

# Automatizando el conteo de células madre

Jerónimo Fotinós

4 de noviembre de 2023

## La imagen

La imagen utilizada corresponde al archivo Sph1 Z Stack – CTRL SOX2.czi. Esta contiene datos de la observación de una tumoresfera por microscopía confocal. La extensión czi es un formato propietario que contiene metadatos que indican que la imagen consta de tres canales, cada uno con una imagen 3D de 9 cortes (píxel) en la dirección  $z$  y 2292 tanto en  $x$  como en  $y$ . El espaciado es de

$$\Delta x = \Delta y = 1,2364633517553391 \times 10^{-7} \text{ m} \approx 0,12 \mu\text{m}$$

$$\Delta z = 1,9999999999999999 \times 10^{-6} \text{ m} \approx 2 \mu\text{m}.$$

Los canales corresponden a la fluorescencia del marcador SOX2, la fluorescencia de los núcleos y al canal óptico.

## Canal SOX2

Queremos borrar la fluorescencia dada por el marcador que difunde, quedándonos con la que corresponde al marcador ligado a las CMC. Para esto, nos quedamos con las áreas más brillantes restando a la imagen una versión “difuminada” de sí misma. Más precisamente, usamos la técnica morfológica de reconstrucción por dilatación (ver [documentación de scikit-image](#)). La Fig. 1 muestra la imagen original a la par de la procesada, mientras que la Fig. 2 muestra la imagen final en mayor tamaño, acompañada por la barra de color que codifica las intensidades de los píxeles.

## Canal de los Núcleos

Podemos usar los núcleos para identificar e individualizar las células. Lo que queremos extraer de este canal son las posiciones de los centros de los núcleos, junto a algunas cantidades geométricas de los núcleos como el área, perímetro, etc. Para esto, haremos segmentación de instancias. Preparamos la imagen para la segmentación del siguiente modo.

1. Ecualización: aumentamos el contraste de la imagen haciendo que el histograma de intensidades ocupe todo el rango disponible. Usamos en particular, usamos Contrast Limited Adaptive Histogram Equalization (CLAHE, ver [documentación](#)), un algoritmo que computa varios histogramas de intensidades en diferentes regiones de la imagen para obtener mejoras locales del contraste, logrando marcar detalles aun dentro de zonas oscuras o claras. Esto es útil, ya que tenemos que hacer todo lo posible para que el algoritmo de segmentación pueda reconocer los bordes de manera clara.
2. Procesamiento morfológico: realizamos una apertura (erosión seguida de dilatación morfológica), para tratar de marcar los bordes, separando los núcleos que están muy pegados, y luego hacemos un cierre de área (ver [documentación](#)), para remover los huecos oscuros dentro de un mismo núcleo.
3. Filtrado Bilateral: para eliminar el ruido en las estructuras, introducido por las operaciones morfológicas, preservando los bordes (ver [documentación](#)).

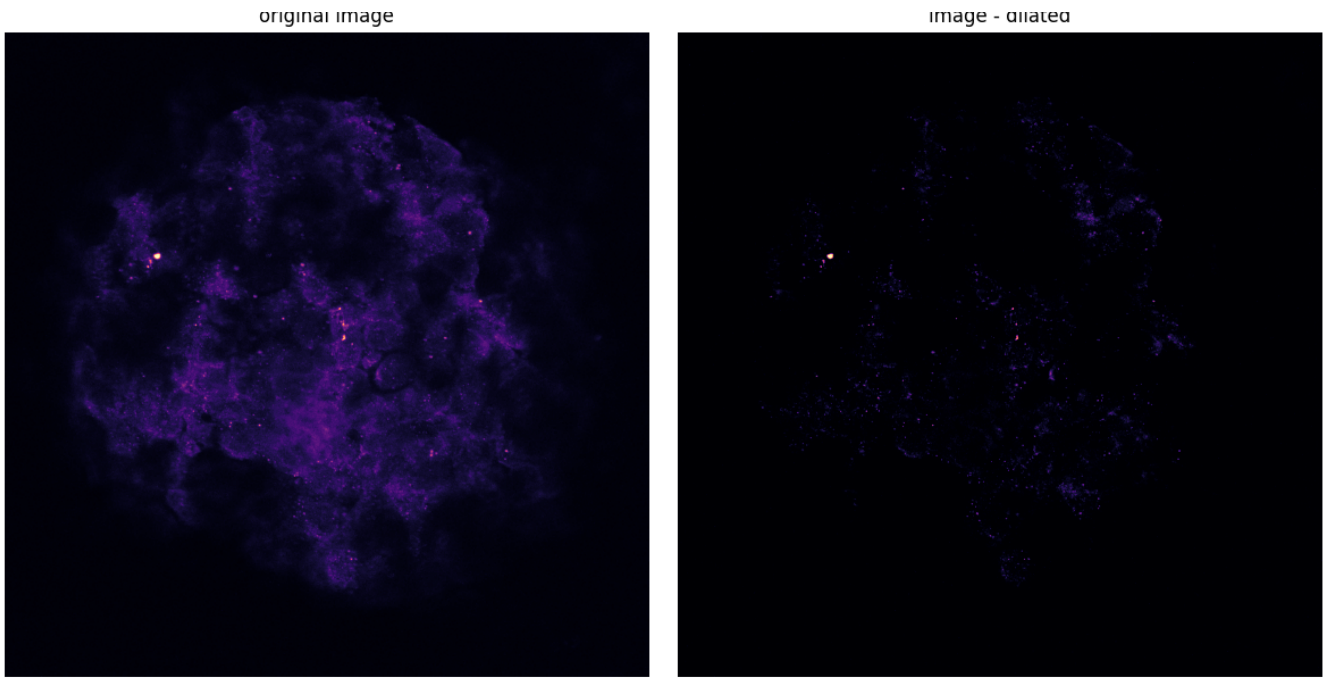


Figura 1: Canal de SOX2, slice 3, antes y después del procesado.

4. Segmentación: se identifican los núcleos mediante una red neuronal de arquitectura U-Net, entrenada para identificar polígonos convexos en la imagen provista (ver [documentación](#)).

El proceso completo puede verse en la Fig. 3. De la segmentación, podemos extraer información geométrica de las estructuras, en particular sus centros. Para dividir el espacio correspondiente a cada célula, hacemos una teselación de Voronoi con los centros como puntos. Esta puede verse en la Fig. 4(a).

Para identificar cuáles de estas células son madre, queremos ver cuál es la intensidad de la fluorescencia del marcador SOX2 por región. Superponiendo la imagen previamente obtenida para dicho canal (Fig. 2) a las regiones, obtenemos la suma de las intensidades para cada región. Esta puede verse en la Fig. 4(b). A fin de visualizar mejor este resultado, superponemos el canal de SOX2, el de los núcleos y la teselación en la Fig. 5.

Para identificar cuáles células son madre, nos falta definir un umbral para la intensidad de la fluorescencia en una región, a partir del cual consideramos que la célula correspondiente es madre. Como esperamos que estas células sean de dos clases (madres y no madres), el histograma de intensidades de fluorescencia por región debería corresponder a una distribución bimodal. Dicho histograma puede verse en la Fig. 6. El criterio para fijar el umbral es importante, ya que los resultados varían fuertemente con el mismo. E.g., sea  $V$  el umbral usado,  $S_V$  el número de células madre obtenido en consecuencia y  $f_V$  la correspondiente fracción de células madre respecto al total, tenemos que

$$V = 25 \implies S_{25} = 139 \implies f_{25} = 0,7354497354497355$$

$$V = 50 \implies S_{50} = 92 \implies f_{50} = 0,48677248677248675$$

$$V = 80 \implies S_{80} = 46 \implies f_{80} = 0,24338624338624337$$

Una manera de tomar un criterio, es utilizar técnicas de agrupamiento fijando el número de grupos a dos. Como una primera aproximación, podemos hacer la hipótesis de que los datos son generados por una mezcla de gaussianas (una corresponde a las células madre y la otra a las diferenciadas). Si filtramos los datos con fluorescencias mayores a 180 (para que no nos desvíen la media) y ajustamos las gaussianas, obtenemos el agrupamiento de la Fig. 7. En este caso, el valor umbral obtenido para la fluorescencia es de aproximadamente 56. Esto es

$$V \approx 56 \implies S_V = 86 \implies f_V = 0,455026455026455$$

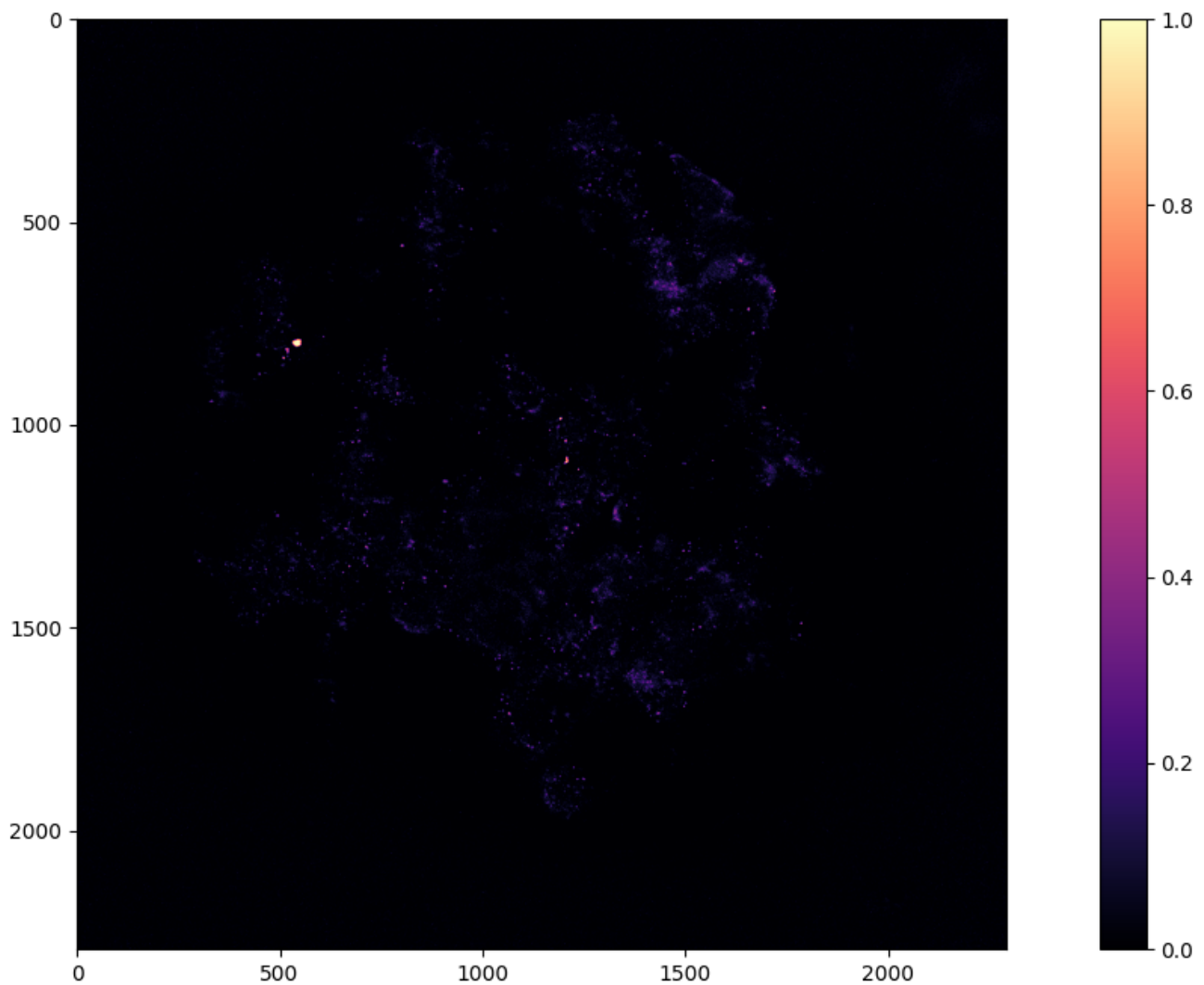


Figura 2: Canal de SOX2, slice 3, imagen procesada con barra de color.

### Observaciones

- Un “gráfico del codo” usando el algoritmo K-medias, muestra que para estos datos usar dos grupos produce la caída más significativa en la varianza de los grupos. Esto indica que para estos datos, 2 es un número adecuado de grupos (como esperábamos).
- El agrupamiento es bastante sensible a los valores inusualmente grandes. Si uno hace el agrupamiento con mezcla de gaussianas considerando todos los datos, el número de células madre cae a 42, con el umbral en la intensidad de fluorescencia cerca de 85.

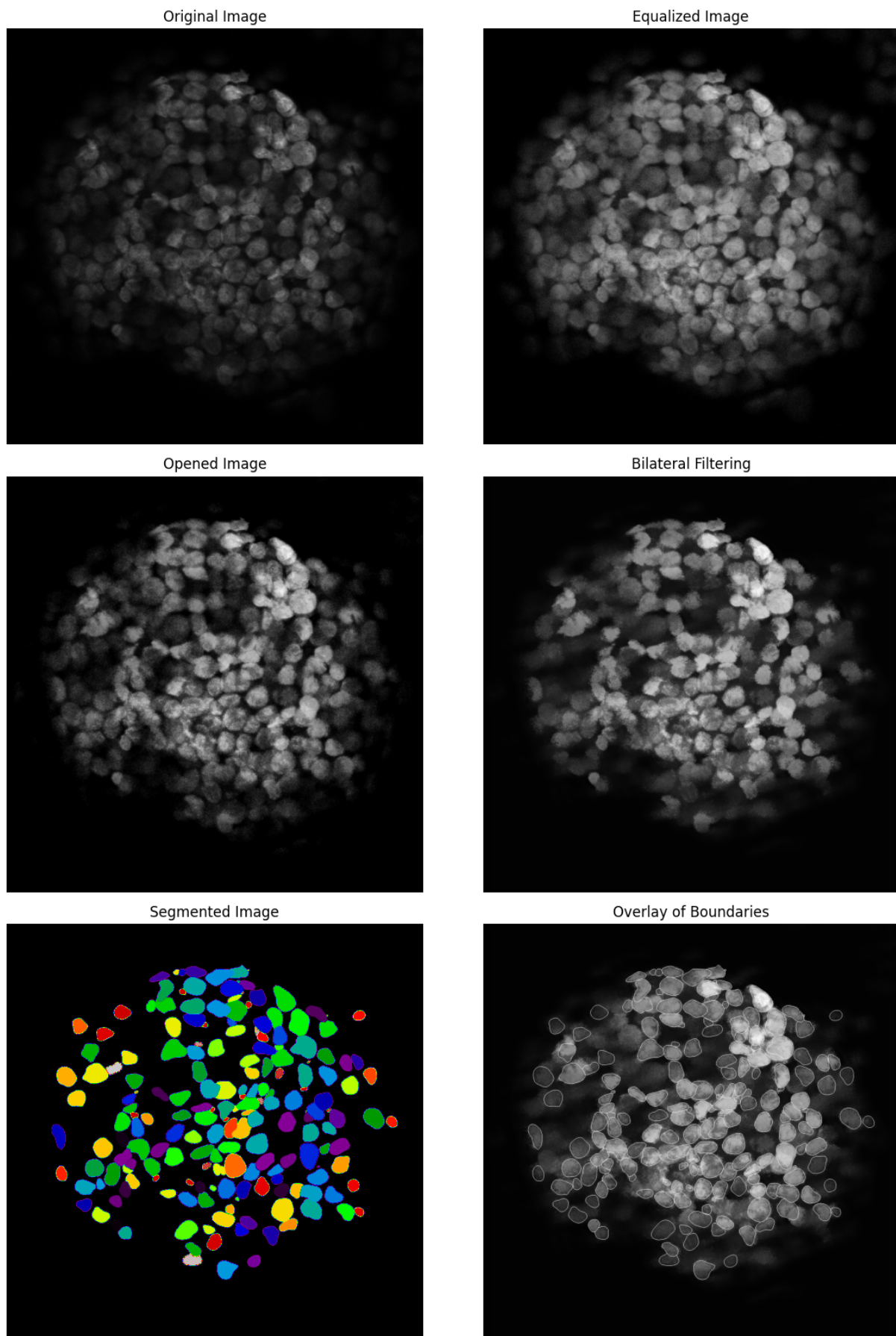
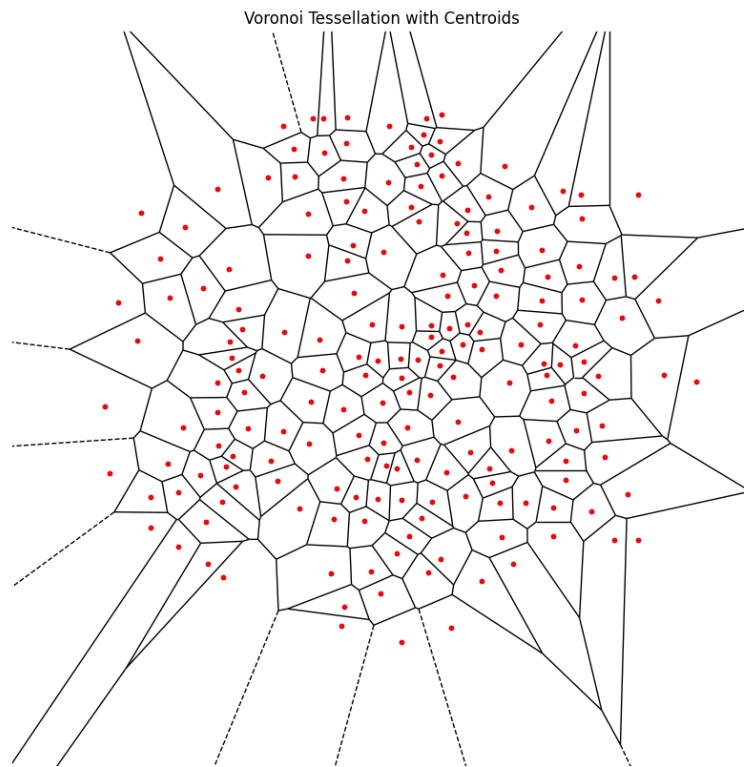
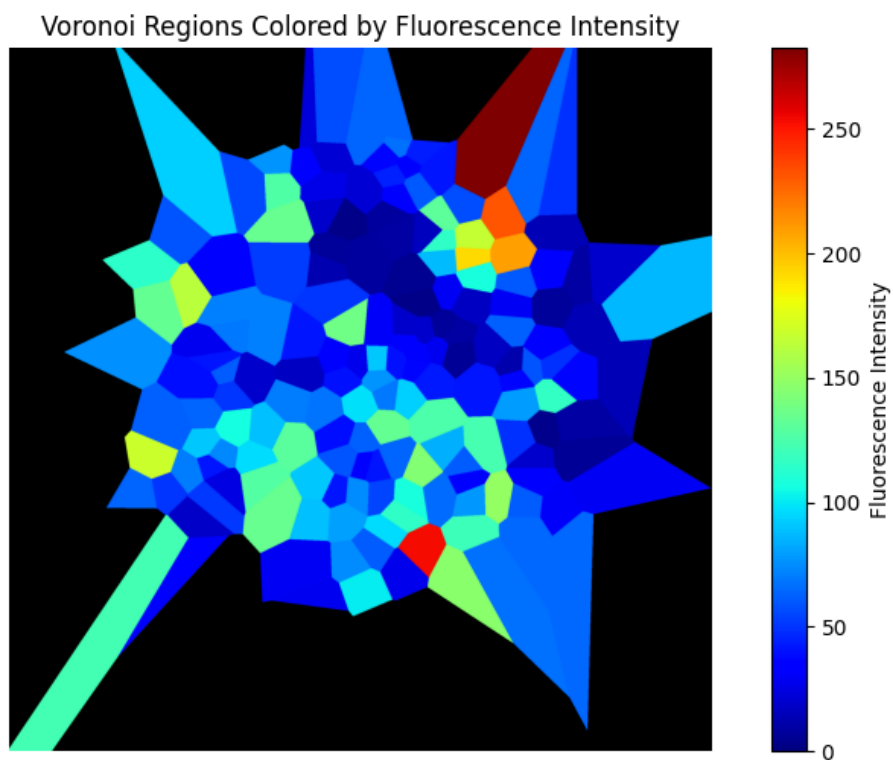


Figura 3: *Procesado del canal de los núcleos.* En la imagen de la esquina inferior derecha, pueden verse las estructuras identificadas sobre la imagen provista. Los núcleos más oscuros y borrosos, que no son reconocidos, corresponden a células que están más al fondo, fuera del plano focal.



(a) Teselación de Voronoi



(b) Regiones por intensidad de fluorescencia SOX2

Figura 4: Teselación de Voronoi: Regiones y Fluorescencia SOX2 por región.



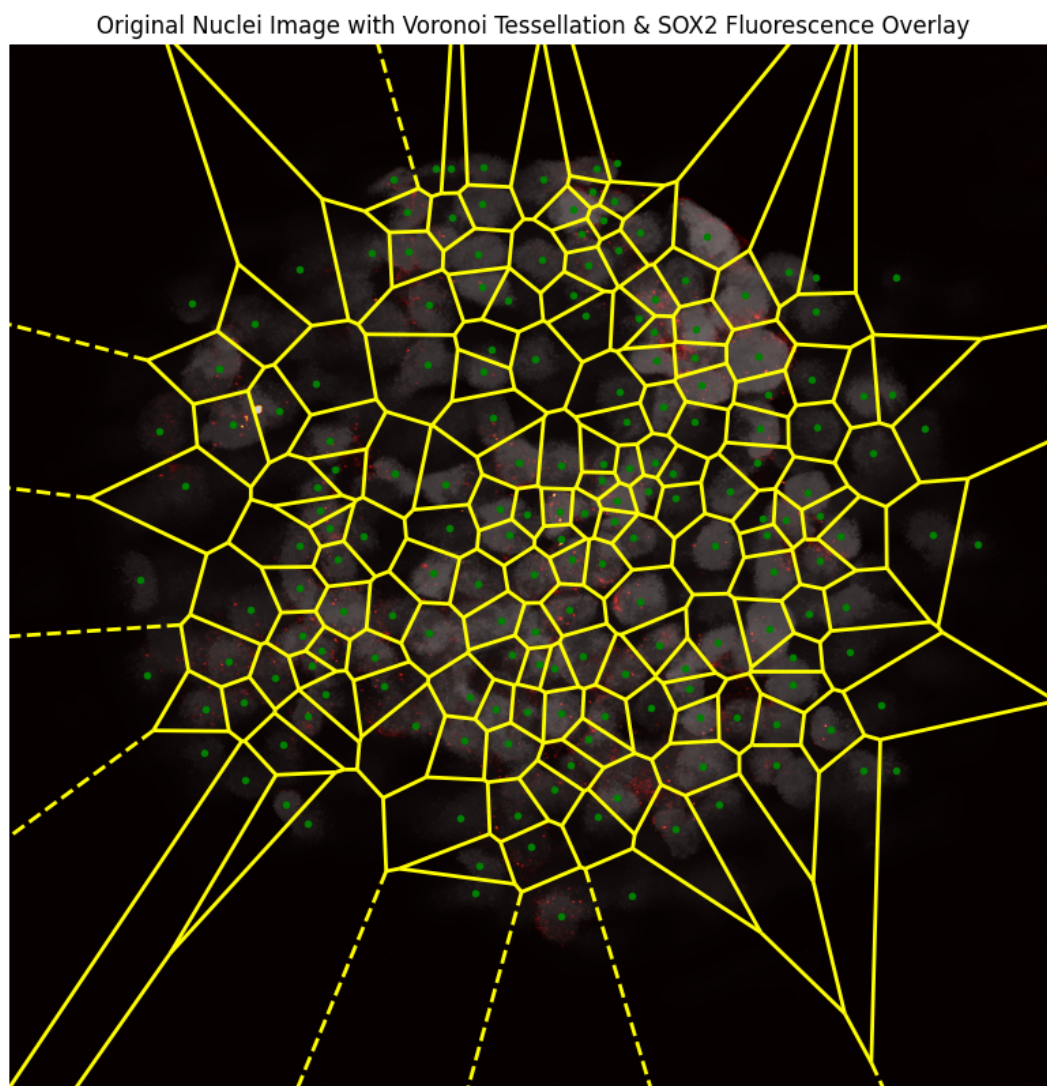


Figura 5: Teselación de Voronoi superpuesta a la imagen de los núcleos y a la fluorescencia de SOX2.

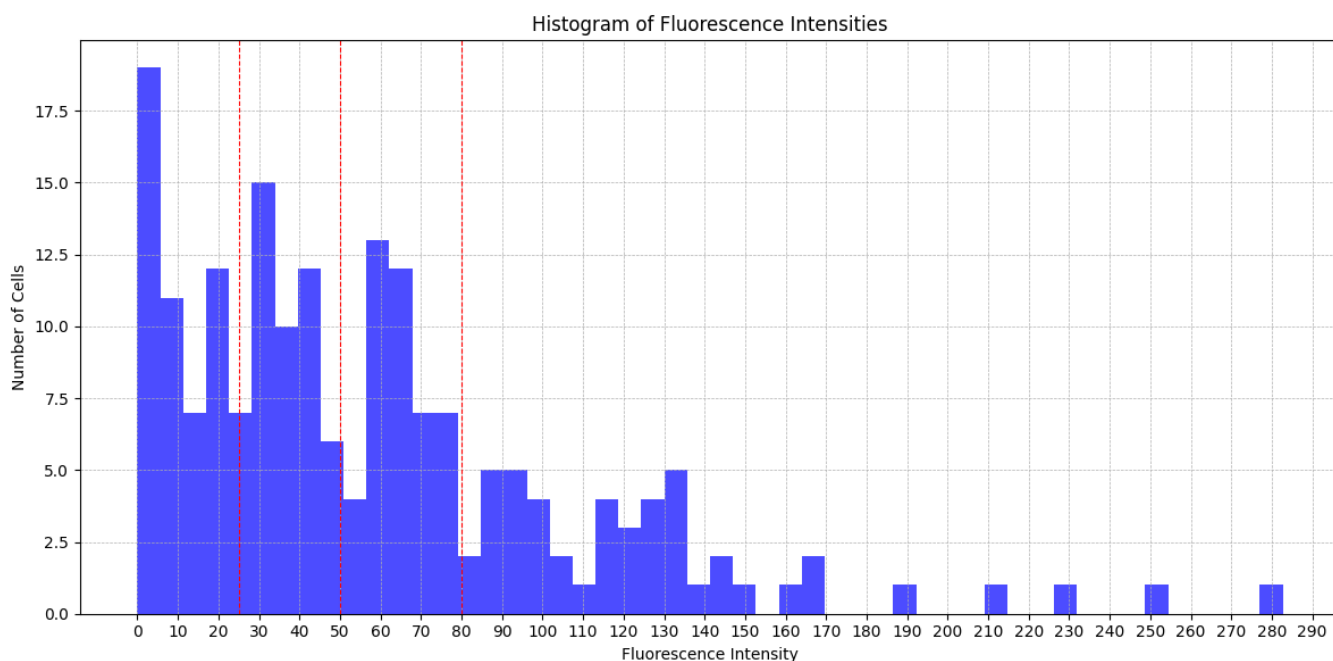


Figura 6: *Histograma de intensidades de fluorescencia de SOX2 por región.* Se marcan en rojo tres posibles umbrales, en las intensidades 25, 50 y 80, a partir de los cuales considerar que una célula es madre.

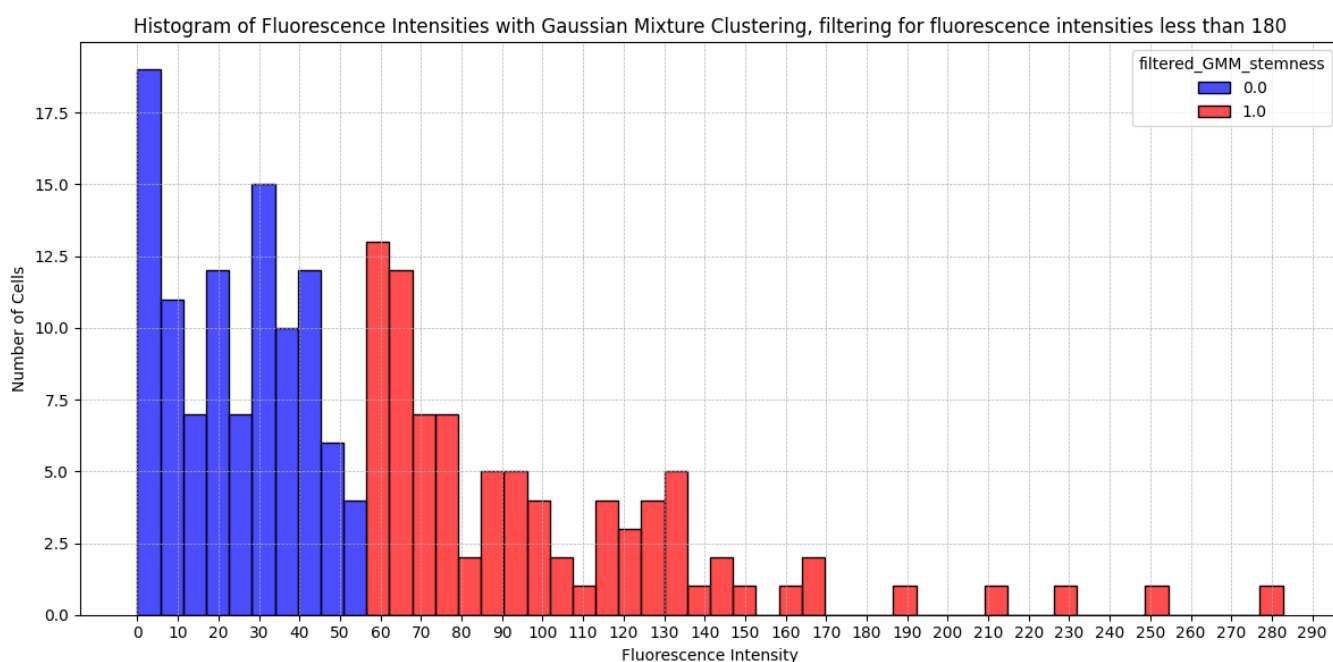


Figura 7: *Histograma de intensidades de fluorescencia de SOX2 por región, agrupadas considerando que son generadas por dos gaussianas.* Los 5 valores más altos se etiquetaron a mano. Para el resto, se usó el algoritmo de agrupamiento de mezcla de gaussianas (sin tener en cuenta los 5 valores mencionados), para formar dos grupos.