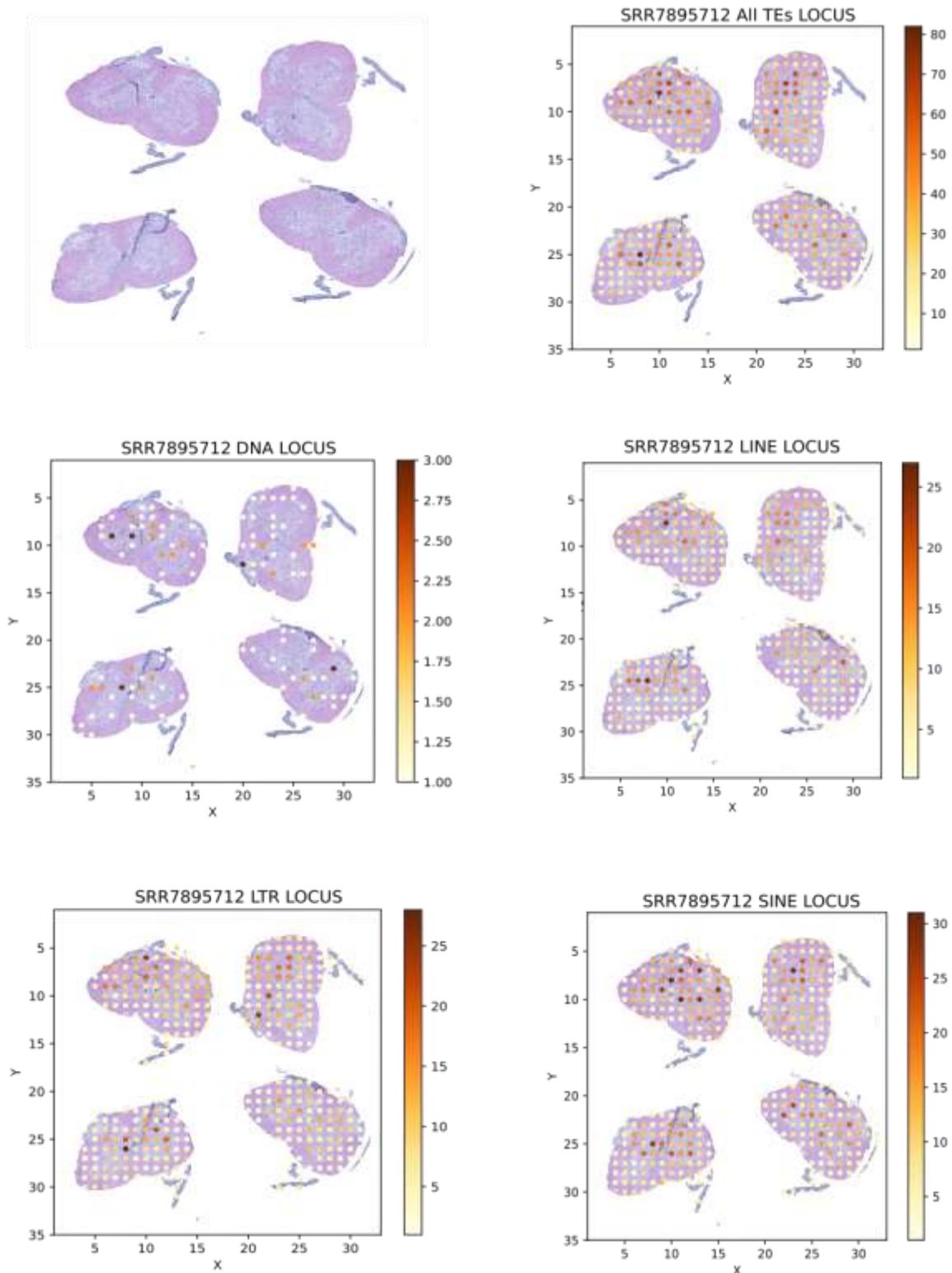


Figura 8. Expresión de TEs en la médula espinal (muestra SRR7895712) usando output a nivel Locus de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

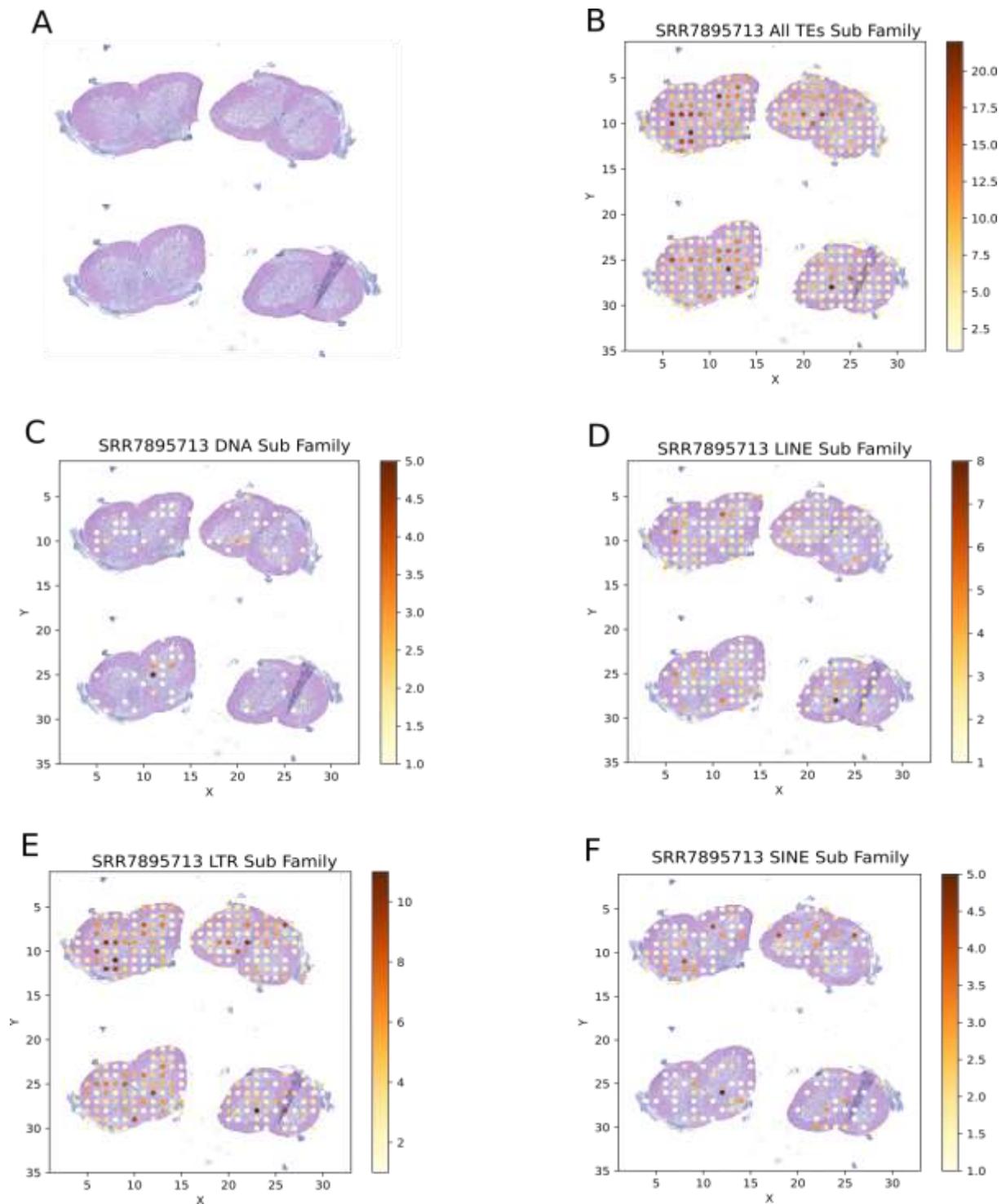


Figura 9. Expresión de TEs en la médula espinal (muestra SRR7895713) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

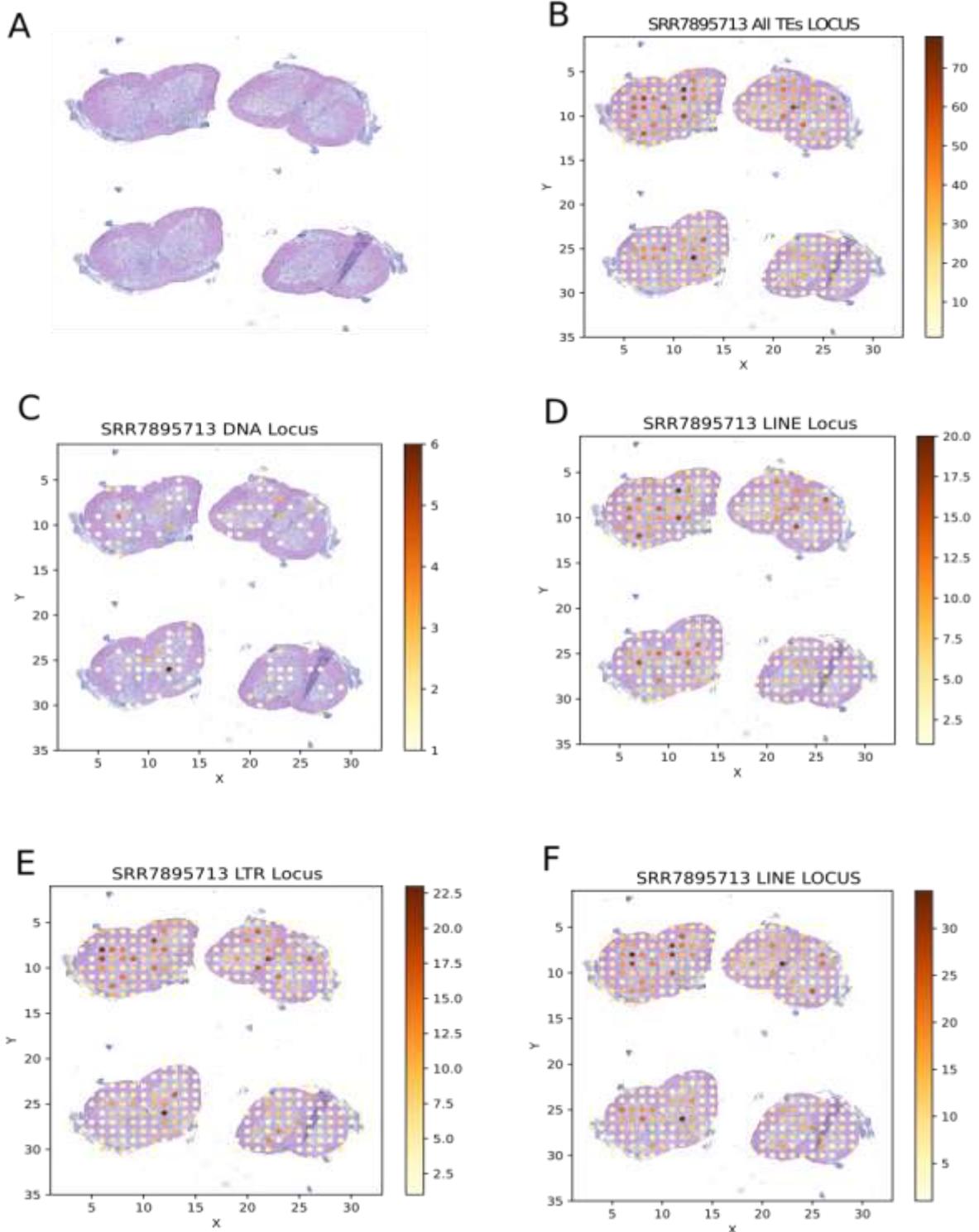


Figura 10. Expresión de TEs en la médula espinal (muestra SRR7895713) usando output a nivel Locus de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

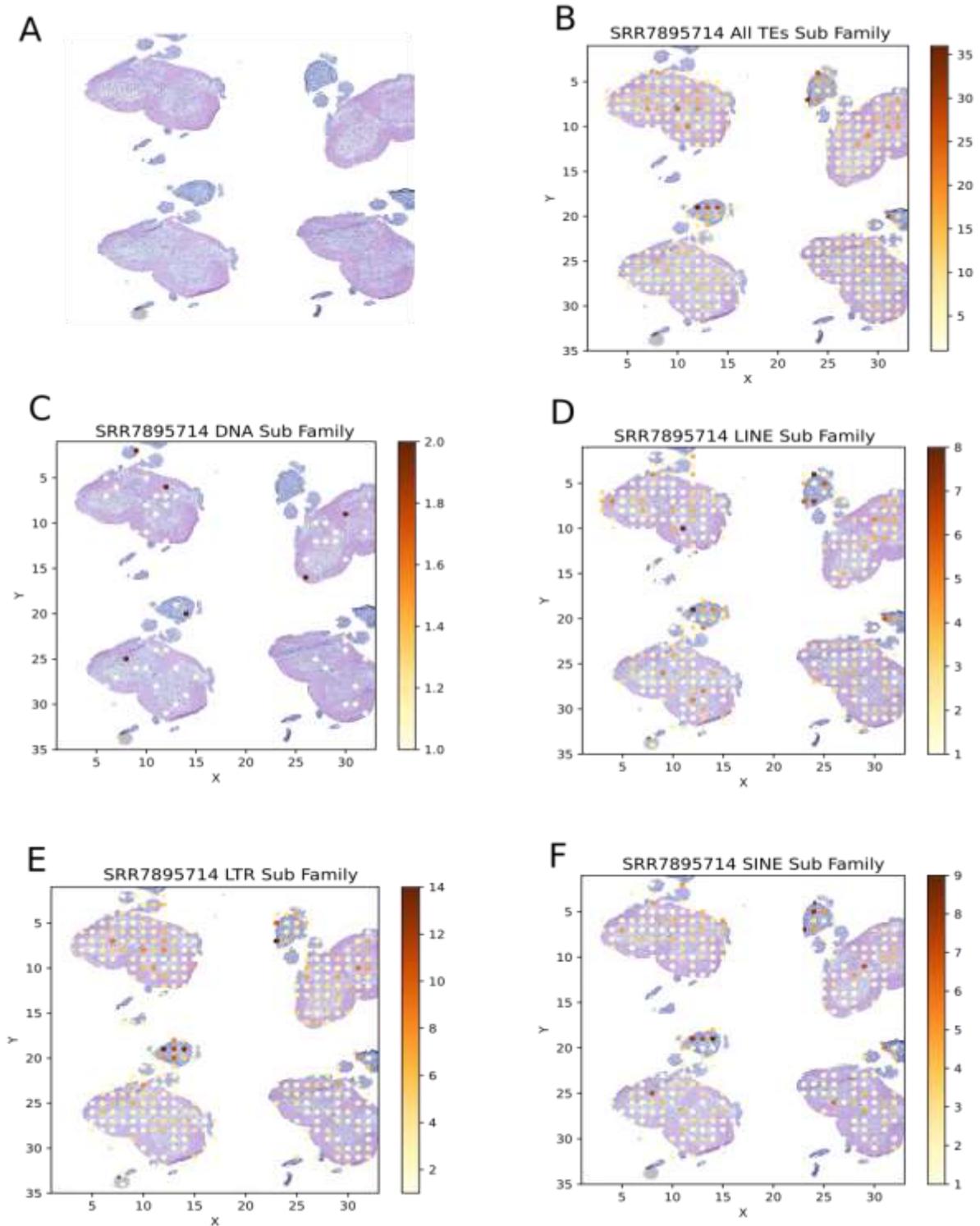


Figura 11. Expresión de TEs en la médula espinal (muestra SRR7895714) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

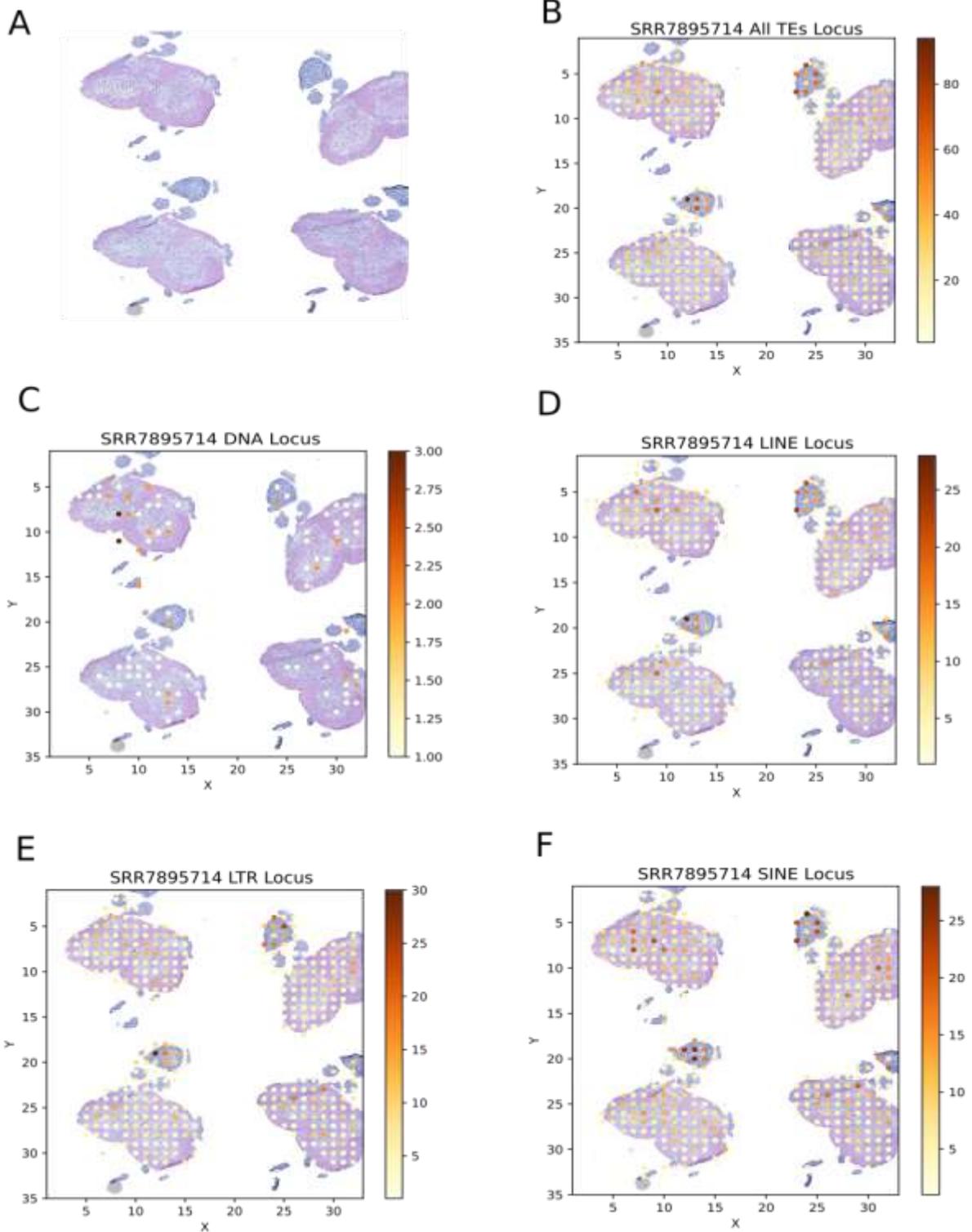


Figura 12. Expresión de TEs en la médula espinal (muestra SRR7895714) usando output a nivel Locus de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

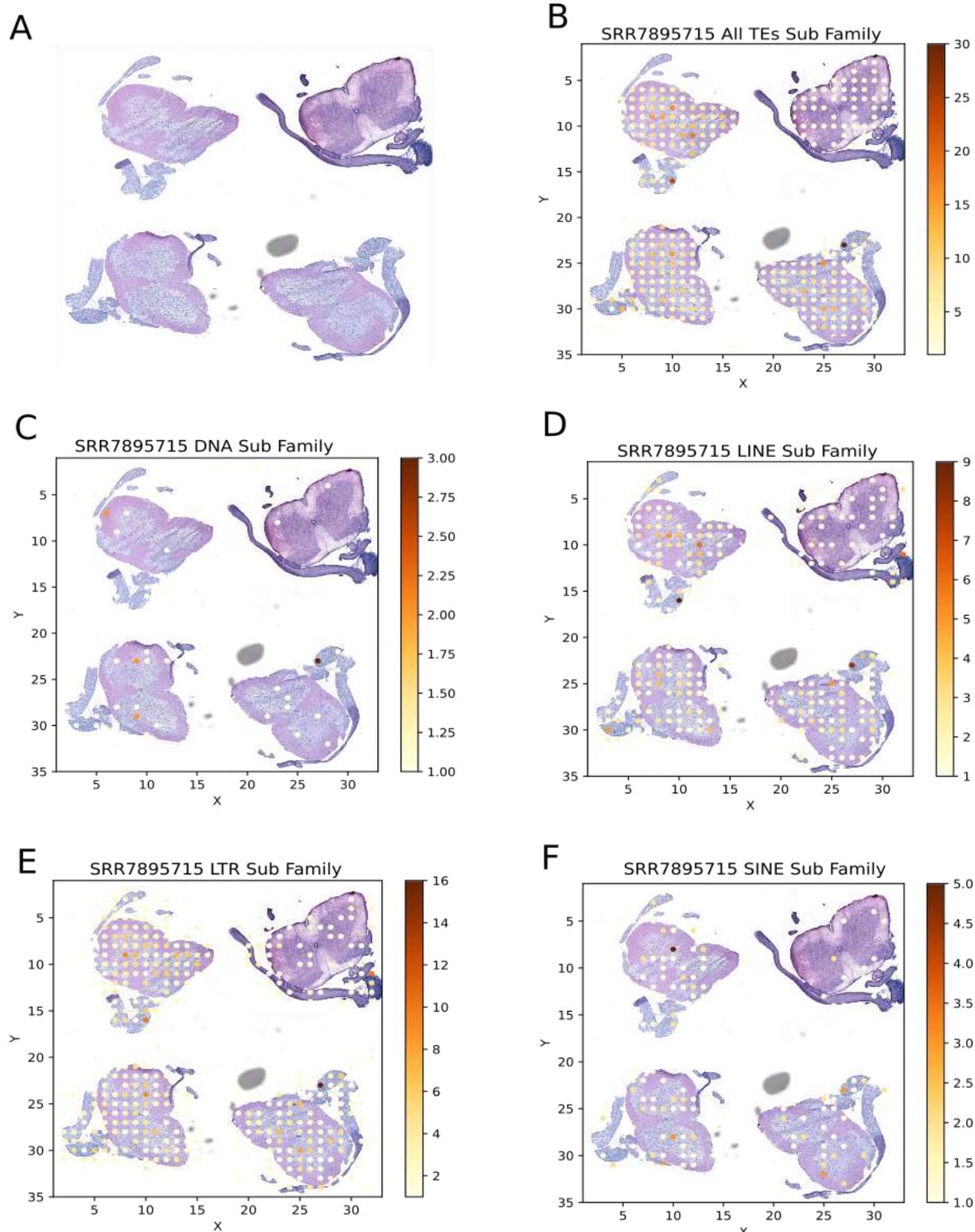


Figura 13. Expresión de TEs en la médula espinal (muestra SRR7895715) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

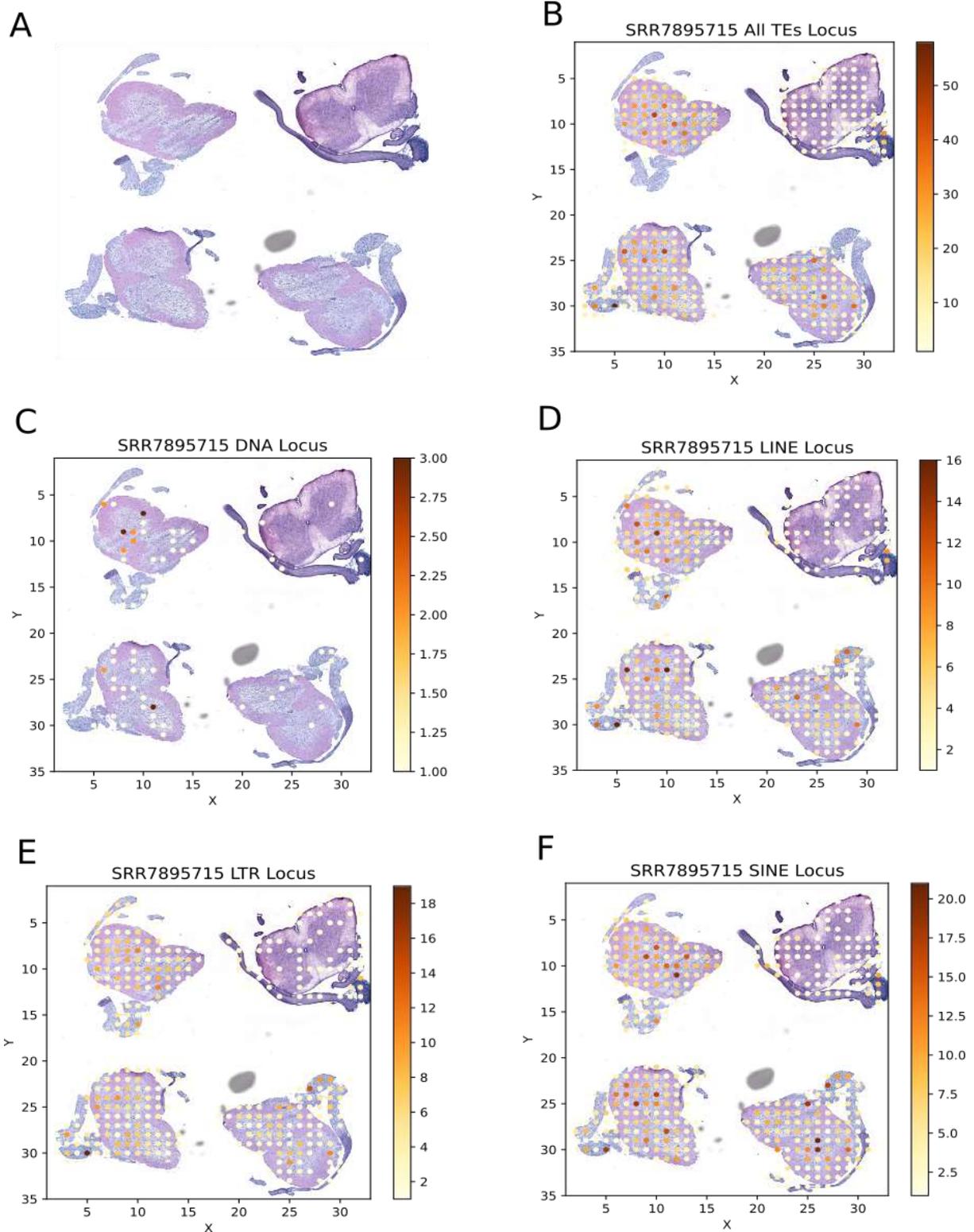


Figura 14. Expresión de TEs en la médula espinal (muestra SRR7895715) usando output a nivel Locus de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

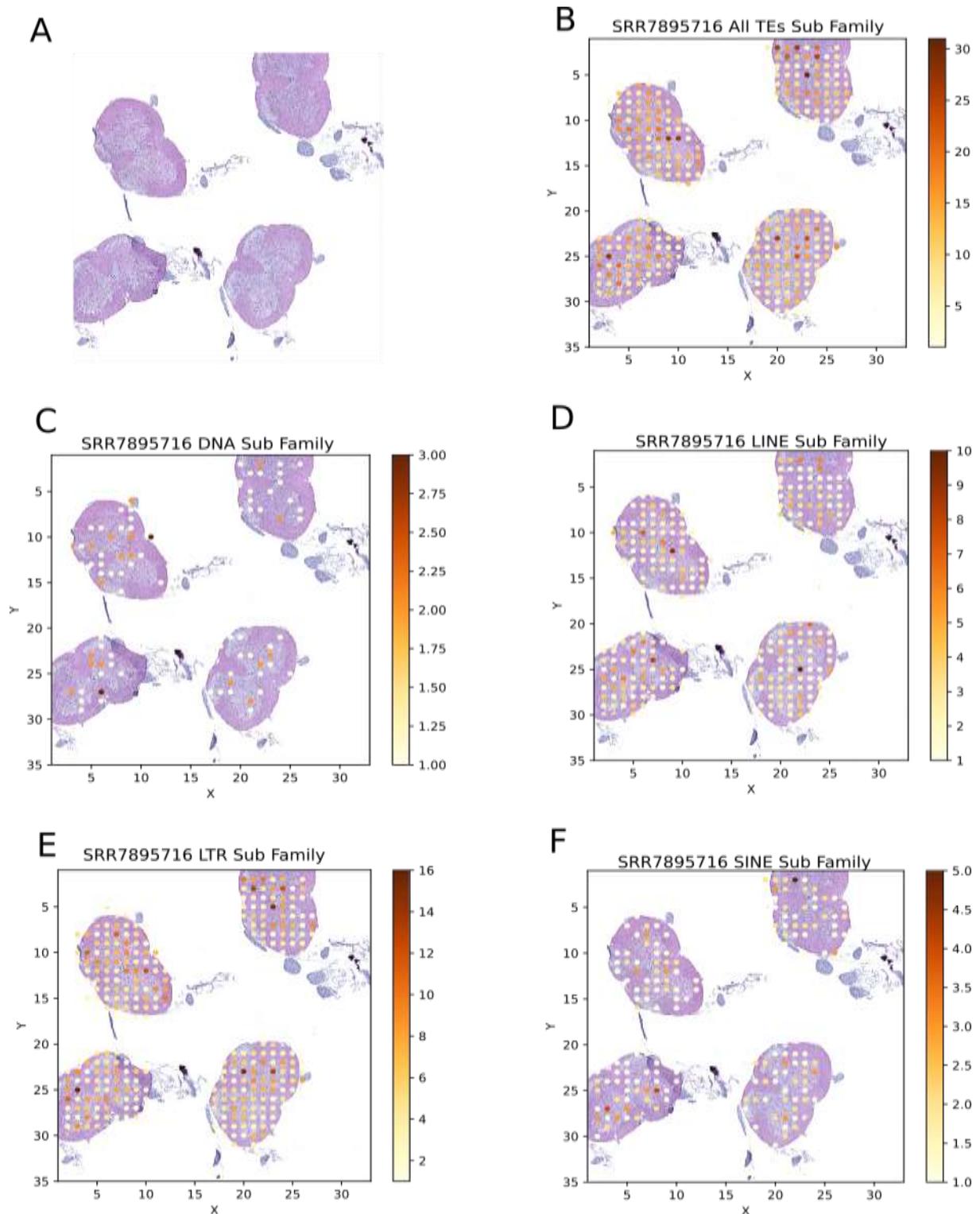


Figura 15. Expresión de TEs en la médula espinal (muestra SRR7895716) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

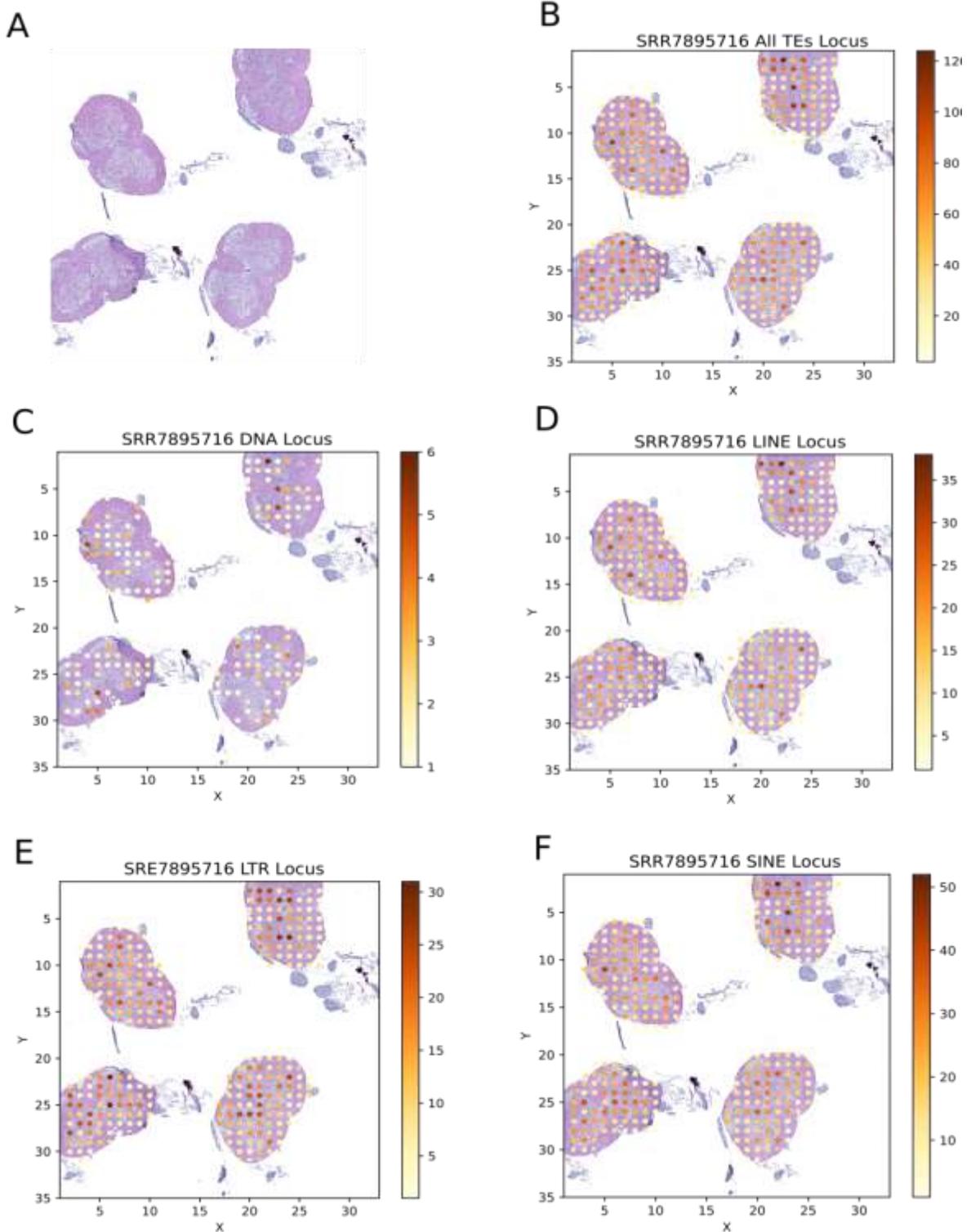


Figura 16. Expresión de TEs en la médula espinal (muestra SRR7895716) usando output a nivel Locus de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

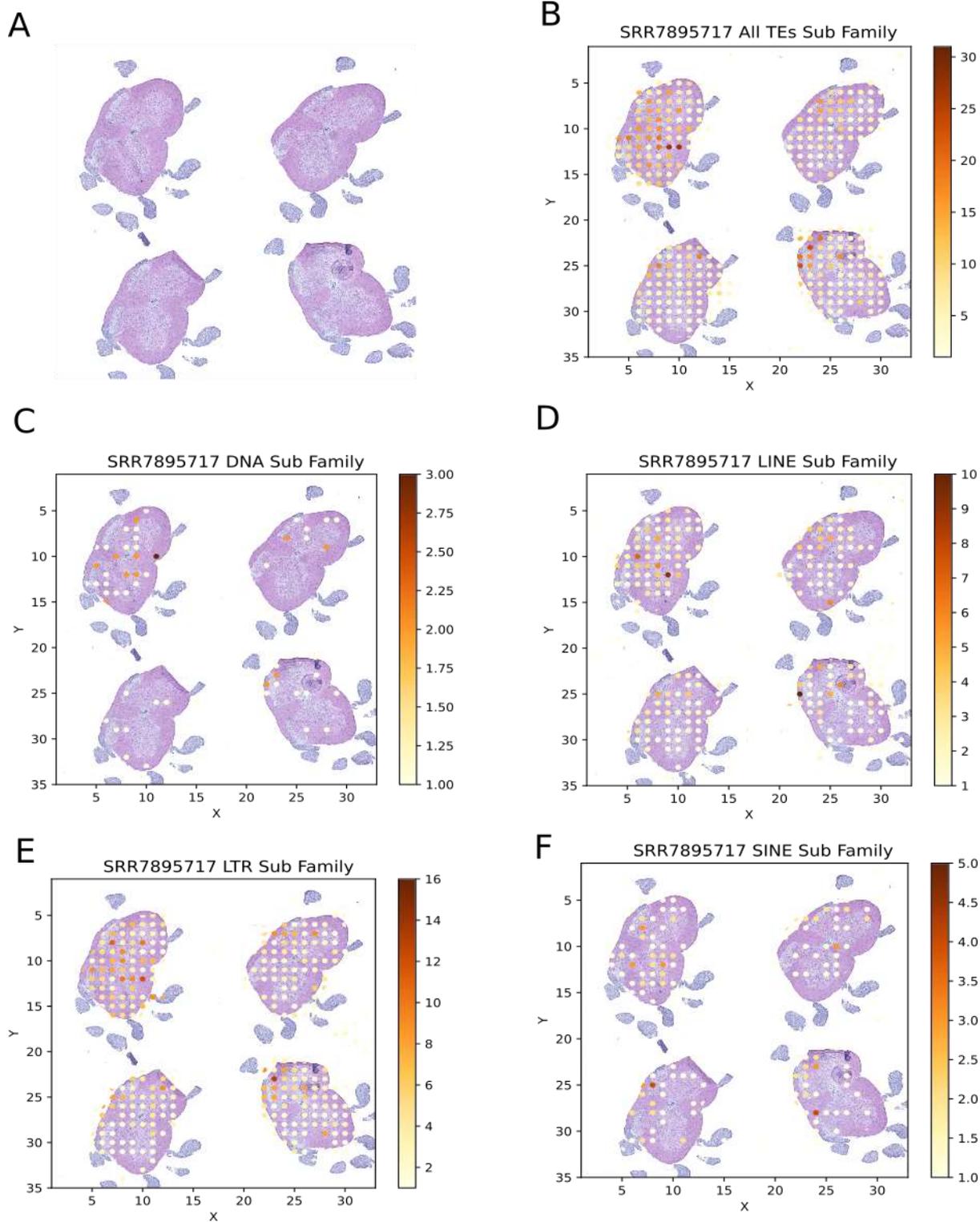


Figura 17. Expresión de TEs en la médula espinal (muestra SRR7895717) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

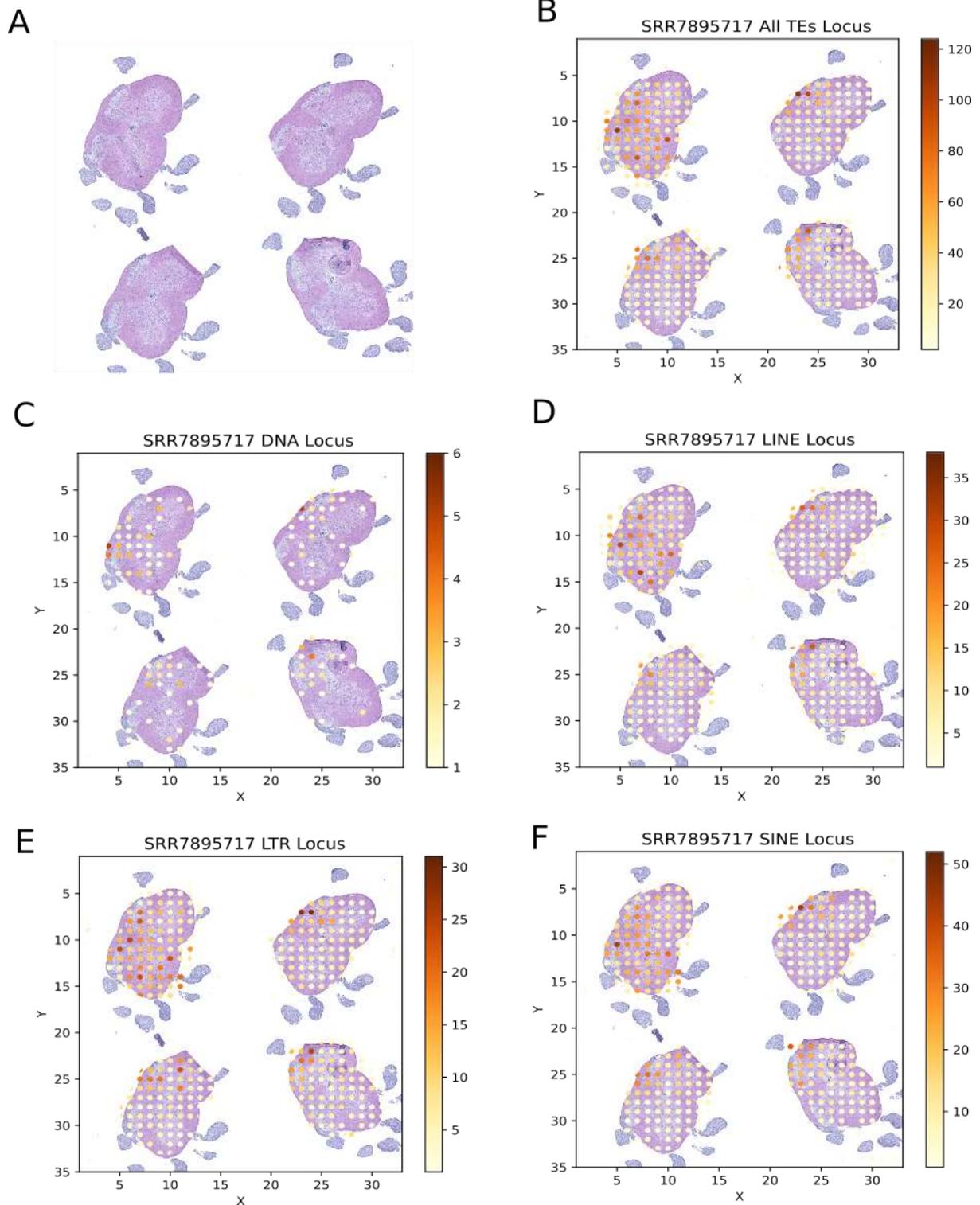


Figura 18. Expresión de TEs en la médula espinal (muestra SRR7895717) usando output a nivel Locus de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

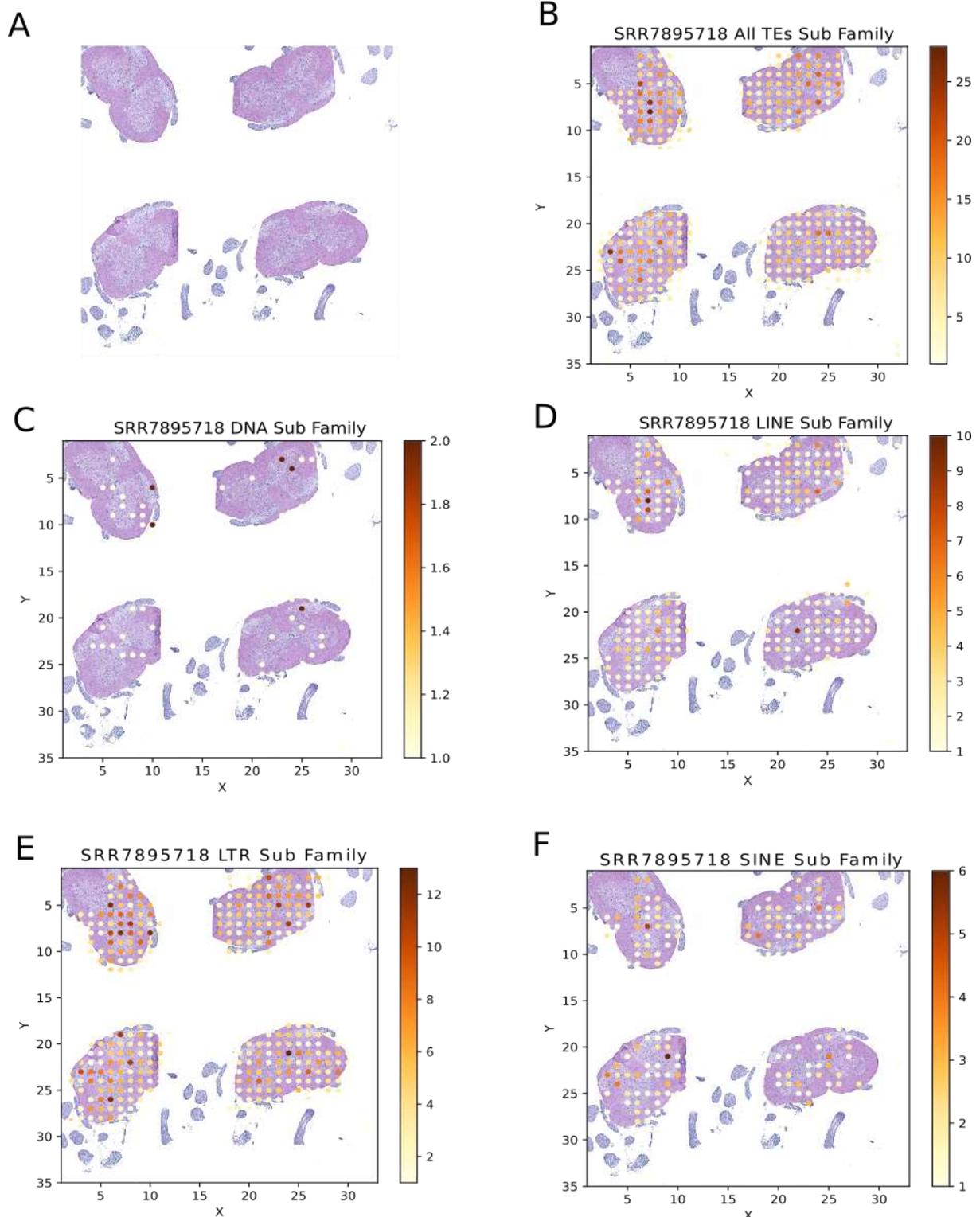


Figura 19. Expresión de TEs en la médula espinal (muestra SRR7895718) usando output a nivel SubFamily de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.

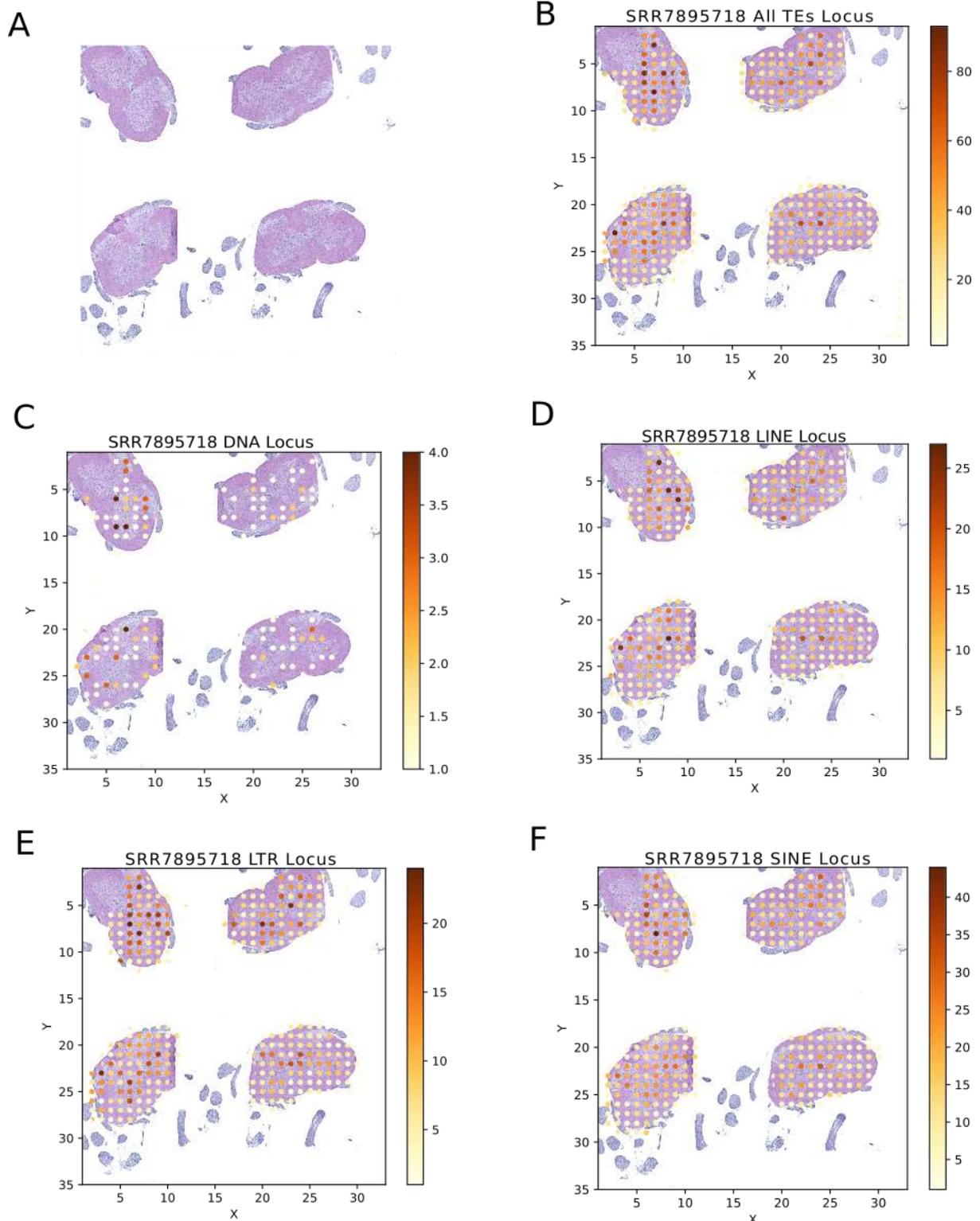


Figura 20. Expresión de TEs en la médula espinal (muestra SRR7895718) usando output a nivel Locus de SpatialTE. A. Médula espinal, sin superposición de expresión de TEs. **B.** Expresión de todos los TEs **C.** Expresión de TEs de tipo DNA. **D.** Expresión de TEs de tipo LINE. **E.** Expresión de TEs tipo LTR. **F.** Expresión de TEs de tipo SINE.