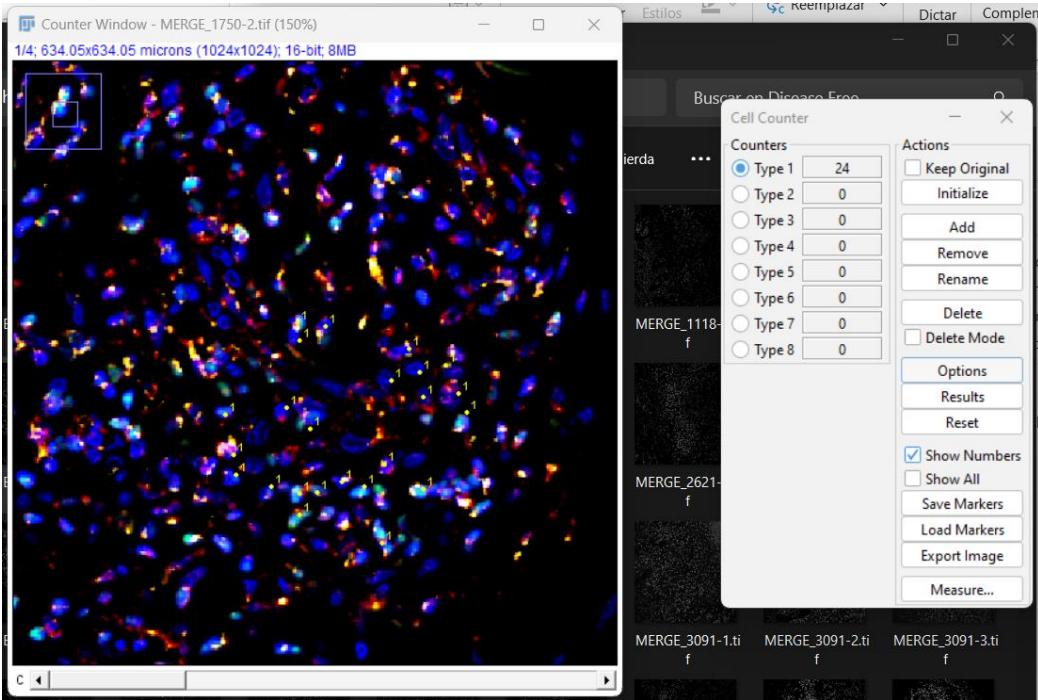


Aprendizaje Automático para Clasificación de Marcadores

M en C Saraí G. De León Rodríguez sara.g.deleonr@gmail.com

Origen del proyecto

¿Una forma de cuantificar de forma objetiva , rápida y reproducible?



Research Article

A Machine Learning Workflow of Multiplexed Immunofluorescence Images to Interrogate Activator and Tolerogenic Profiles of Conventional Type 1 Dendritic Cells Infiltrating Melanomas of Disease-Free and Metastatic Patients

Saraí G. De León Rodríguez ,^{1,2} Paúl Hernández Herrera ,³ Cristina Aguilar Flores ,⁴ Vadim Pérez Koldenkova ,⁵ Adán Guerrero ,³ Alejandra Mantilla ,⁶ Ezequiel M. Fuentes-Pananá,⁷ Christopher Wood ,³ and Laura C. Bonifaz ,^{1,8}

¹UMAE Hospital de Especialidades, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Unidad de Investigación Médica en Inmunología, Mexico City, Mexico

²Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Mexico City, Mexico

³Laboratorio Nacional de Microscopía Avanzada, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

⁴UMAE Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Unidad de Investigación Médica en Inmunología, Mexico City, Mexico

⁵Laboratorio Nacional de Microscopía Avanzada-IMSS, División de Desarrollo de La Investigación, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico

⁶Servicio de Patología, Hospital de Oncología Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico

⁷Unidad de Investigación en Virología y Cáncer, Hospital Infantil de México Federico Gómez, Mexico City, Mexico

⁸Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico



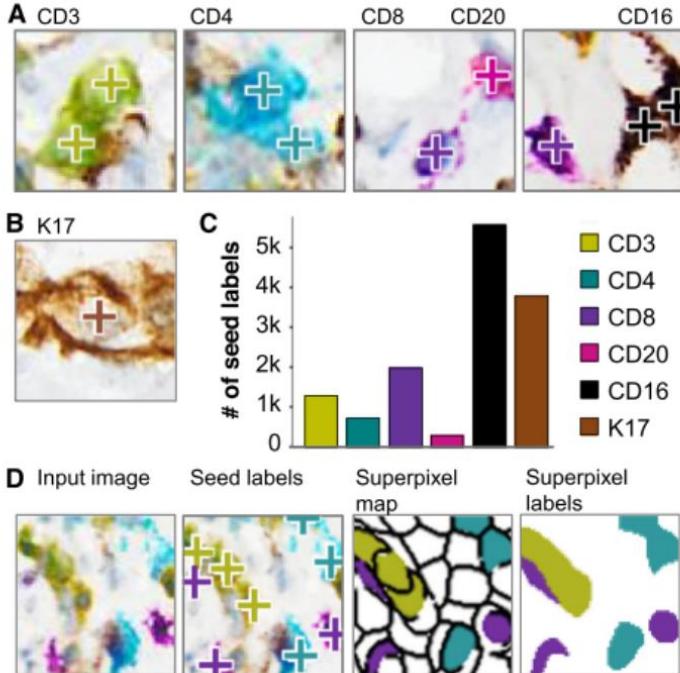
Clasificadores Machine-Learning

Clasificación: El proceso de predecir a que clase pertenece un objeto

Clase: objeto o categoría.

Ejemplo: Clasificar un correo como spam o no spam

Uso de clasificadores



Clasificadores para IHC

Clasificadores imagenología médica

Clasificadores patrones





¿Qué necesito para entrenar?

Habilidad para clasificar manualmente tinciones.

Conocer el problema a resolver

Datos de calidad (imágenes). Imágenes con marcas poco homogéneas son complicadas

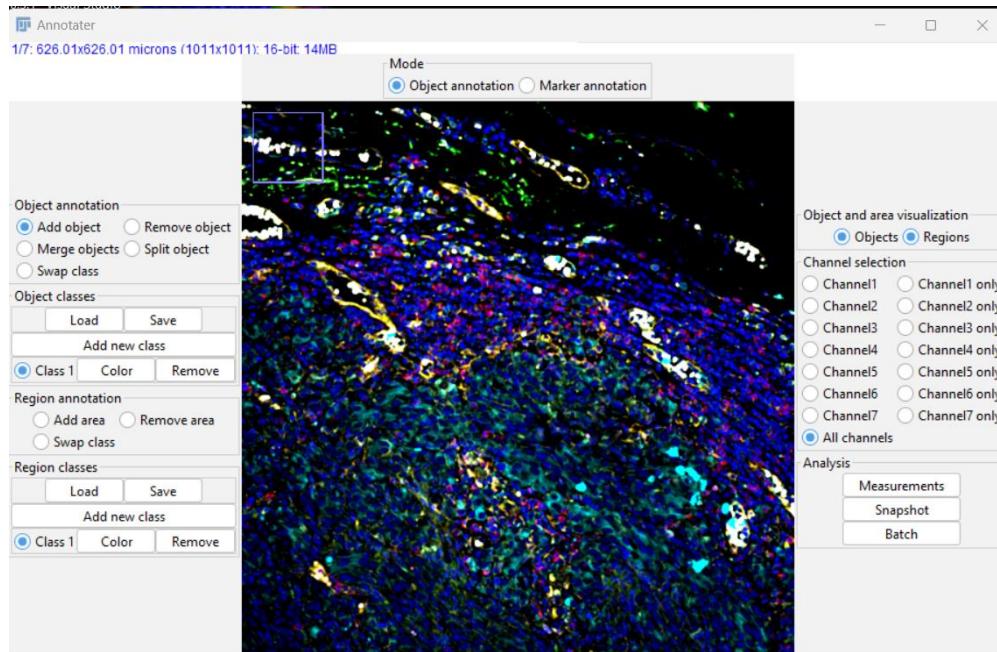
Actitud. Lidiar con la frustración

Annotater

Herramienta para anotación y entrenamiento de clasificadores

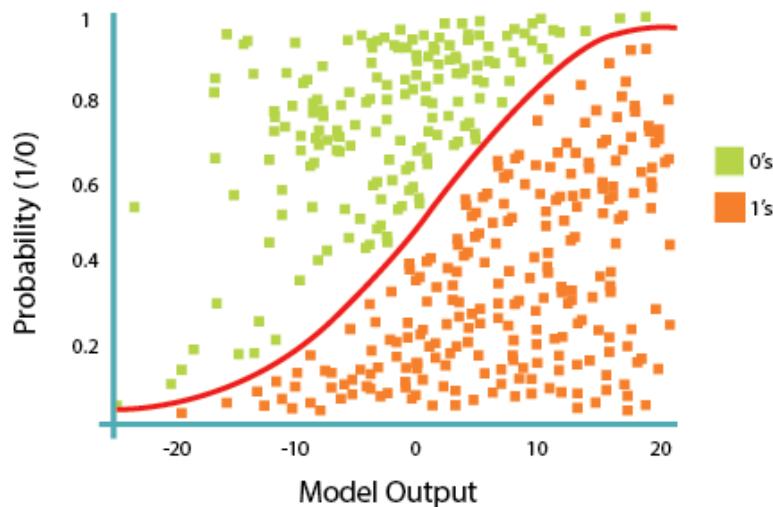


Thierry Pecot



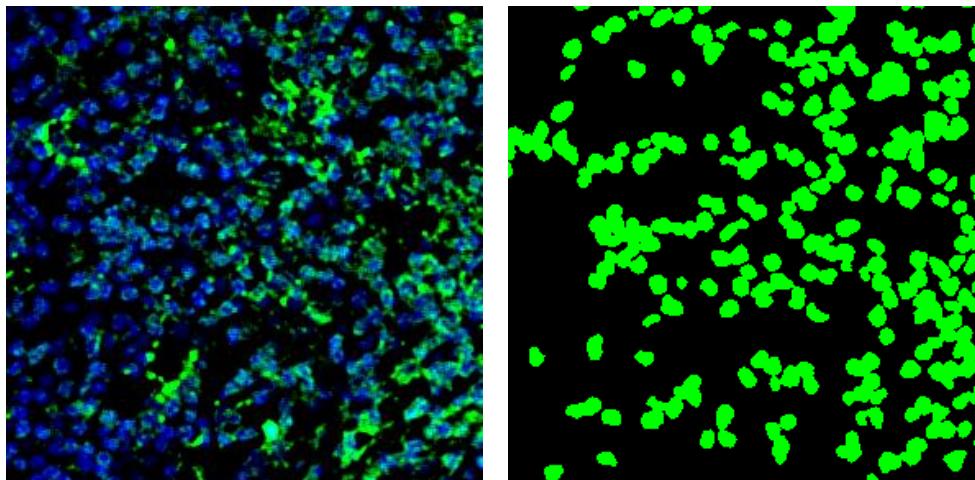
Regresión logística

La regresión logística es un algoritmo de clasificación que se utiliza para predecir la probabilidad de una variable dependiente categórica. En la regresión logística, la variable dependiente es una variable binaria que contiene datos codificados como 1 – 0, sí – no, abierto – cerrado, etc.



¿Qué características aprende?

- Intensidad de fluorescencia promedio
- Desviación estandar de la intensidad.
- Circularidad de núcleos.
- Tamaño





Ventajas sobre análisis manual

Eficiencia: menos tiempo de análisis

Objetividad

Reproducibilidad

Información nueva

Análisis de expresión específico

Conocer coordenadas



Cronograma

SESION1

- Explicación del experimento
- Alineación de imágenes
- Segmentación de núcleos
- Creando clasificadores

SESION 2

- Resolución dudas.
- Evaluación de clasificadores
- Extracción de datos
- Análisis downstream. ¿más de 7 canales?



Explicación del experimento



International Journal of
Molecular Sciences



Article

Acral Melanoma Is Infiltrated with cDC1s and Functional Exhausted CD8 T Cells Similar to the Cutaneous Melanoma of Sun-Exposed Skin

Saraí G. De Leon-Rodríguez ^{1,2,†}, Cristina Aguilar-Flores ^{3,†}, Julián A. Gajón ^{1,4}, Alejandra Mantilla ⁵, Raquel Gerson-Cwilich ⁶, José Fabián Martínez-Herrera ^{6,7}, Benigno E. Rodríguez-Soto ⁸, Claudia T. Gutiérrez-Quiroz ⁹, Vadim Pérez-Koldenkova ¹⁰, Samira Muñoz-Cruz ¹, Laura C. Bonifaz ^{1,11,*} and Ezequiel M. Fuentes-Pananá ^{12,*}

¹ UMAE Hospital de Especialidades, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Unidad de Investigación Médica en Inmunoquímica, Mexico City 06720, Mexico

² Posgrado en Ciencias Biológicas, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico

³ UMAE Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City 06720, Mexico

⁴ Posgrado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico

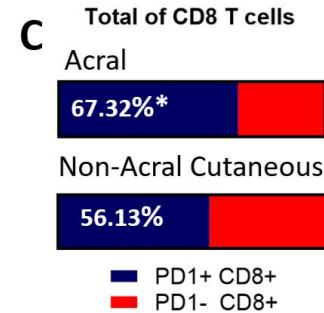
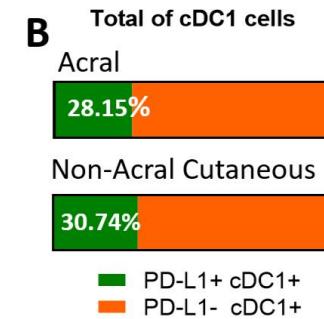
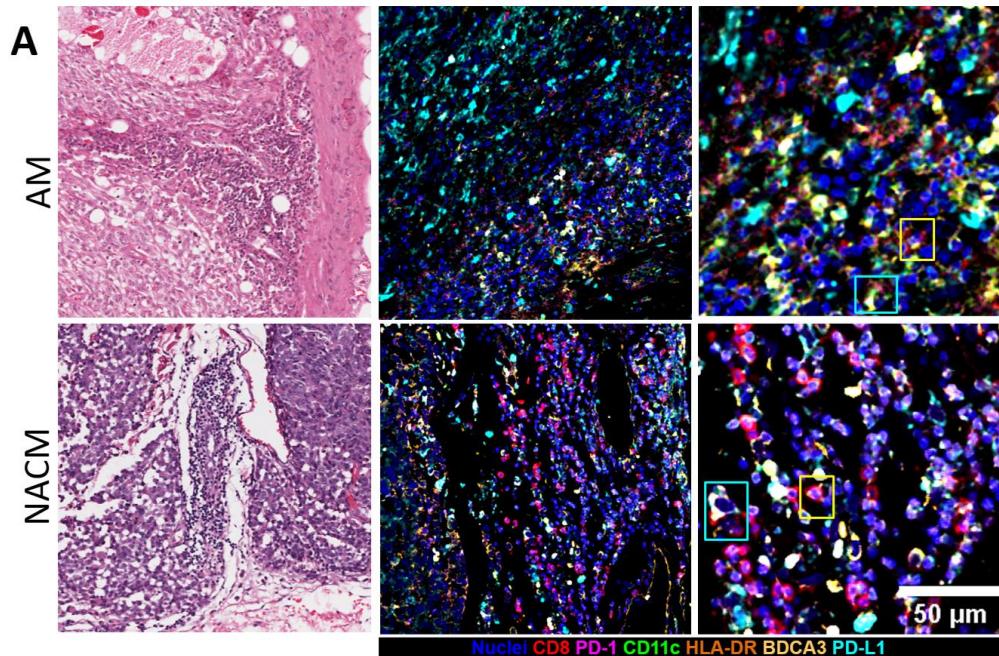
⁵ Servicio de Patología, Hospital de Oncología Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City 06720, Mexico

⁶ Cancer Center, Medical Center American British Cowdray, Mexico City 01120, Mexico

⁷ Latin American Network for Cancer Research (LAN-CANCER), Lima 11702, Peru

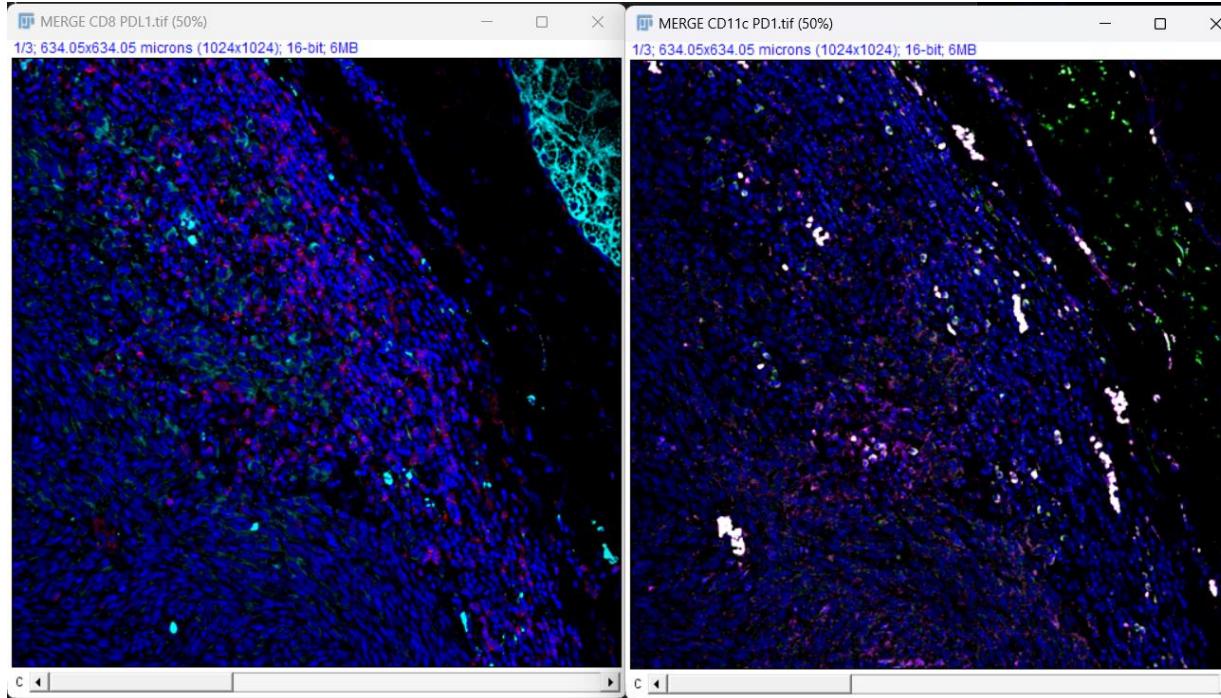


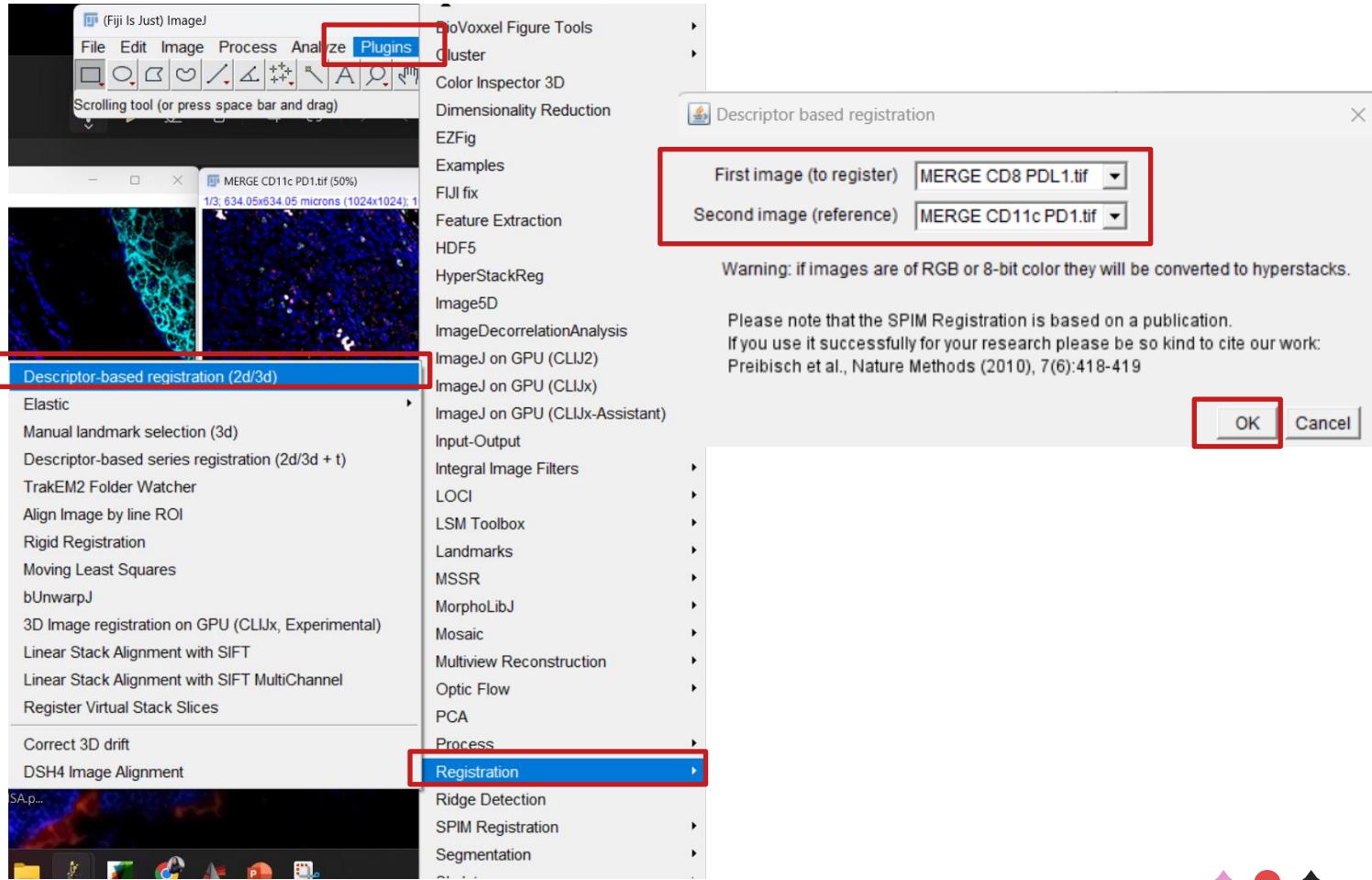
Evaluación de fenotipos

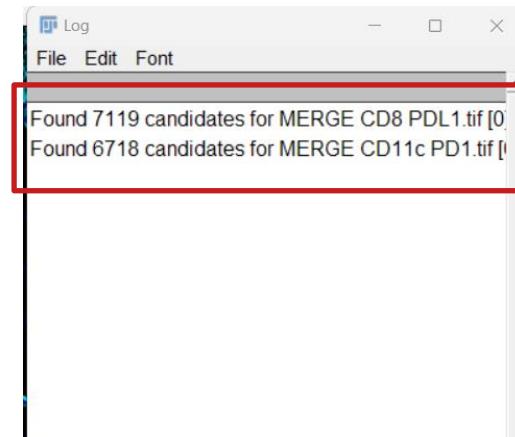
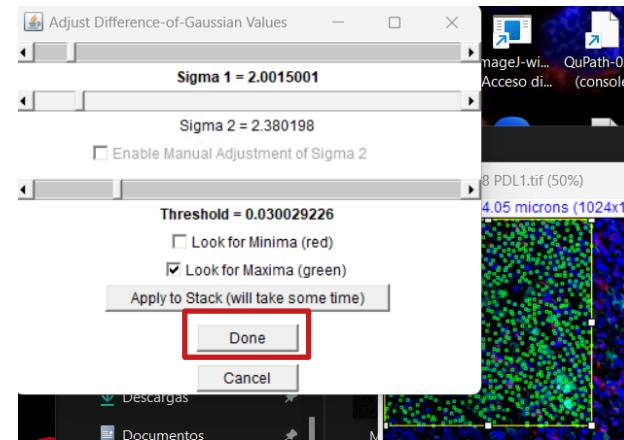
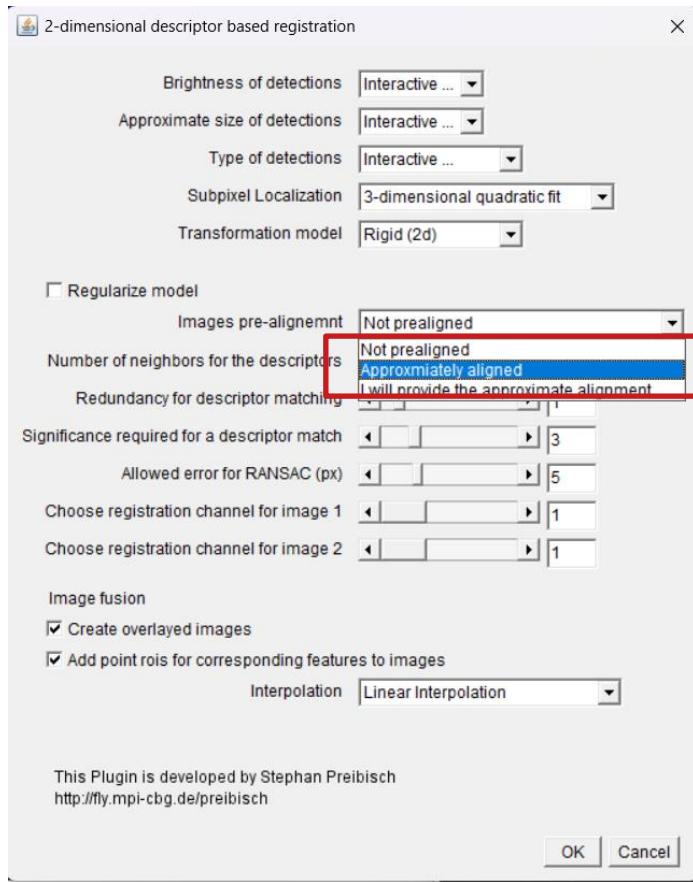


Alineación de imágenes

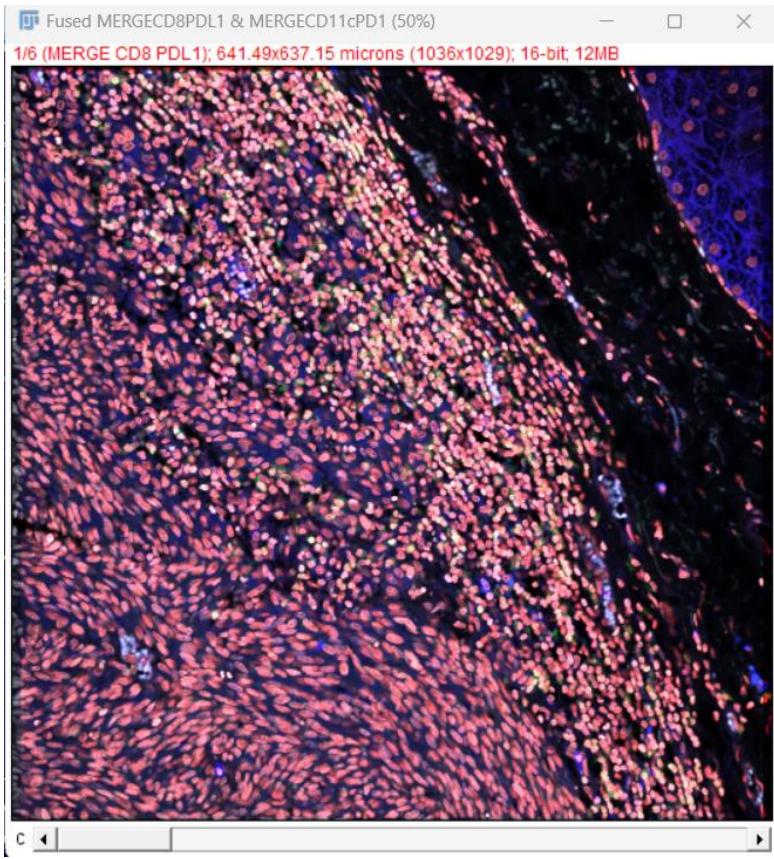
Abrir imágenes a alinear



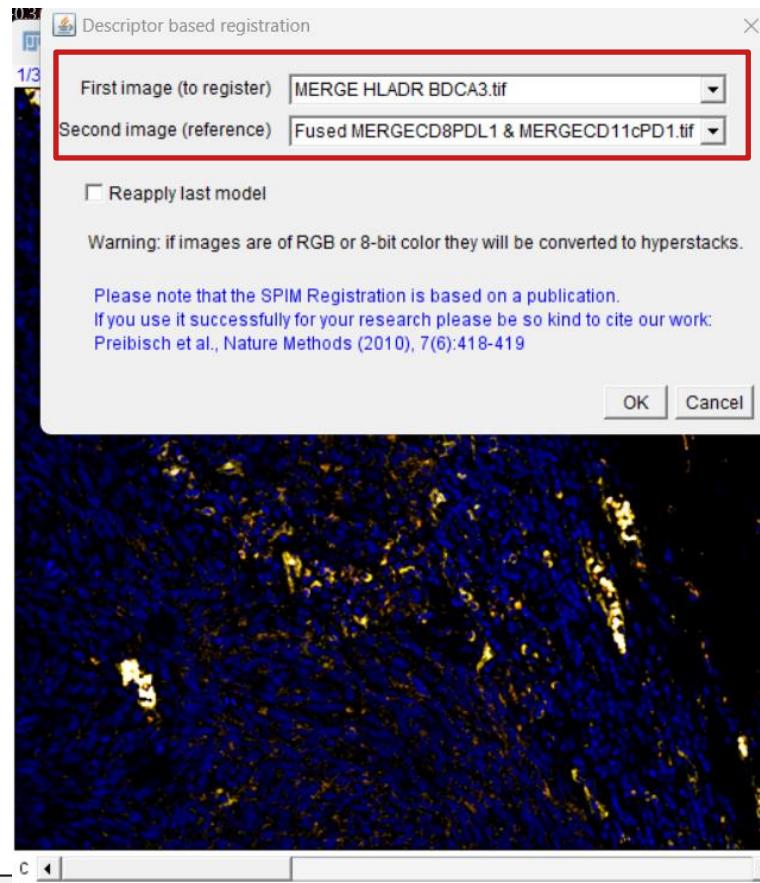




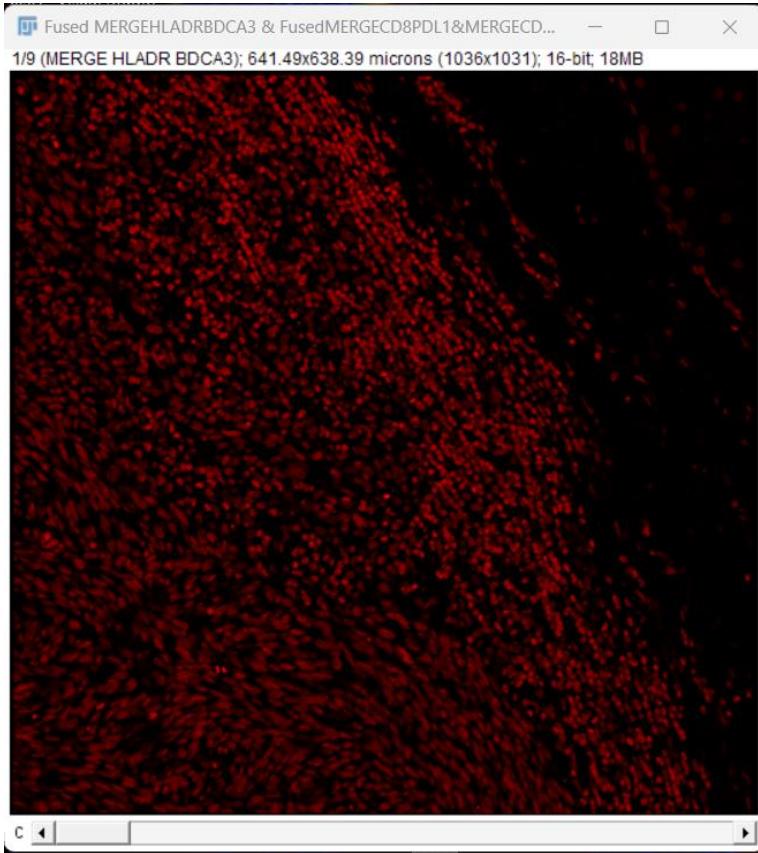
Se generara una imagen alineada



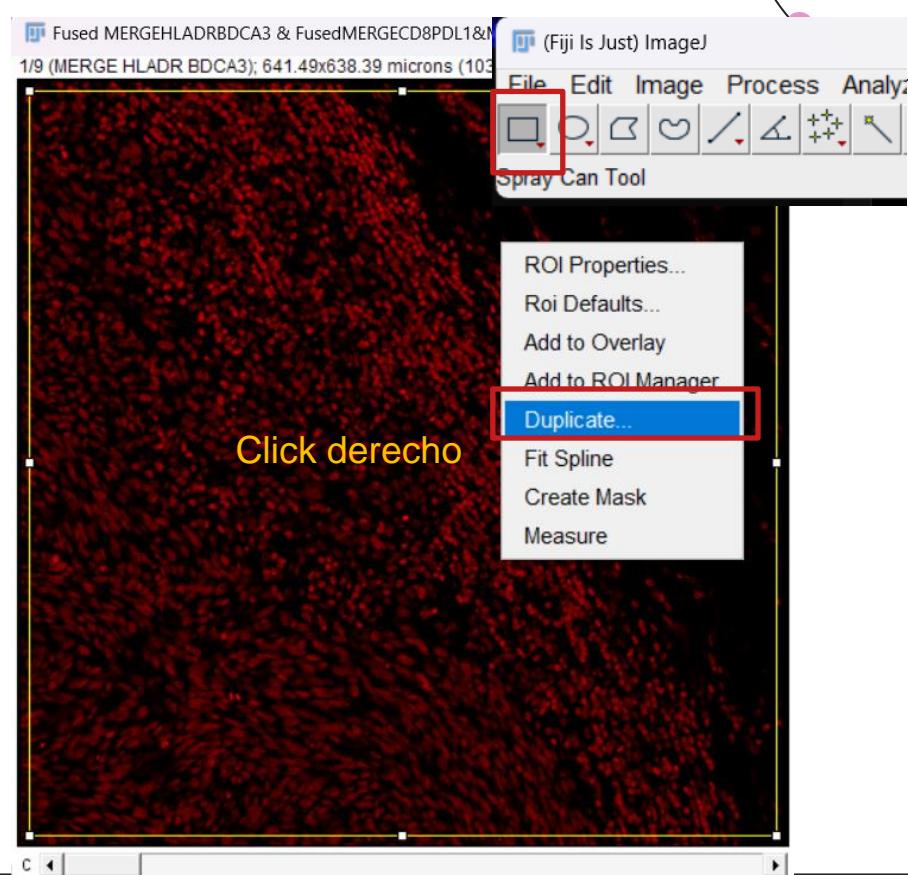
Abrir la tercera imagen a alinear y repetir el proceso



Stack de 9 canales



Antes de separarlo recortar al tamaño adecuado



Organización de los canales

Es importante determinar el orden y color que cada marcador tendrá de la composición, esta asignación debe ser la misma para todas las imágenes

Canal 1-Núcleos azul

- Green
- Red
- Yellow
- Cyan
- Magenta
- Yellow hot
- Sepia
- Grays
- Orange hot
- Magenta hot

Canal 2-CD8 rojo

Canal 3-PD1 magenta

Canal 4-CD11c verde

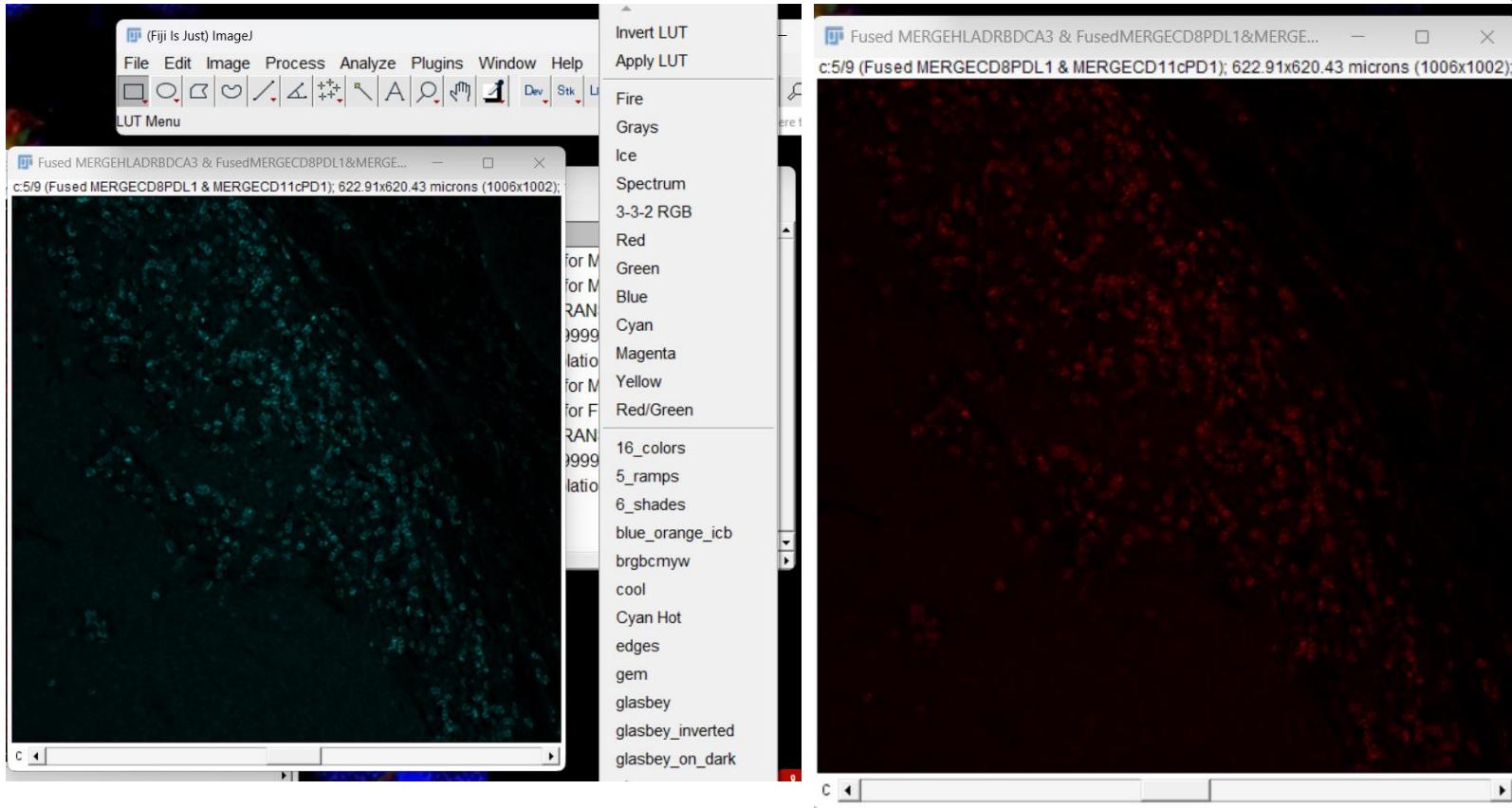
Canal 5-HLADR Orange hot

Canal 6 BDCA3 Yellow hot

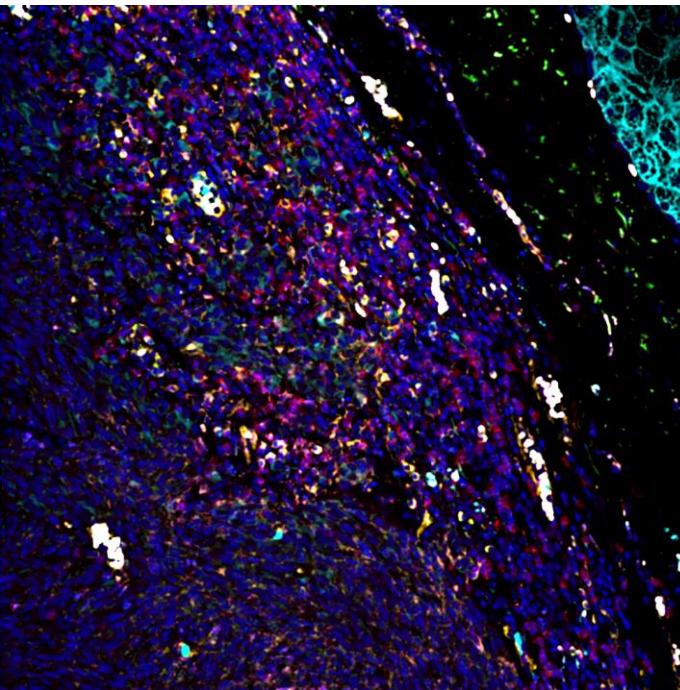
Canal 7-PDL1 Cyan



Asignar lut a cada canal



Separar los canales y hacer la composición final

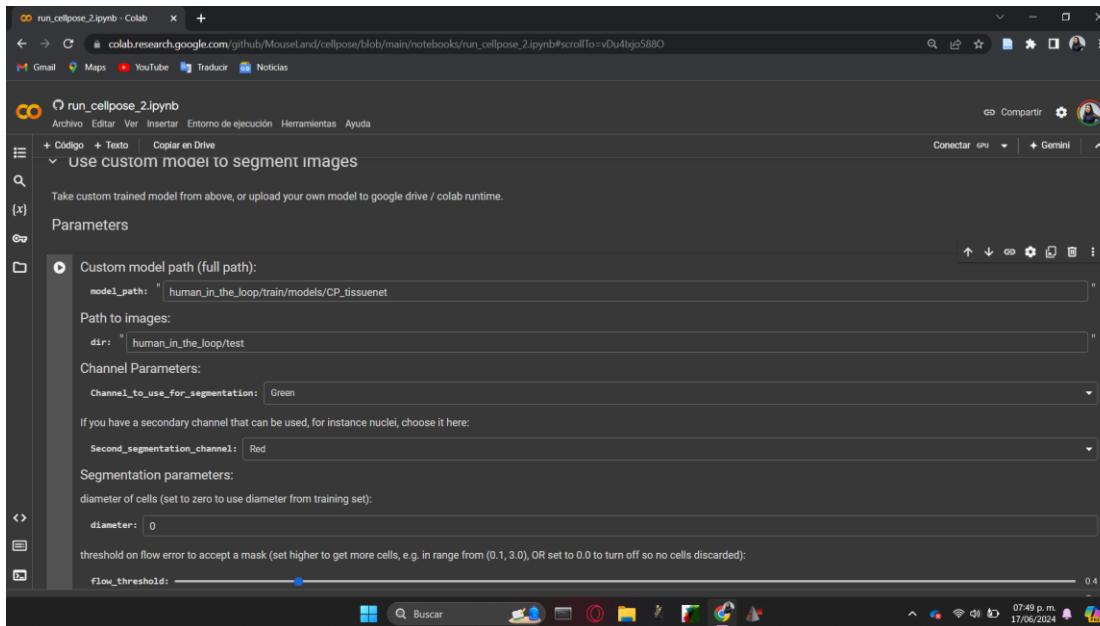


Segmentación de núcleos

Google collab

Separar los núcleos y subirlos a una carpeta de drive

Google Colaboratory



Entrenamiento de clasificadores

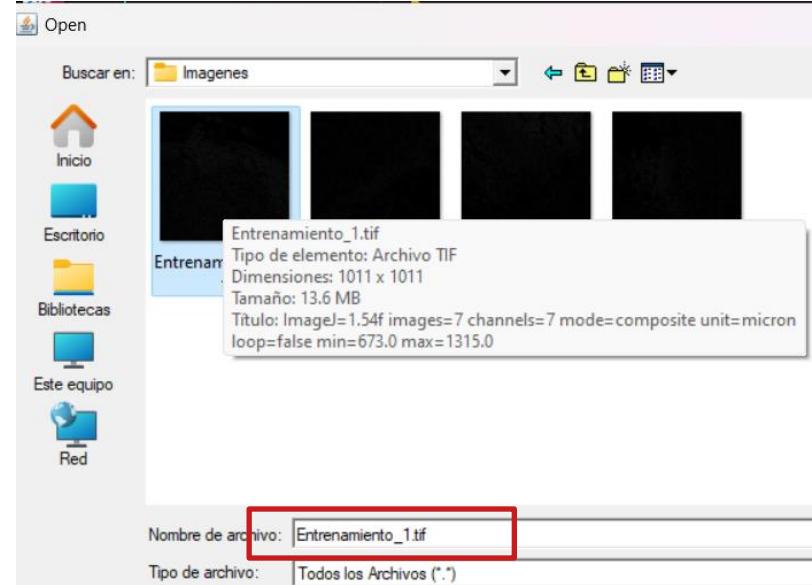
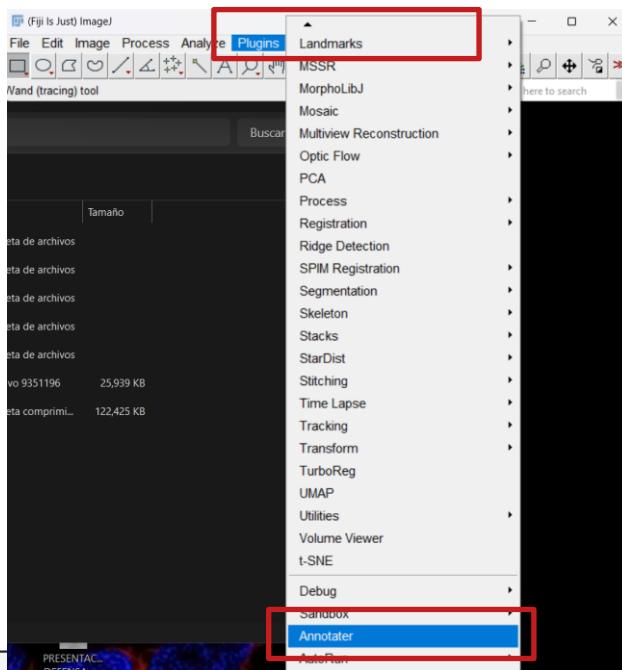
De nuestro experimento tenemos dos conjuntos de imágenes

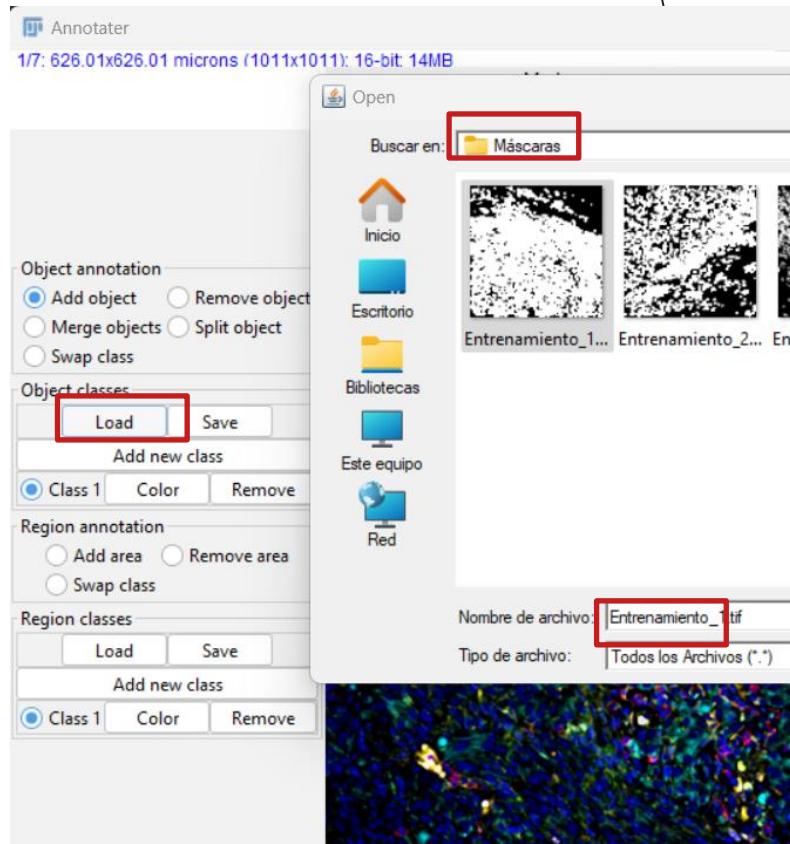
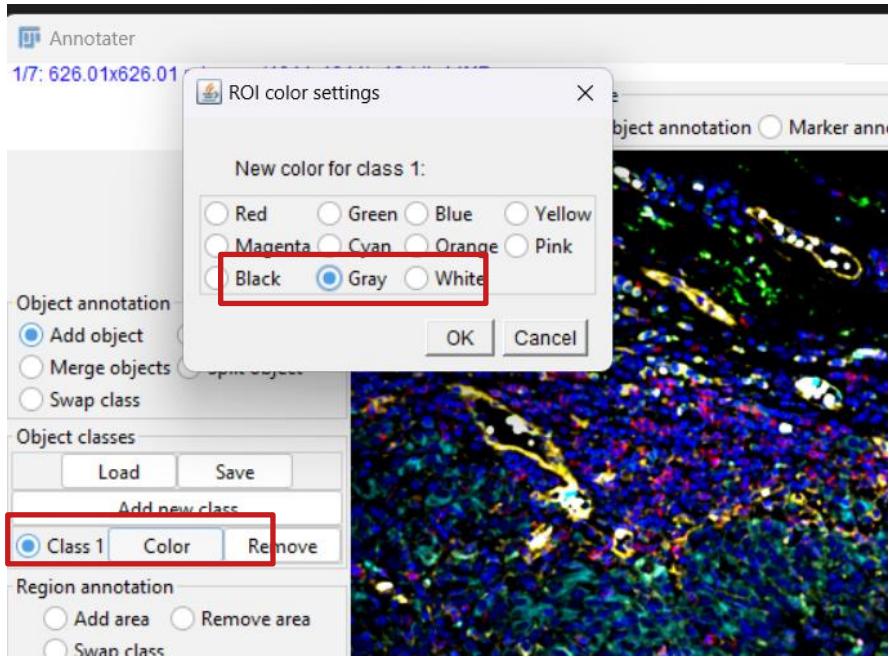
1. Datos para entrenamiento. Este dividirse en aprox 80% para entrenamiento y 20% validad
2. Datos problema/ Datos a evaluar. Este conjunto no debe incluir el primer conjunto idealmente.

Entrenamiento de clasificadores

Abrir imagen de entrenamiento en annotater

Empalmar con su máscara de núcleos segmentados





Annotater

1/7: 626.01x626.01 microns (1011x1011): 16-bit: 14MB

Mode

Object annotation

Marker annotation

Object annotation

- Add object
- Remove object
- Merge objects
- Split object
- Swap class

Object classes

Add new class

- Class 1
-
-

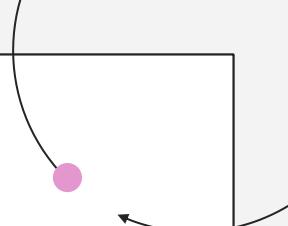
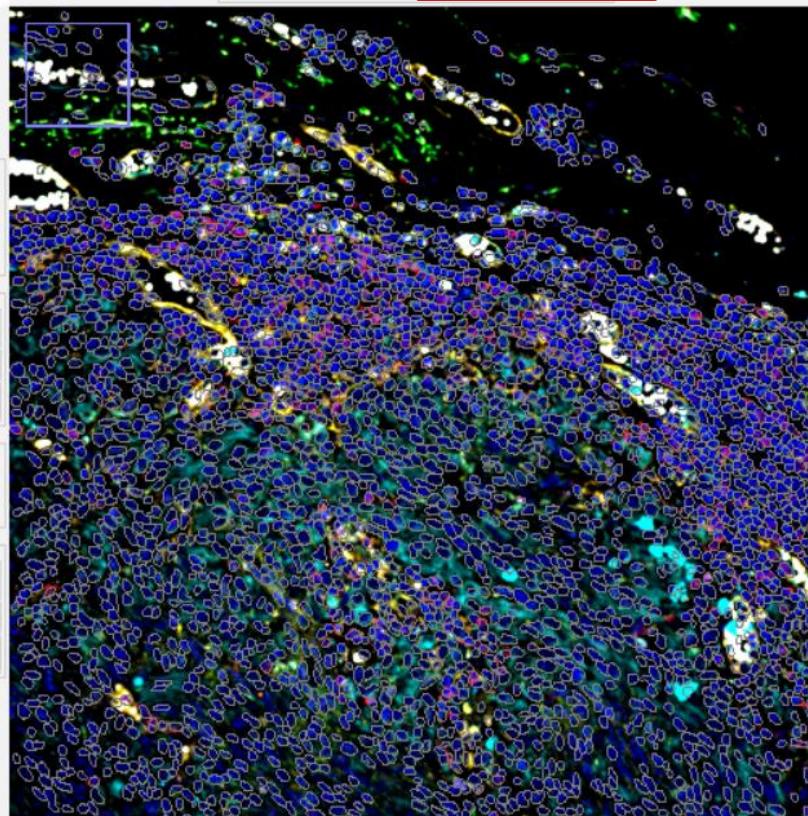
Region annotation

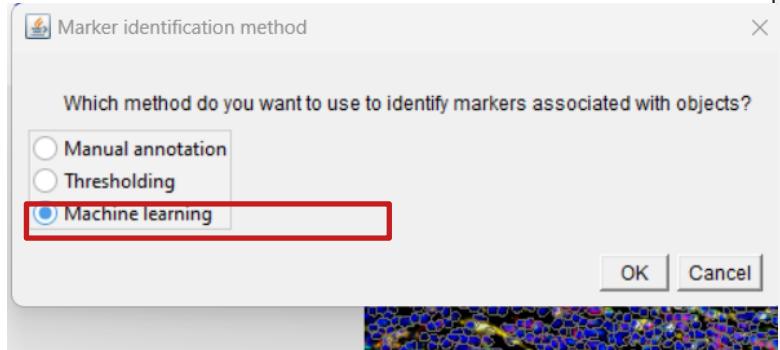
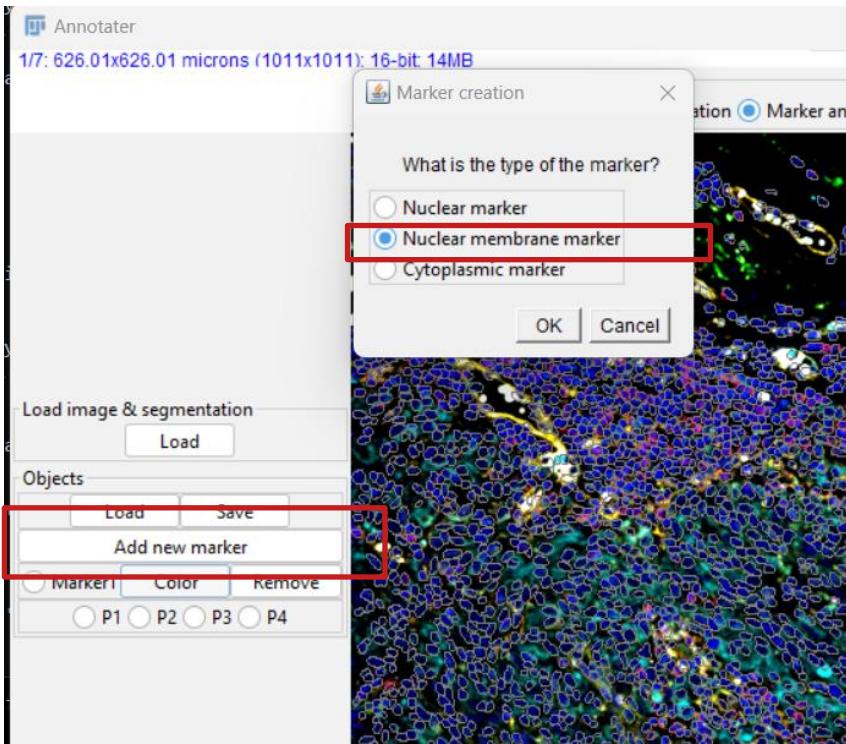
- Add area
- Remove area
- Swap class

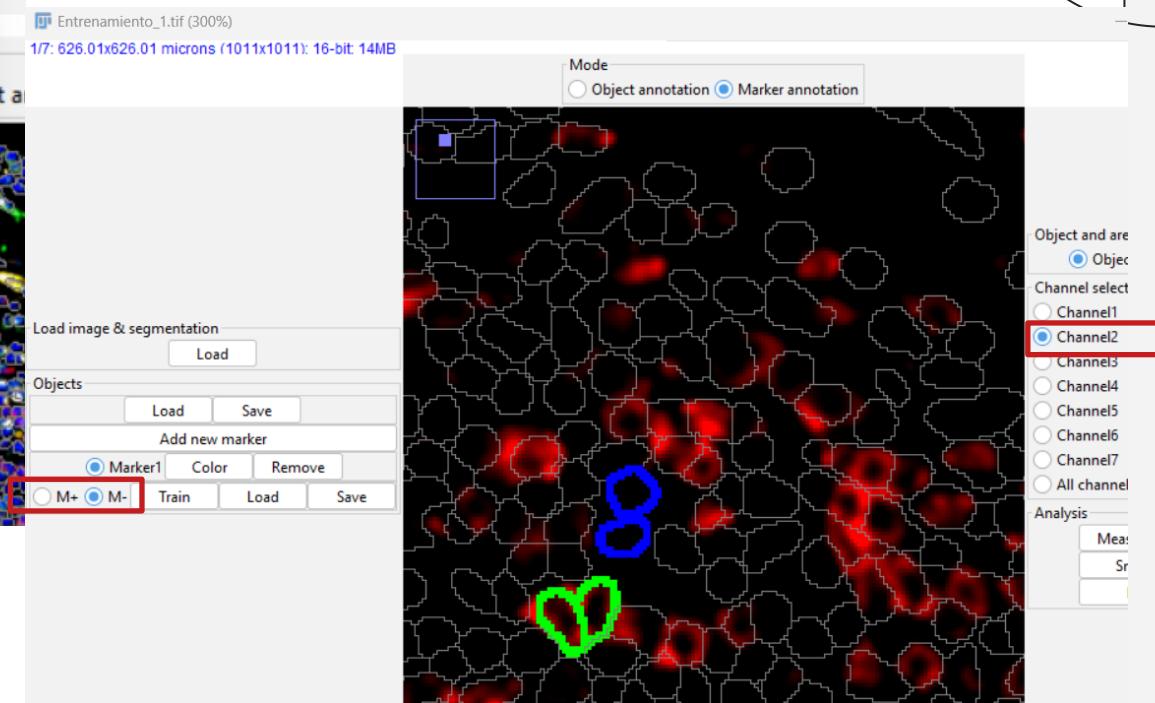
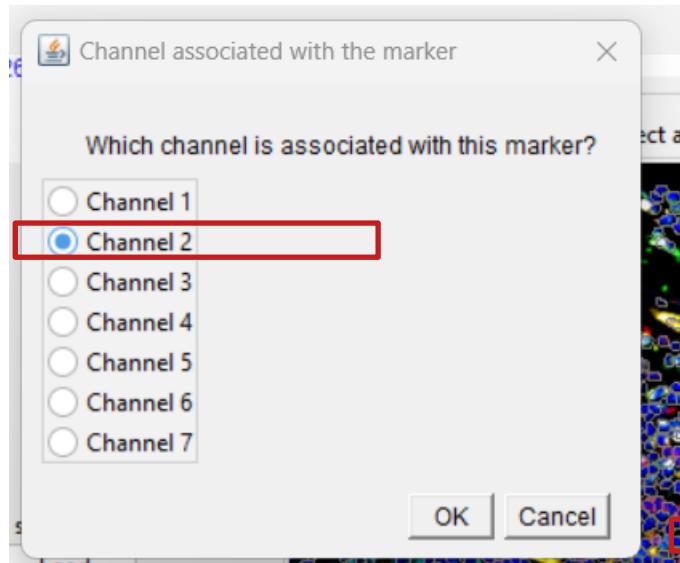
Region classes

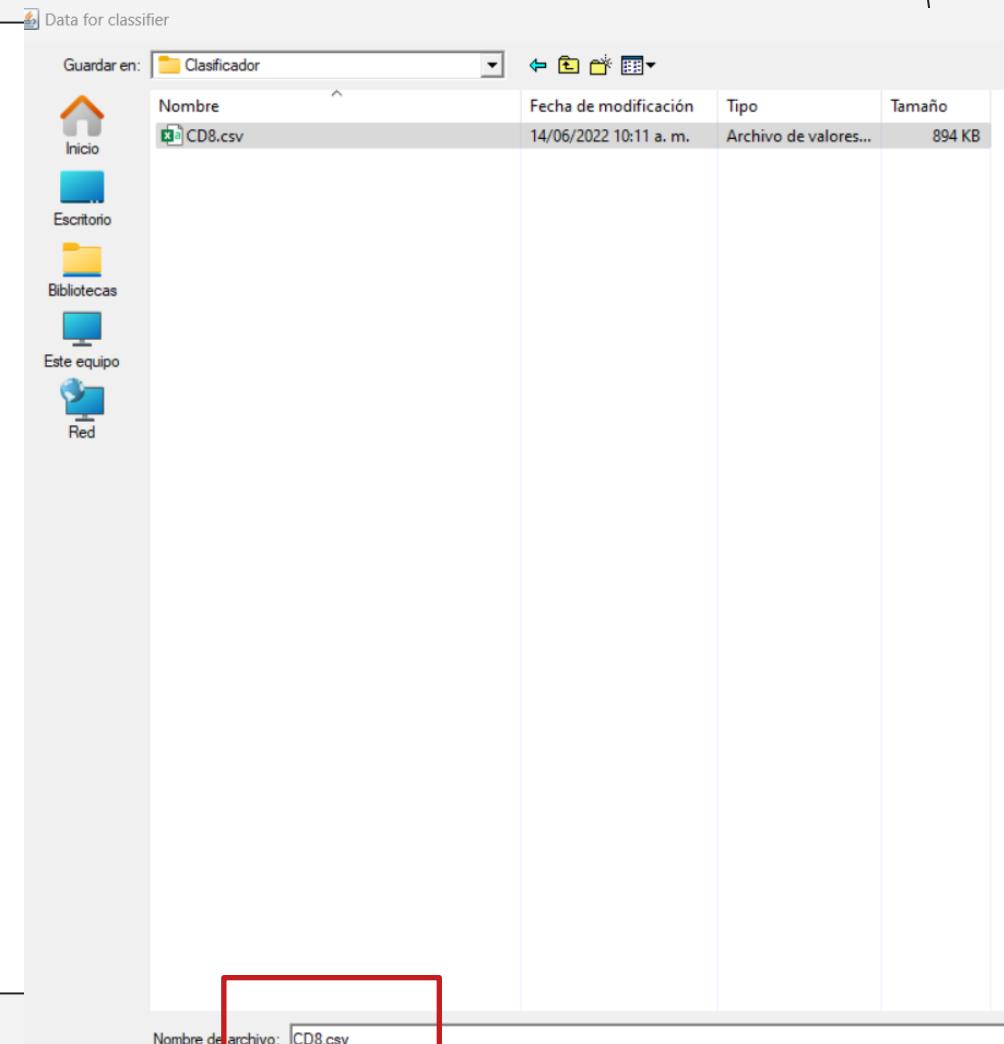
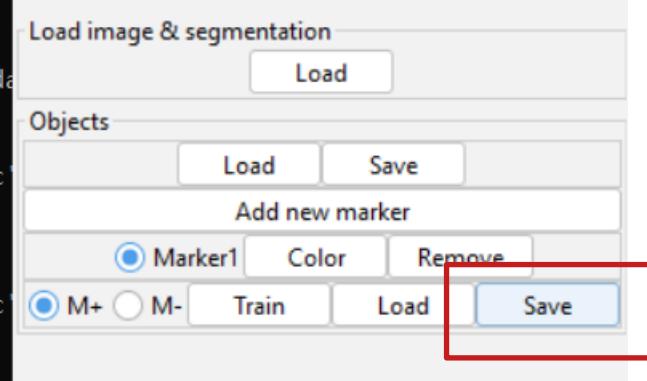
Add new class

- Class 1
-
-



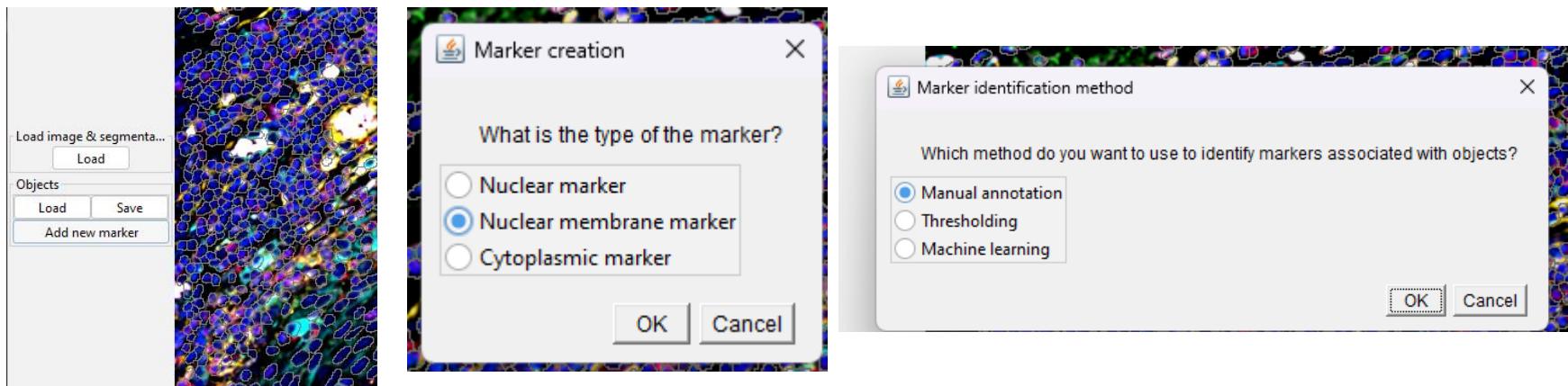






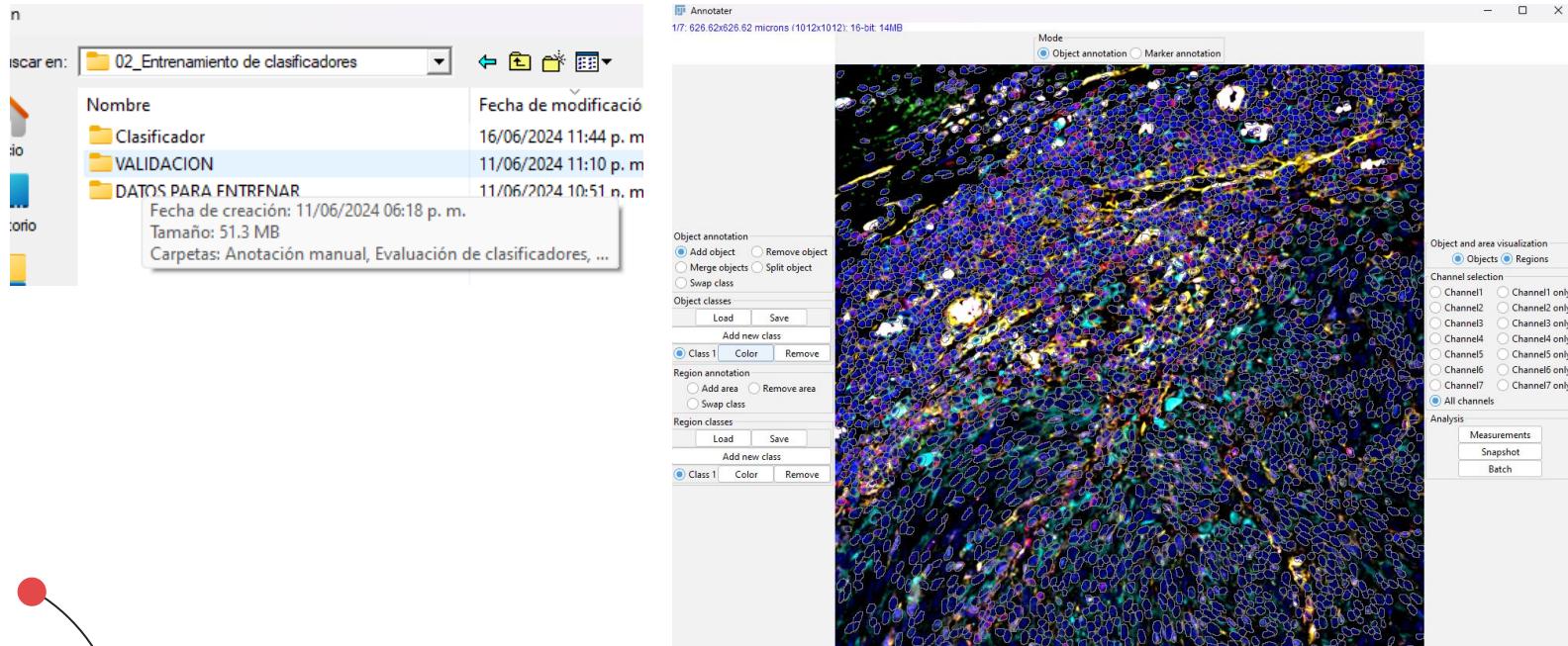
Evaluación de mi clasificador

2. En área de marker annotation añadir marcador en modo manual



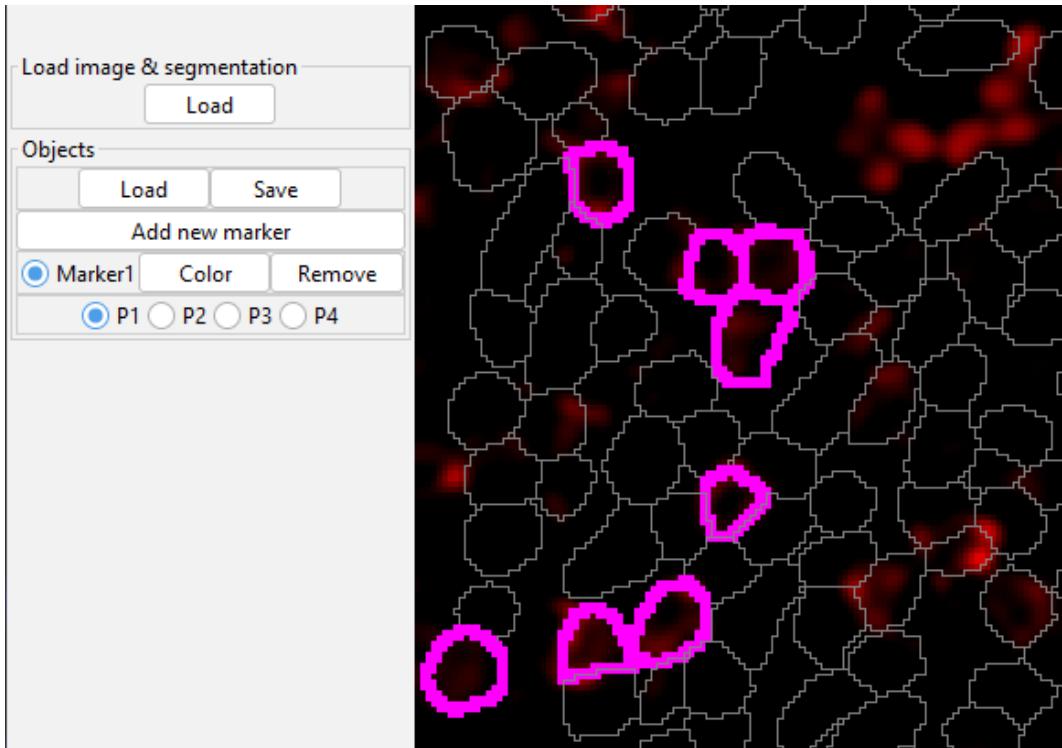
Evaluación de mi clasificador

1. Abrir imagen de validación en annotater y empalmar con máscara



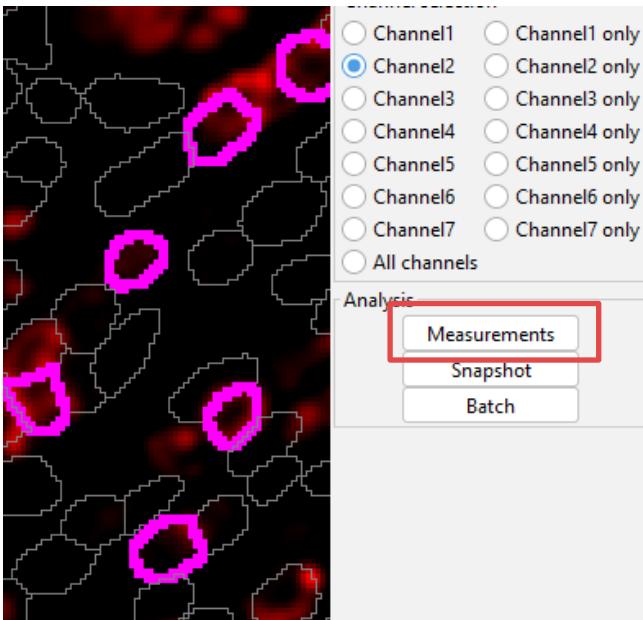
Evaluación de mi clasificador

3. Anotar manualmente TODA la imagen con las células positivas



Evaluación de mi clasificador

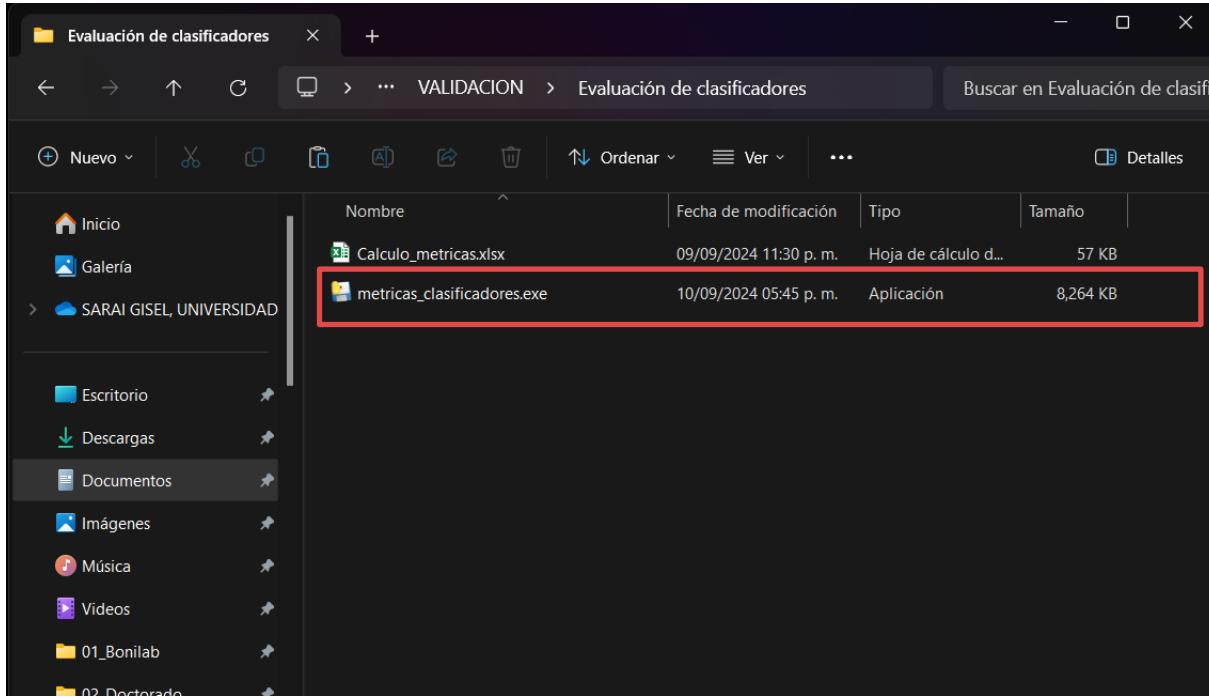
3. Guardar hoja .csv en imagen anotada y .csv de predicción con clasificador



1. Anotación manual.csv
2. Predicción clasificador

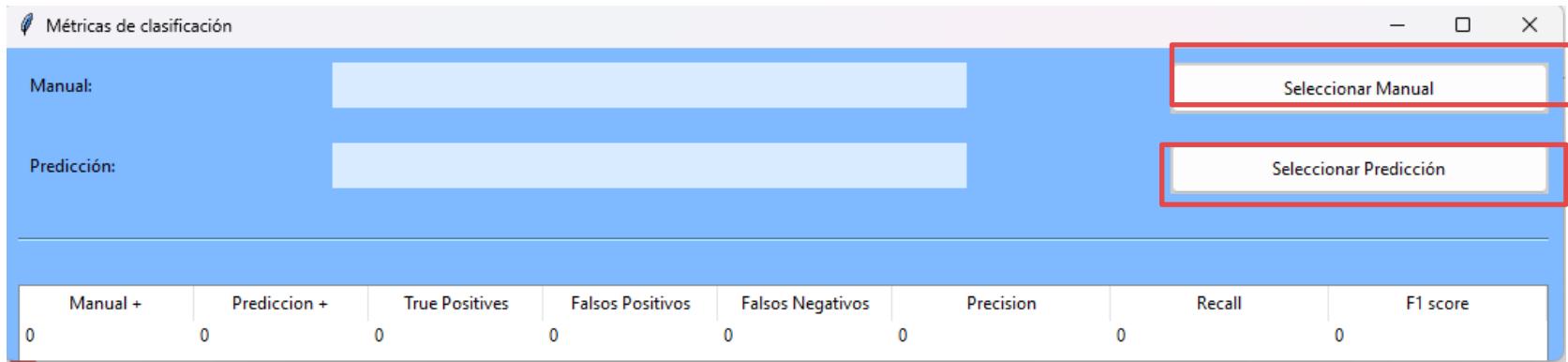
Evaluación de mi clasificador

4. Abrir interfaz para sacar métricas



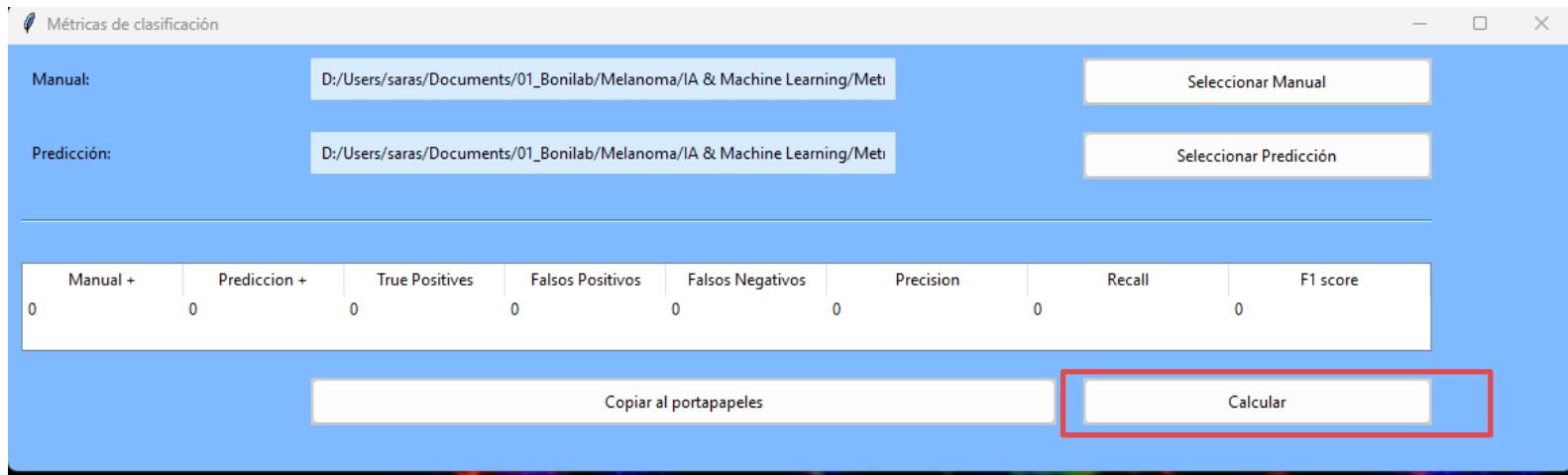
Evaluación de mi clasificador

5. Seleccionar las hojas de anotación manual y predicción



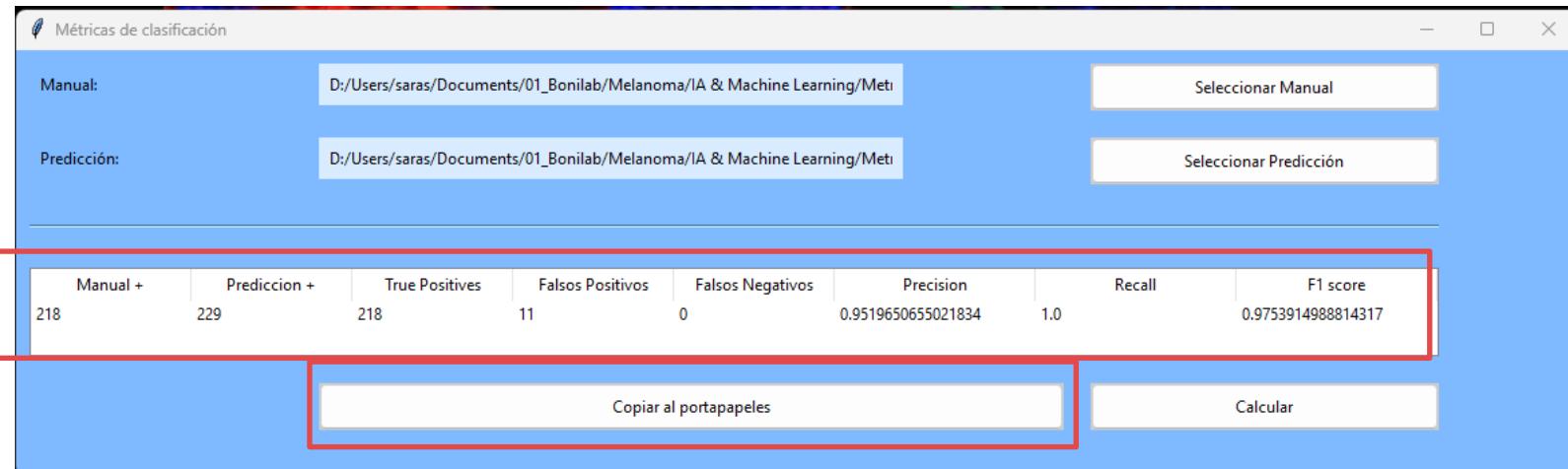
Evaluación de mi clasificador

6. Dar click en “calcular”



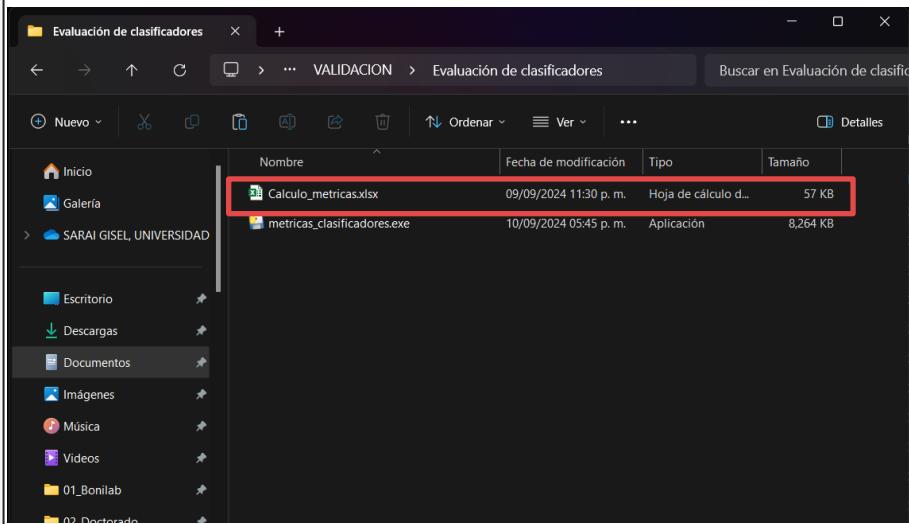
Evaluación de mi clasificador

6. Dar click en “copiar al portapapeles”



Evaluación de mi clasificador

7. Pegar en la hoja de cálculo "Calculo_Metricas"



Columna1	False positives	False negatives	True Positives	Precision	Recall	F1 score
Imagen 1	30	66	316	0.9132948	0.82722513	0.8681319
Imagen 2	33	10	948	0.96636086	0.98956159	0.9778236
Imagen 3	42	1	461	0.91650099	0.9978355	0.9554404

Precis: $\frac{TP}{TP + FN}$

$$F1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Recall: $\frac{TP}{TP + FN}$

Cálculo de métricas

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{Average Precision} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}},$$

$$F1 = 2 \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

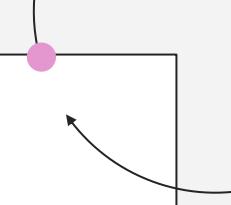
$$TP = PREDICTED - FP$$

$$FP = PREDICTED - MANUAL$$

$$FN = MANUAL - PREDICTED$$



Metricas



Cuando cada uno de los clasificadores a utilizar cumpla con buenas métricas podemos pasar al análisis del todo el segmento de datos

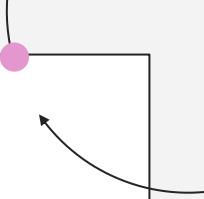
Organización de los datos

Nombre de los archivos

Acral_clavedepx-#decampo.tif

Análisis en batch

📁 CLASIFICADORES	17/06/2024 09:44 p. m.	Carpeta de archivos
📁 EXCEL RESULTS	10/06/2024 03:51 p. m.	Carpeta de archivos
📁 IMAGENES	17/06/2024 08:30 p. m.	Carpeta de archivos
📁 MARCADORES	10/06/2024 03:51 p. m.	Carpeta de archivos
📁 NUCLEOS	10/06/2024 08:13 p. m.	Carpeta de archivos
📁 NUCLEOS SEGMENTADOS	17/06/2024 08:33 p. m.	Carpeta de archivos



Resultados

Archivo .txt por cada imagen analizada

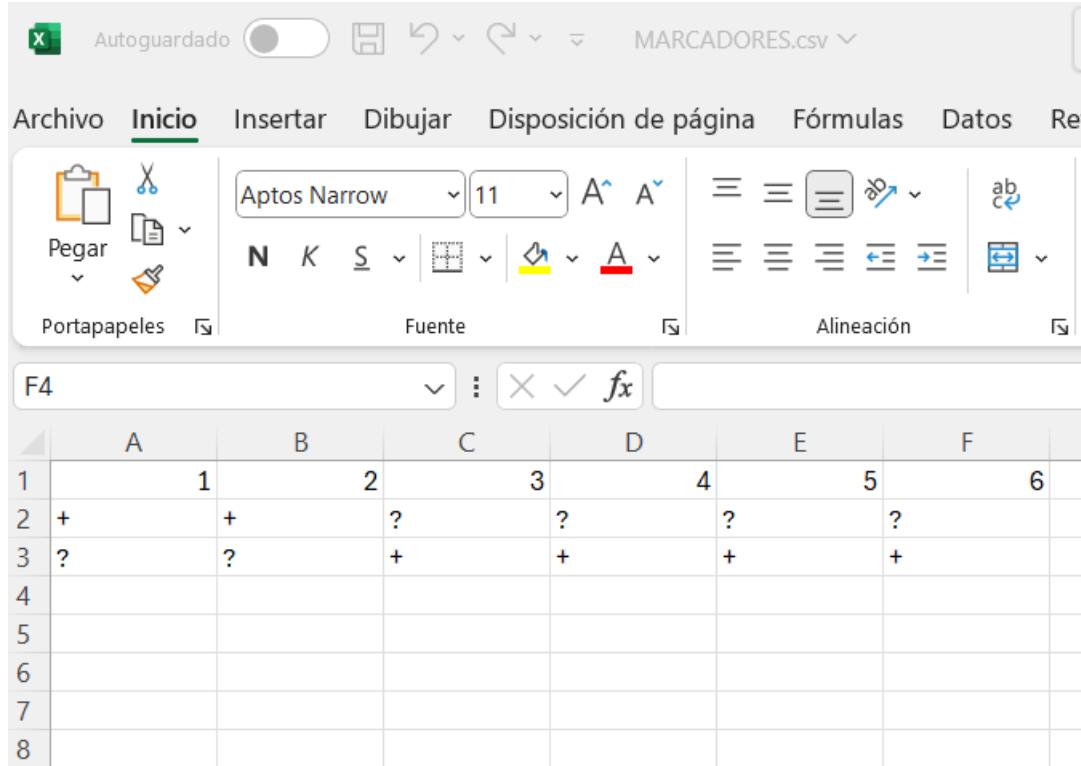
1 fila = 1 célula 1 columna = 1 característica

Lenguaje

1,2,3,4,5

+ - ?

1p2p3p

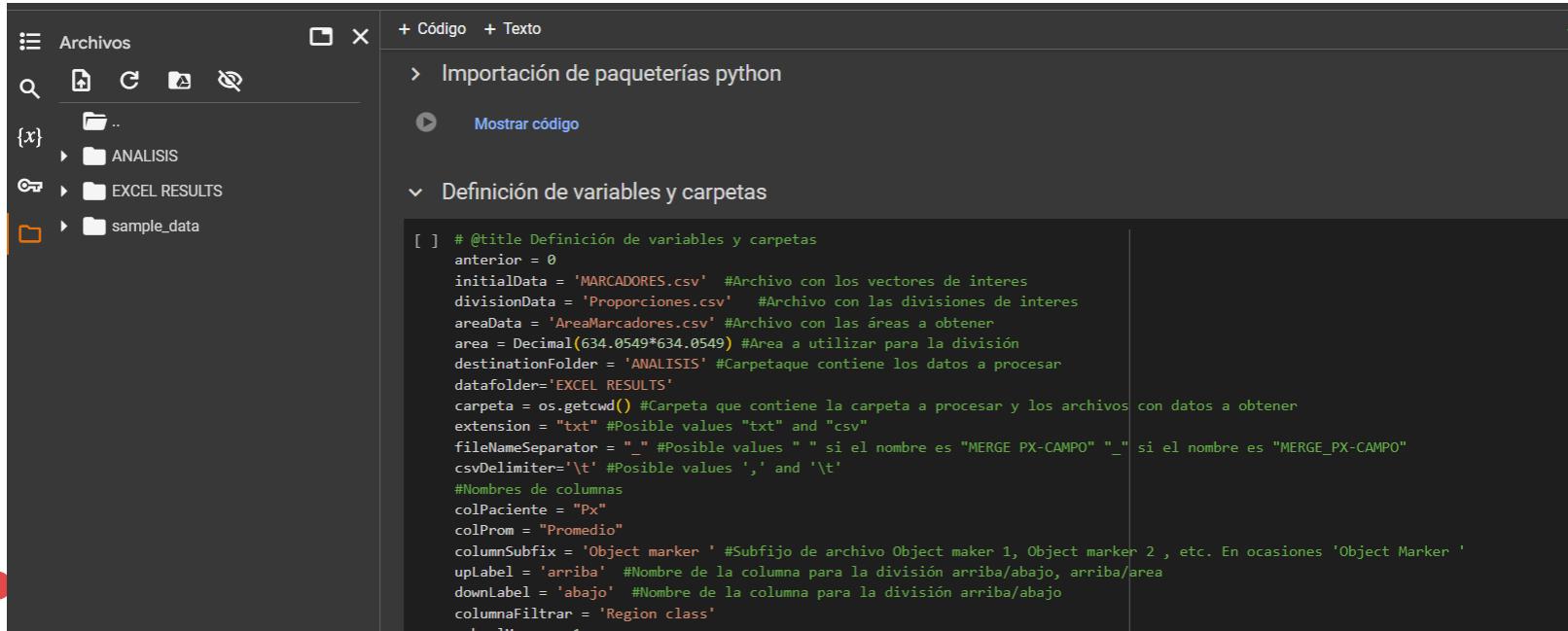


The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	
1	1	2	3	4	5	6	
2	+	+	?	?	?	?	
3	?	?	+	+	+	+	
4							
5							
6							
7							
8							

Extracción de datos

Abrir liga



The screenshot shows a Jupyter Notebook interface with a dark theme. On the left is a file browser pane titled 'Archivos' showing a directory structure with folders like ANALYSIS, EXCEL RESULTS, and sample_data. The main pane displays a code cell under the heading 'Importación de paquetes python'. Below it, another code cell is expanded, titled 'Definición de variables y carpetas'. The code in this cell is as follows:

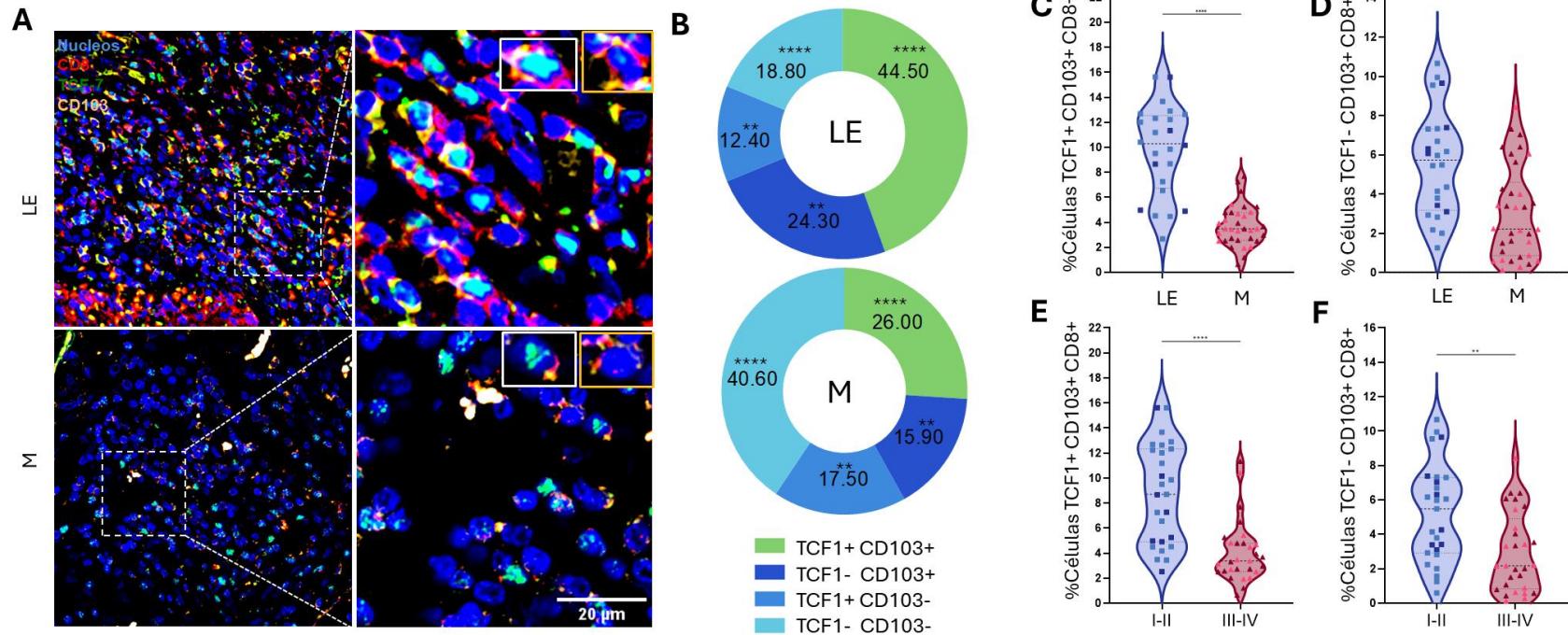
```
# @title Definición de variables y carpetas
anterior = 0
initialData = 'MARCADORES.csv' #Archivo con los vectores de interes
divisionData = 'Proporciones.csv' #Archivo con las divisiones de interes
areaData = 'AreaMarcadores.csv' #Archivo con las áreas a obtener
area = Decimal(634.0549*634.0549) #Area a utilizar para la división
destinationFolder = 'ANALISIS' #Carpeta que contiene los datos a procesar
dataFolder='EXCEL RESULTS'
carpeta = os.getcwd() #Carpeta que contiene la carpeta a procesar y los archivos con datos a obtener
extension = "txt" #Possible values "txt" and "csv"
fileNameSeparator = "_" #Possible values "_" si el nombre es "MERGE_PX-CAMPO" "_" si el nombre es "MERGE_PX-CAMPO"
csvDelimiter='\t' #Possible values ',' and '\t'
#Nombres de columnas
colPaciente = "Px"
colProm = "Promedio"
columnSubfix = 'Object marker ' #Subfijo de archivo Object marker 1, Object marker 2 , etc. En ocasiones 'Object Marker '
upLabel = 'arriba' #Nombre de la columna para la división arriba/abajo, arriba/area
downLabel = 'abajo' #Nombre de la columna para la división arriba/abajo
columnaFiltrar = 'Region class'
```

Análisis downstream

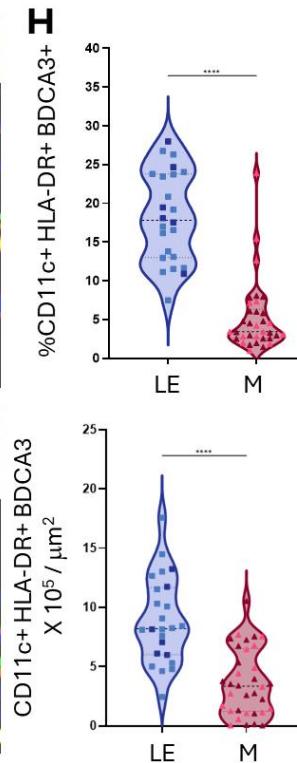
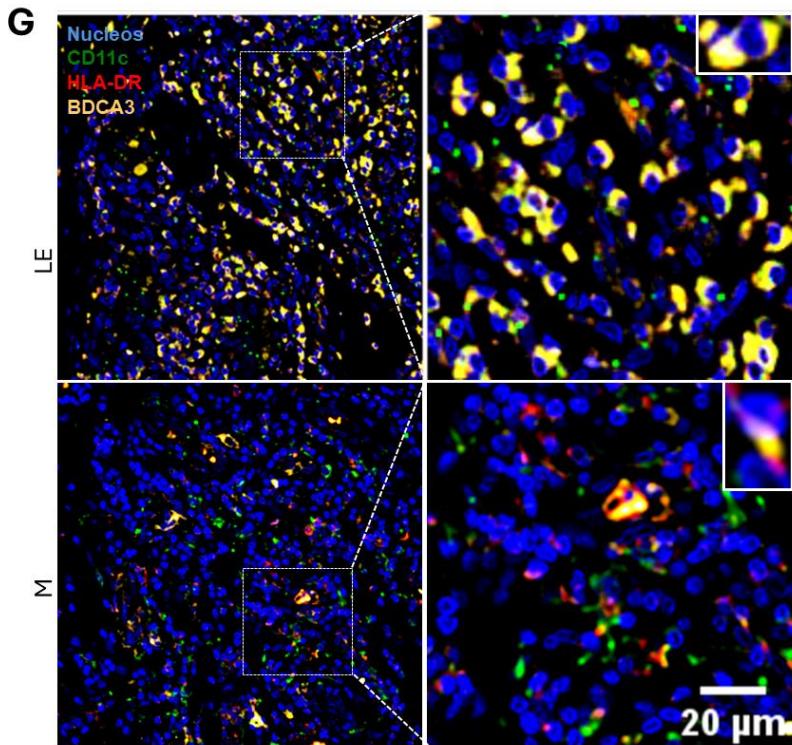
TCF1-positive and TCF1-negative TRM CD8 T cell subsets and cDC1s orchestrate melanoma protection and immunotherapy response

Saraí G De León-Rodríguez  ¹, Cristina Aguilar-Flores, ² Julián A Gajón, ³
Ángel Juárez-Flores, ⁴ Alejandra Mantilla, ⁵ Raquel Gerson-Cwilich, ⁶
José Fabian Martínez-Herrera, ^{6,7} Diana Alejandra Villegas-Osorno, ⁶
Claudia T Gutiérrez-Quiroz, ⁸ Sergio Buenaventura-Cisneros, ⁹
Mario Alberto Sánchez-Prieto, ¹⁰ Edmundo Castelán-Maldonado, ⁹
Samuel Rivera Rivera, ¹⁰ Ezequiel Moises Fuentes-Pananá  ¹¹,
Laura Bonifaz  ¹²

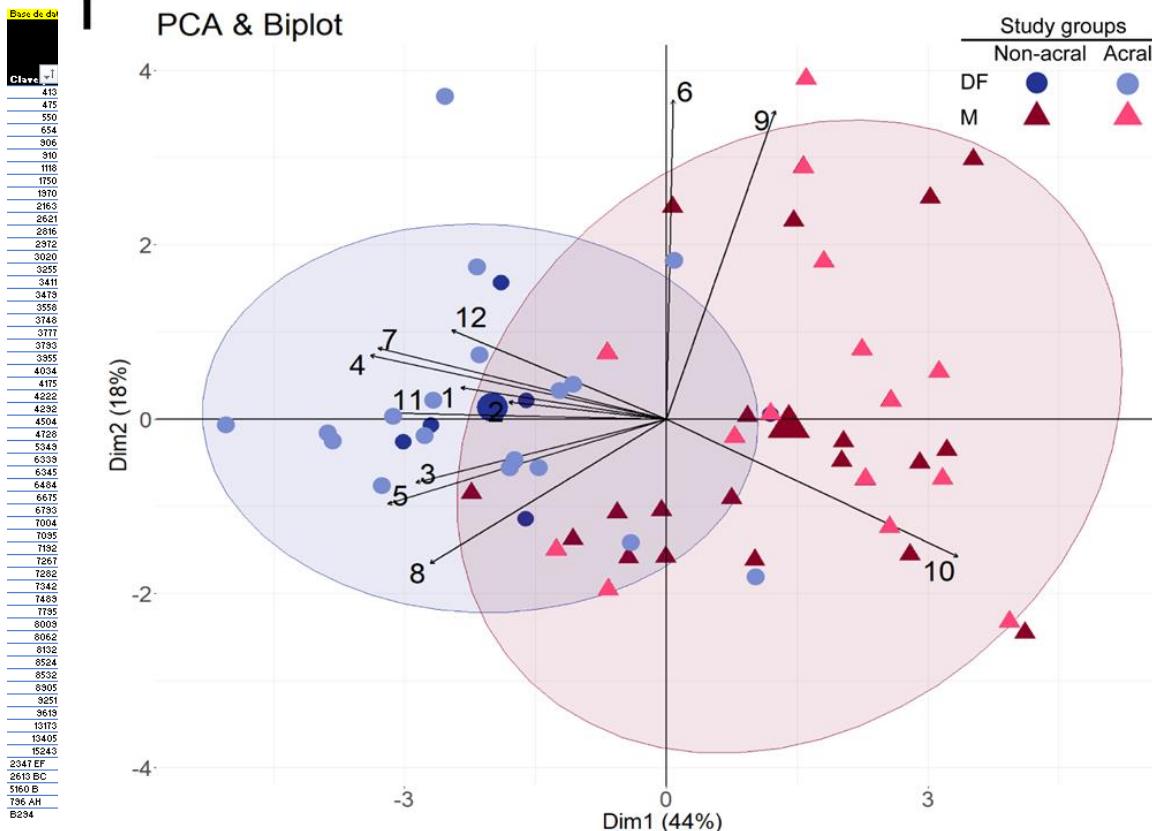
Linfocitos CD8 Residentes de tejido TCF7+ y TCF7- relacionados al control de la enfermedad



Linfocitos CD8 Residentes de tejido TCF7+ y TCF7- relacionados al control de la enfermedad



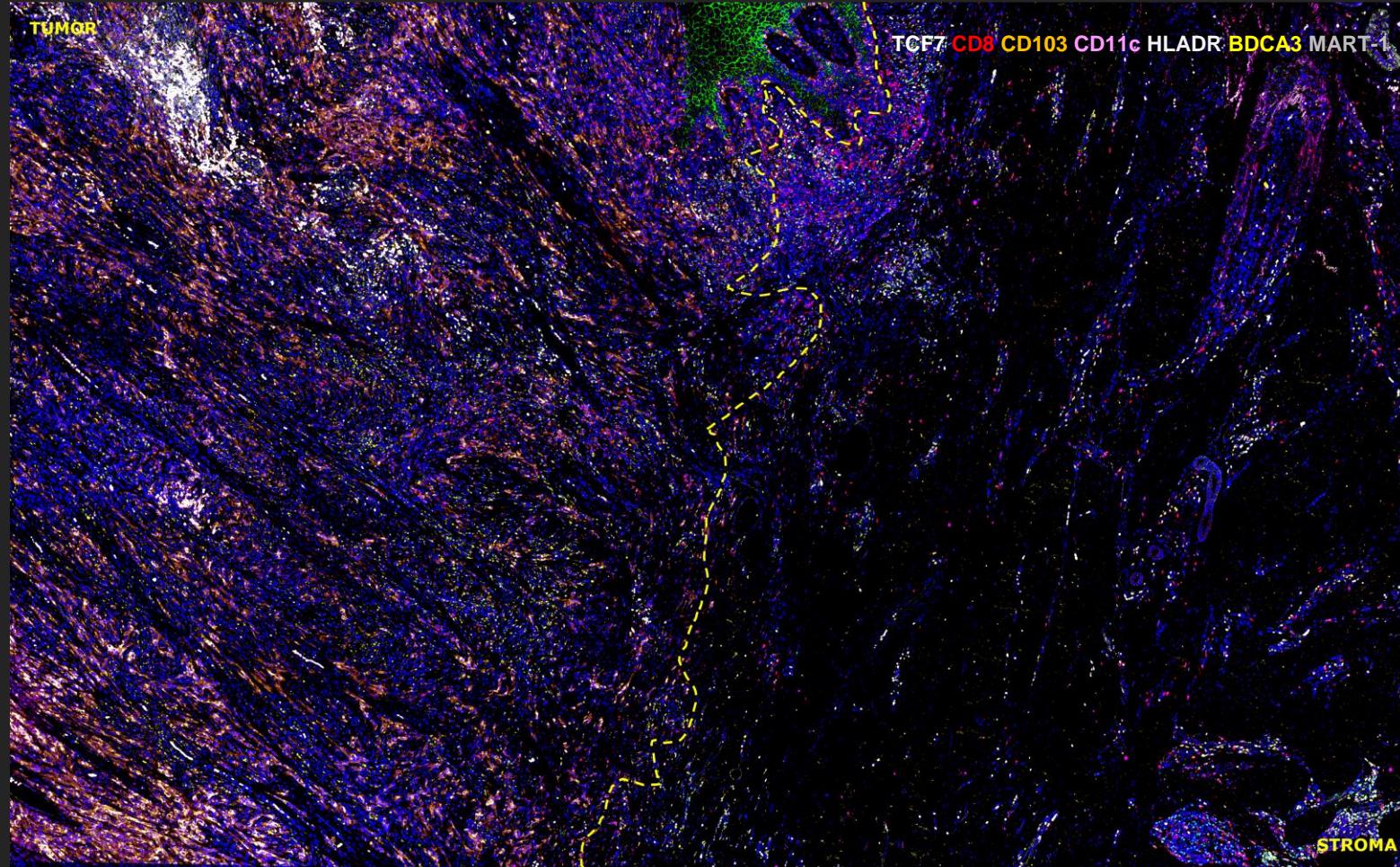
Linfocitos CD8 Residentes de tejido TCF7+ y TCF7- relacionados al control de la enfermedad



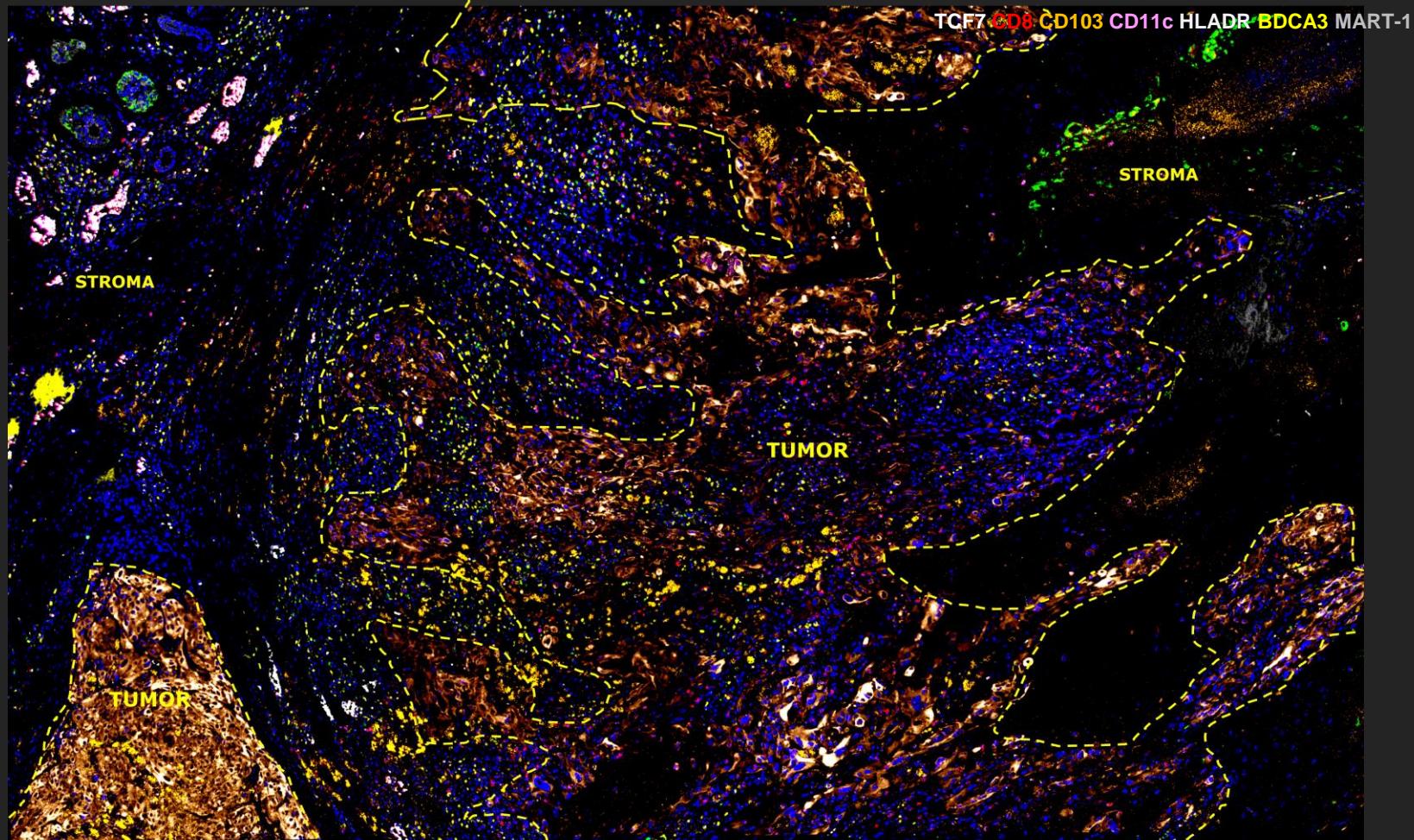
1c

vectorial
31.965048408
33.209178955
36.52234476
12.45064503
7.82003951
35.34935102
35.37845443
35.364127398
37.66355705
45.34530412
37.80189838
8.62205143
14.0073514
18.20620318
16.31586833
38.62565122
34.60664218
33.56091264
14.42467023
6.754593383
18.21320072
25.50565433
23.91719292
44.3782541
45.34530412
37.52371508
18.33906733
34.63373806
36.44556721
7.860454678
36.1206554
42.44462812
15.6957403
43.14410311
36.44462812
40.3455601
26.2455376
37.41674537
21.36684065
35.12239356
37.18153333
42.31940563
25.65630623
38.77711323
13.30101832
44.81509451
10.43139313
26.80307002
28.0017062
23.6432235
24.3194274
28.83750148
27.22255683
31.26870243
38.01057241
27.27375306
30.52430547

Escaneos del tejido tumoral y estromal: Paciente Libre de Enfermedad

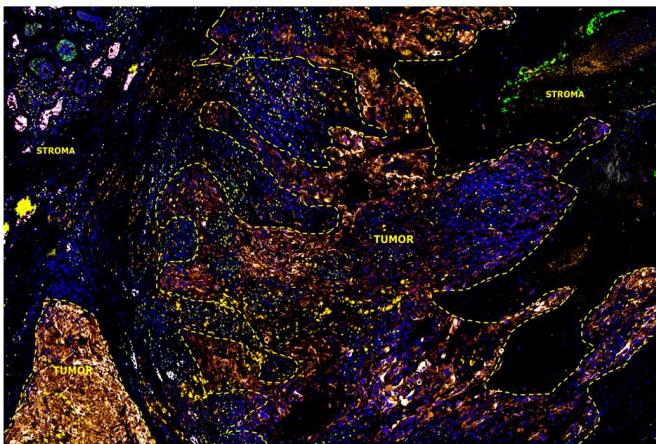
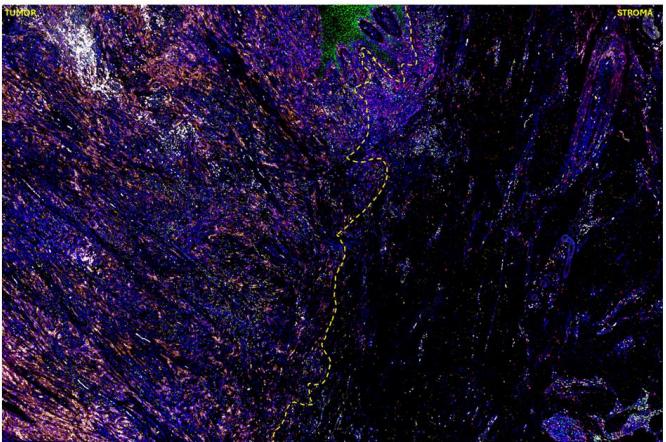


Escaneos del tejido tumoral y estromal: Paciente Metastásico

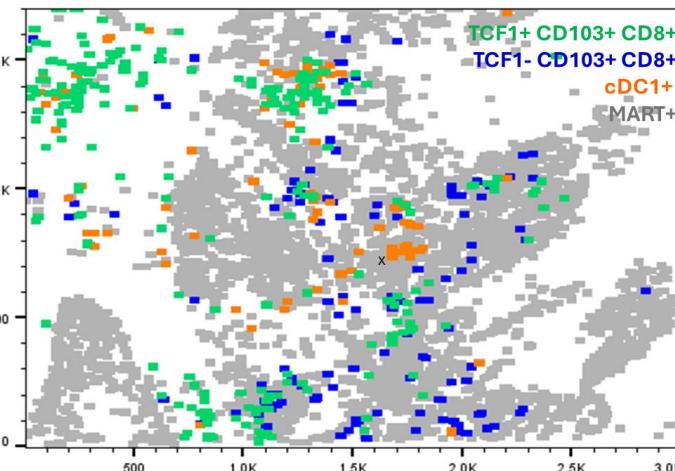
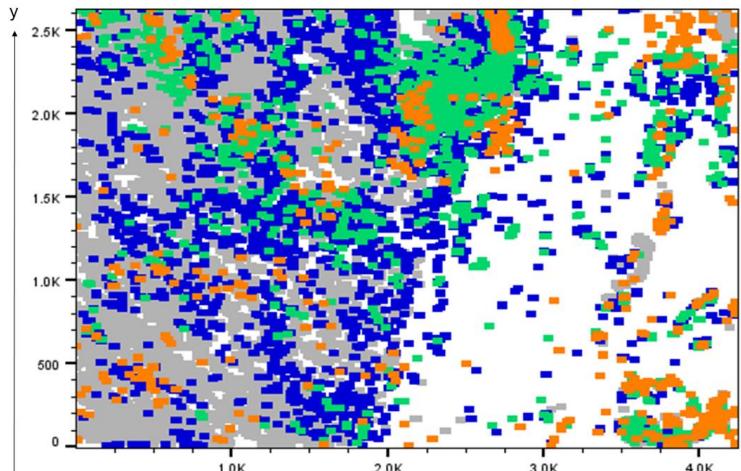


Correlación espacial de las poblaciones inmunes

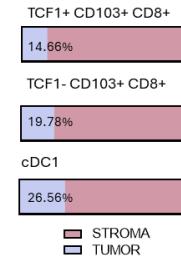
D

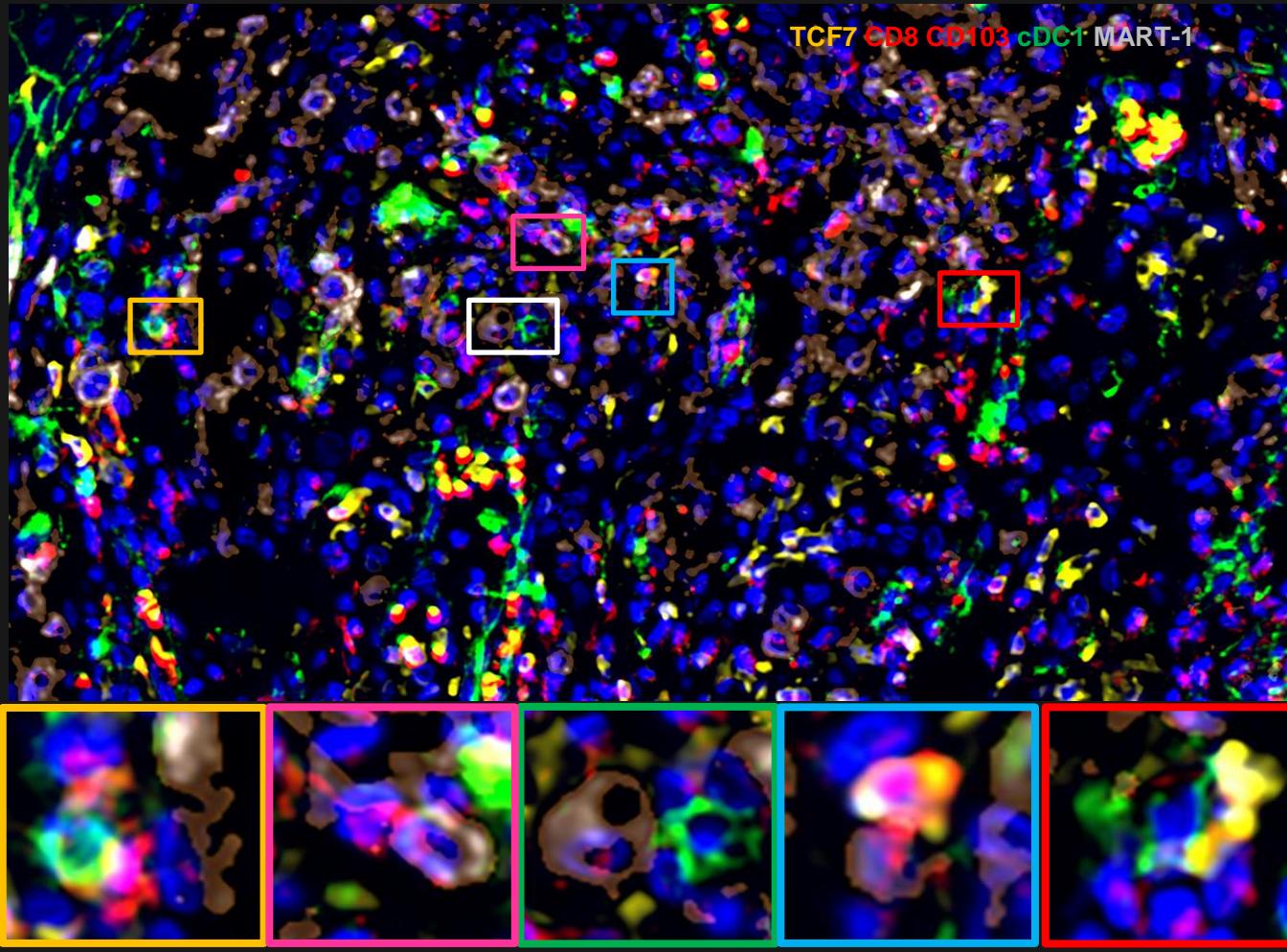


y



E



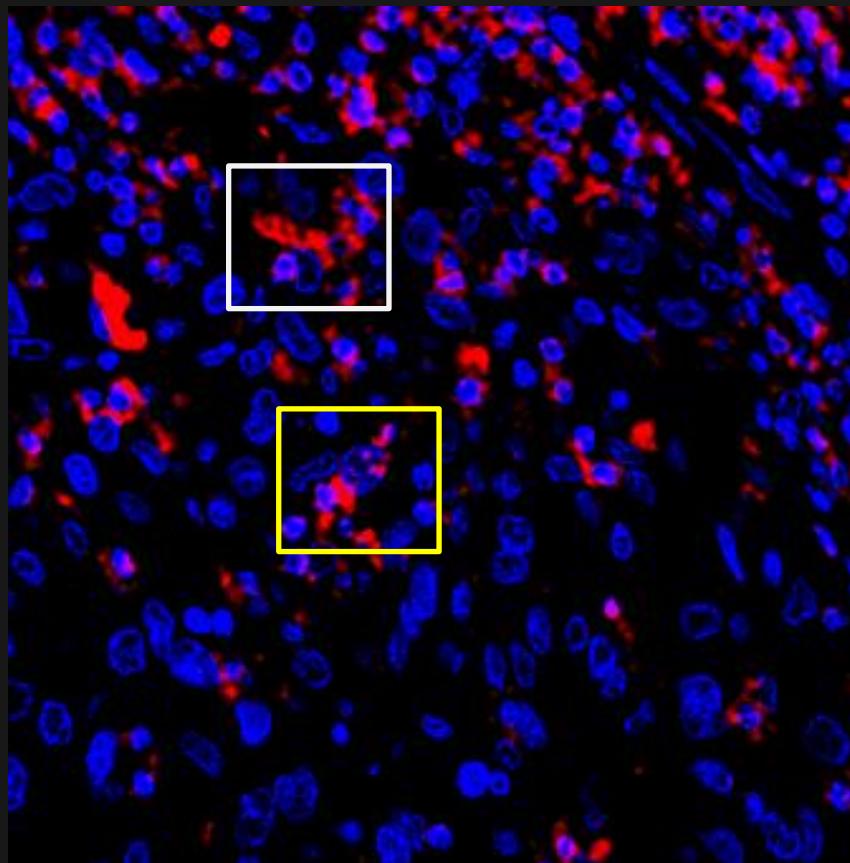
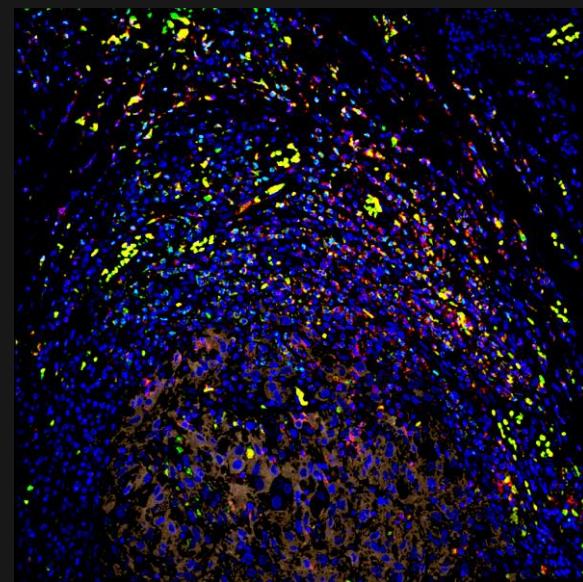


Libre de enfermedas

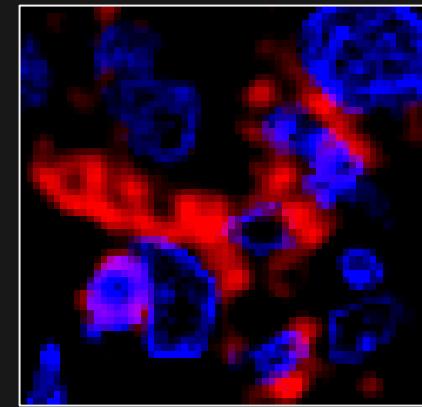
CD8 CD103 PD1 GZB LAMP1

TUMOR CASP3

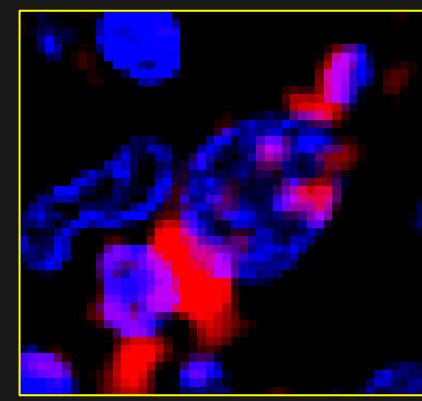
TCF7



TCF7+



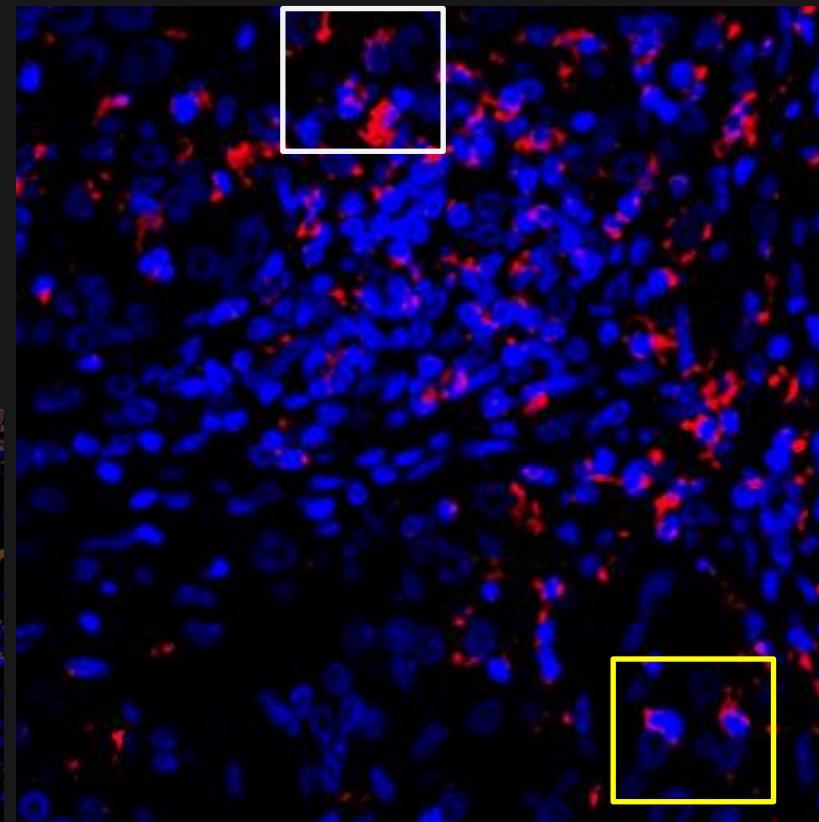
TCF7-



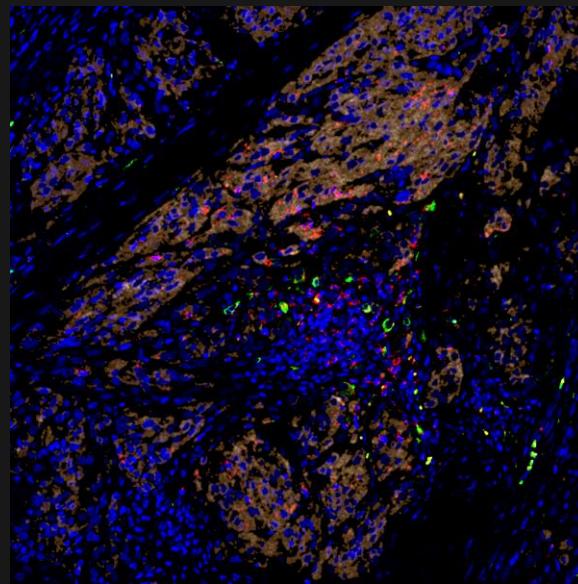
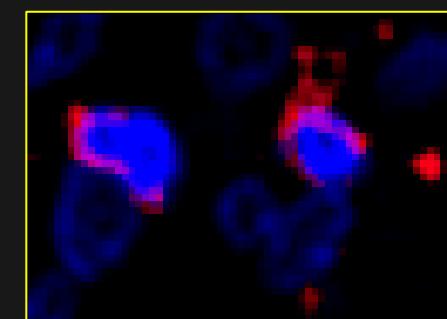
Metastásico
CD8 CD103 PD1 GZB LAMP1

TUMOR CASP3
TCF7

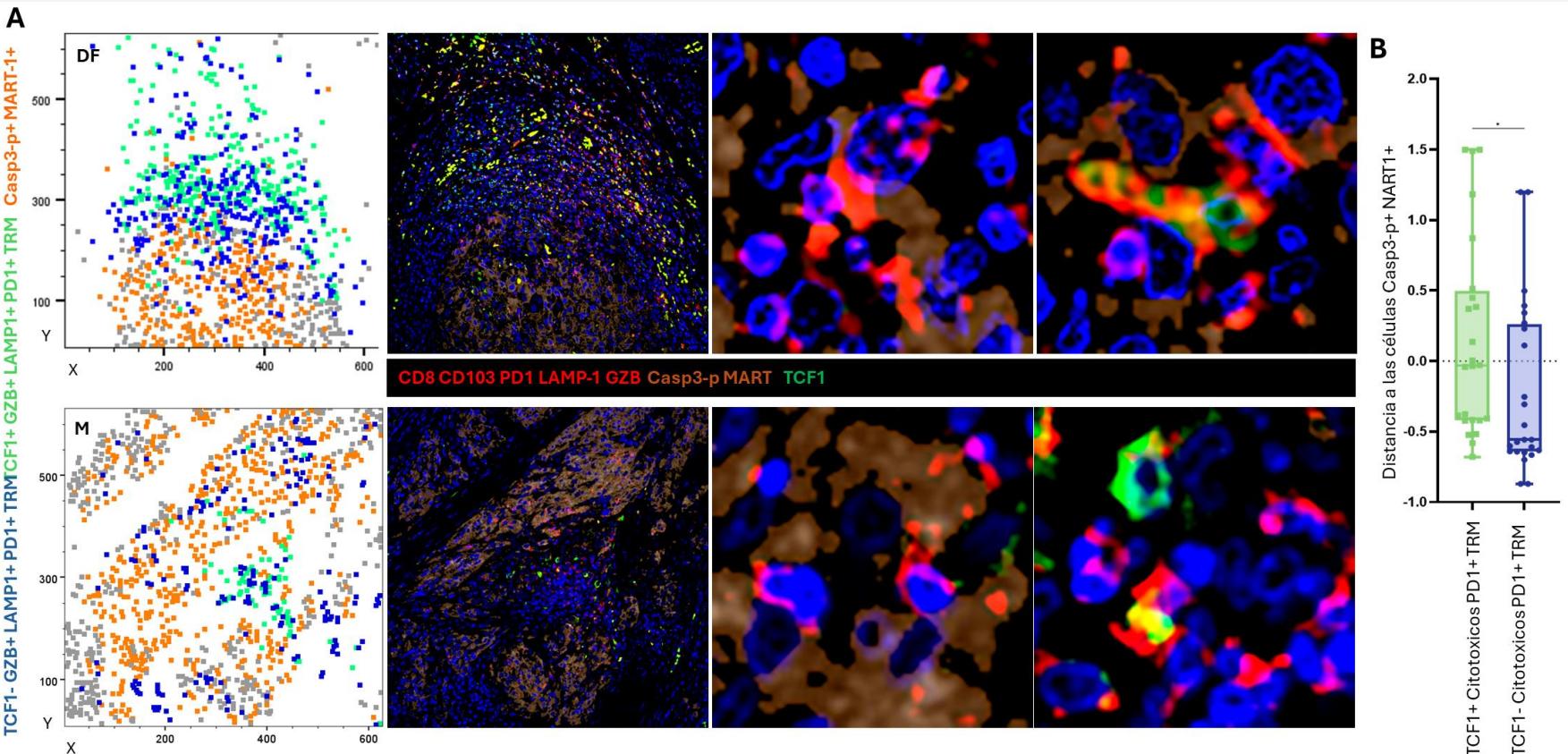
TCF7+



TCF7-

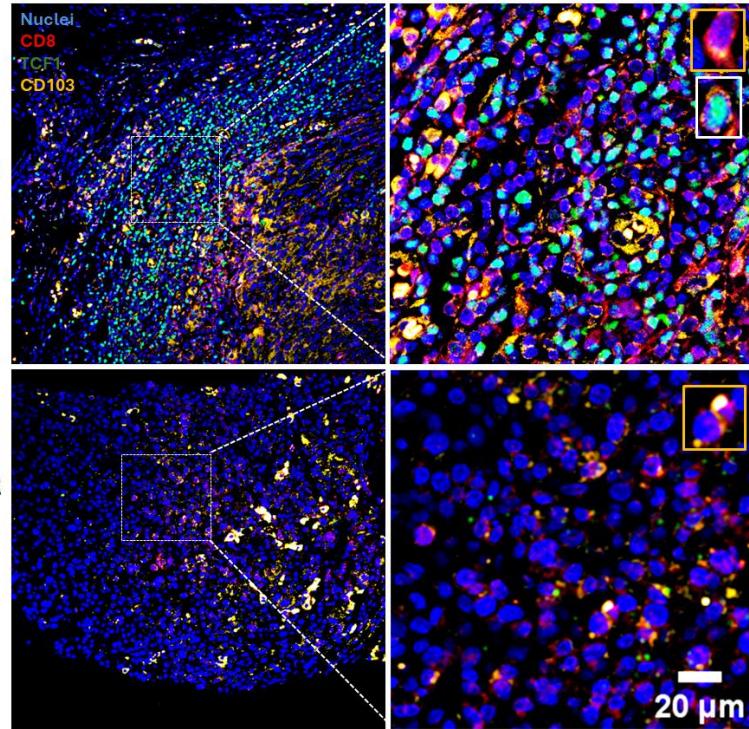


Los linfocitos TCF1- TRM citotóxicos se encuentran más cerca de las células MART-1+ Casp-3-p+

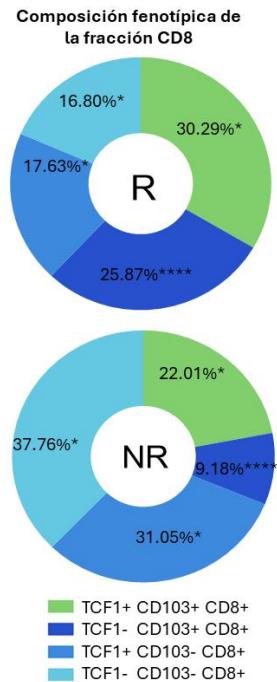


Linfocitos residentes de tejido que expresan TCF7+ y TCF7- están relacionados a la respuesta a CBI

A

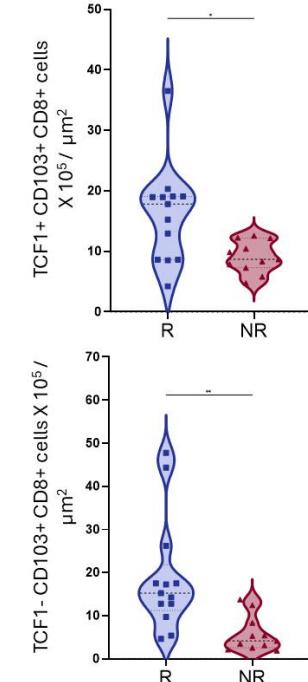
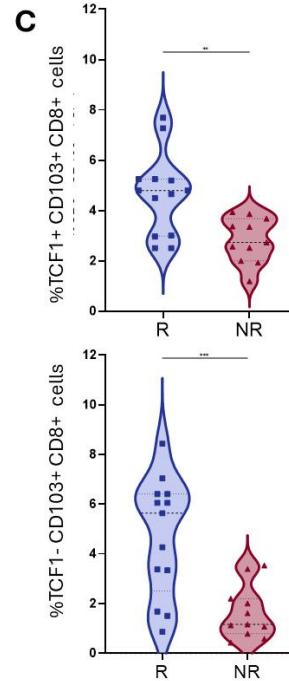


B

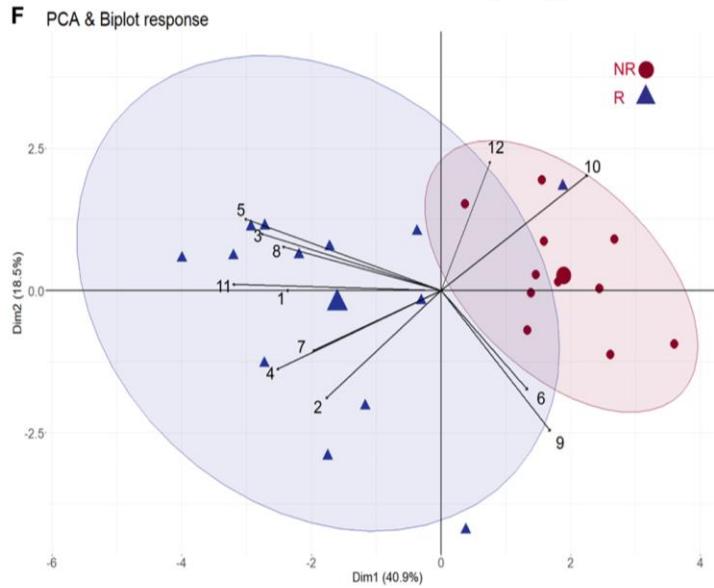
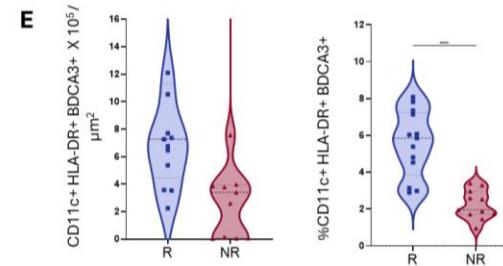
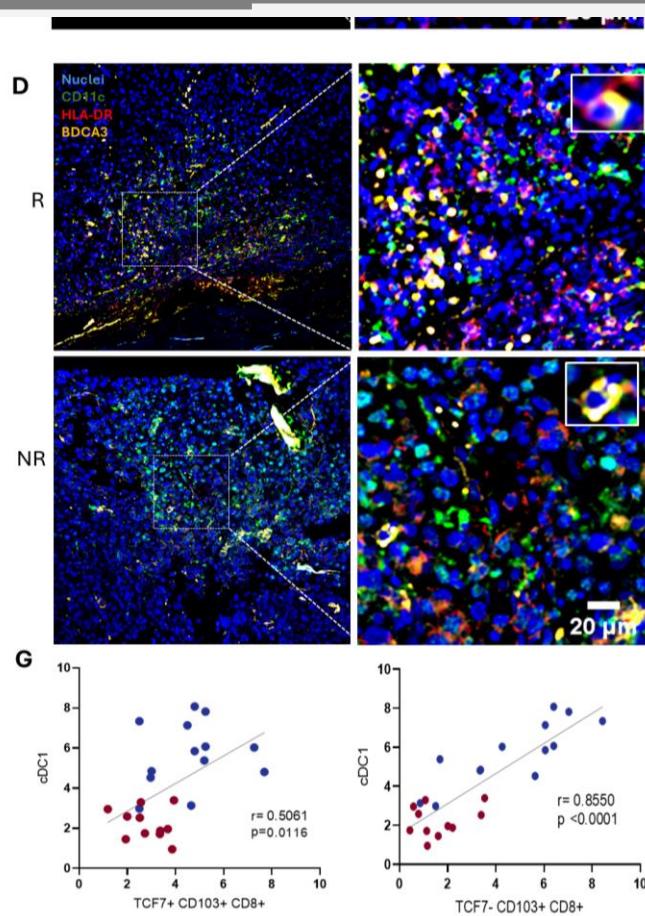


F

16-



La abundancia de cDC1 es mayor en pacientes que responden a la immunoterapia.



Gracias!

Contacto

sara.g.deleonr@gmail.com
5568745474