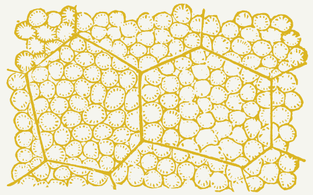
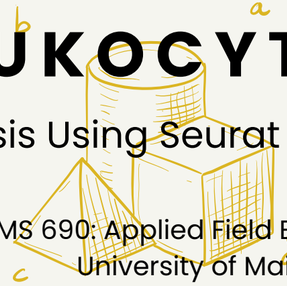


MALIGNANT MELANOCYTES SHOW SIGNIFICANT UPREGULATION OF CLASS 2 HUMAN LEUKOCYTIC ANTIGENS

<scRNA-seq Analysis Using Seurat and 10X Genomics Data>



BMS 690: Applied Field Experience
University of Maine

MICHAEL GAGE

THE DATA/EXPERIMENTAL DESIGN

Species: Human

Anatomical Entity: Blood

Disease State: Melanoma

Preservation: Fresh Frozen

Cell Types: Dissociated Tumor Cells (DTCs)

Cell Count: 10,645 (biological replicates)

Samples: 1 (non-multiplexed analysis)

Donor Count: 1

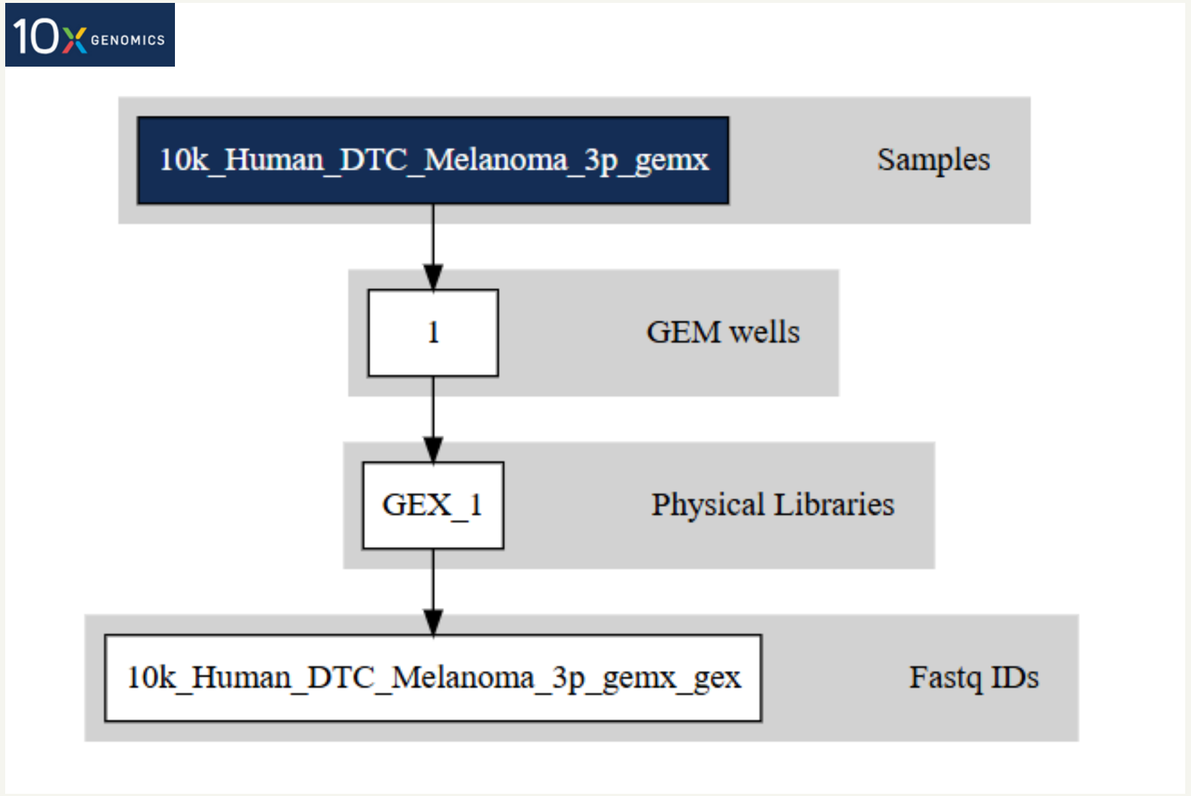
Donor Sex: Female

Donor Age: 50-60

Paired or Single: Paired End Reads

Confidently Mapped Reads: 96.57%

Sequencer: Illumina NovaSeq 6000

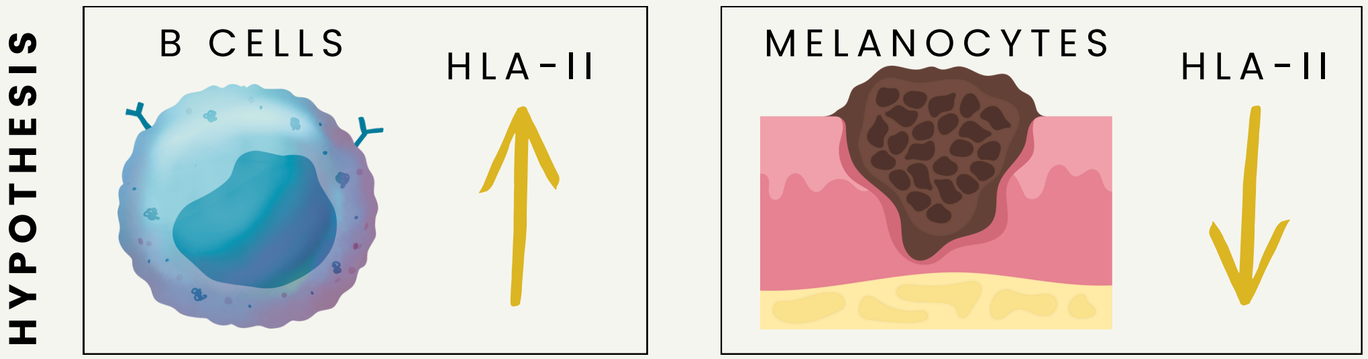


GOAL OF THE ANALYSIS

Research Question: Do 10X melanoma cells express Class 2 Human Leukocytic Antigens (HLA class II)?

Hypothesis: In this scRNA-seq dataset, B-cell clusters will show clear upregulation of HLA class II genes (e.g., HLA-DOA/DOB/DPA/DPB/DQA/DQB/DRA/DRB, with HLA-DMA/DMB), whereas the melanoma (melanocyte) clusters will not exhibit a coordinated HLA-II program.

Goal: To determine the presence of HLA class II genes across clusters and to discover what cell types are associated with these clusters.



BACKGROUND

What are Human Leukocytic Antigens (HLAs)?

Cell-surface glycopeptides that present endo- and exogenous antigens to T lymphocytes, enabling the immune system to characterize 'self' from 'non-self'. These are also known as human major histocompatibility complexes (MHCs) and provide the foundation for infection and disease immunity.

HLA CLASS I

Expressed by almost all nucleated cells, presenting antigens to CD8+ T cells.

HLA-A	HLA-J	HLA-T
HLA-B	HLA-K	HLA-U
HLA-C	HLA-L	HLA-V
HLA-E	HLA-N	HLA-W
HLA-F	HLA-P	HLA-X
HLA-G	HLA-R	HLA-Y
HLA-H	HLA-S	HLA-Z

HLA CLASS II

Expressed in dendritic cells, macrophages, and B cells, presenting antigens to CD4+ T cells.

HLA-D →

HLA-DOA	HLA-DRB
HLA-DOB	HLA-DMA
HLA-DPA	HLA-DMB
HLA-DPB	
HLA-DQA	
HLA-DQB	
HLA-DRA	

(Ryan & Cobb, 2012); (Marsh et al., 2010)

SUMMARY OF METHODS

LOAD PACKAGES & SETUP THE SEURAT OBJECT (R)

Packages: dplyr, Seurat, patchwork, hdf5r, SingleR

- .h5 (HDF5) file was used to conserve space locally
- Read10x_h5() function was used to load data
- A Seurat object was made with non-normalized data
- 29,104 features across 10,367 'samples' (cells)

STANDARD PRE-PROCESSING

- Percent of reads mapped to mitochondrial genome (-MT) was extracted and added to the Seurat object.
  - MT data was visualized in a feature scatter plot, showing a 0.85 correlation between nFeature and nCount.
- The Seurat object was subset to a feature count between 200 and 2500, with a %MT below 12.

NORMALIZING THE DATA

- By default, Seurat employs a global-scaling normalization method that normalizes the feature expression measurements for each cell by the total expression, multiplies this by a scale factor (10,000 by default), and log-transforms the result.
- NormalizeData()

FEATURE SELECTION

- A subset of features with high cell-to-cell variation was calculated (2000)
  - FindVariableFeatures()
- Top ten variable features were plotted and labeled
- VariableFeaturePlot()

*code for each step can be found in the associated RMarkdown file

SUMMARY OF METHODS CONT.

SCALING THE DATA

- Data was scaled by applying a linear transformation prior to dimensional reduction
  - ScaleData()

LINEAR DIMENSIONAL REDUCTION

- Principle Component Analysis was performed, and results were examined
  - RunPCA()
- PCA results were visualized in a Heatmap to qualitatively determine which PC's had the best bimodal patterns
- DimHeatmap()

DETERMINE DIMENSIONALITY

- A ranking of principle components based on the percentage of variance
  - An elbow plot was developed to visualize the ranking, showing majority of true signal in the first 15 PCs
- ElbowPlot()

CELL CLUSTERING (UMAP/tSNE)

- A K-nearest neighbor (KNN) graph was constructed and Jaccard similarity was utilized to develop clusters for the first 15 PCs
- FindNeighbors()
- Modularity optimization (Louvain algorithm) was applied
- FindClusters()
- UMAP and tSNE created

*code for each step can be found in the associated RMarkdown file

SUMMARY OF METHODS CONT.

FIND DIFFERENTIALLY EXPRESSED FEATURES

- Markers (features) were determined for every cluster compared to all remaining cells (differential gene expression)
  - FindAllMarkers()
- Violin and Feature plots were used to visualize select marker genes from each cluster
- Top ten markers per cluster were visualized with DoHeatmap()

CLUSTER IDENTIFICATION

- Clusters were identified automatically with SingleR and manually. Both results were compared
  - Manual research tools: *Human Protein Atlas, PanglaoDB, and CellMarker2.0*
- RenameIdents() was used to assign manual labels to each cluster

VISUALIZING HLA CLASS II GENES

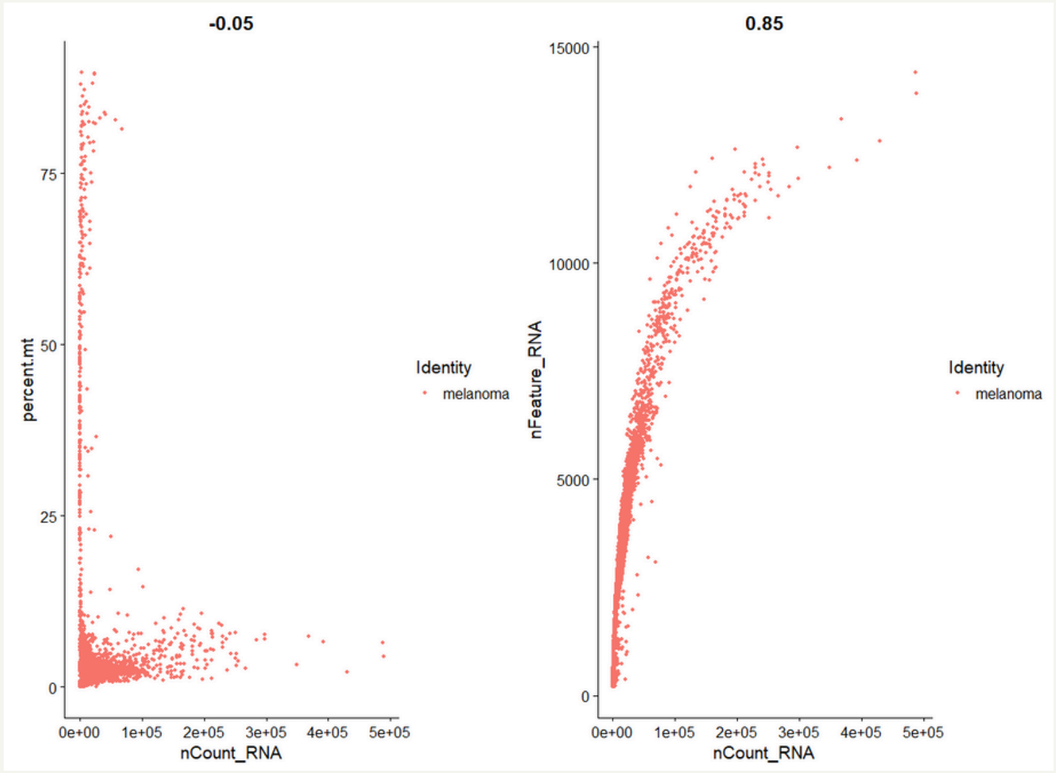
- HLA Class II (HLA-D) genes were visualized across clusters
  - VlnPlot(), FeaturePlot(),

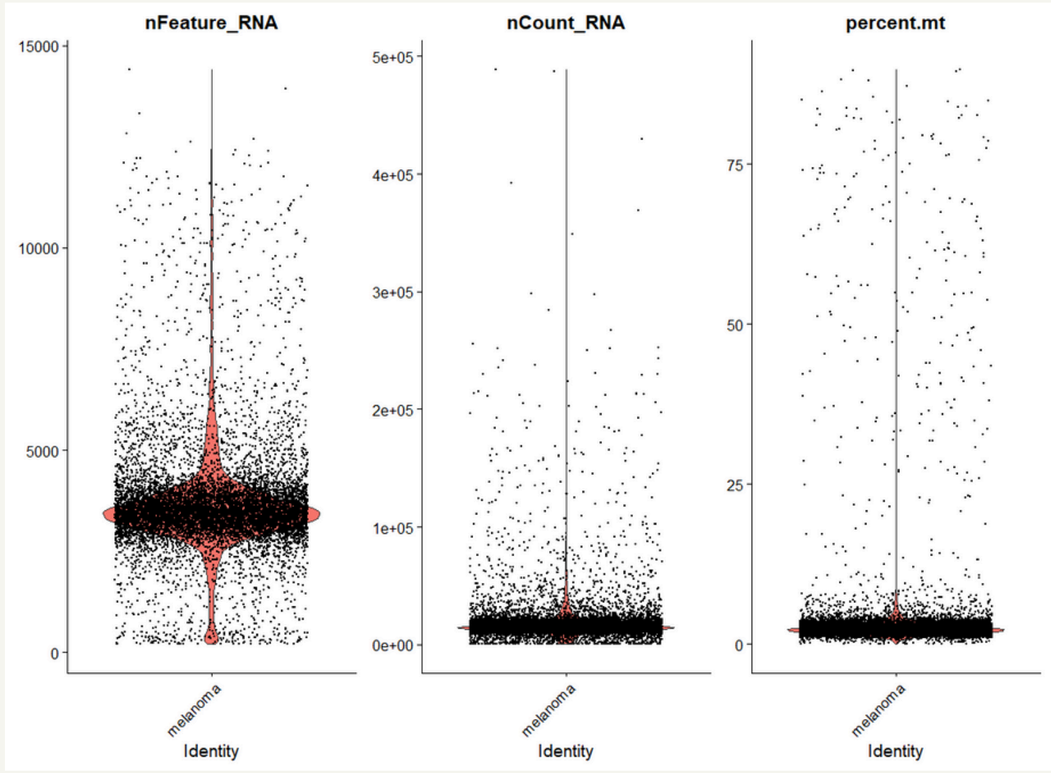
BIOLOGICAL INTERPRETATION OF RESULTS

- Distribution of HLA class II gene up-regulation across clusters was compared to current research for interpretation

*code for each step can be found in the associated RMarkdown file

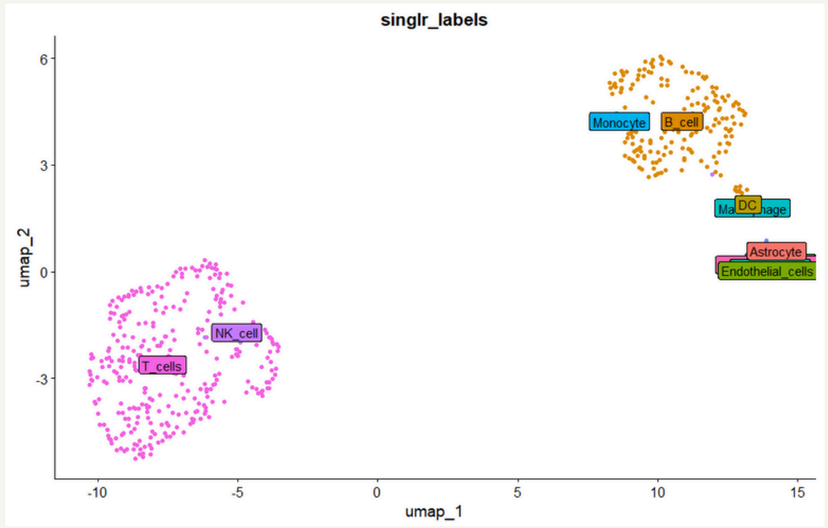
QUALITY OF DATA

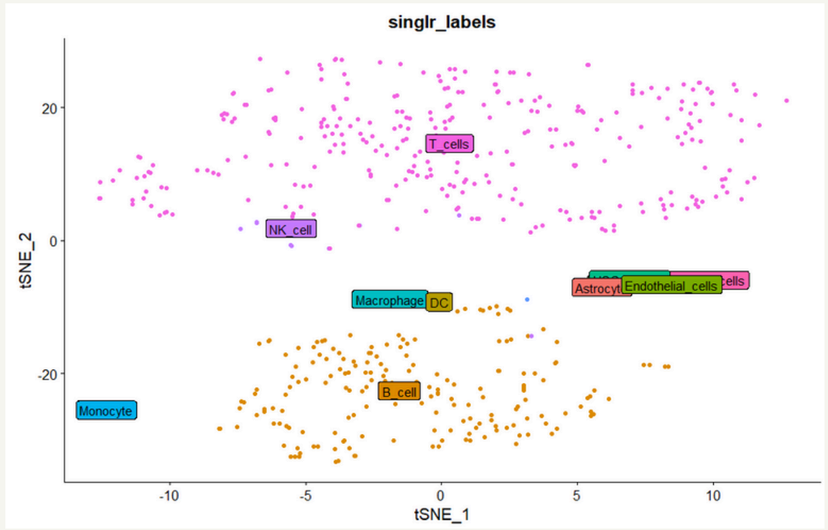




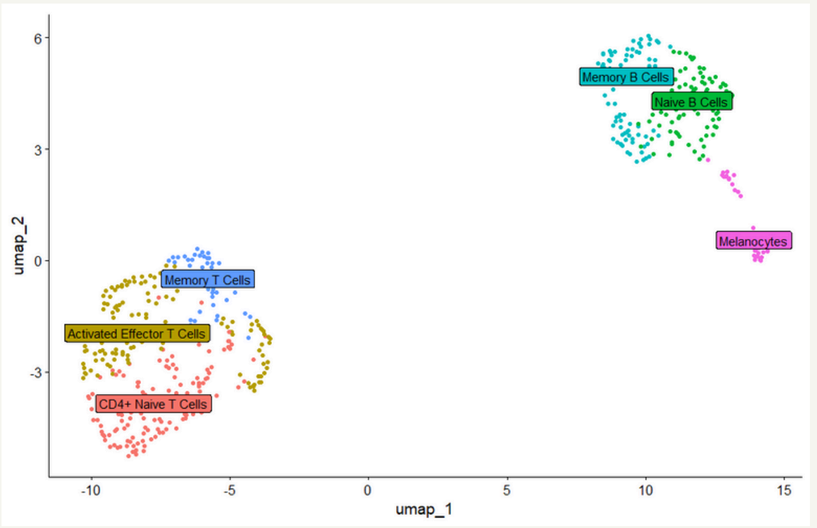
UMAP AND TSNE

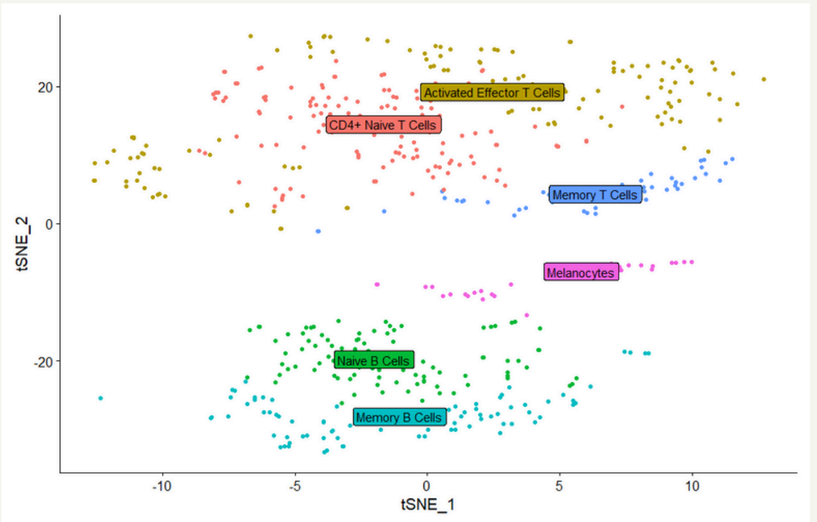
SINGLER





MANUAL





LIST OF DIFFERENTIALLY EXPRESSED FEATURES

TOP TEN ACROSS ALL CLUSTERS

Gene	Log2FC	padj	Cluster
IGHM	4.35	3.43E-58	2
IGHD	4.12	1.02E-49	2
TCL1A	4.94	1.76E-46	2
COL19A1	3.05	4.89E-46	2
EBF1	3.71	2.30E-45	3
FCRL1	3.17	3.66E-45	2
HLA-DMB	3.27	3.82E-44	2
CD79A	3.30	4.48E-43	2
MS4A1	2.83	1.23E-42	2
HLA-DQB1	2.92	1.51E-42	2

TOP TEN FOR EACH CLUSTER

Gene	Log2FC	padj	Cluster
RPS3	1.63	8.50E-39	0
IL7R	1.94	1.33E-37	0
RPS14	1.66	2.24E-37	0
RPS15A	1.51	8.91E-37	0
CD3D	2.56	3.80E-36	0
CD3E	2.21	1.06E-34	0
RPS12	1.44	4.63E-34	0
TPT1	1.21	1.59E-33	0
RPS3A	1.40	2.39E-33	0
RPL7	1.46	4.72E-32	0

Gene	Log2FC	padj	Cluster
PRKCH	1.83	2.11E-23	1
FYN	2.01	3.35E-21	1
PRKCA	1.96	9.41E-21	1
MIR23AHG	2.53	3.82E-20	1
ZSWIM6	1.80	4.19E-20	1
ABCC1	2.06	1.49E-19	1
PBX4	1.95	3.15E-19	1
INTS6L	1.91	7.12E-19	1
BCL11B	1.53	1.02E-18	1
HIVEP2	1.74	4.83E-18	1

Gene	Log2FC	padj	Cluster
IGHM	4.35	3.43E-58	2
IGHD	4.12	1.02E-49	2
TCL1A	4.94	1.76E-46	2
COL19A1	3.05	4.89E-46	2
FCRL1	3.17	3.66E-45	2
HLA-DMB	3.27	3.82E-44	2
CD79A	3.30	4.48E-43	2
MS4A1	2.83	1.23E-42	2
HLA-DQB1	2.92	1.51E-42	2
HLA-DRA	2.96	2.07E-41	2

Gene	Log2FC	padj	Cluster
EBF1	3.71	2.30E-45	3
LYN	3.17	2.45E-42	3
ARHGAP24	3.38	2.87E-38	3
PLEKHG1	3.30	1.01E-37	3
DENND3	3.95	4.45E-37	3
BANK1	2.73	3.01E-35	3
CYRIA	3.83	8.07E-35	3
CDK14	2.98	1.63E-31	3
RUBCNL	2.68	7.85E-30	3
ADAM28	2.28	6.73E-27	3

Gene	Log2FC	padj	Cluster
TXK	3.11	7.65E-19	4
LINC00243	4.05	1.58E-15	4
MALAT1	1.01	3.54E-11	4
LINC00861	3.28	4.27E-11	4
CCND3	1.85	2.31E-10	4
LINC01550	3.77	4.60E-09	4
FGD3	2.38	5.91E-09	4
DCP1B	3.44	7.16E-09	4
LINC02325	4.00	2.75E-08	4
PRMT2	2.21	2.90E-08	4

Gene	Log2FC	padj	Cluster
SPARC	8.09	1.75E-29	5
IGFBP2	10.92	8.45E-24	5
CRYAB	7.60	2.68E-19	5
S100B	7.77	2.89E-19	5
GPNMB	8.14	4.15E-19	5
FXYD3	6.93	5.31E-19	5
S100A1	8.55	1.04E-18	5
CST3	7.54	1.49E-18	5
QPCT	8.23	5.12E-18	5
ENSG00000291105	8.33	3.00E-17	5

ASSIGNING CELL TYPES TO CLUSTERS

- SingleR was used primarily
  - The HumanPrimaryCellAtlas from celldex was used as reference data
  - SingleR recategorized clusters
  - Several cell types were misidentified (i.e. astrocytes and endothelial cells)

- Manual methods were utilized for a more fine-tuned approach
  - The Human Protein Atlas, PanglaoDB, and CellMarker2.0
- The list of differentially expressed features for each cluster contained canonical gene markers used for cluster assignment
- Neighboring clusters were determined to be subtypes of the same general cell type

CANONICAL MARKERS AND ASSOCIATED CELL TYPE

CLUSTER 0

IL7R

↓

CD4+ Naive T Cells

CLUSTER 1

IL2RA, CD28

↓

Activated Effector T Cells

CLUSTER 2

TCL1A, IGHM

↓

Naive B Cells

CLUSTER 3

TNFRSF13B (TACI), CR1

↓

Memory B Cells

CLUSTER 4

LEF1, GIMAP5

↓

Memory T cells

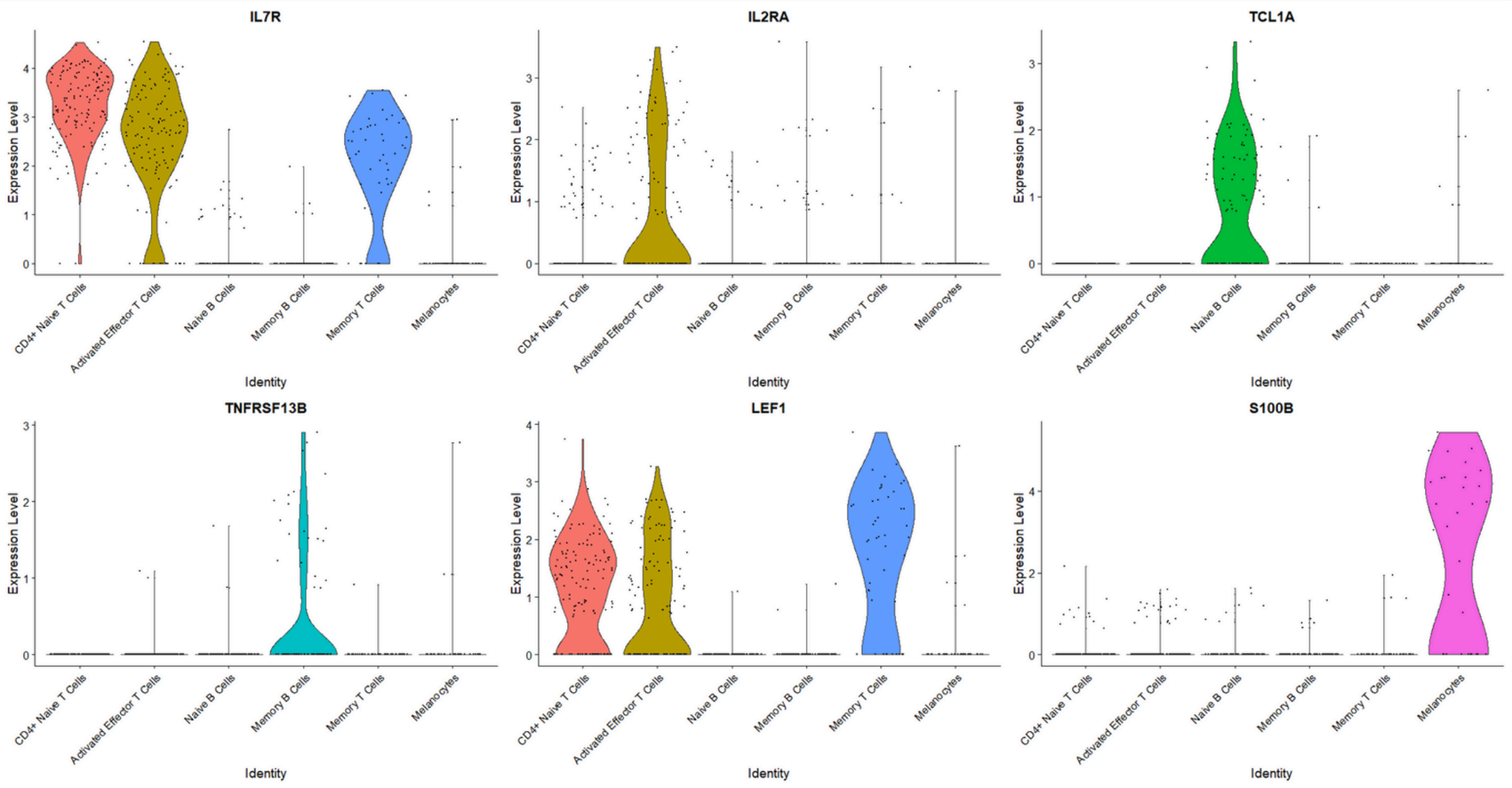
CLUSTER 5

S100B, MLANA

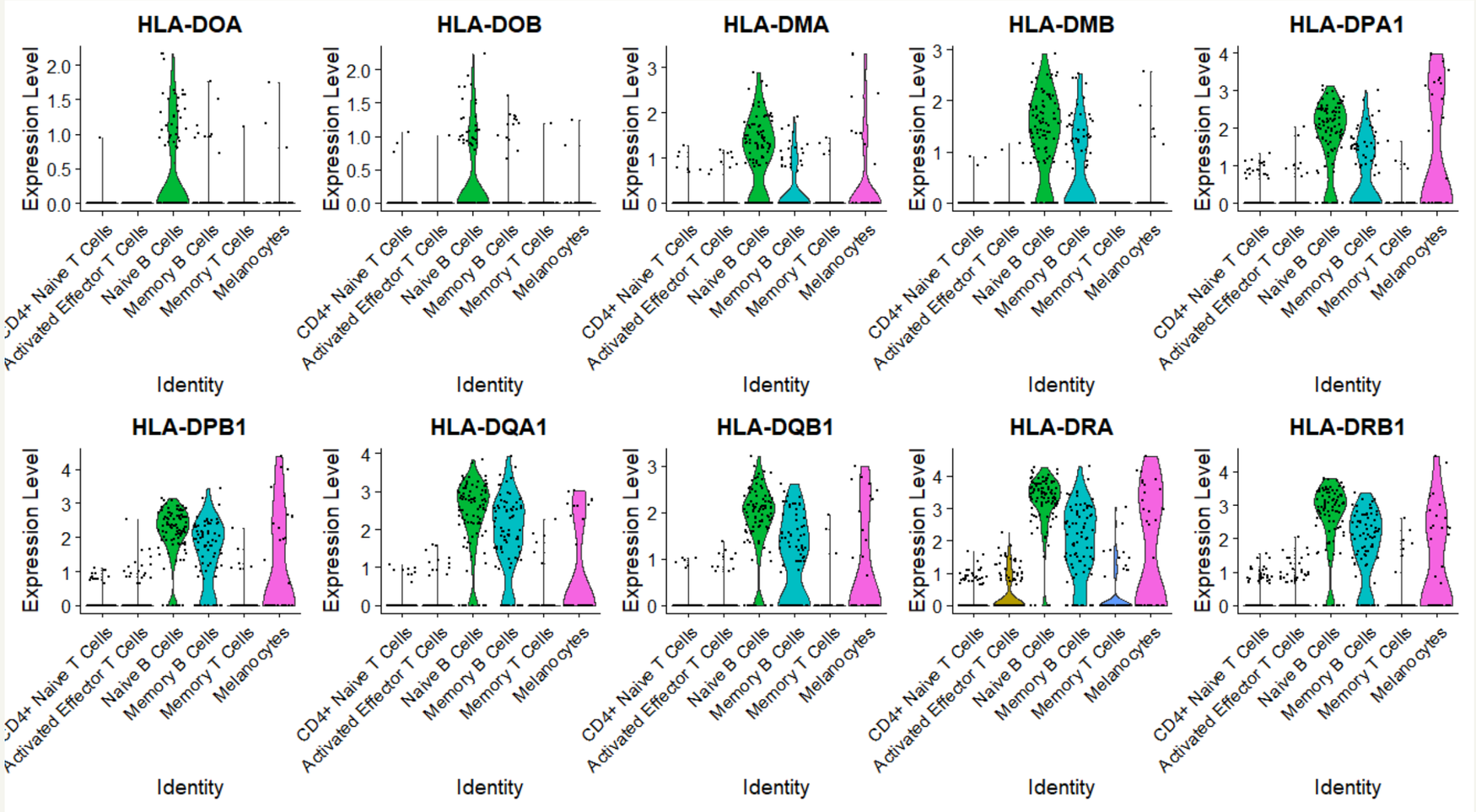
↓

Melanocytes

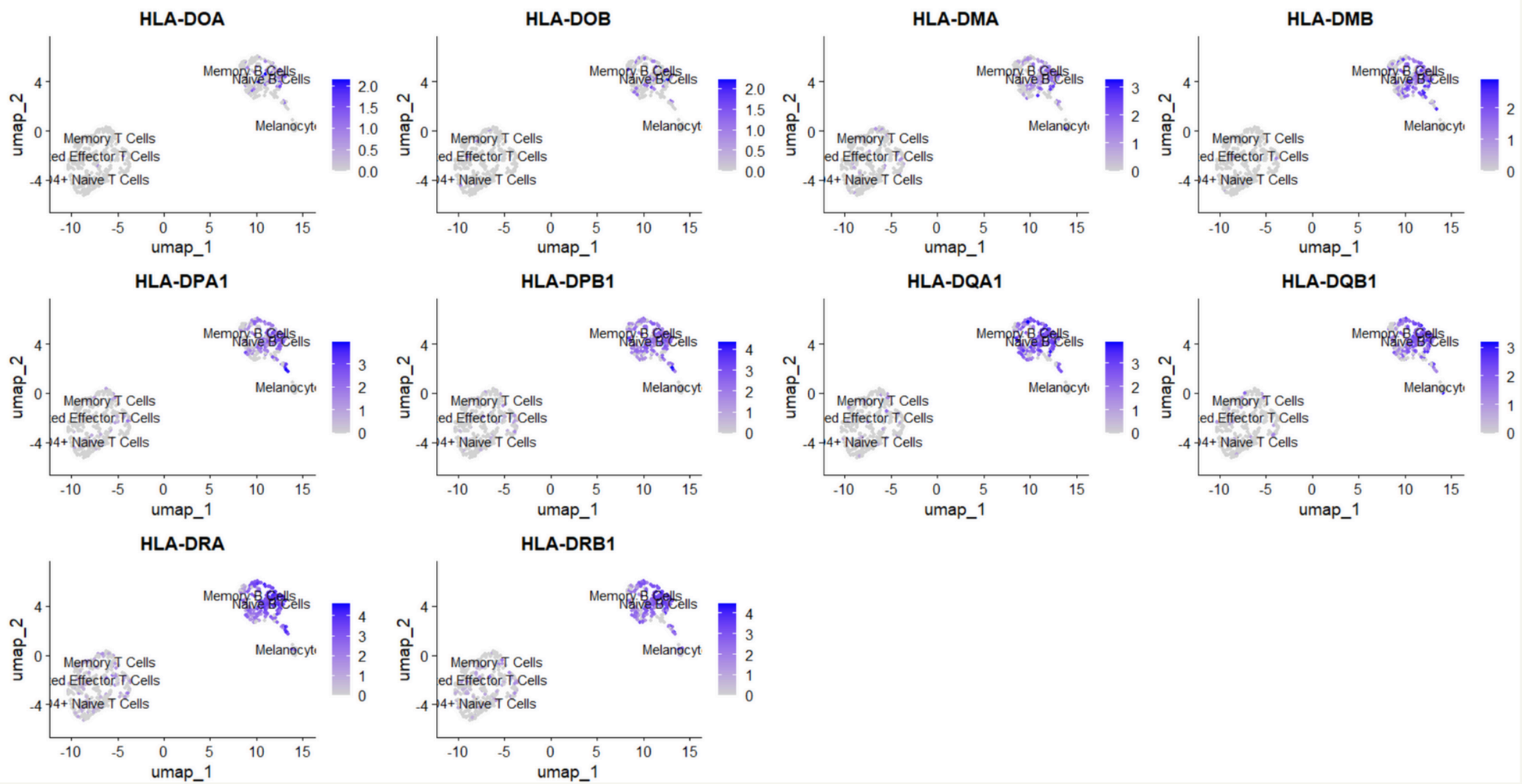
CANONICAL MARKER CLUSTER DISTRIBUTION



ANALYSIS RESULTS



ANALYSIS RESULTS CONT.



ANALYSIS RESULTS CONT.

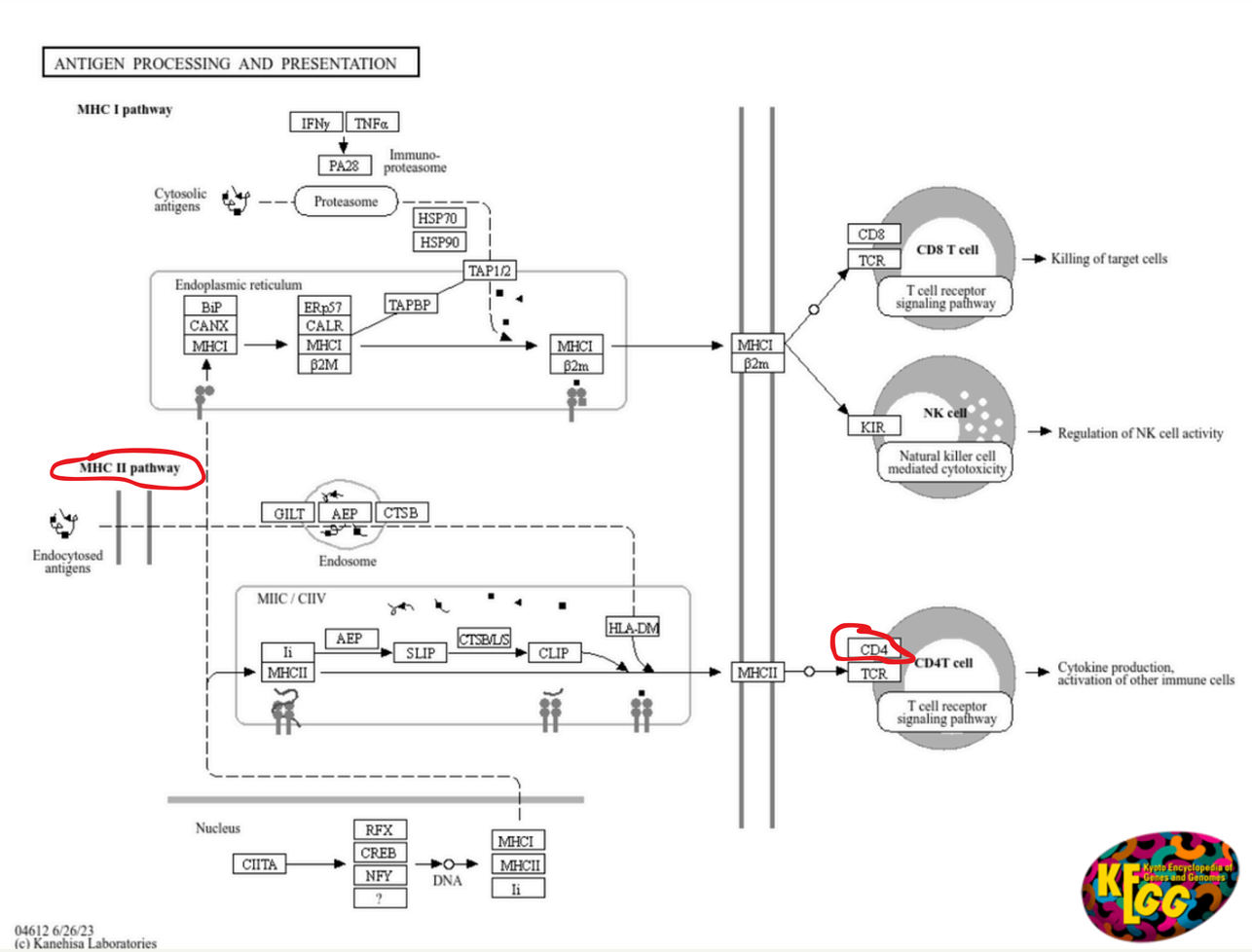
HLA Class II genes were significantly upregulated in B cells *and* melanocytes → **results provide partial support for the hypothesis.**

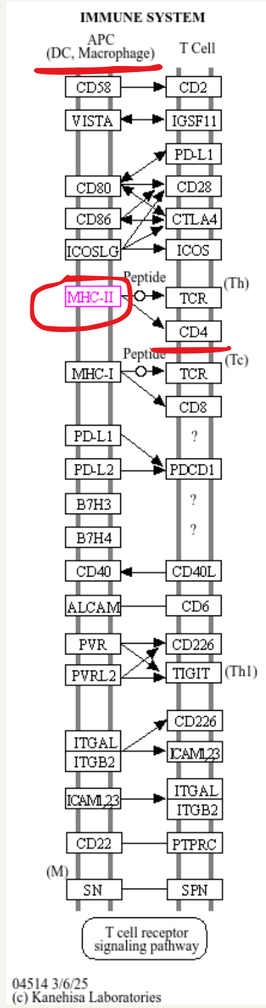
In this dataset, malignant melanocytes express multiple core HLA class II genes (DRA/DRB1, DPA1/DPB1, DQB1) with DMA/DMB, **indicating a tumor-cell HLA-II program**.

This suggests that this patient **may have improved response to anti-PD-1 therapy** with overall survival (Johnson, et al., 2016).

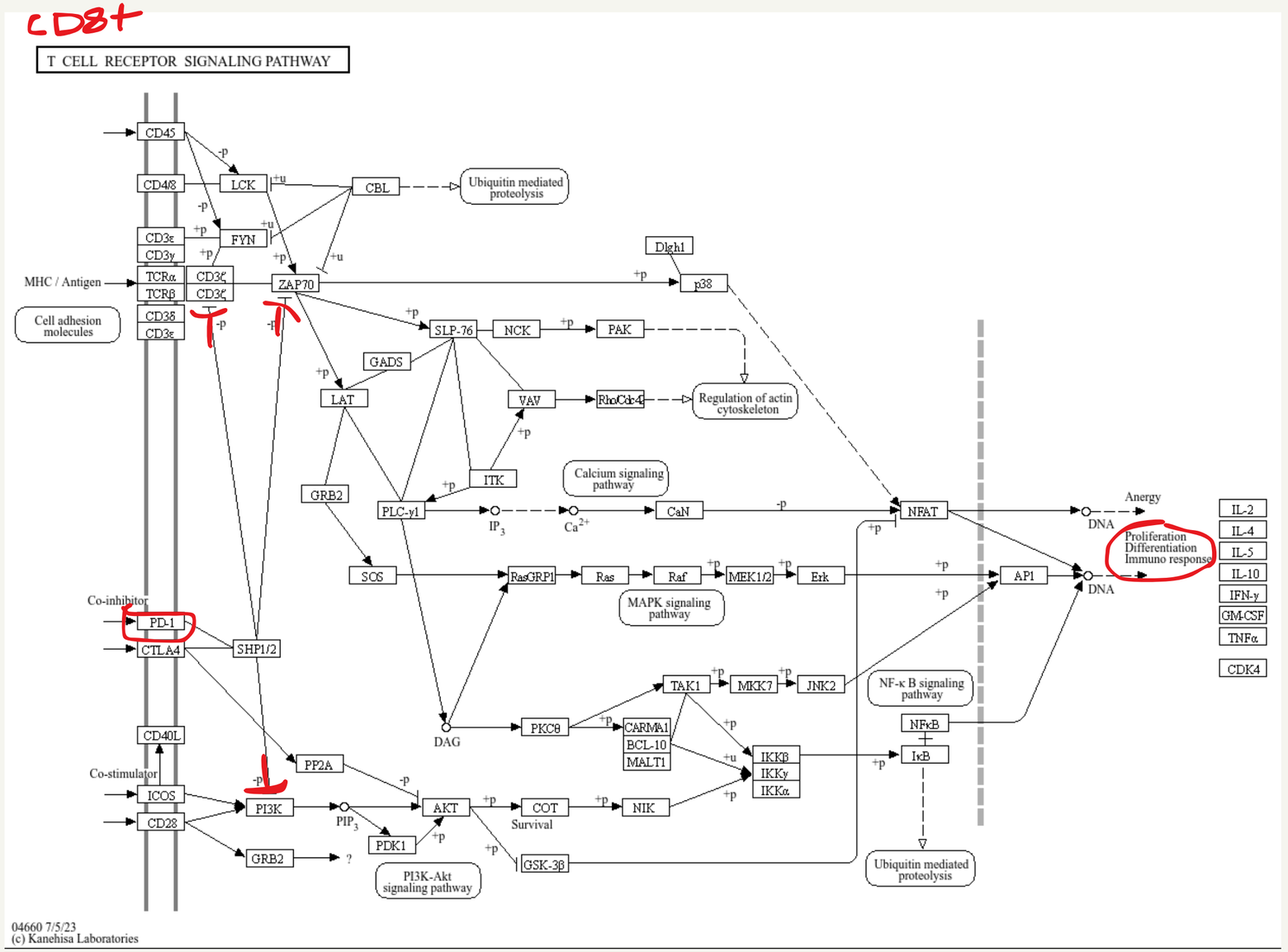
Normally, HLA-II genes encode MHC-II molecules in antigen presenting cells (APCs) (Chen, et al., 2019). These cells are typically restricted to mononuclear cells- specifically macrophages, dendritic cells, and B cells (Payne, 2023).

ANALYSIS RESULTS CONT.





ANALYSIS RESULTS CONT.



SUMMARY AND FUTURE DIRECTIONS

Analysis showed 6 cell type clusters: CD4+ Naive T Cells, Activated Effector T Cells, Naive B Cells, Memory B Cells, Memory T Cells, and Melanocytes.

HLA Class II genes were significantly upregulated in B cells *and* melanocytes.

In this dataset, malignant melanocytes expressed multiple core HLA class II genes (DRA/DRB1, DPA1/DPB1, DQB1) with DMA/DMB, indicating a tumor-cell HLA-II program that is specific to Antigen Presenting Cells, not tumor cells. This program enhances patient outcomes on anti-PD-1 therapy drugs.

It would be interesting to collect samples from this patient after treatment of anti-PD-1 therapy to determine if the presence of HLA-II genes could be counted as biomarkers for improved treatment response. Though not possible in this case, a similar study may be able to accomplish this.

REFERENCES

Payne, S. (2023). Viruses. In Elsevier eBooks. https://doi.org/10.1016/c2020-0-03715-1

Chen, Y., Chang, W., Lin, E., Chen, Y., & Kuo, P. (2019). Expressions of HLA Class II Genes in Cutaneous Melanoma Were Associated with Clinical Outcome: Bioinformatics Approaches and Systematic Analysis of Public Microarray and RNA-Seq Datasets. Diagnostics, 9(2), 59. https://doi.org/10.3390/diagnostics9020059

Johnson, D. B., Estrada, M. V., Salgado, R., Sanchez, V., Doxie, D. B., Opalenik, S. R., Vilgelm, A. E., Feld, E., Johnson, A. S., Greenplate, A. R., Sanders, M. E., Lovly, C. M., Frederick, D. T., Kelley, M. C., Richmond, A., Irish, J. M., Shyr, Y., Sullivan, R. J., Puzanov, I., . . . Balko, J. M. (2016). Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy. Nature Communications, 7(1). https://doi.org/10.1038/ncomms10582

Ryan, S. O., & Cobb, B. A. (2012). Roles for major histocompatibility complex glycosylation in immune function. Seminars in Immunopathology, 34(3), 425–441. https://doi.org/10.1007/s00281-012-0309-9

Marsh, S. G. E., Albert, E. D., Bodmer, W. F., Bontrop, R. E., Dupont, B., Erlich, H. A., Fernández-Viña, M., Geraghty, D. E., Holdsworth, R., Hurley, C. K., Lau, M., Lee, K. W., Mach, B., Maiers, M., Mayr, W. R., Müller, C. R., Parham, P., Petersdorf, E. W., Sasazuki, T., . . . Trowsdale, J. (2010). Nomenclature for factors of the HLA system, 2010. Tissue Antigens, 75(4), 291–455. https://doi.org/10.1111/j.1399-0039.2010.01466.x

Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y., & Ishiguro-Watanabe, M. (2024). KEGG: biological systems database as a model of the real world. Nucleic Acids Research, 53(D1), D672–D677.

THANK YOU