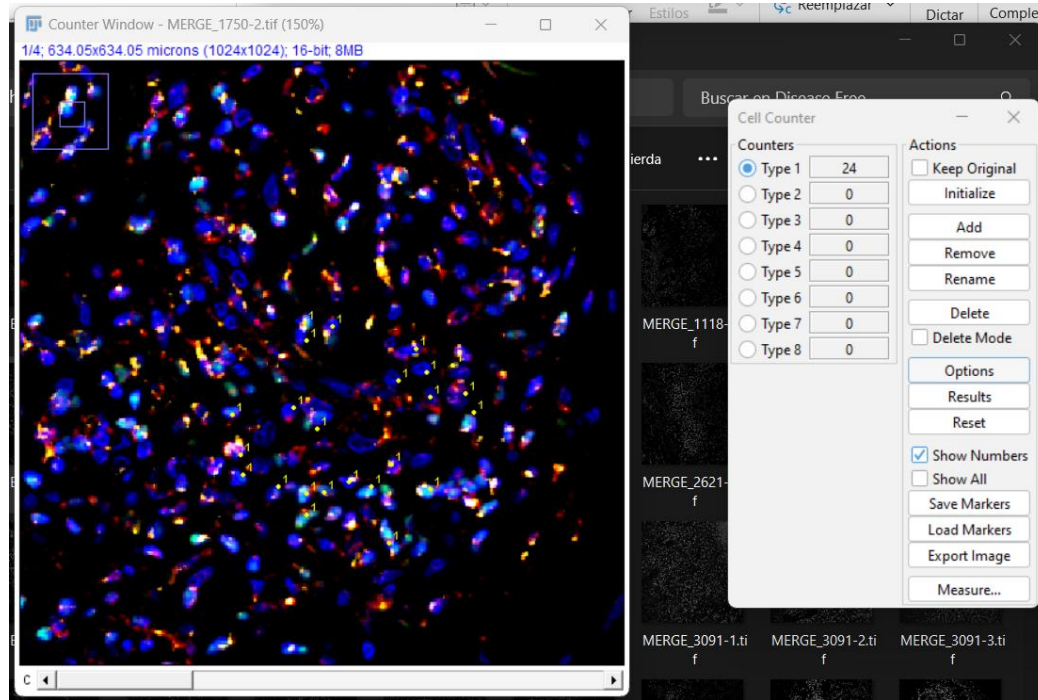


Aprendizaje Automático para Clasificación de Marcadores

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-Pananá⁷ **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

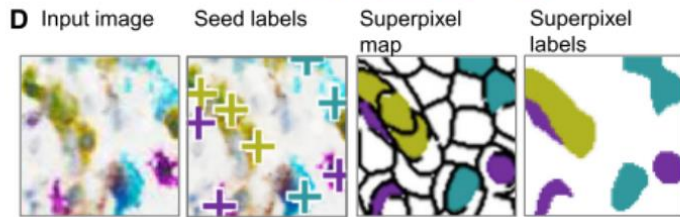
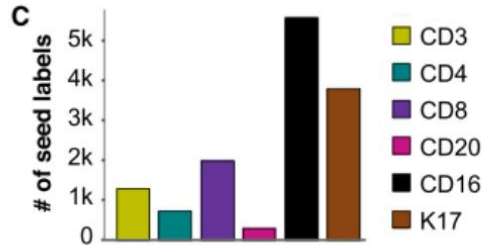
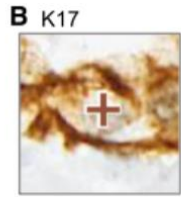
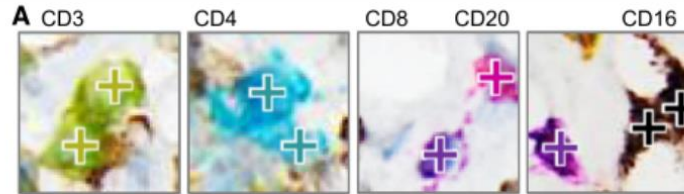
Clasificación: El proceso de predecir a que clase pertenece un objeto

Clase: objeto o categoría.

Ejemplo: Clasificar un correo como spam o no spam



Uso de clasificadores



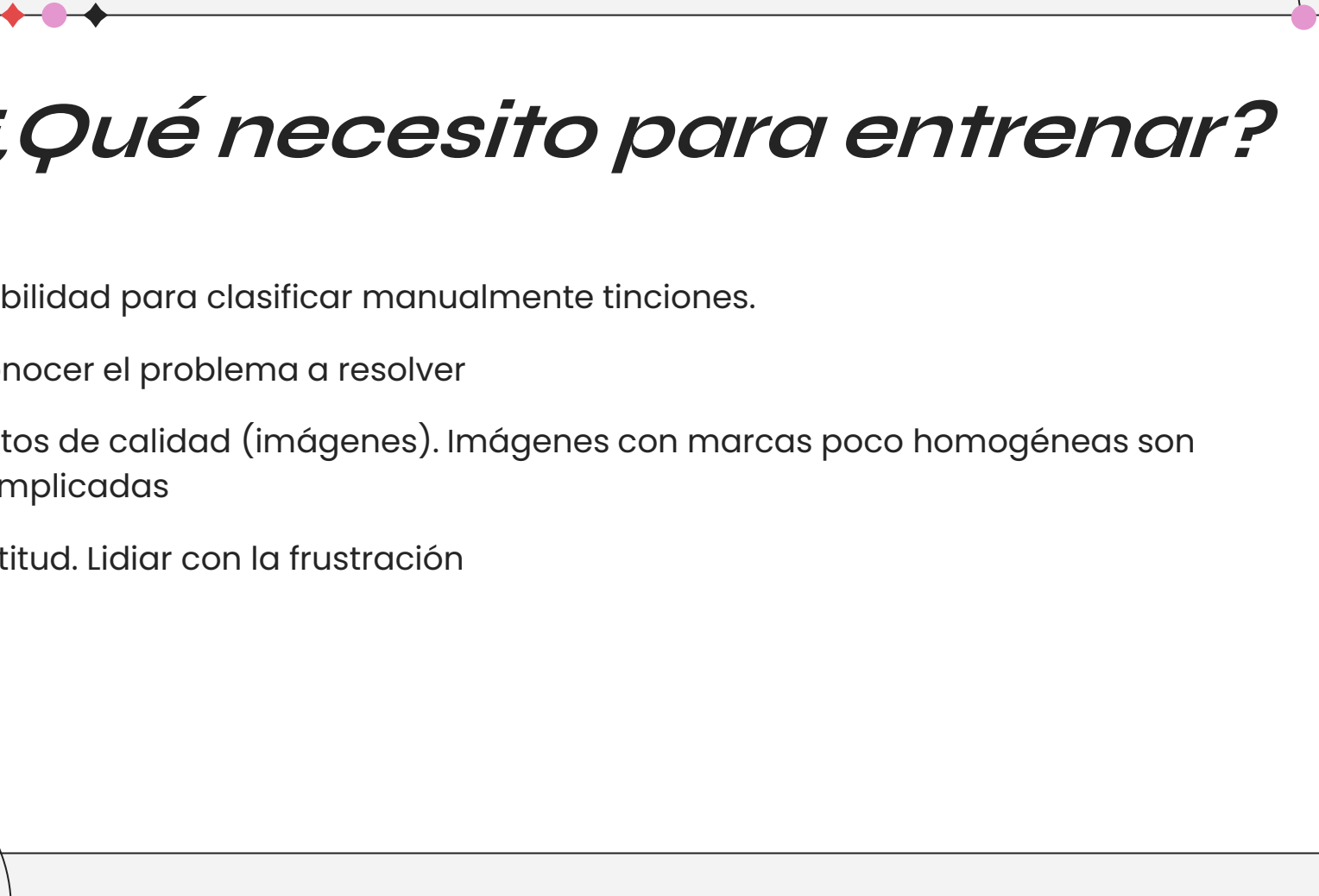
Clasificadores para IHC

Clasificadores imagenología
médica

Clasificadores patrones



Illustration of a histology slide with a seed label and a visualization of a non-invasive training data. A: Examples of CD3, CD4, CD8, and CD20.



¿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

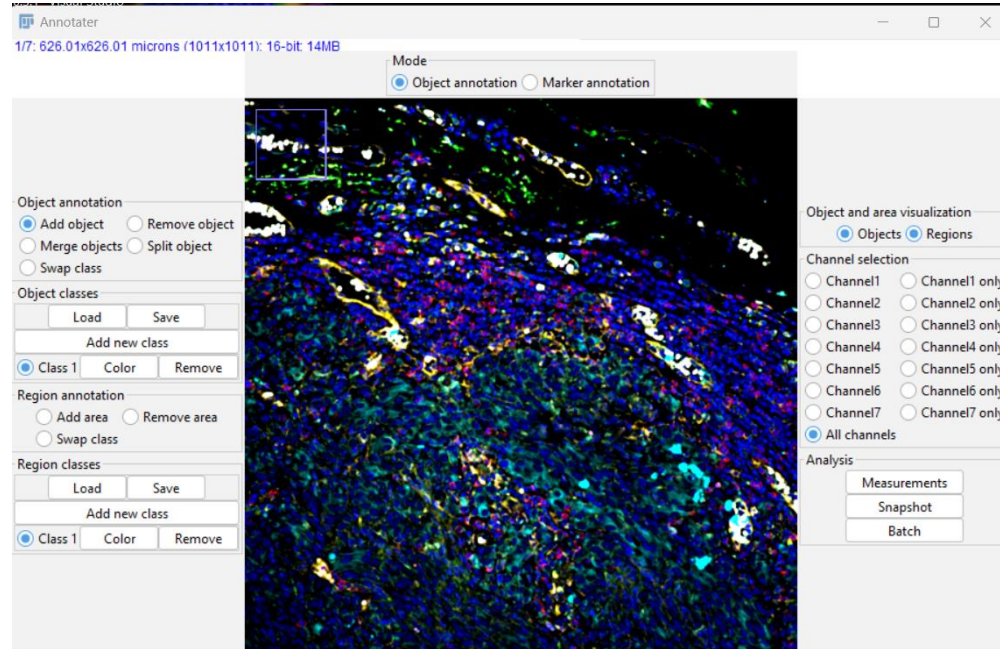
Actitud. Lidar con la frustración

Annotater

Herramienta para anotación y entrenamiento de clasificadores

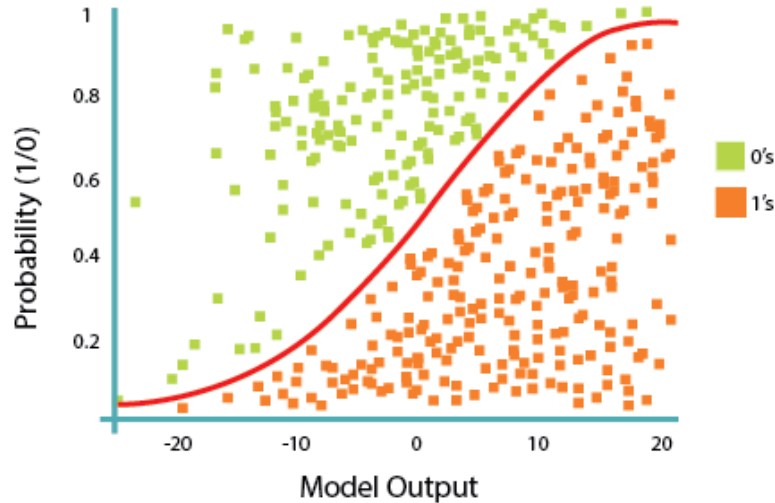


Thierry Pecot



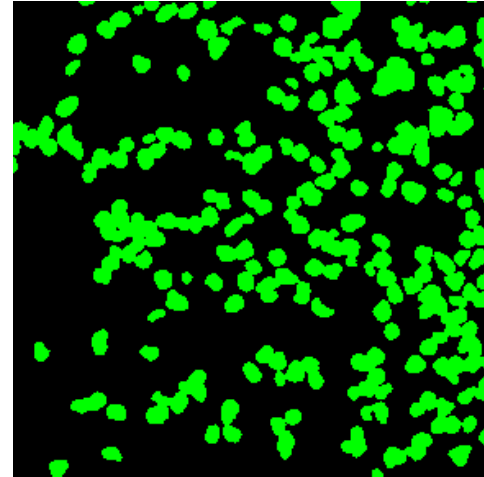
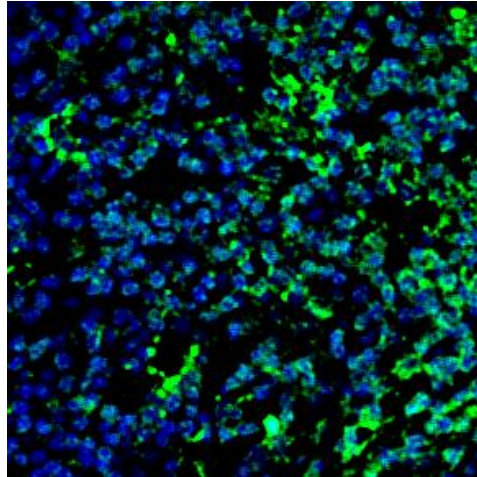
Regresión logística

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

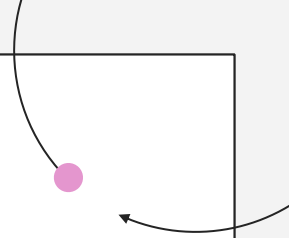
Reproducibilidad

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

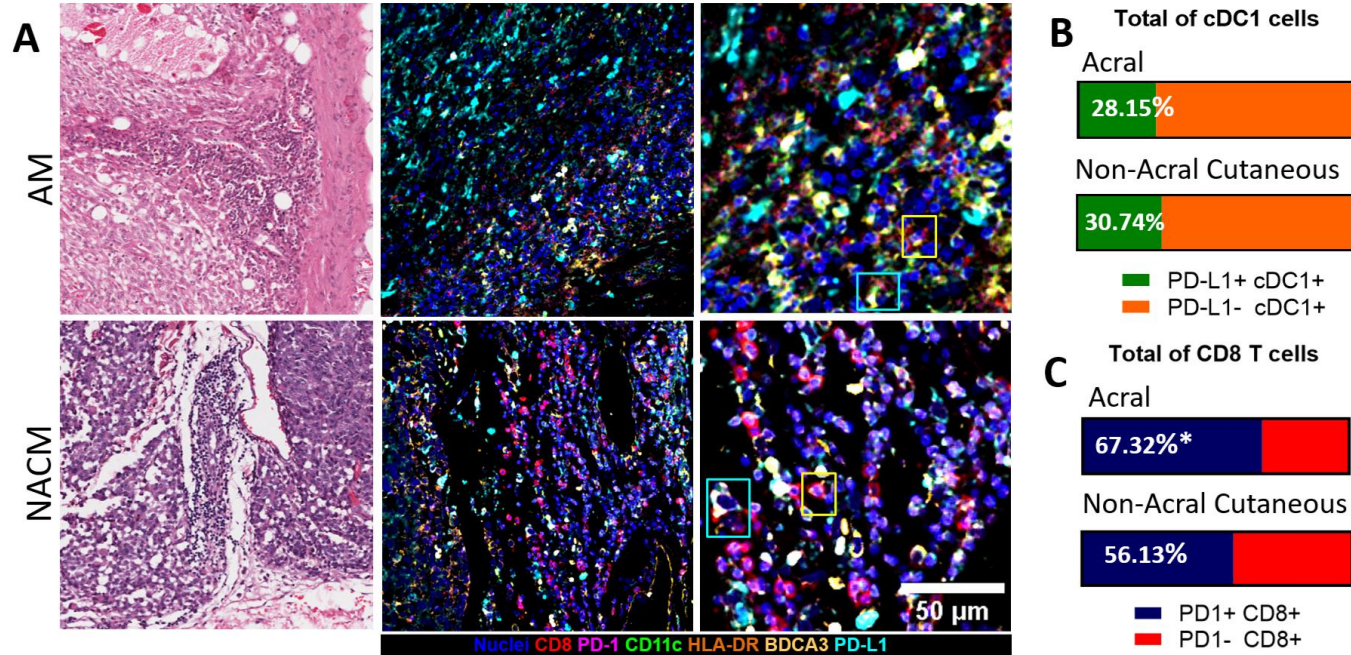
⁶ Cancer Center, Medical Center American British Cowdray, Mexico City 01120, Mexico

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

[†]

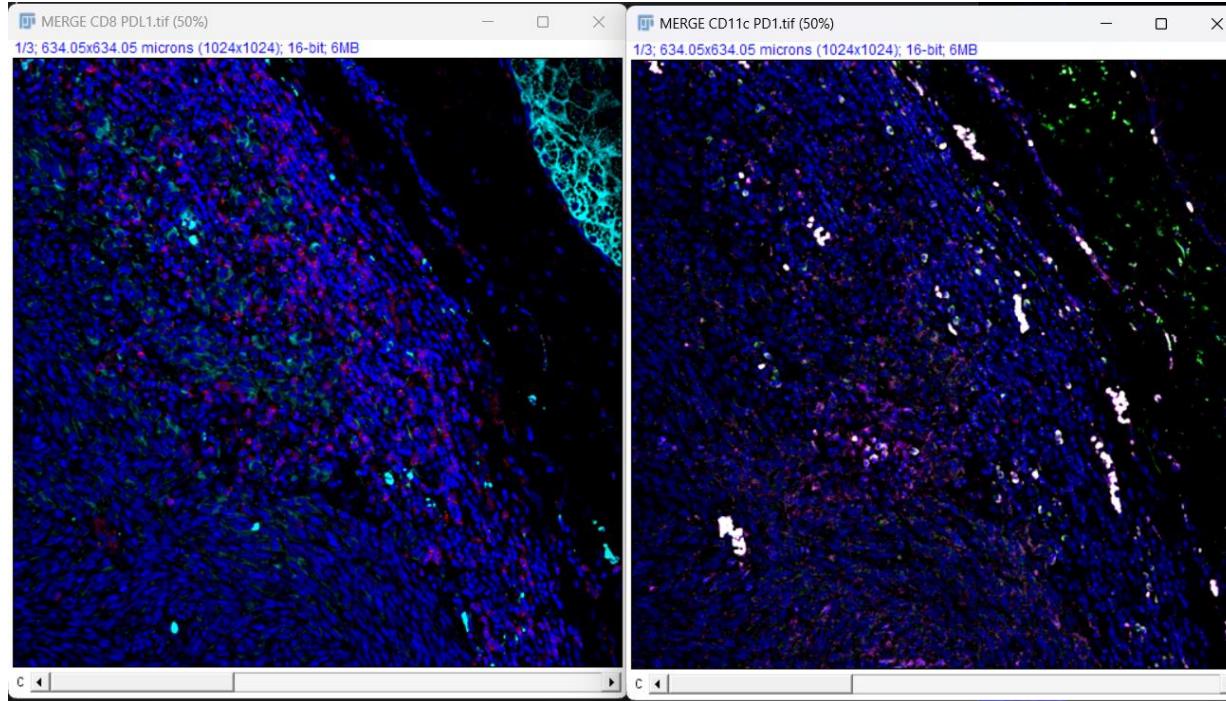


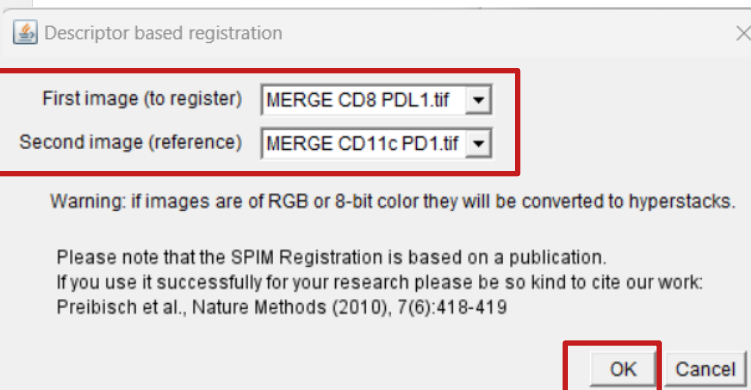
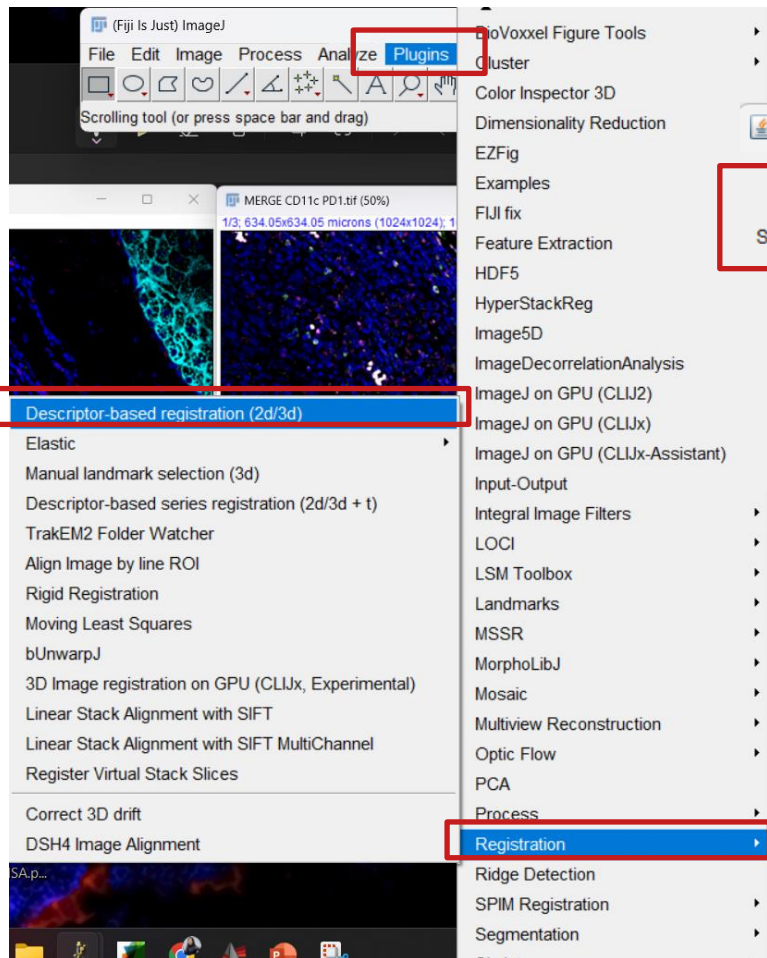
Evaluación de fenotipos

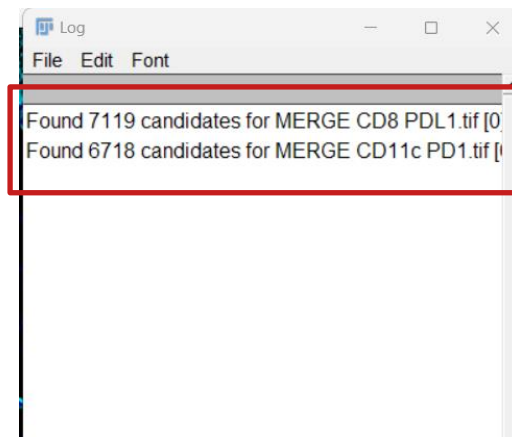
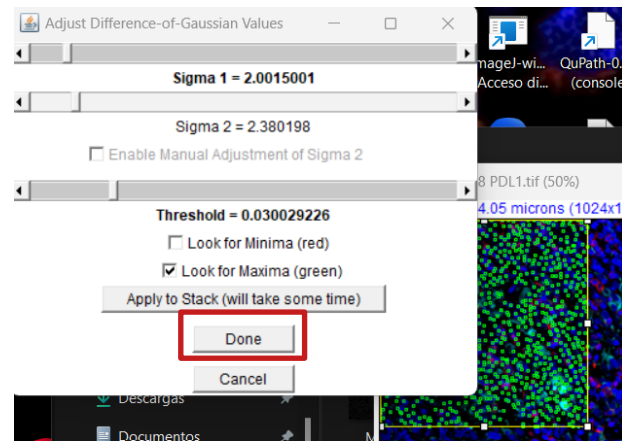
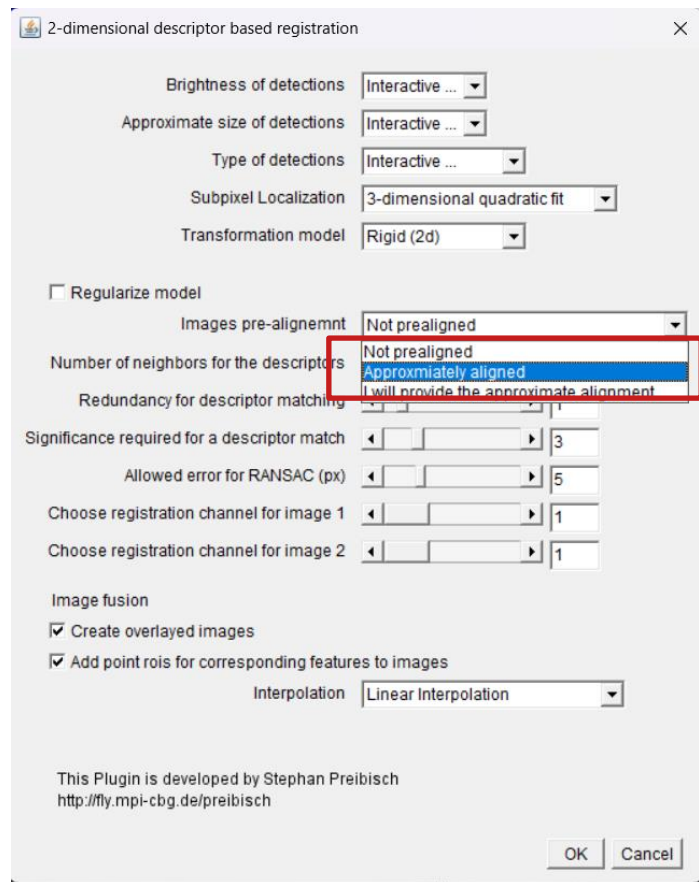


Alineación de imágenes

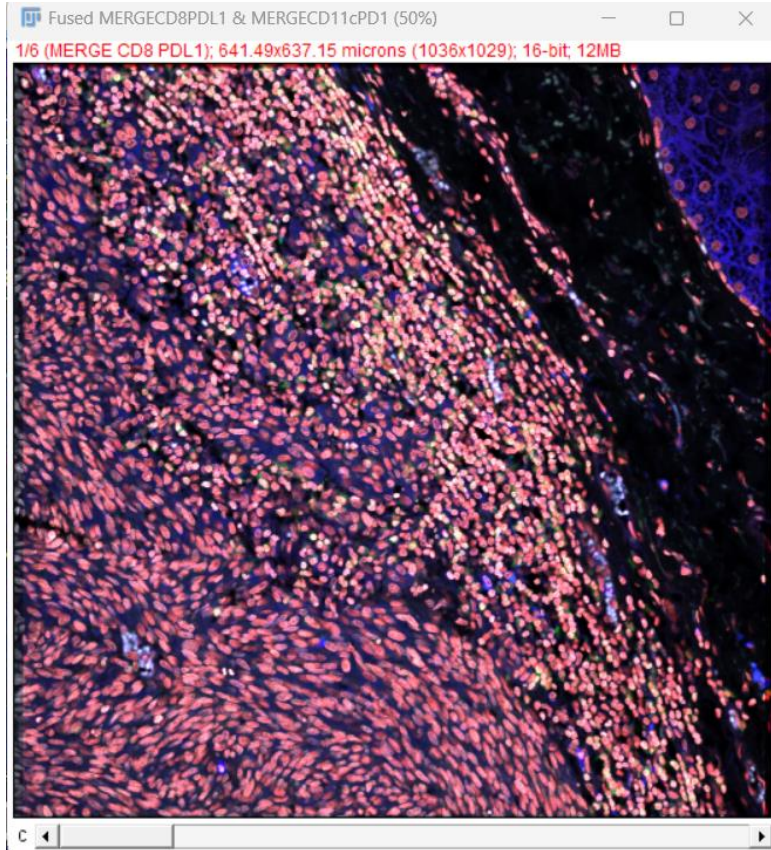
Abrir imágenes a alinear



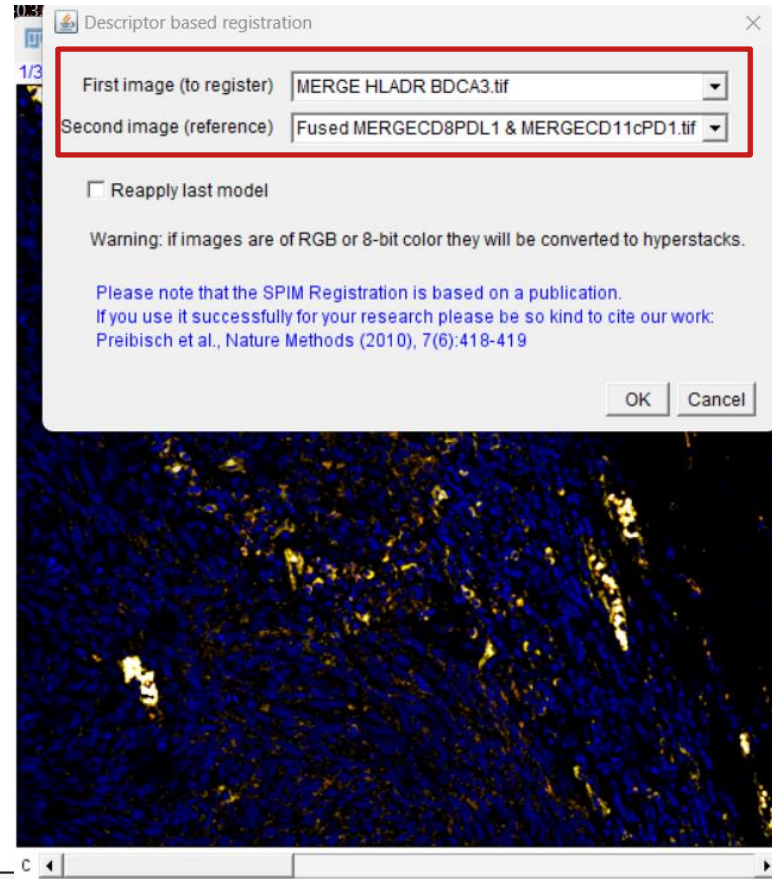




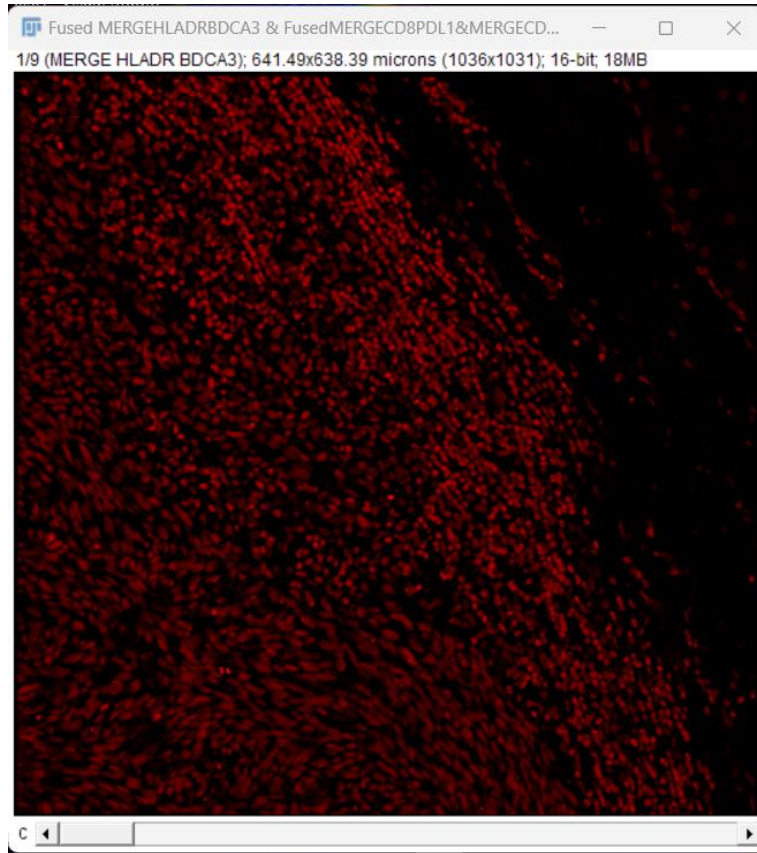
Se generara una imagen alineada



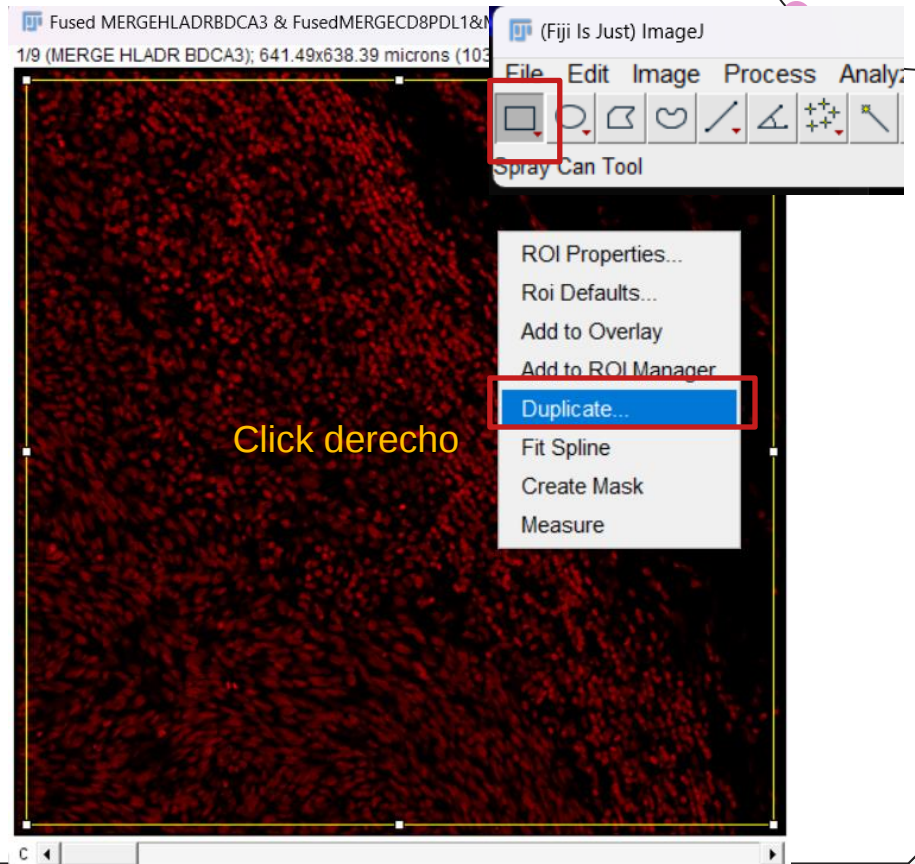
Abrir la tercera imagen a alinear y repetir el proceso



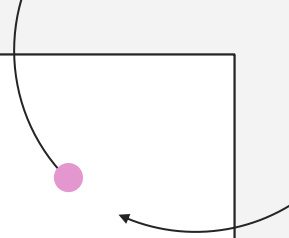
Stack de 9 canales



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

Canal 2-CD8 rojo

Canal 3-PD1 magenta

Canal 4-CD11c verde

Canal 5-HLADR Orange hot

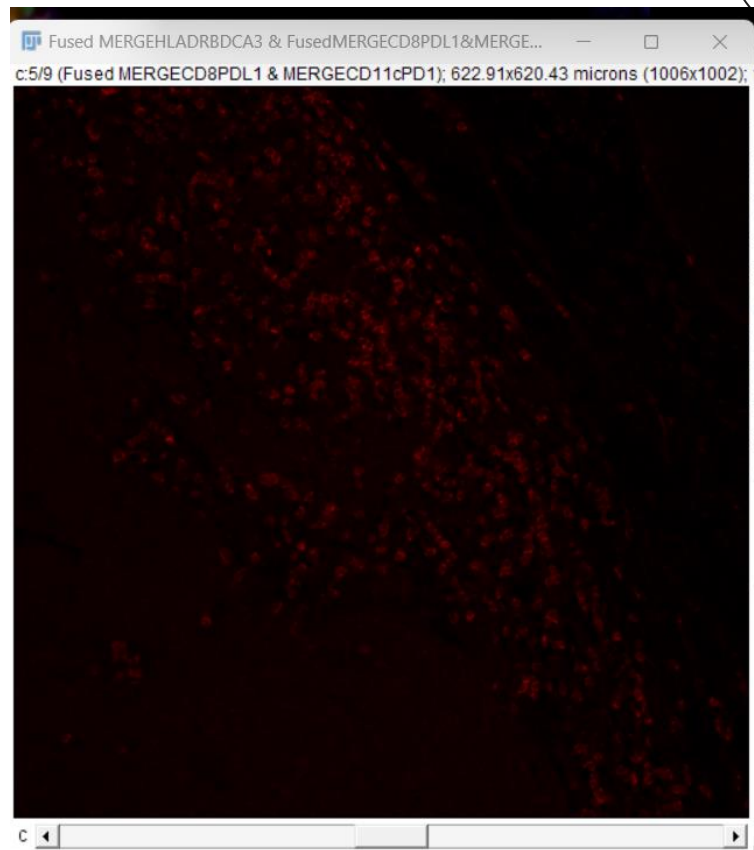
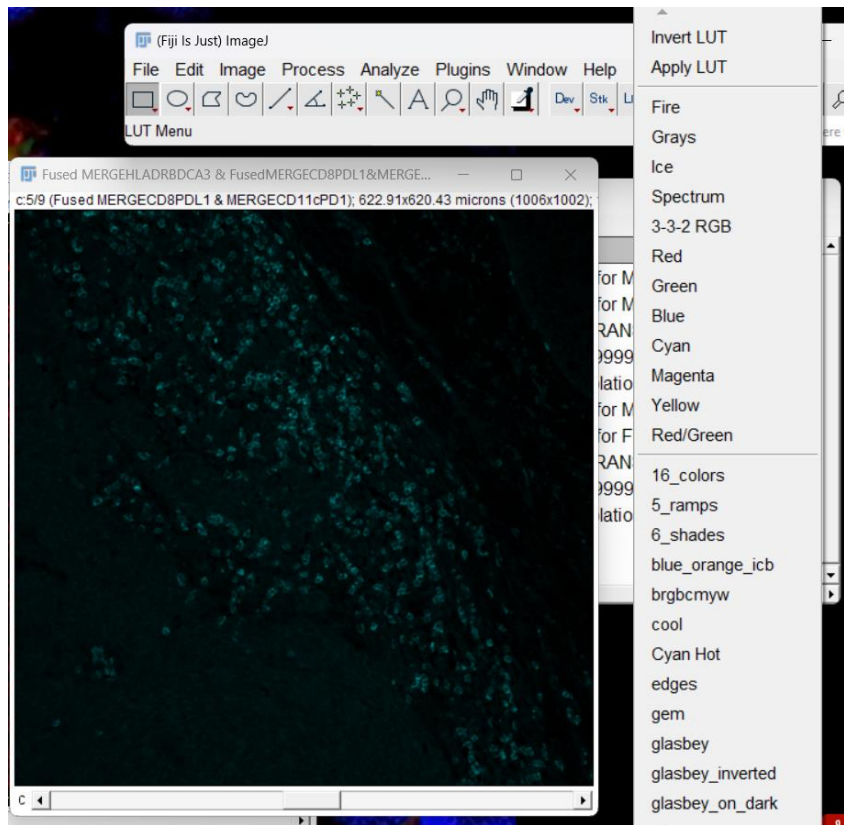
Canal 6 BDCA3 Yellow hot

Canal 7-PDL1 Cyan

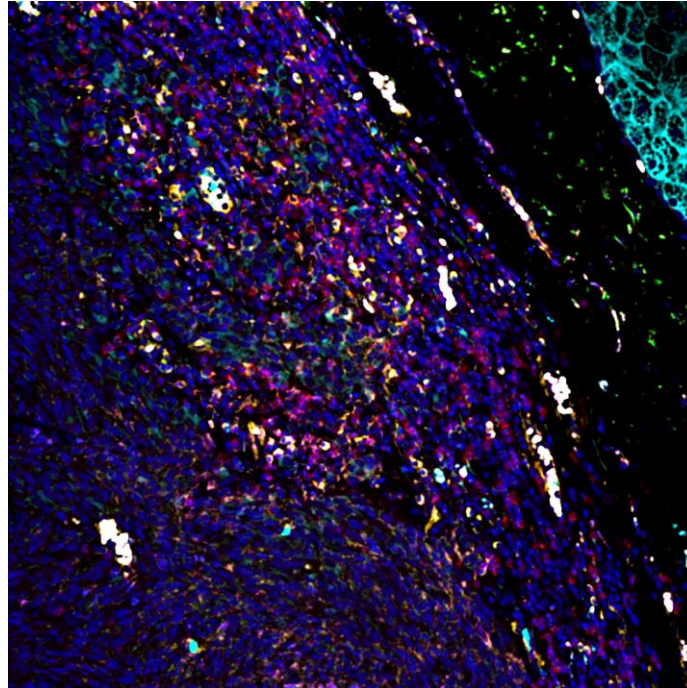
- Green
- Red
- Yellow
- Cyan
- Magenta
- Yellow hot
- Sepia
- Grays
- Orange hot
- Magenta hot



Asignar lut a cada canal



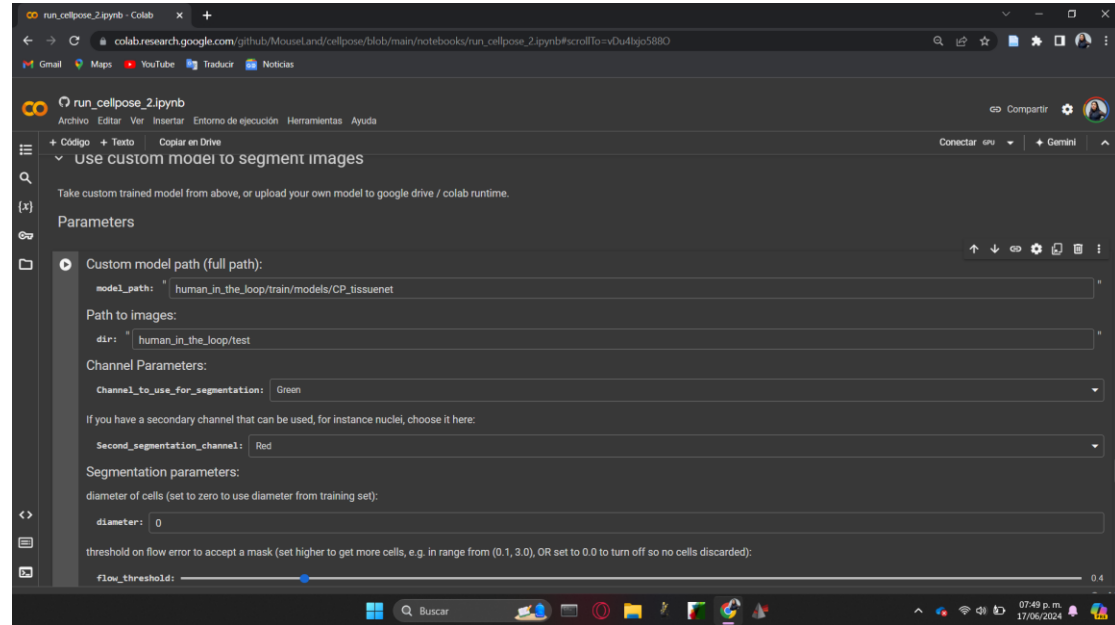
Separar los canales y hacer la composición final



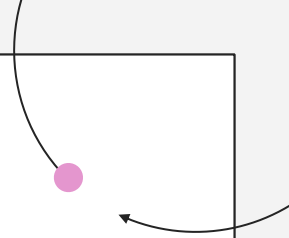
Segmentación de núcleos

Google collab

Separar los núcleos y subirlos a una carpeta de drive



Entrenamiento de clasificadores



De nuestro experimento tenemos dos conjuntos de imágenes

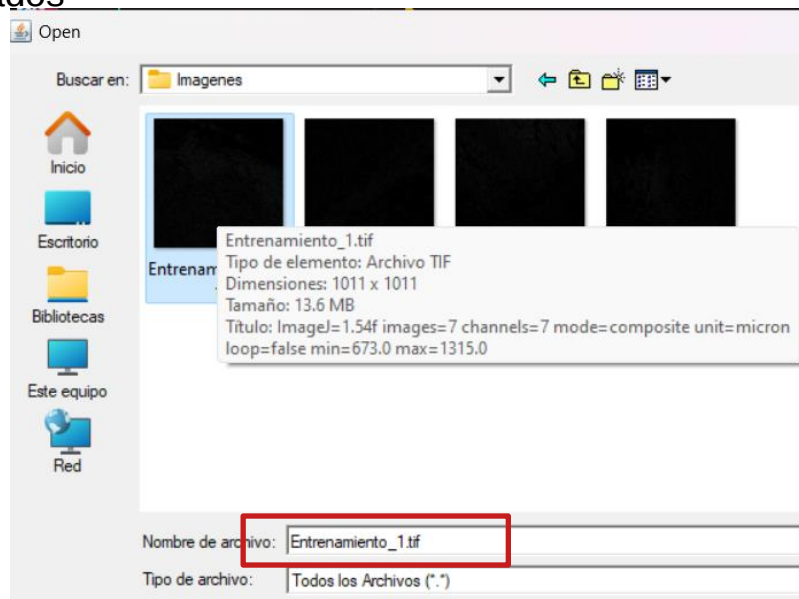
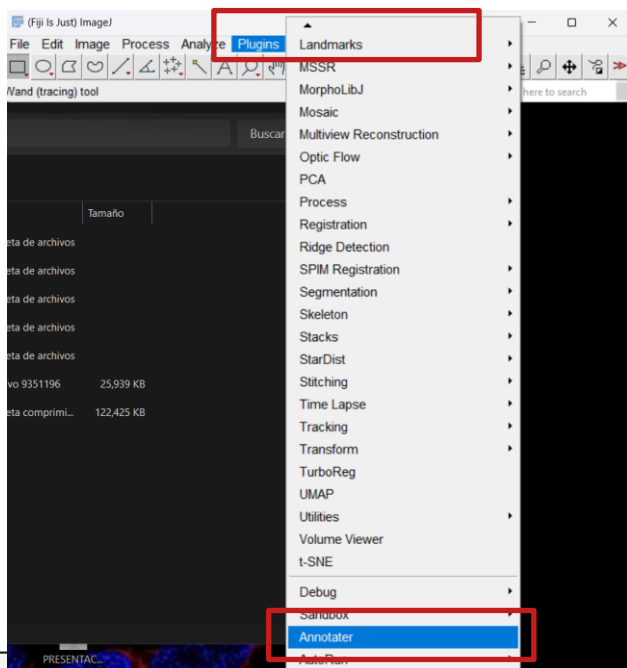
1. Datos para entrenamiento. Este dividirse en aprox 80% para entrenamiento y 20% validad
2. Datos problema/ Datos a evaluar. Este conjunto no debe incluir el primer conjunto idealmente.

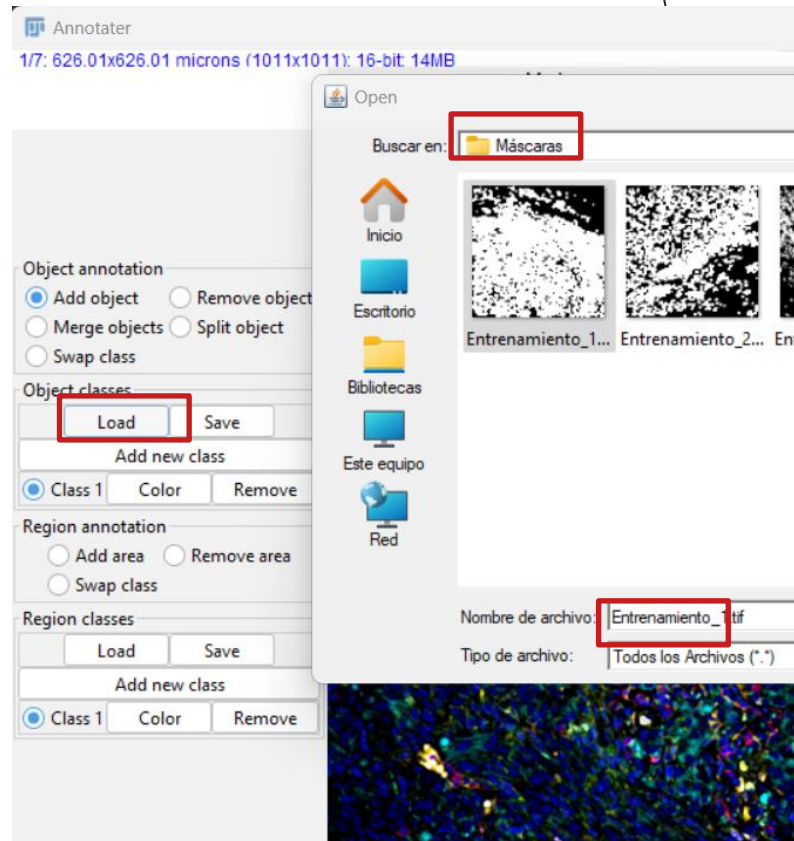
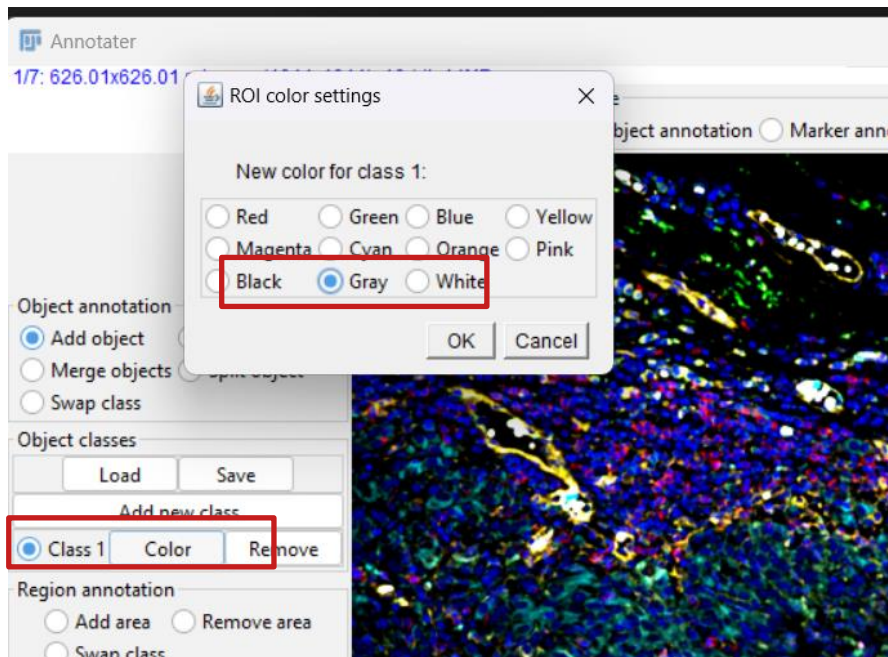


Entrenamiento de clasificadores

Abrir imagen de entrenamiento en annotater

Empalmar con su máscara de núcleos segmentados





Annotator

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

Load

Save

Add new class

☒ Class 1

Color

Remove

Region annotation

☐ Add area ☐ Remove area

☐ Swap class

Region classes

Load

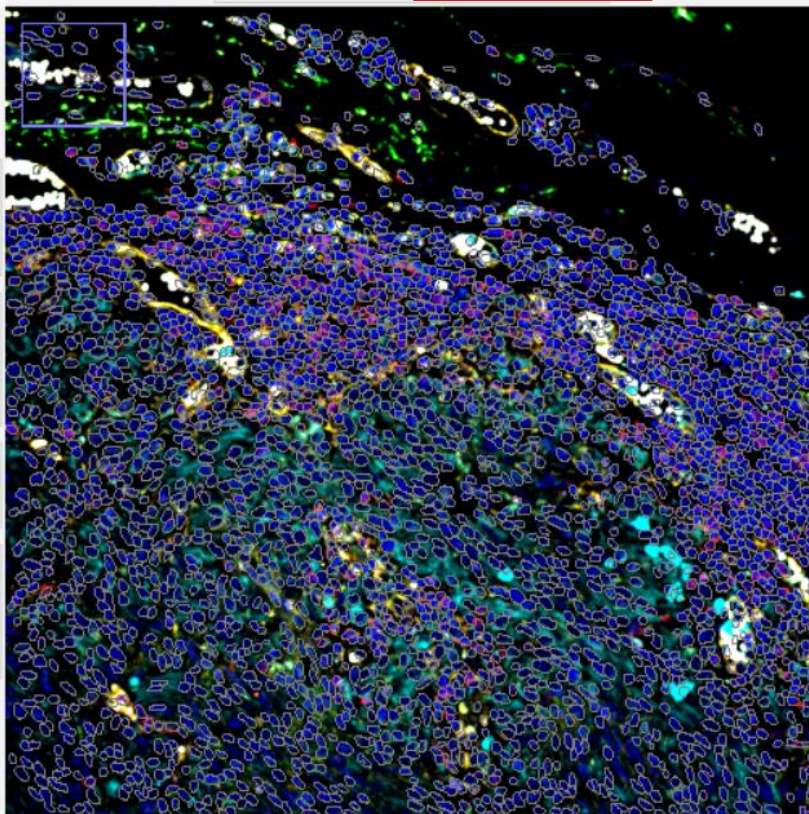
Save

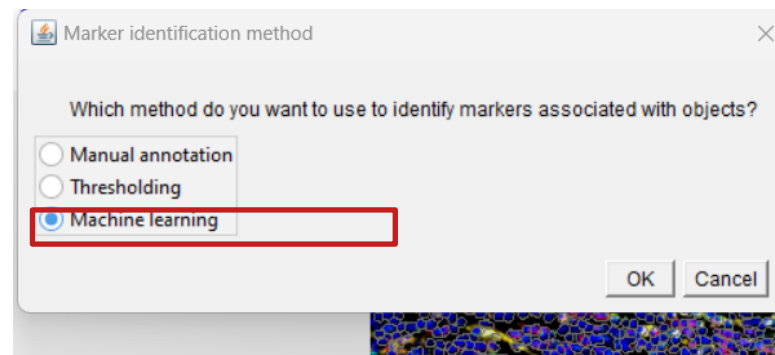
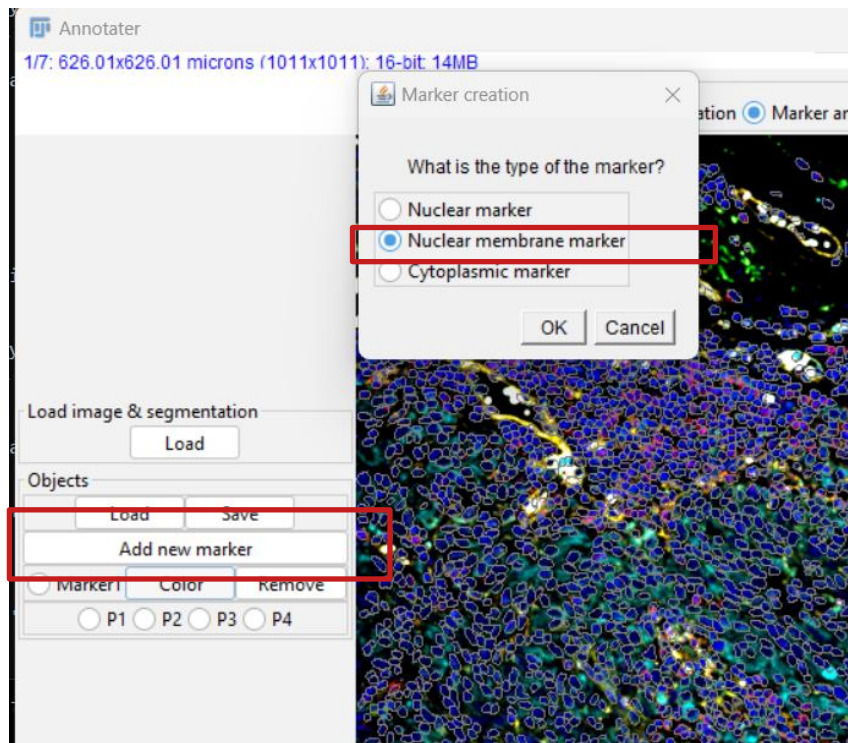
Add new class

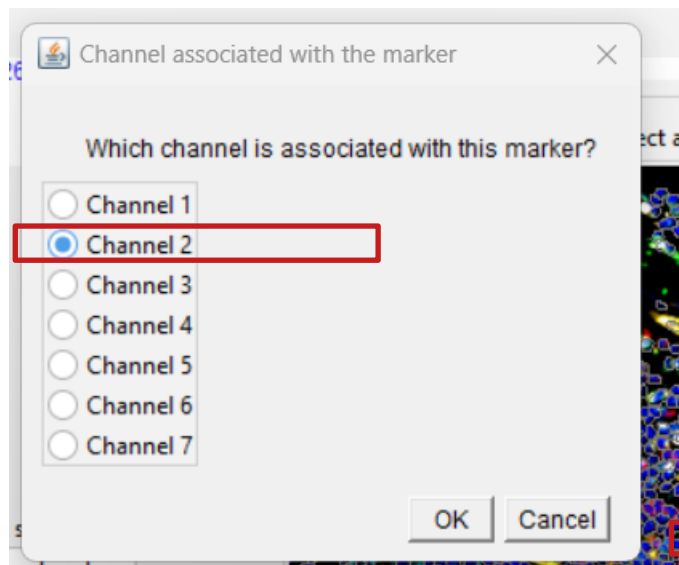
☒ Class 1

Color

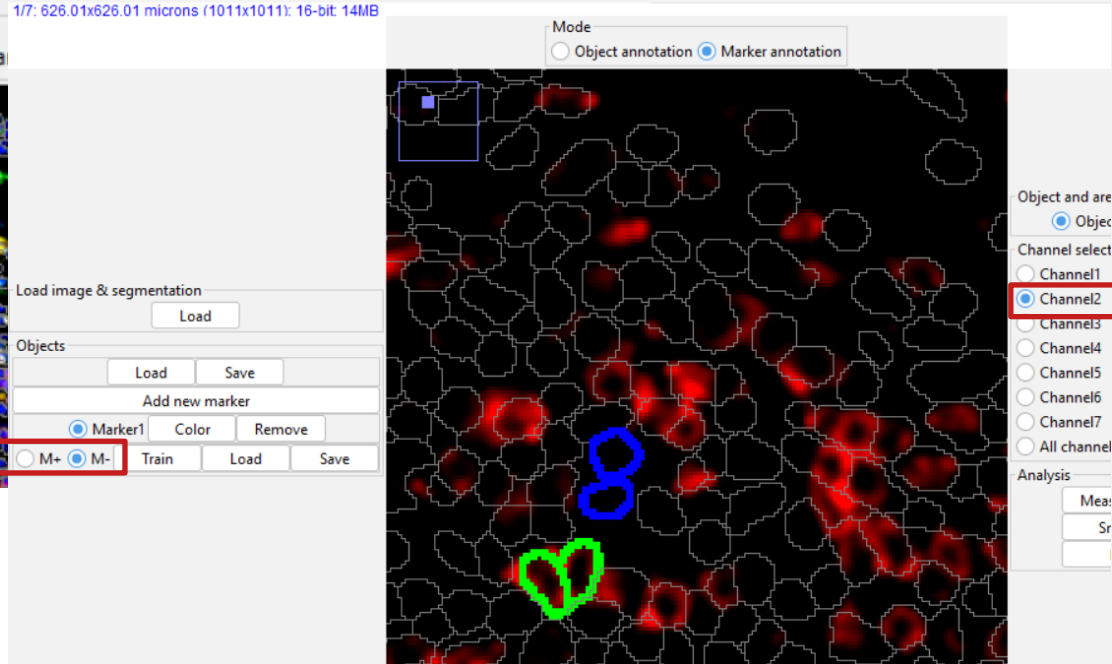
Remove







Entrenamiento_1.tif (300%)
1/7: 626.01x626.01 microns (1011x1011): 16-bit: 14MB



Load image & segmentation

Load

Objects

Load

Save

Add new marker



Marker1

Color

Remove



M+



M-

Train

Load

Save

Data for classifier

Guardar en: Clasificador



Inicio



Escritorio



Bibliotecas



Este equipo



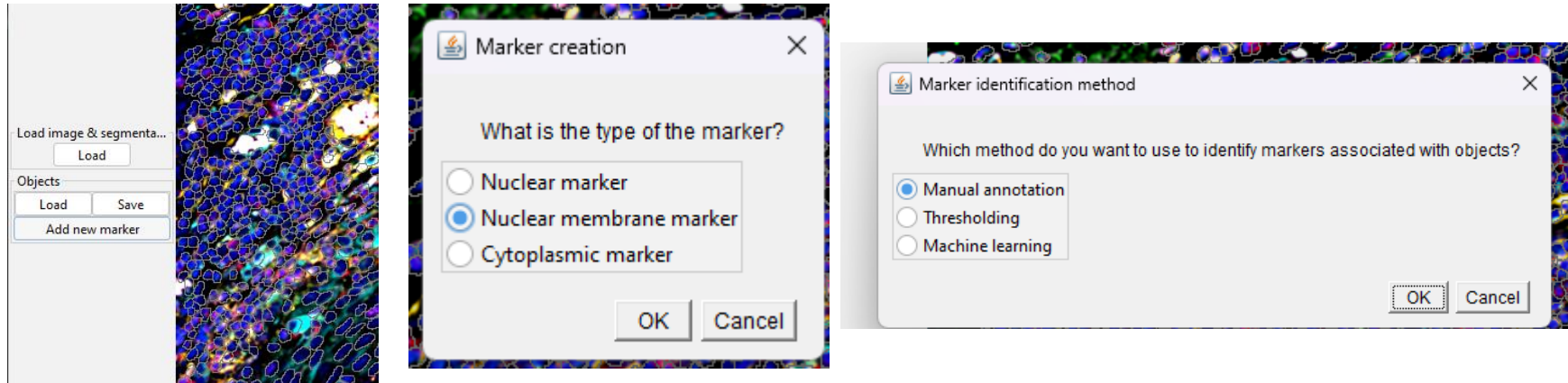
Red

Nombre	Fecha de modificación	Tipo	Tamaño
CD8.csv	14/06/2022 10:11 a. m.	Archivo de valores...	894 KB

Nombre de archivo: CD8.csv

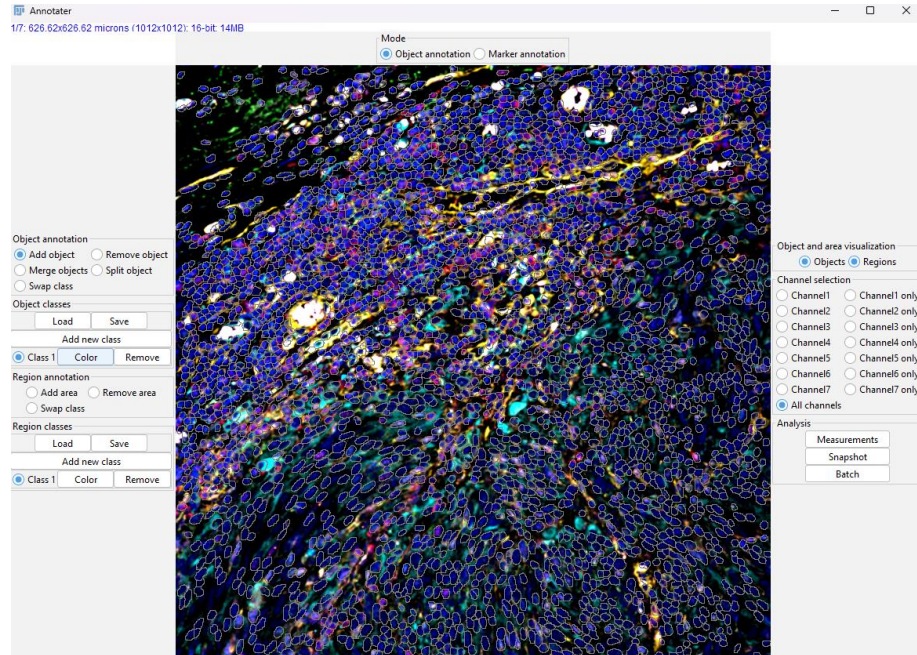
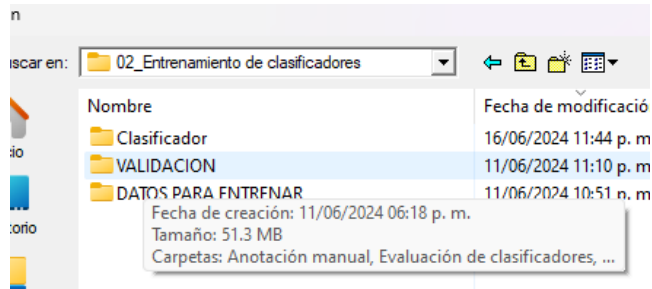
Evaluación de mi clasificador

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



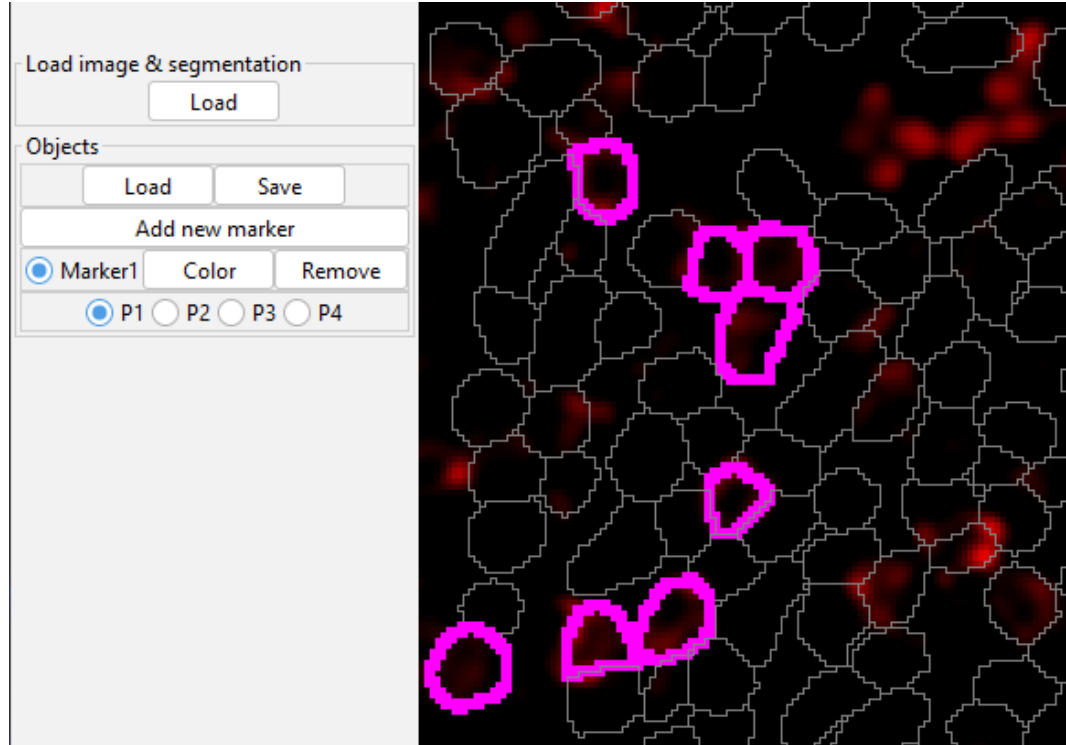
Evaluación de mi clasificador

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



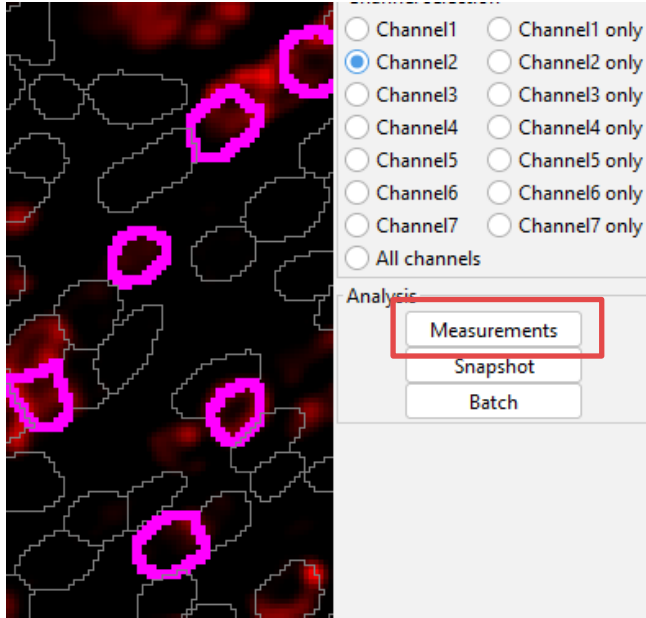
Evaluación de mi clasificador

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

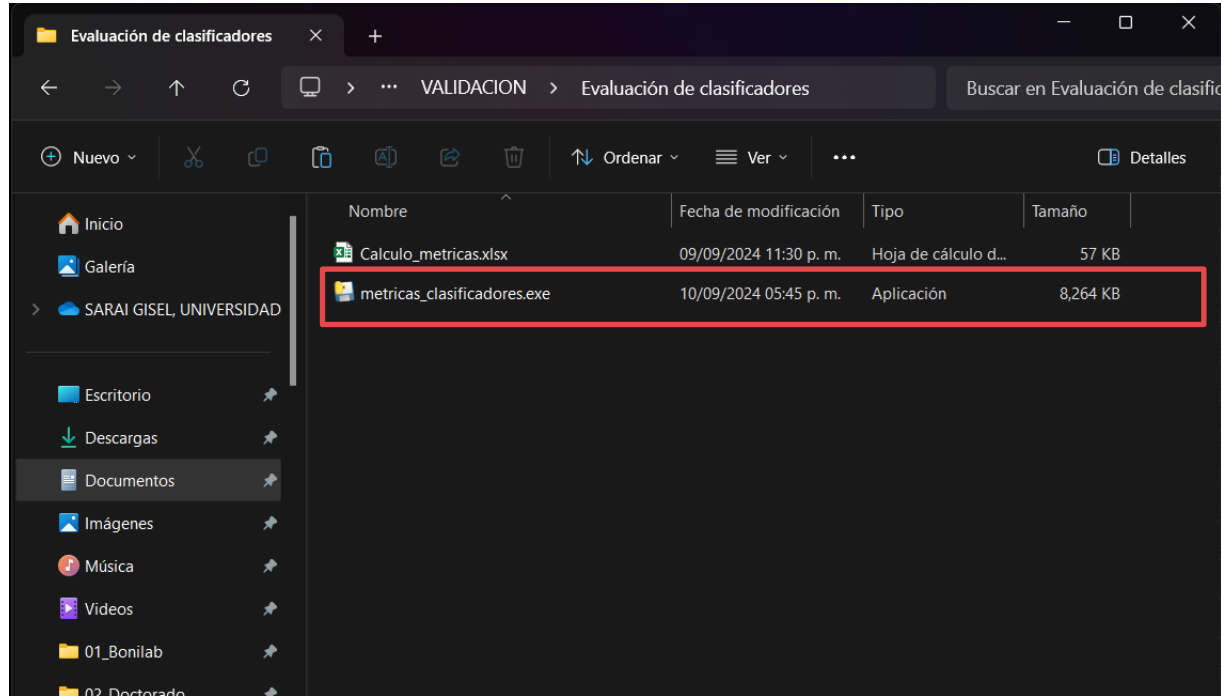


1. Anotación manual.csv

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

Métricas de clasificación

Manual:

Seleccionar Manual

Predicción:

Seleccionar Predicción

Manual +	Prediccion +	True Positives	Falsos Positivos	Falsos Negativos	Precision	Recall	F1 score
0	0	0	0	0	0	0	0

Evaluación de mi clasificador

6. Dar click en “calcular”

Métricas de clasificación

Manual: D:/Users/saras/Documents/01_Bonilab/Melanoma/IA & Machine Learning/Meti Seleccionar Manual

Predicción: D:/Users/saras/Documents/01_Bonilab/Melanoma/IA & Machine Learning/Meti Seleccionar Predicción

Manual +	Predicción +	True Positives	Falsos Positivos	Falsos Negativos	Precision	Recall	F1 score
0	0	0	0	0	0	0	0

Copiar al portapapeles Calcular

Evaluación de mi clasificador

6. Dar click en “copiar al portapapeles”

Métricas de clasificación

Manual: D:/Users/saras/Documents/01_Bonilab/Melanoma/IA & Machine Learning/Meti Seleccionar Manual

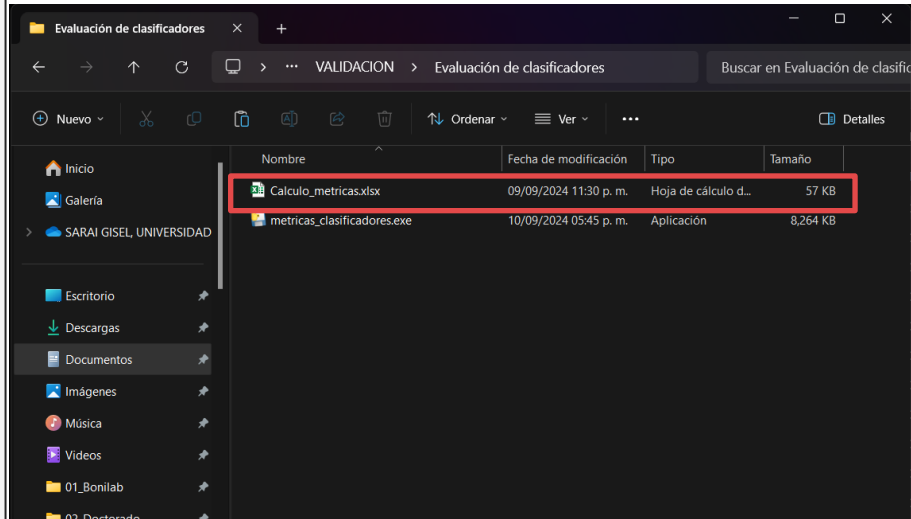
Predicción: D:/Users/saras/Documents/01_Bonilab/Melanoma/IA & Machine Learning/Meti Seleccionar Predicción

Manual +	Prediccion +	True Positives	Falsos Positivos	Falsos Negativos	Precision	Recall	F1 score
218	229	218	11	0	0.9519650655021834	1.0	0.9753914988814317

Copiar al portapapeles Calcular

Evaluación de mi clasificador

7. Pegar en la hoja de calculo "Calculo_Metricas"



Excel spreadsheet titled 'Calculo_metricas.xlsx' showing a table with columns: Imagen, False positives, False negatives, True Positives, Precision, Recall, and F1 score. The data for three images is shown, with the first row highlighted in red.

Imagen	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

Below the table, the formulas for Precision, Recall, and F1 score are shown:

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

Cálculo de métricas

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

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

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

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

$$TP = PREDICTED - FP$$

$$FP = PREDICTED - MANUAL$$

$$FN = MANUAL - PREDICTED$$



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

cell c	MaxCh6 cell c	MaxCh7 cell c	X	Y	Circ.	AR	Round	Solidity	Object marker	Object marker	Object marker	Object marker	Object marker	Object marker
584	619	244	129.56	15.297	0.874	1.614	0.619	0.906	0	0	0	0	0	0
496	1048	1707	70.732	15.627	0.875	1.67	0.599	0.899	0	0	0	0	0	0
581	611	265	251.516	16.367	0.897	1.932	0.518	0.886	0	0	0	1	0	0
311	789	3359	44.102	14.966	0.854	1.442	0.694	0.906	0	0	0	0	0	0
239	682	2474	27.859	19.153	0.701	2.653	0.377	0.847	0	0	0	0	0	0
410	458	434	146.756	16.617	1	1.142	0.876	0.907	1	0	0	0	0	0
506	934	225	108.129	21.312	0.836	1.83	0.546	0.865	0	0	0	0	0	0
484	972	542	78.562	22.652	0.919	1.351	0.74	0.899	0	0	0	0	0	0
606	846	195	349.716	22.168	1	1.373	0.728	0.926	0	0	0	0	0	0
614	919	1735	298.558	25.641	0.872	1.296	0.772	0.878	0	0	0	0	1	0
533	470	253	271.366	25.806	0.846	1.147	0.872	0.896	0	0	0	0	0	0
485	728	723	286.617	30.804	1	1.139	0.878	0.926	0	0	0	0	0	0
449	670	342	91.081	33.29	1	1.333	0.75	0.917	0	0	0	0	0	0
701	702	251	149.435	36.25	0.941	1.204	0.83	0.911	0	0	0	0	0	0
715	670	336	198.728	35.783	0.746	1.146	0.873	0.862	0	0	0	0	0	0
398	1228	3586	39.663	41.591	0.887	1.427	0.701	0.909	0	0	0	0	1	0
513	934	256	263.402	42.443	0.886	1.527	0.655	0.883	0	0	0	1	0	0
586	880	193	269.242	40.481	0.878	1.543	0.648	0.877	0	0	0	0	1	0
428	638	498	69.806	43.251	0.968	1.049	0.953	0.929	0	0	0	0	0	0
468	607	241	87.532	45.389	0.777	1.398	0.715	0.88	0	0	0	0	0	0
457	377	351	221.224	47.691	1	1.096	0.913	0.92	0	0	0	0	0	0
462	813	3392	317.952	51.986	0.801	2.035	0.491	0.915	1	0	0	0	1	0
423	713	1705	92.622	53.197	0.967	1.596	0.626	0.929	0	0	0	0	0	0

Lenguaje

1,2,3,4,5

+ - ?

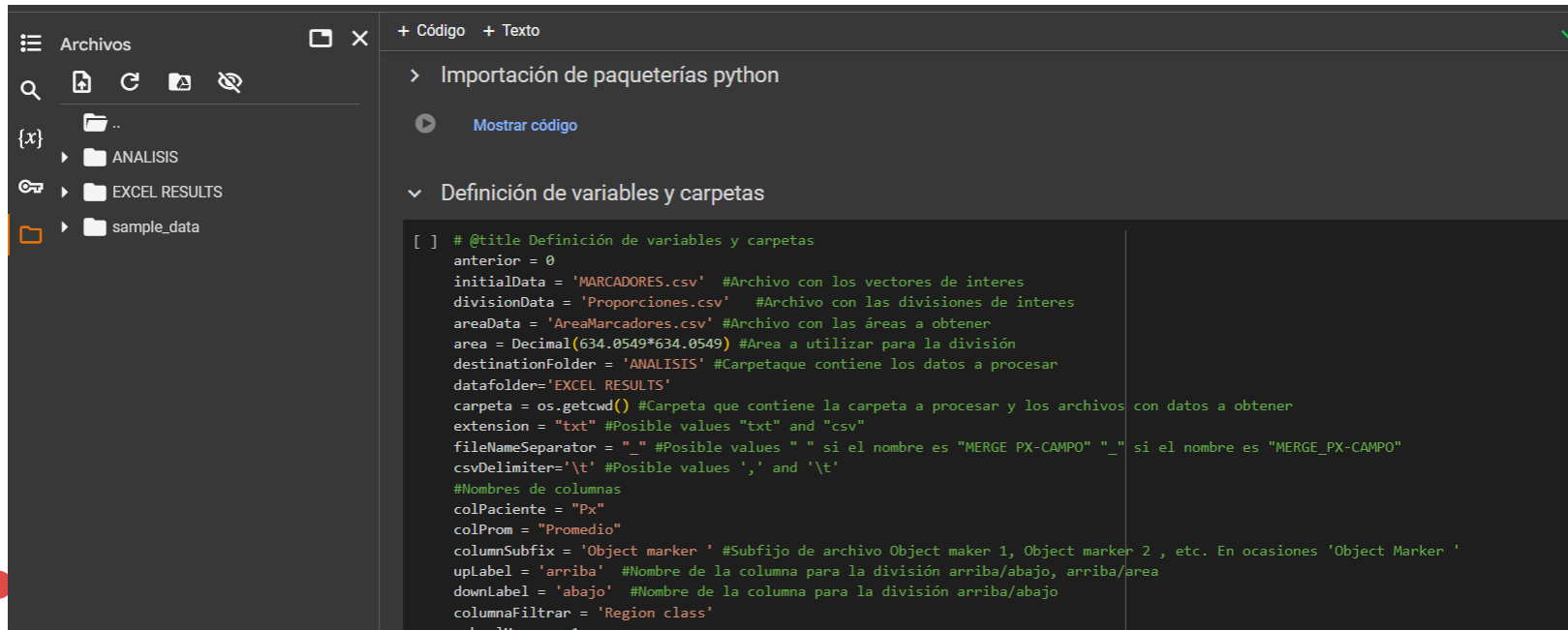
1p2p3p

The screenshot shows the Microsoft Excel interface. The ribbon is set to 'Inicio' (Home). The 'Fuente' (Font) group shows 'Aptos Narrow' font, size 11, with bold (N), italic (K), and underline (S) options. The 'Alineación' (Alignment) group shows text alignment options. The formula bar shows 'F4' and a function icon (fx). The grid below has columns A through F and rows 1 through 8. The data in the grid is as follows:

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

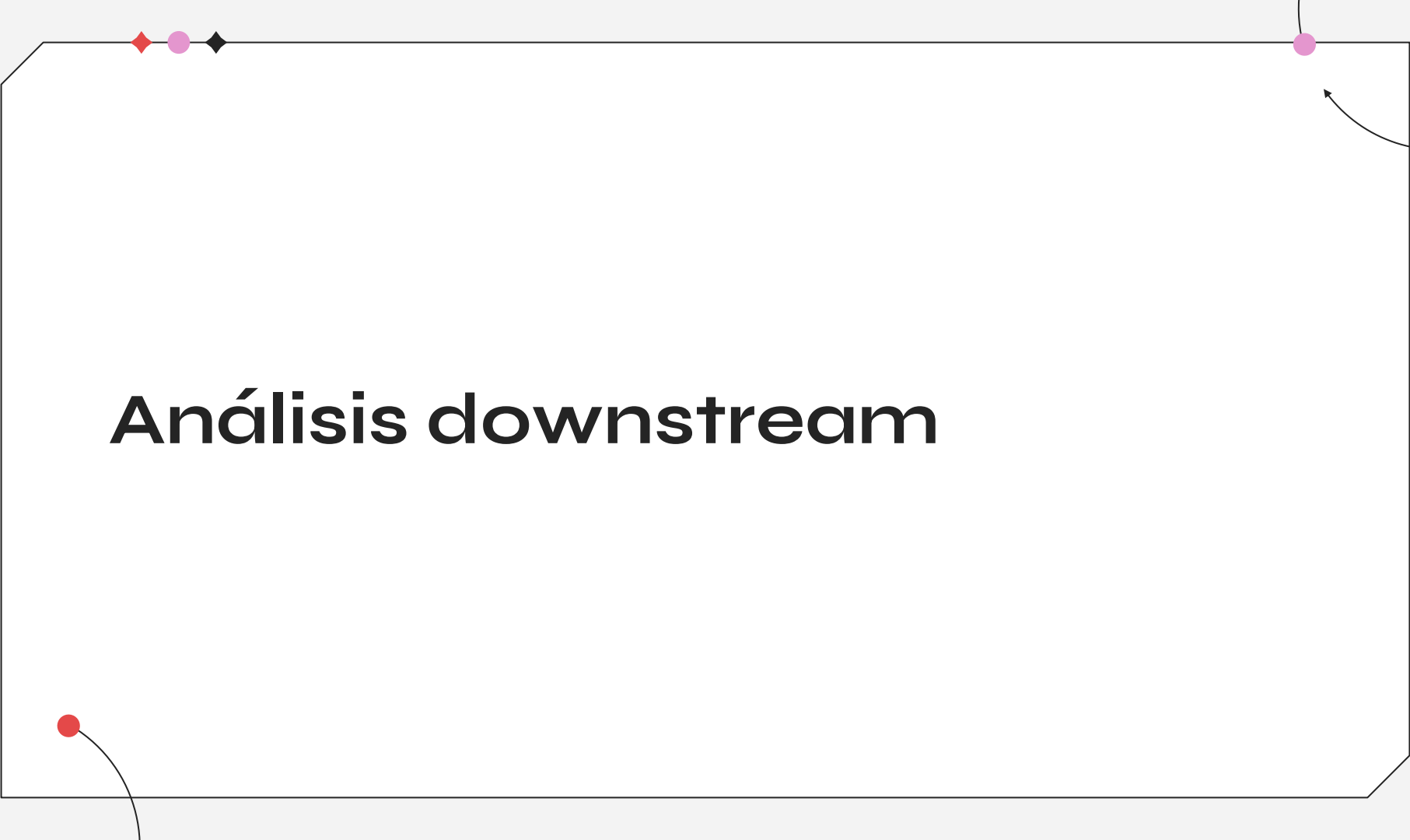
Extracción de datos

Abrir liga



The screenshot shows a code editor with a dark theme. On the left, a file explorer sidebar shows a project structure with folders 'ANALISIS', 'EXCEL RESULTS', and 'sample_data'. The main editor area has a tab titled '+ Código' and contains a Python script. The script is divided into two sections: 'Importación de paqueterías python' and 'Definición de variables y carpetas'. The code defines various variables for file paths, folder names, and column identifiers, using comments to explain their purpose.

```
[ ] # @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' #Carpetaque 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" #Posible values "txt" and "csv"
fileNameSeparator = "_" #Posible values " " si el nombre es "MERGE PX-CAMPO" " _" si el nombre es "MERGE_PX-CAMPO"
csvDelimiter='\t' #Posible values ',' and '\t'
#Nombres de columnas
colPaciente = "Px"
colProm = "Promedio"
columnSubfix = 'Object marker ' #Subfijo de archivo Object maker 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'
columnaFiltrar = 1
```

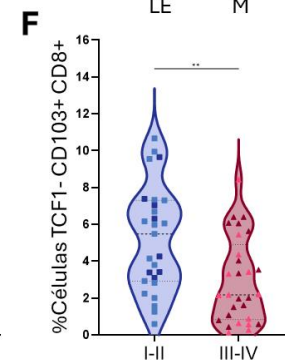
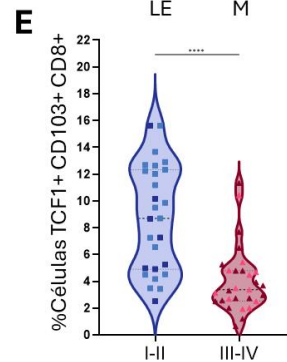
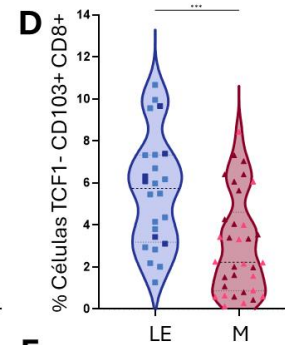
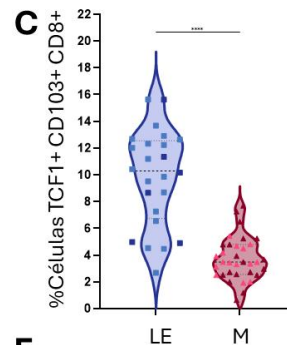
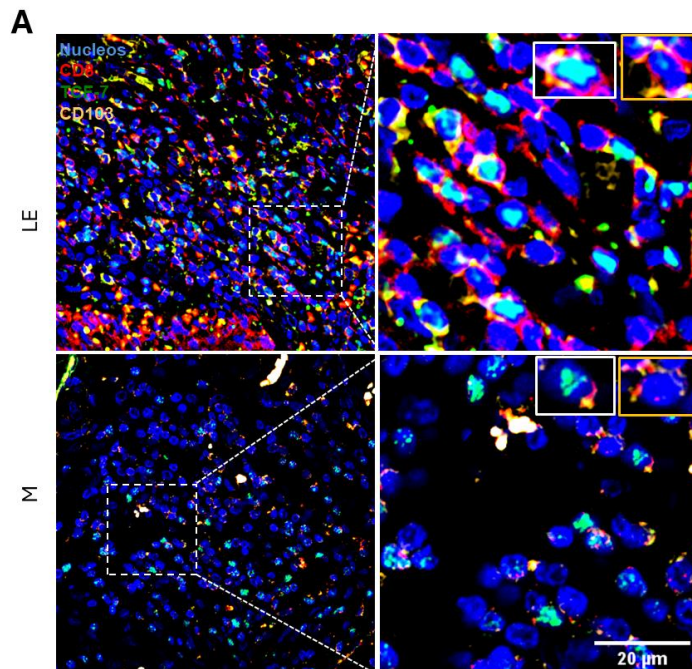


Análisis downstream

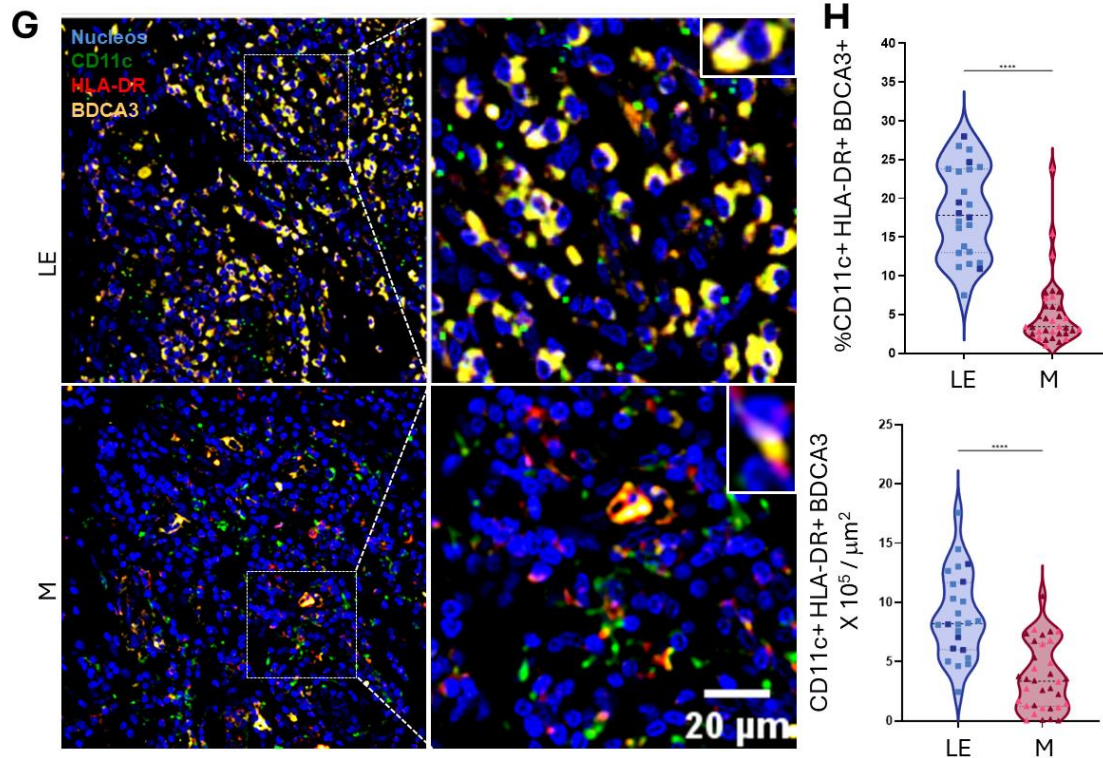
TCF1-positive and TCF1-negative TRM CD8 T cell subsets and cDC1s orchestrate melanoma protection and immunotherapy response

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



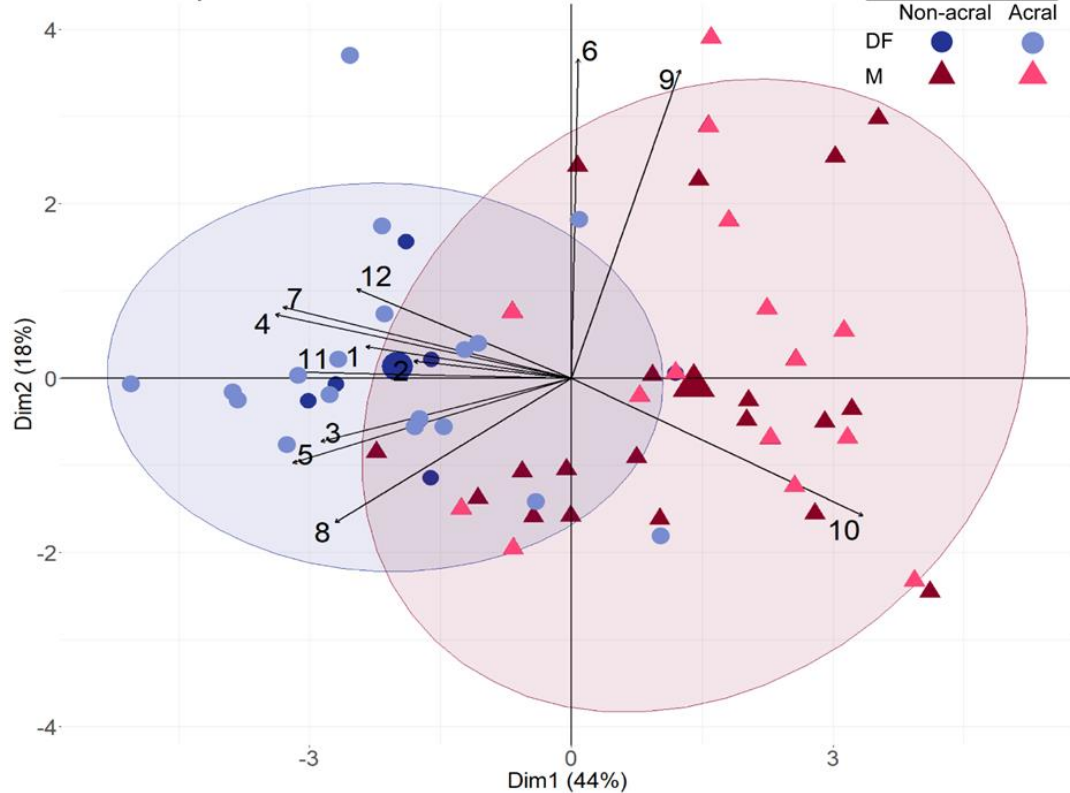
Linfocitos CD8 Residentes de tejido TCF7+ y TCF7- relacionados al control de la enfermedad

Base de datos

Clave

4170
4175
550
654
906
910
1118
11750
1910
2163
2521
2816
2972
3020
3255
3411
3479
3558
3748
3771
3783
3955
4034
4175
4222
4232
4504
4728
5343
6339
6345
6484
6675
6793
7004
7095
7192
7267
7282
7342
7493
7795
8003
8062
8152
8524
8532
8905
9251
9619
13173
13405
15243
2347 EF
2613 BC
5160 B
736 AH
B294

PCA & Biplot

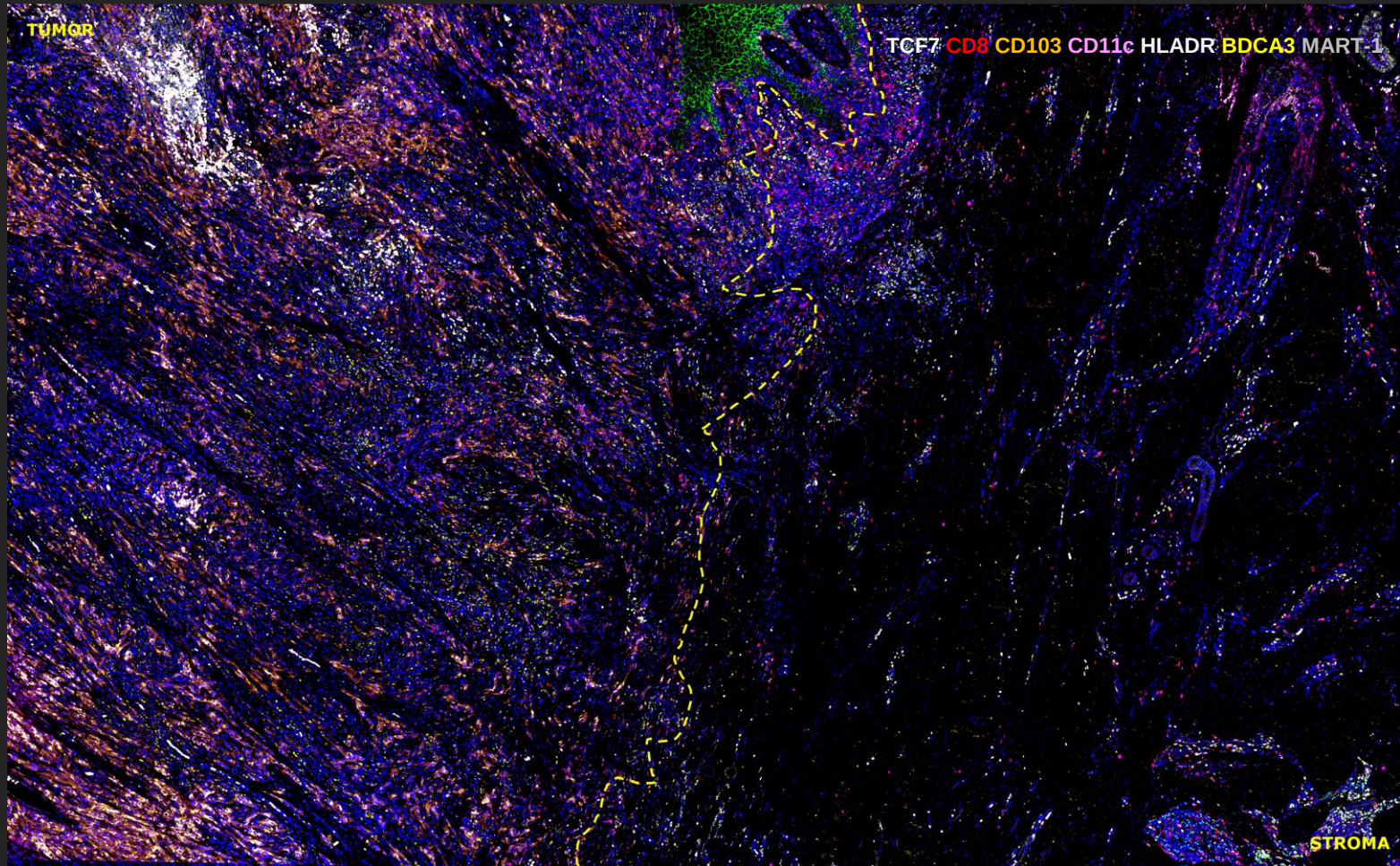


Base de datos

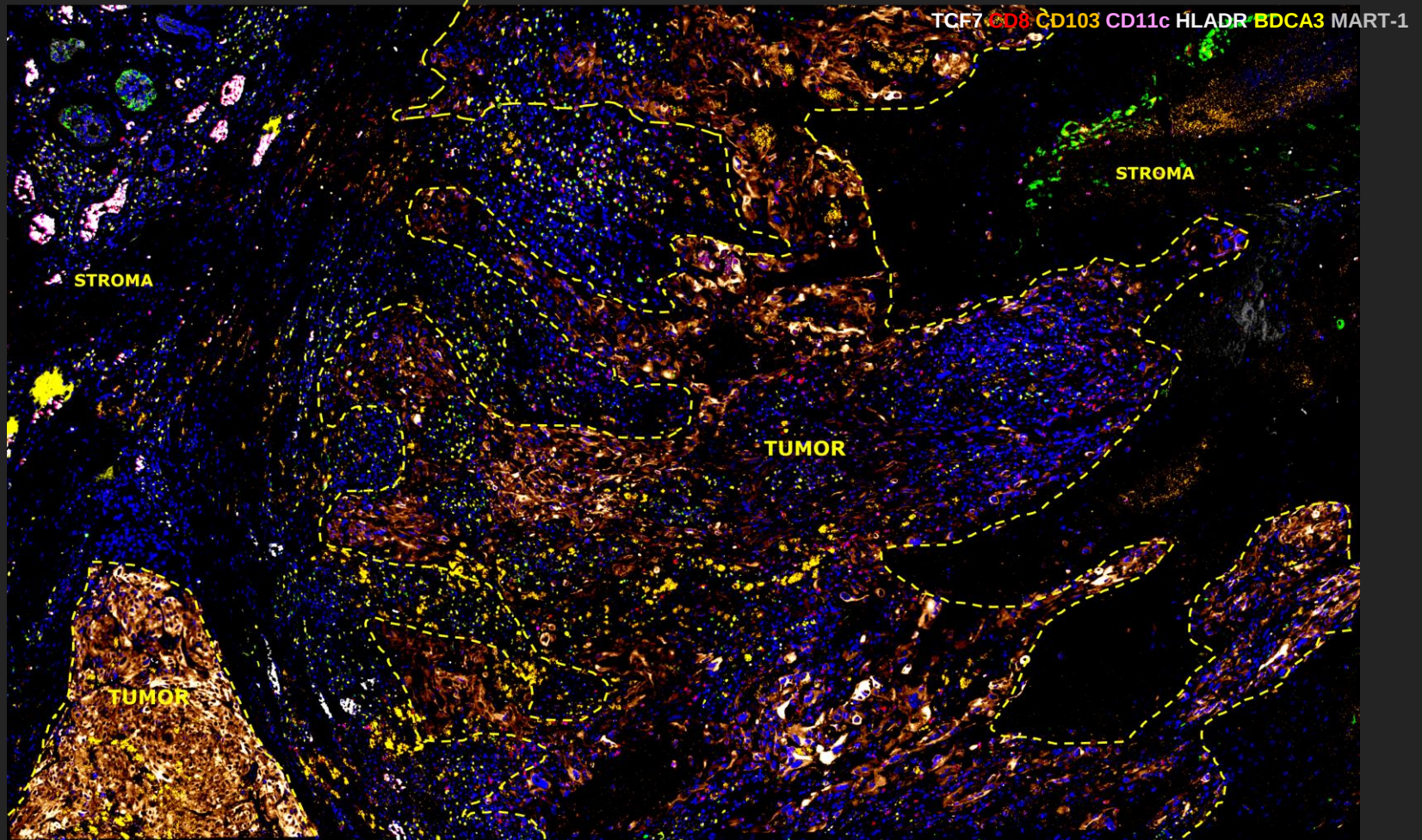
Clave

37.36504608
35.00617365
30.52224476
13.46064603
7.822009557
35.34351502
35.37845443
33.64127398
37.66355735
43.63450412
37.80193636
8.62205143
41.65574633
14.0059574
19.20623019
116.3758633
38.62565122
34.60664218
33.56091264
14.42467023
5.794535383
14.21520072
25.50565433
23.51792912
44.57822541
40.16610163
37.52971508
18.33906799
34.63373806
36.14356727
7.860454576
36.71206554
42.44482812
15.05957403
43.14410717
36.24483531
40.78563601
26.24595376
37.41634537
21.36684065
35.18223594
37.18133399
42.31940583
25.63630823
39.17717523
31.30101832
44.81503451
23.34319319
28.00177082
28.00177082
23.14542295
24.38194274
28.83780148
27.22255883
31.26810243
38.01037241
27.27375306
30.52430547

Escaneos del tejido tumoral y estromal: Paciente Libre de Enfermedad

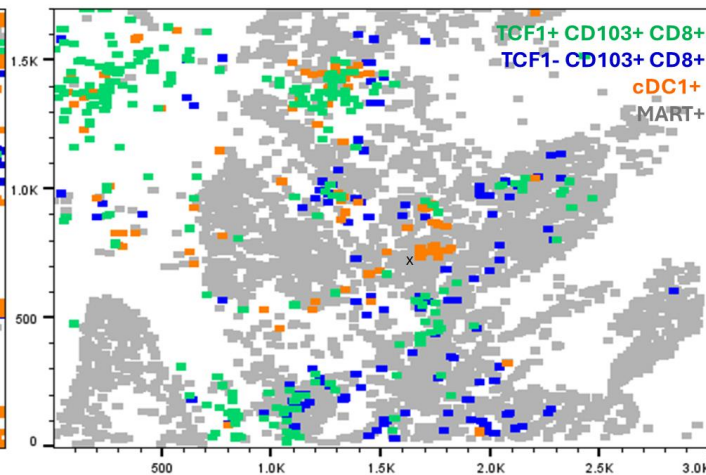
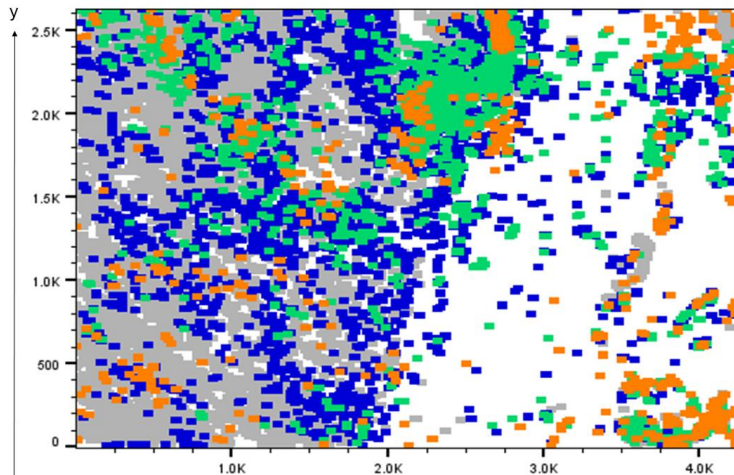
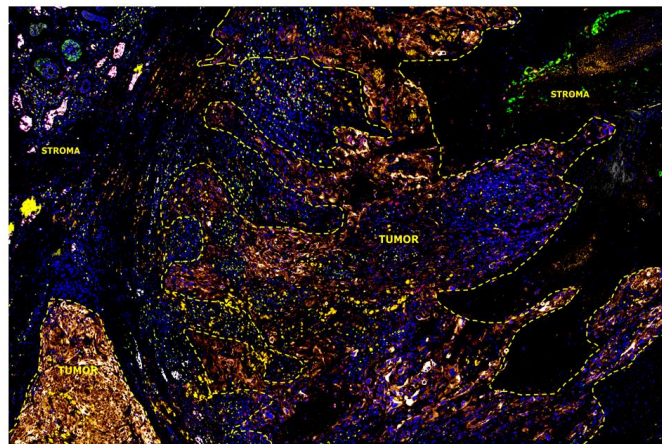
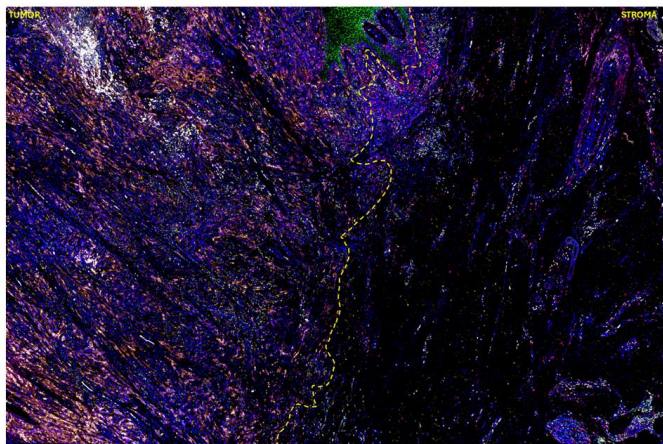


Escaneos del tejido tumoral y estromal: Paciente Metastásico



Correlación espacial de las poblaciones inmunes

D



E

TCF1+ CD103+ CD8+

14.66%

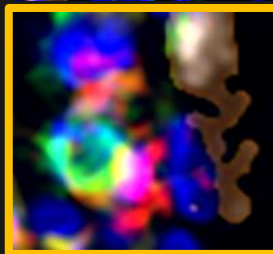
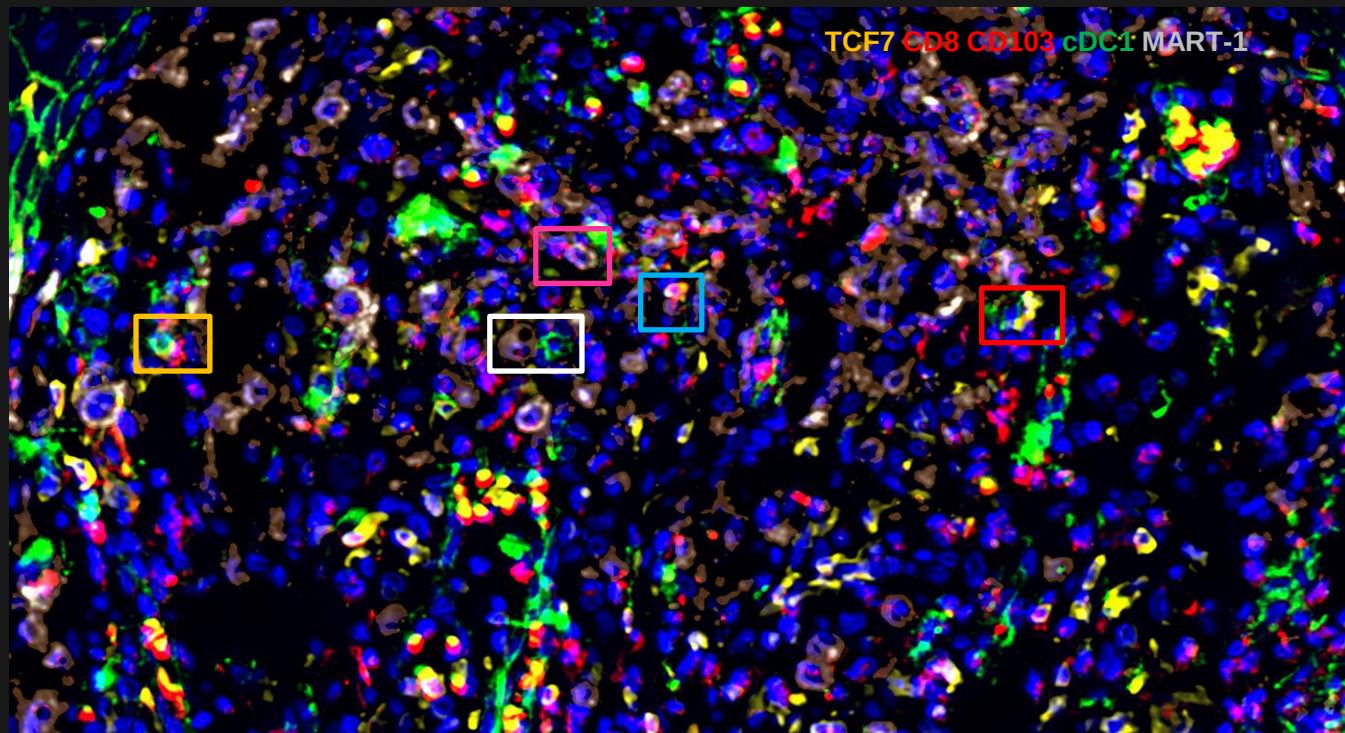
TCF1- CD103+ CD8+

19.78%

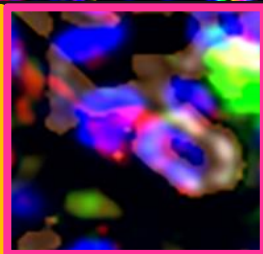
cDC1

26.56%

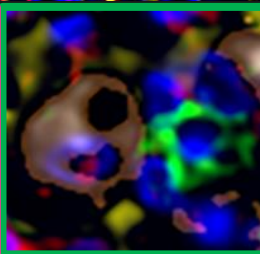
STROMA
TUMOR



cDC1 vs TCF1- CD103+
CD8+



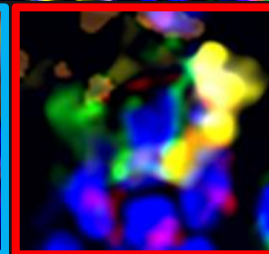
TCF1- CD103+ CD8+ VS
MART-1+



cDC1 VS MART-1+



TCF1+ CD103+ CD8+ VS
MART-1+



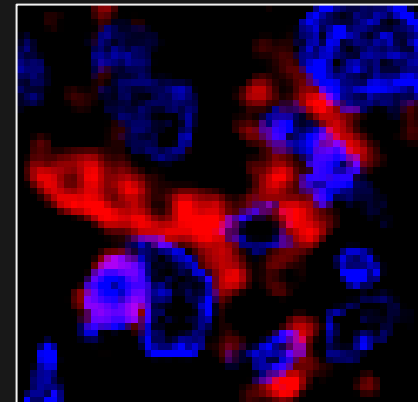
cDC1 vs TCF1+ CD103+
CD8+

Libre de enfermedades
CD8 CD103 PD1 GZB LAMP1

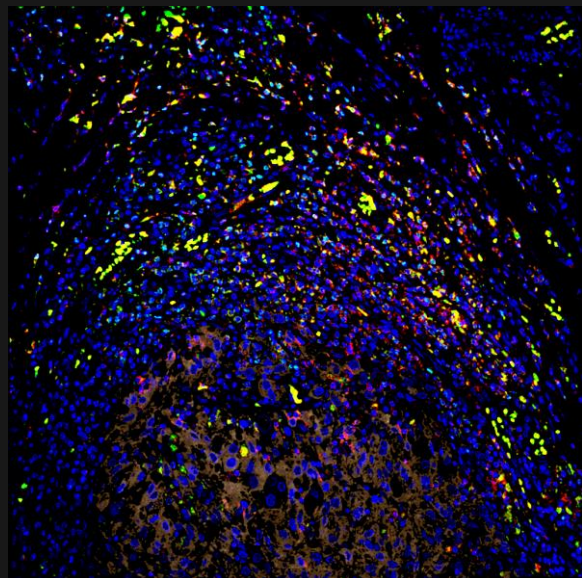
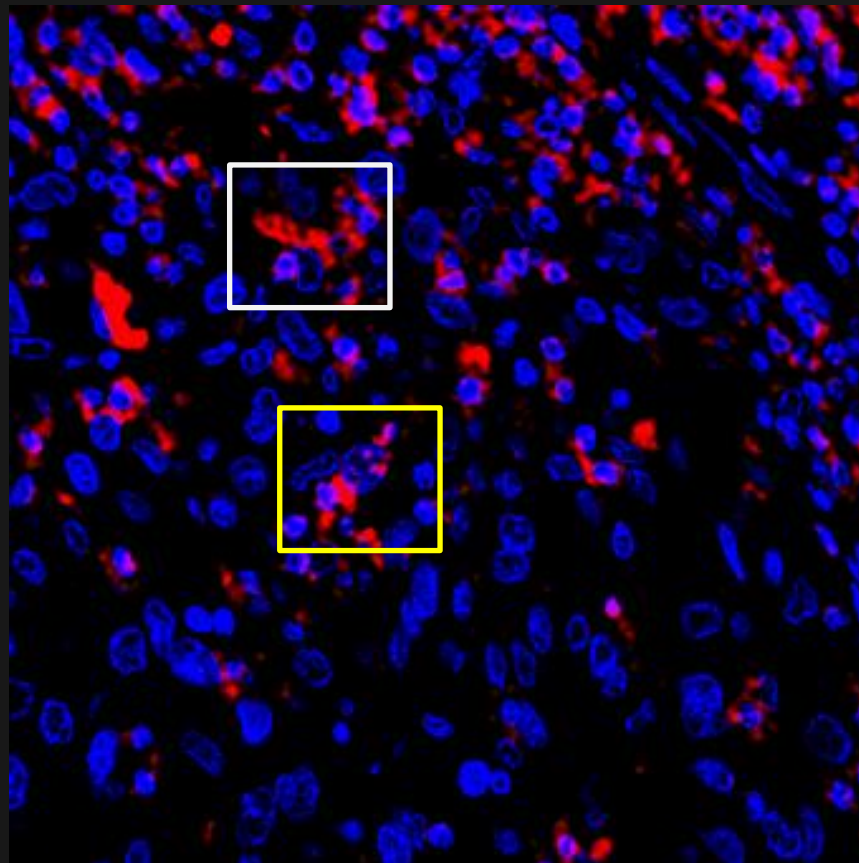
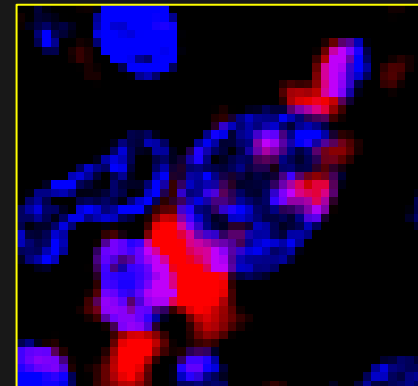
TUMOR CASP3

TCF7

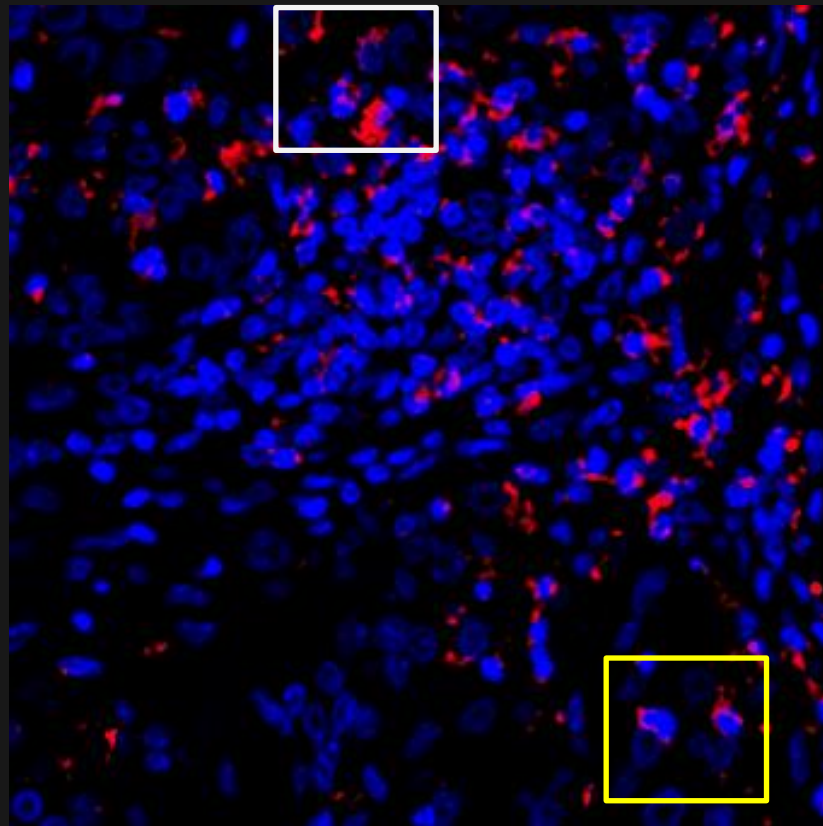
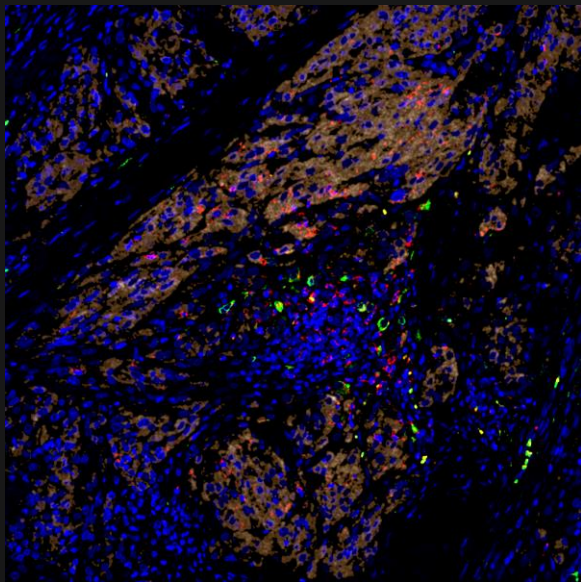
TCF7+



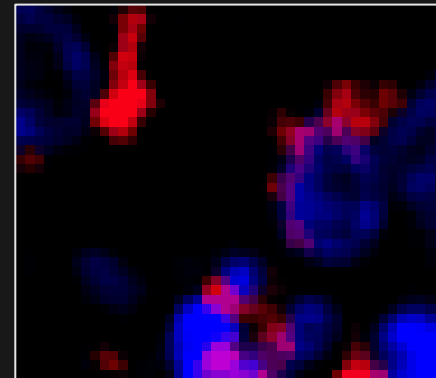
TCF7-



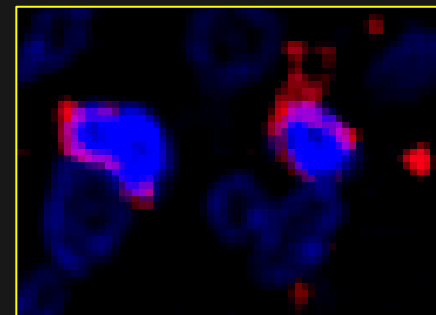
Metastásico
CD8 CD103 PD1 GZB LAMP1
TUMOR CASP3
TCF7



TCF7+

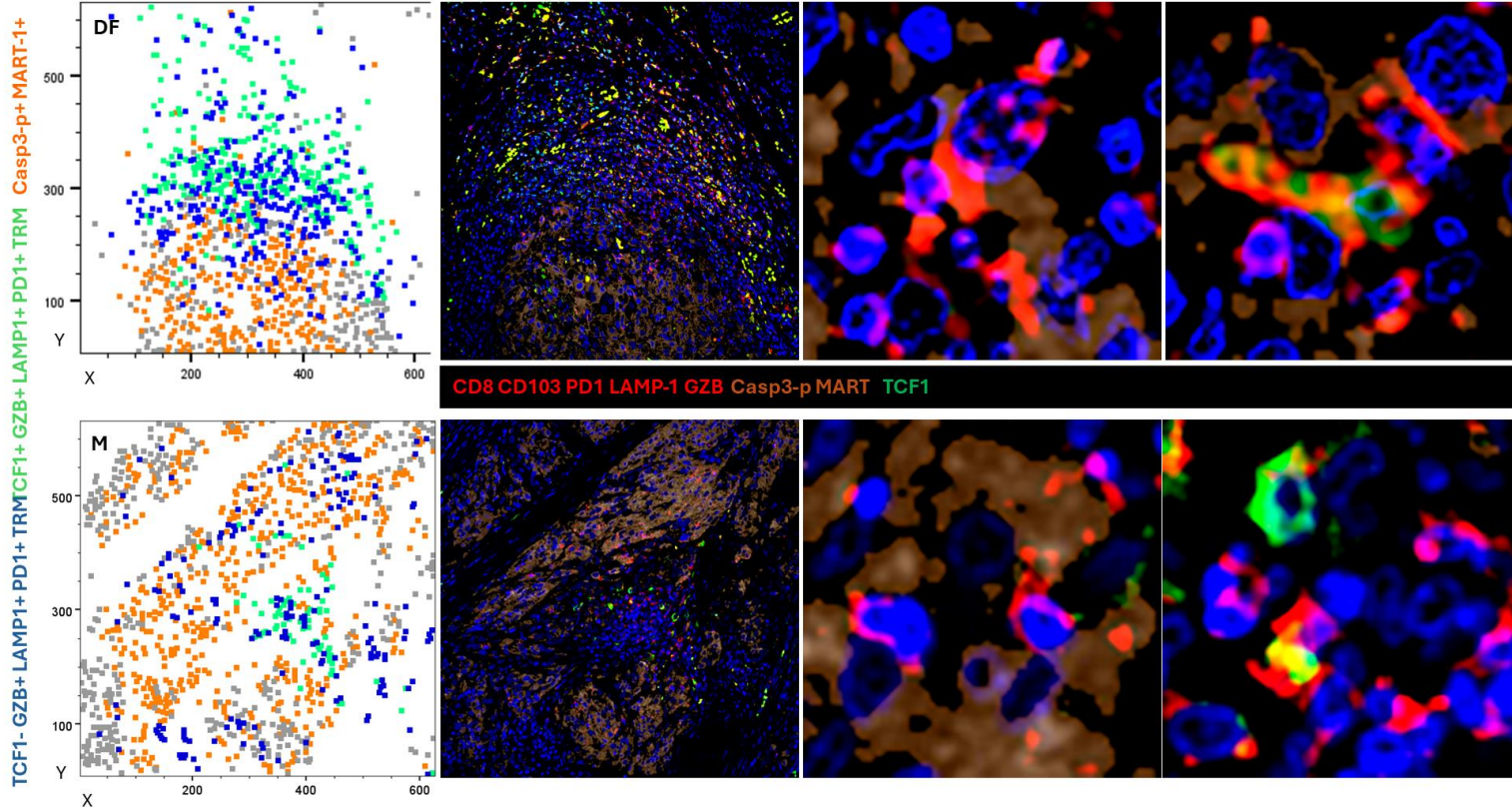


TCF7-

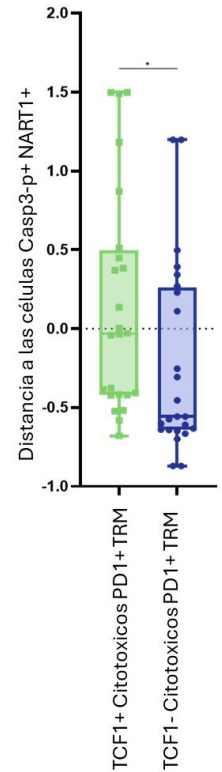


Los linfocitos TCF1- TRM citotóxicos se encuentran más cerca de las células MART-1+ Casp-3-p+

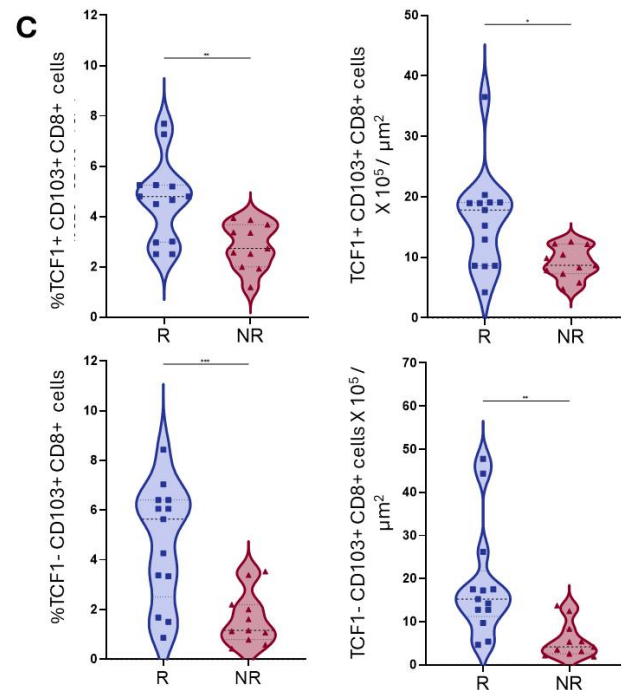
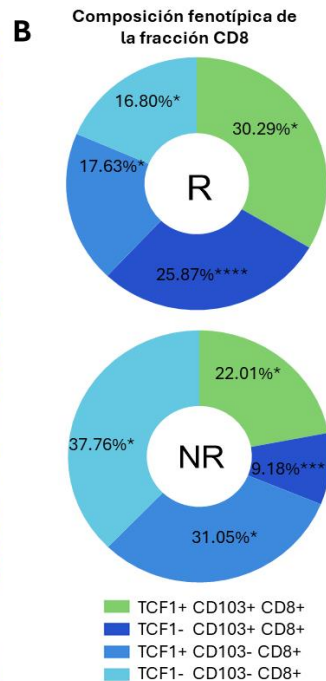
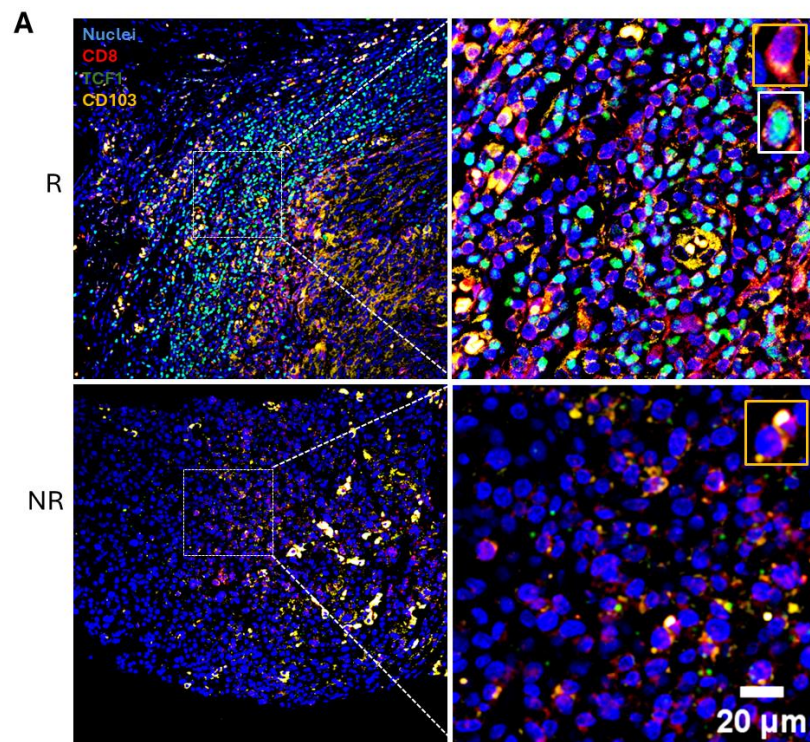
A



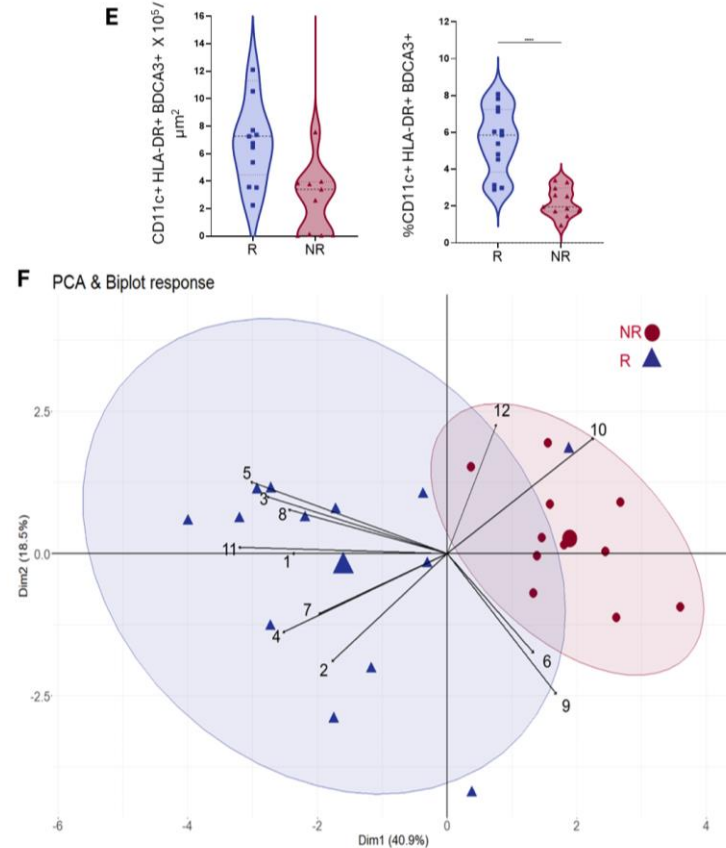
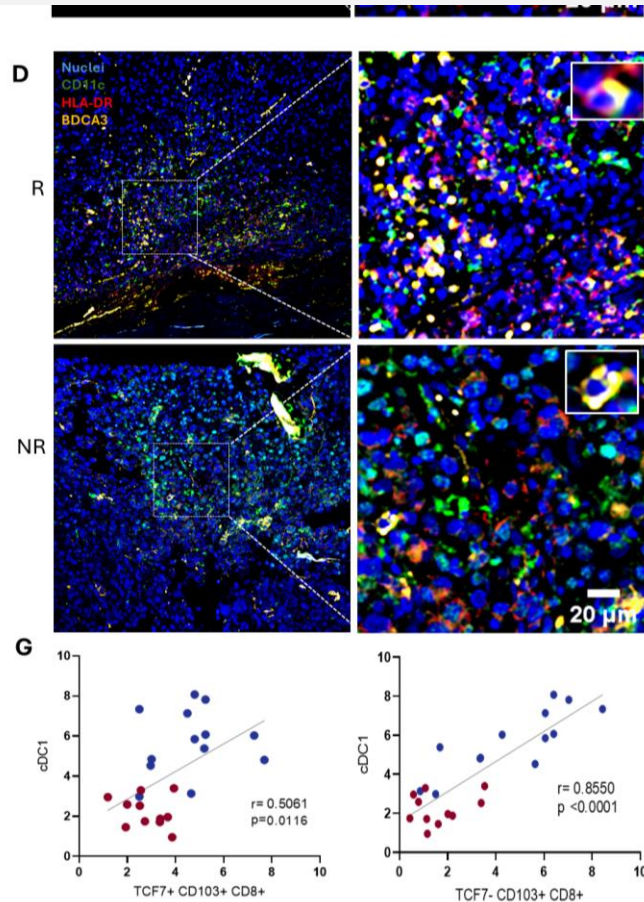
B

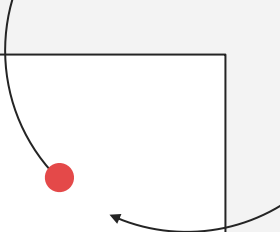


Linfocitos residentes de tejido que expresan TCF7+ y TCF7- están relacionados a la respuesta a CBI



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





Gracias!

Contacto

sara.g.deleonr@gmail.com

5568745474

