Identification of High-Risk Cells in Single-Cell Spatially Resolved Transcriptomics Data Using DEGAS Spatial Smoothing

Debolina Chatterjee[1] Justin L. Couetil[1] Ziyu Liu[2] Kun Huang[1]
Chao Chen[3] Jie Zhang[1] Michael A. Kalwat[4] Travis S.Johnson[1]

[1]Indiana University School Of Medicine, Indianapolis, IN
[2]Purdue University, West Lafayette, IN
[3]Stony Brook University, Stony Brook, NY
[4] Indiana Biosciences Research Institute, Indianapolis, IN

Abstract

Summary: The examination of high-risk cells and regions in tissue samples from spatially resolved transcriptomics platforms offers meaningful insights into specific disease processes. For existing methods, while cell types or clusters can be identified and associated with disease attributes, individual cells are unable to be associated in the same manner. **Method:** Diagnostic Evidence GAuge of Single-cells and spatial transcriptomics (DEGAS), solves the above problem by employing latent representations of gene expression data and domain adaptation to transfer disease attributes from patients to individual cells from single-cell RNA sequencing datasets. In this research, we present and evaluate DEGAS's versatility in adapting to data arising from various single-cell spatially resolved transcriptomics (scSRT) platforms. DEGAS successfully identified high-risk cells and regions in liver hepatocellular carcinoma and skin cutaneous melanoma, which were validated through known markers. Additionally, DEGAS was applied to our newly generated Type II Diabetes Xenium dataset revealing high-risk cells within the tissue samples.
Availability and Implementation: The DEGAS software can be accessed at https://github.com/tsteelejohnson91/DEGAS. For the updated smoothing functions and associated codes, visit https://github.com/dchatter04/DEGAS-Spatial-Smoothing. Sources for the datasets reviewed are detailed in their respective sections. A description of some datasets, along with extra tables and figures, is provided in the Supplementary Materials file. Our newly generated Xenium data for Type II Diabetes can be found at https://doi.org/10.7303/syn68699752.

1 Introduction

Single-cell spatially resolved transcriptomics (scSRT) merges sequencing and imaging techniques to study tissue samples with high precision, offering crucial insights into disease mechanisms. Over recent years, scSRT data have evolved to include both spatial and gene expression details, along with related histological images of tissue specimens. Some of the popular platforms that provide such datasets are 10X Genomics' Xenium [8], MERFISH [3], segFISH+ [5] and Nanostring Technologies' CosMx [7], among others. Although these data give detailed information about single cells, patient-level attributes such as the overall survival or other physiological characteristics can not be directly attributed to individual cells. Furthermore, these datasets are limited by sample size, making it difficult to generalize findings. On the other hand, bulk RNA-seq datasets from platforms such as the Gene Expression Omnibus (GEO) or The Cancer Genome Atlas (TCGA) reveal information on

diverse disease characteristics for patients, such as diagnostic details like disease subtype, status, prognostic information like survival, and treatment responses, but they lack the granular single-cell details. This limitation hinders the identification of cell subsets linked to disease characteristics, especially when disease-related cells are mixed with non-disease-related cells. Thus, there is a pressing need for methods to translate information from scSRT data to the patient level. Using the rationale that single-cell RNA-seq (scRNA-seq) data and patient-level transcriptomic data (such as RNA-seq with clinical annotations) share the same gene set and, therefore, a common feature space, there is a natural link allowing efficient information transfer between these data types to uncover associations between patients and cells. Diagnostic Evidence Gauge of Single-cells and spatial transcriptomics (DEGAS) [9] uses transfer learning methods like domain adaptation [10] and multitask learning [2] to link individual cells to disease risk for scRNA-seq data. In this article, we demonstrate how DEGAS was modified to utilize the spatial information from scSRT datasets. In this article, we demonstrate how DEGAS was modified to utilize the spatial information from scSRT datasets and its utility in liver hepatocellular carcinoma (LIHC), skin cutaneous melanoma (SKCM), and Type II Diabetes (T2D).

2 Methodology

The DEGAS framework (Fig. 1 A) is designed using Python's TensorFlow library and provides an interface in R. In this article we skip the detailed explanation of the framework as it can be found in the original publication [9].

The DEGAS model preprocessing and training pipeline is largely the same between scRNA-seq and scSRT data. After training the model (Supplementary Algorithm 1), for each single cell, the patient-level disease association (DA) scores are predicted. Higher DA scores can be interpreted as higher-risk of the disease. The key modification required to adapt to spatial datasets lies in the post-processing steps.

2.1 Postprocessing

To adapt to spatial platforms, we use localized smoothing on the predicted DA scores to preserve the spatial context of the tissue and prevent oversmoothing. Let us denote the DA scores, which are scaled within $[-1, 1]$ by S^{scaled}. We define windows based on pixel coordinates from the metadata and locally smooth the DA scores by the Spatial Window Smoothing algorithm below:

SWiS Algorithm

Let $\mathcal{C} = \{c_1, c_2, \ldots, c_N\}$ denote the collection of N cells. Each cell c_i is associated with spatial coordinates $(x_i, y_i) \in \mathbb{R}^2$, and the DA acores $S_i^{\text{scaled}} \in \mathbb{R}$, where $\mathbb{R}$ is the set of real numbers. Let $\mathcal{D}$ be set of all cells, where each cell is represented by its spatial coordinates and DA scores, i.e., $\mathcal{D} = \{(x_i, y_i, S_i^{\text{scaled}})\}_{i=1}^N$, and focus on a small patch of the tissue that spans a 2-dimensional window with $x_{\min} = \min_i x_i, \quad x_{\max} = \max_i x_i, \quad y_{\min} = \min_i y_i, \quad y_{\max} = \max_i y_i$. A window starting at position (x, y) defines a rectangular patch $W_{x,y} = \left\{(x_i, y_i, S_i^{\text{scaled}}) \in \mathcal{D} \mid x < x_i < x + a,\ y < y_i < y + b\right\}$, where a is window width (X-axis range), b: window height (Y-axis range), c is step size (stride along X and Y) (as displayed in Fig. 1 B). To ensure the sliding window works well will adequate number of cells (at least $n_{\min}$), while defining the window, we exclude cells that are too close to the boundary of the window, we use an internal padding p to only consider the cells which are in the *core* of it. The *core* of the window exclusing marginal cells is defined as $W_{x,y}^{\text{core}} = \{(x_i, y_i, S_i^{\text{scaled}}) \in W_{x,y} \mid x + p < x_i < x + a - p,\ y + p < y_i < y + b - p\}$.

Starting at the top-left, the window moves rightward until it covers the X-range, then steps down

by c along the Y-axis, repeating the process until all cells are included. Therefore, the horizontal shift is $x \to x + c$ until $x + a > x_{\max}$, and vertical shift is $y \to y + c$, resetting $x = x_{\min}$. This generates a collection of windows $\{W_j\}_{j=1}^{J}$, each with a core region W_j^{core}.

Within each window W_j^{core}, define the set of core cells $\mathcal{C}_j = \left\{ i \mid (x_i, y_i) \in W_j^{\text{core}} \right\}$. The Algorithm 1 provides a step-by-step description of the SWiS algorithm.

Cells not included in any window can either be excluded from the final output or only their raw scores be considered. A cell can have a list of smoothed DA scores due to overlapping windows and are averaged over all overlapping windows. This ensures full slide coverage while preserving local expression through spatially aware smoothing.

FoVS Algorithm For datasets that are naturally divided into Fields of View (FOVs), instead of defining the sliding window we use the naturally defined FOVs, while other computations remain the same as the SWiS Algorithm 1. Each FOV has its own spatial coordinates and is treated as a discrete region. After smoothing within individual FOVs, the results are combined to reconstruct the full tissue. Next, we present the applicability of DEGAS+SWiS/FoVS in identifying high-risk cells across various diseases and data-types.

3 DEGAS identifies high-risk pancreatic cells associated with Type II diabetes using analysis of Xenium data

DEGAS preprocessing steps were conducted on the newly generated Xenium Type II Diabetes (T2D) data (described in Section 2 of the Supplementary material), including filtering for high-quality cells and normalizing gene expression to account for technical biases and ensuring robust downstream analysis. For the patient-level bulk RNA-seq data, clinical metadata obtained from the GEO dataset GSE159984 [11], which includes data on 58 non-diabetic (ND) and 27 Type II Diabetic (T2D), was used. This dataset provides important information on patient characteristics such as body mass index (BMI), sex, and T2D status.

There were no cell-type labels in the scSRT data and the bulk RNA-seq data included T2D status classes (T2D vs. ND), we therefore employed the 5x BAg 3-layer *BlankClass* DEGAS model. This model predicted the likelihood of each individual cell being categorized as associated with T2D or ND. We applied SWiS on the DEGAS predicted T2D association scores, with width and height 1000, stride 500, padding 100, 50 minimum cells, and 5 nearest neighbours.

The spatial plot of the DEGAS T2D association scores highlighted individual cells as low T2D-association (blue) or high T2D-association (red) (Fig. 1 C-F). The Wilcoxon rank-sum test revealed that the average T2D-association scores for T2D samples were significantly higher than those of the ND samples ($P < 0.001$, Fig. 1 G,H). This shows the viability of the DEGAS+SWiS approach.

We define low-risk cells as those with predicted T2D association scores below the 25th percentile, and high-risk cells as those with scores above the 75th percentile. A threshold of 0.5 is used for the average $\log_2$-fold change, and only genes expressed in at least 10% of cells are included in the analysis.

Supplementary Tables 2-3 list genes that are significantly upregulated or downregulated (i.e., Benjamini-Hochberg adjusted P-value ≤ 0.05) in the high-risk regions of the Xenium samples. These tables also report the $\log_2$-fold changes in expressionof these genes between high-risk, low-risk, and other cells. Supplementary Figures 1-4 present the corresponding volcano plots. Some of the genes that were found to be significantly up/downregulated in the high-risk cells and their association with T2D have been studied, include *PTPRC* [12], *AMY2A* [15, 1] and *CFTR*[13, 6]. Observing these genes in the high-risk cells identified through our algorithm further demonstrates its effectiveness in identifying T2D risk associated cells.

Algorithm 1 SWiS: Spatial Window Smoothing

Inputs:

- Cell data $\mathcal{C} = \{(x_i, y_i, S_i^{\text{scaled}})\}_{i=1}^{N}$
- Parameters: window width a, height b, stride c, padding p
- Thresholds: $n_{\min}$ (min cells per core), k (nearest neighbors)

Initialize: For each cell i, initialize list of smoothed scores $\mathcal{S}_i \leftarrow \emptyset$ (where $\emptyset$ is the empty set)

Sliding Window Loop:
for The jth window $j = 1, 2, \cdots, J$ with step size c **do**
  Define window W_j: cells with $x < x_i < x + a$, $y < y_i < y + b$
  Define core $W_{x,y}^{core}$: cells with $x + p < x_i < x + a - p$, $y + p < y_i < y + b - p$
  if number of cells in $W_j^{\text{core}} < n_{\min}$ **then**
    Skip this window
  end if
  for each core cell c_i in $W_{x,y}^{core}$ **do**
    Identify k nearest neighbors of cell c_i in the jth window as $\mathcal{N}_k^k(c_i)$ within W_j^{core} (excluding c_i)
    Compute local smoothed value:

$$\tilde{S}_i^{(j)} = \frac{1}{k} \sum_{c_l \in \mathcal{N}_j^k(c_i)} S_l^{\text{scaled}}$$

    Append $\tilde{S}_i^{(j)}$ to $\mathcal{S}_i$
  end for
end for

Final Averaging:
for each cell c_i **do**
  if $\mathcal{S}_i \neq \emptyset$ **then**
    Compute average over all smoothed estimates:

$$\tilde{S}_i = \frac{1}{|\mathcal{S}_i|} \sum_{\tilde{S}_i^{(j)} \in \mathcal{S}_i} \tilde{S}_i^{(j)}$$

  else
    Set $\tilde{S}_i = S_i^{\text{scaled}}$ ▷ If cell was not included in any core
  end if
end for

Output: Final smoothed values $\tilde{S}_i$ for all $i = 1, 2, \cdots, N$

4 DEGAS identifies high-risk cell types in CosMx Hepatocellular carcinoma samples

CosMx data is characterized by multiple FOVs that collectively cover the entire tissue area. Within each FOV, individual cells are mapped with precise FOV-specific spatial coordinates and the locations of individual RNA molecules within those cells. We obtained this publicly available Liver Hepatocellular Carcinoma (LIHC) data (see Section 3 of Supplementary material for details). The bulk RNA-seq data with clinical metadata were acquired from the (TCGA LIHC study). This dataset included data from 377 human individuals along with their clinical information, including sex, age, overall survival status, and time.

The gene expression count data for each of the two CosMx samples was individually preprocessed according to DEGAS's specified preprocessing steps mentioned in the Supplementary Algorithm 1. However, there were no labels distinguishing individual single cells as tumor or normal. Thus, following the identification of a common gene set between the scSRT and the bulk RNA-seq data, the 5x BAg 3-layer *BlankCox* model of DEGAS was utilized. For CosMx data, due to the presence of well-defined FOVs and FOV-specific pixel coordinates, we used the naturally defined FOVs to smooth the LIHC-associations, with 50 nearest neighbours. The spatial map of DEGAS+FoVS LIHC-association scores (Fig. 1 I,J) reveals high LIHC-association regions (red) and low LIHC-association regions (blue) for all cells within a field of view in both the non-carcinoma and LIHC samples. As expected, there is a highly significant difference in the LIHC-association scores between the non-carcinoma and LIHC samples ($P < 0.001$, Fig.1 K). We consider low-risk cells as those with predicted LIHC association scores below the 25th percentile, and high-risk cells as those with scores above the 75th percentile. In the non-LIHC sample, 83219 low- and 83219 high-LIHC association cells were identified, while in the LIHC sample, 115750 low LIHC-association cells and 344691 high LIHC-association cells were identified. This result reveals a significantly higher number of high LIHC-association cells in the cancer sample ($\chi^2 = 34873.64$, $P < 0.001$) compared to the non-carcinoma sample. For patient-level T2D association predictions using the DEGAS model, we utilized binary 3-year survival information from the TCGA dataset. Overall survival time was used as the ground truth, and we excluded censored patients to ensure reliable labeling. Patients who survived beyond 3 years were labeled as 0, while those who experienced an event (death) within 3 years were labeled as 1. The Receivic operating characteristic area under the curve (ROC-AUC) was calculated considering the predicted DEGAS patient level LIHC-association scores. The DEGAS model demonstrated strong discriminatory performance, achieving a high ROC-AUC of 0.7686 (Fig. 1 L).

To further investigate the LIHC-association score for individual cell types, we plotted the overall LIHC-association scores for each cell type within the liver tissues (Fig. 1 M). In the non-carcinoma sample, we observed that stellate cells, and macrophages, representing stromal and immune cells—exhibited elevated scores. In the LIHC (carcinoma) sample, tumor cells, non-inflammatory macrophages, and cholangiocytes showed higher average scores compared to endothelial or gamma delta T cells (Fig. 1 M). Notably, regions in non-carcinoma sample exhibiting high hazard scores for LIHC were enriched with stellate cells, macrophages, mature B cells, and erythroid cells (see Supplementary Figure 5). Macrophages, in particular, are known to contribute to tumor growth by suppressing immune surveillance through the inhibition of cytotoxic T cells [14]. A similar pattern was observed in a DEGAS-based study of prostate cancer, where specific normal tissues were linked to cancer progression [4]. In that study, normal regions with high disease-association scores were found to be infiltrated by macrophages, detectable in both transcriptomic data and histopathological images [4]. In our study, we observed a comparable trend.

5 DEGAS identifies high-risk cells in Skin Cutaneous Melanoma samples

For this analysis, the bulk human Skin Cutaneous Melanoma (SKCM) data were acquired from TCGA SKCM study. This reference data contained 293 patients and their overall survival in months. The Xenium scSRT data was sourced from a publicly available 10X Xenium FFPE Human Skin dataset (https://www.10xgenomics.com/) which consists of a single tissue sample. There were no known cell-labels for scSRT data and bulk RNA-seq data has survival times, we applied the *BlankCox* DEGAS model, with overall survival as the patient-level outcome. The SKCM-association scores were calculated, and the SWiS algorithm was applied to find the smoothed DEGAS hazard scores. The SWiS SKCM-association scores were overlaid with the tissue's spatial coordinates, highlighting areas of higher SKCM association within the tissue (Fig.1 N). Our results demonstrate the effectiveness of DEGAS+SWiS in identifying high-risk regions within Xenium tissue samples (Fig. 1 O).

6 Summary

This article demonstrated DEGAS's seamless application across multiple scSRT platforms and diseases, including 10X Genomics Xenium and Nanostring's CosMx and various disease types. Additionally, we implemented smoothing techniques, i.e., SWiS and FoVS, that account for spatial localization and created a novel pipeline for scSRT datasets. The application of DEGAS+SWiS/FoVS led to meaningful findings, such as identifying high-risk cell types in liver cancer (LIHC). Many genes associated with Type II Diabetes were found to be significantly upregulated in high-risk cells identified by DEGAS+SWiS. These discoveries position DEGAS with spatial smoothing as a powerful and efficient tool for advancing translational research involving scSRT, offering significant potential for early diagnosis, prevention, and therapeutic interventions.

Author Contributions

D.C. -Conceptualization, Analysis, Software, Writing; J.L.C., Z.L -Analysis; K.H. -Conceptualization, Review; C.C., J.Z. -Conceptualization, Review, Funding; M.A.K. -Data curation, Review; T.S.J. -Conceptualization, Software, Supervision, Review, Funding.

Acknowledgement

Sequencing was performed by the Center for Medical Genomics (CMG) at Indiana University School of Medicine (IUSM), which is partially supported by the Indiana University Grand Challenges Precision Health Initiative. This work was supported by the NIH-NCI (R21CA264339) funding received by T.S.J and J.Z., and NIH-NIGMS (1R01GM148970) received by T.S.J., J.Z., and C.C.. General funding support from the IU Precision Health Initiative and AnalytixIN are also acknowledged.

References

[1] Mutiat A Abdulkareem, Bunmi A Owolabi, Emmanuel S Saheed, Remilekun F Aromolaran, Rukayat M Bashiru, Toheeb A Jumah, Doris U Chijioke, Onyinyechi J Amaechi, Fehintoluwa C Adeleke, Omiyale O Charles, et al. Genetic factors and the role of pancreatic amylase in the pathogenesis of type 2 diabetes. *Egyptian Journal of Medical Human Genetics*, 25(1):33, 2024.

[2] Rich Caruana. Multitask learning. *Machine learning*, 28:41–75, 1997.

[3] Kok Hao Chen, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang, and Xiaowei Zhuang. Spatially resolved, highly multiplexed rna profiling in single cells. *Science*, 348(6233):aaa6090, 2015.

[4] Justin L Couetil, Ziyu Liu, Ahmed K Alomari, Jie Zhang, Kun Huang, and Travis S Johnson. Diagnostic evidence gauge of spatial transcriptomics (degas): Using transfer learning to map clinical data to spatial transcriptomics in prostate cancer. *bioRxiv*, pages 2023–04, 2023.

[5] Chee-Huat Linus Eng, Michael Lawson, Qian Zhu, Ruben Dries, Noushin Koulena, Yodai Takei, Jina Yun, Christopher Cronin, Christoph Karp, Guo-Cheng Yuan, et al. Transcriptome-scale super-resolved imaging in tissues by rna seqfish+. *Nature*, 568(7751):235–239, 2019.

[6] Junzhong Gu, Weiwei Zhang, Lida Wu, and Yuchun Gu. Cftr deficiency affects glucose homeostasis via regulating glut4 plasma membrane transportation. *Frontiers in Cell and Developmental Biology*, 9:630654, 2021.

[7] Shanshan He, Ruchir Bhatt, Carl Brown, Emily A Brown, Derek L Buhr, Kan Chantranuvatana, Patrick Danaher, Dwayne Dunaway, Ryan G Garrison, Gary Geiss, et al. High-plex imaging of rna and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nature biotechnology*, 40(12):1794–1806, 2022.

[8] Amanda Janesick, Robert Shelansky, Andrew D Gottscho, Florian Wagner, Stephen R Williams, Morgane Rouault, Ghezal Beliakoff, Carolyn A Morrison, Michelli F Oliveira, Jordan T Sicherman, et al. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nature communications*, 14(1):8353, 2023.

[9] Travis S Johnson, Christina Y Yu, Zhi Huang, Siwen Xu, Tongxin Wang, Chuanpeng Dong, Wei Shao, Mohammad Abu Zaid, Xiaoqing Huang, Yijie Wang, et al. Diagnostic evidence gauge of single cells (degas): a flexible deep transfer learning framework for prioritizing cells in relation to disease. *Genome medicine*, 14(1):11, 2022.

[10] Wouter M Kouw and Marco Loog. A review of domain adaptation without target labels. *IEEE transactions on pattern analysis and machine intelligence*, 43(3):766–785, 2019.

[11] Lorella Marselli, Anthony Piron, Mara Suleiman, Maikel L Colli, Xiaoyan Yi, Amna Khamis, Gaelle R Carrat, Guy A Rutter, Marco Bugliani, Laura Giusti, et al. Persistent or transient human β cell dysfunction induced by metabolic stress: specific signatures and shared gene expression with type 2 diabetes. *Cell Reports*, 33(9), 2020.

[12] NCBI. PTPRC protein tyrosine phosphatase receptor type C [Homo sapiens (human)], 2024. Accessed: 2025-07-02.

[13] Thierry Ntimbane, Blandine Comte, Geneviève Mailhot, Yves Berthiaume, Vincent Poitout, Marc Prentki, Rémi Rabasa-Lhoret, and Emile Levy. Cystic fibrosis-related diabetes: from cftr dysfunction to oxidative stress. *The Clinical Biochemist Reviews*, 30(4):153, 2009.

[14] Stefano Ugel, Francesco De Sanctis, Susanna Mandruzzato, Vincenzo Bronte, et al. Tumor-induced myeloid deviation: when myeloid-derived suppressor cells meet tumor-associated macrophages. *The Journal of clinical investigation*, 125(9):3365–3376, 2015.

[15] Rakhee Yadav, Jai Prakash Bhartiya, Sunil Kumar Verma, and Manoj Kumar Nandkeoliar. The evaluation of serum amylase in the patients of type 2 diabetes mellitus, with a possible correlation with the pancreatic functions. *Journal of clinical and diagnostic research: JCDR*, 7(7):1291, 2013.

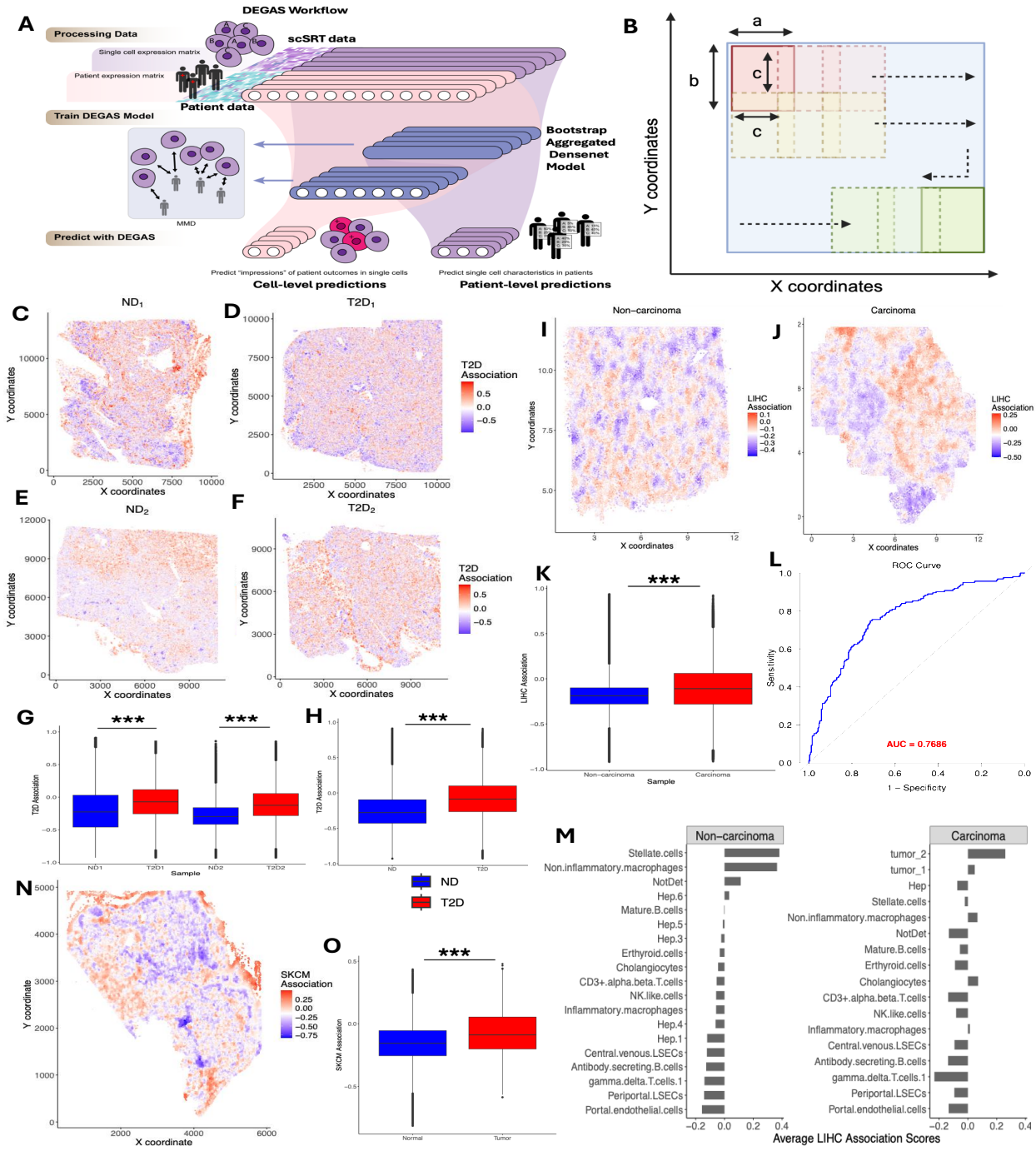
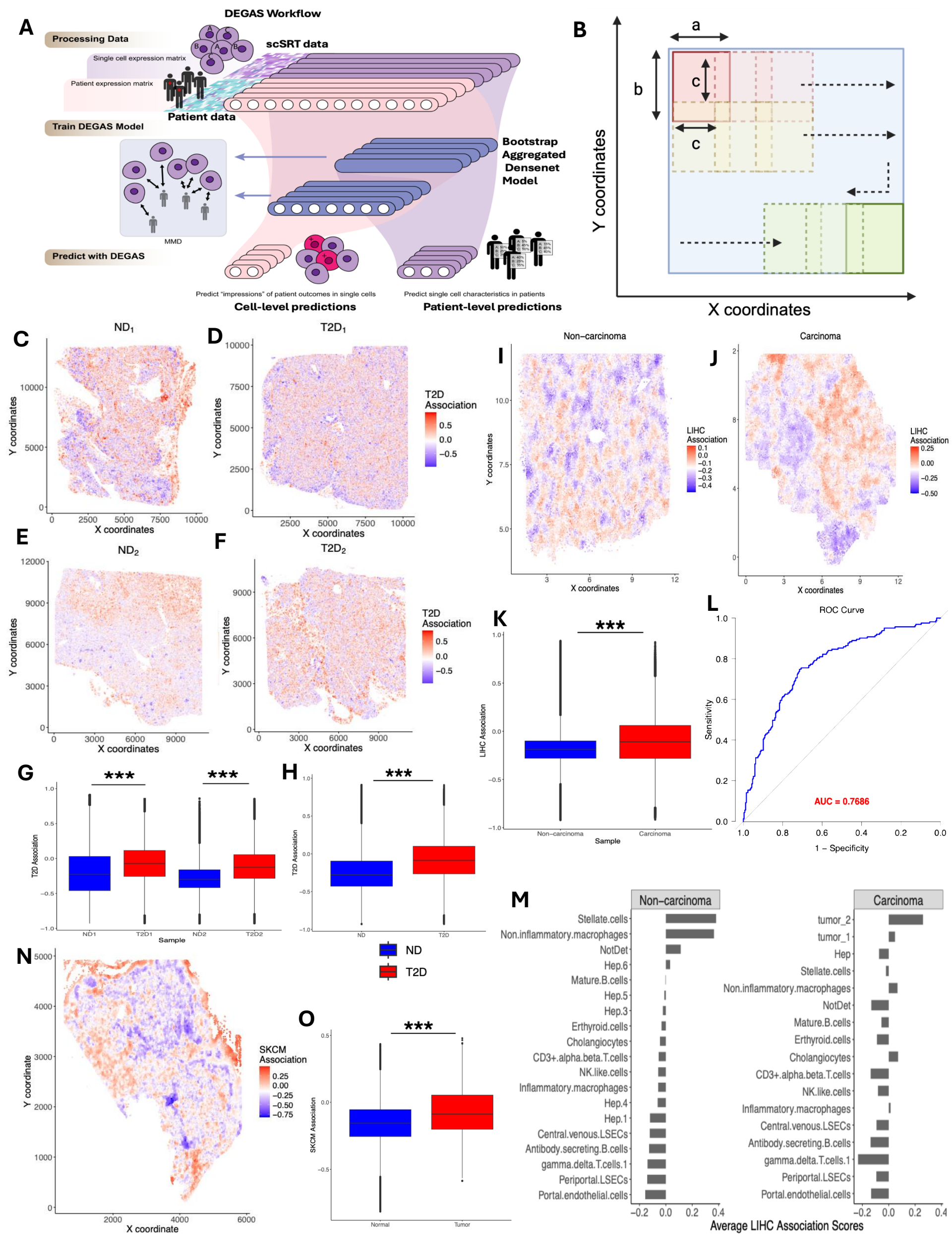
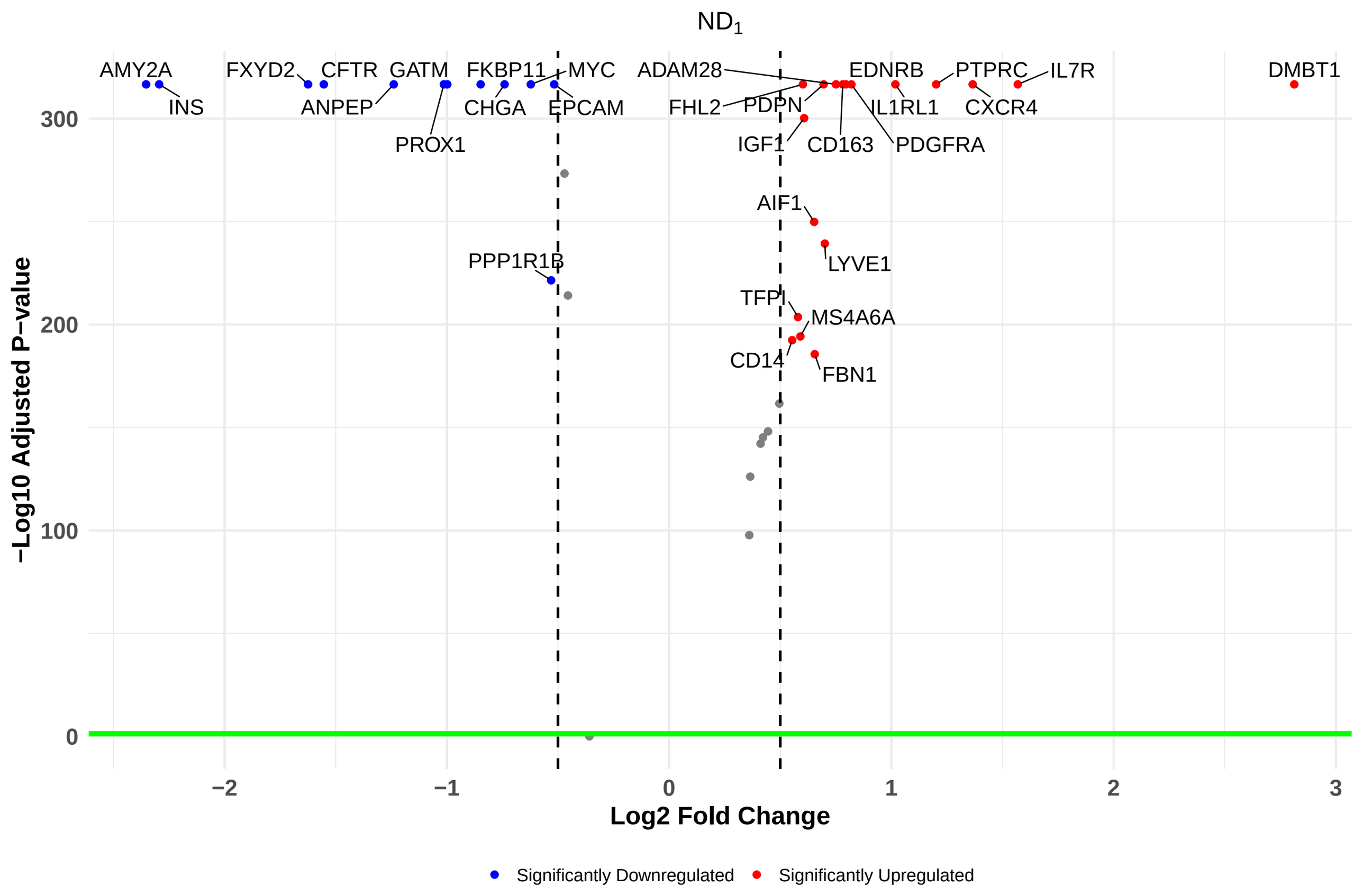
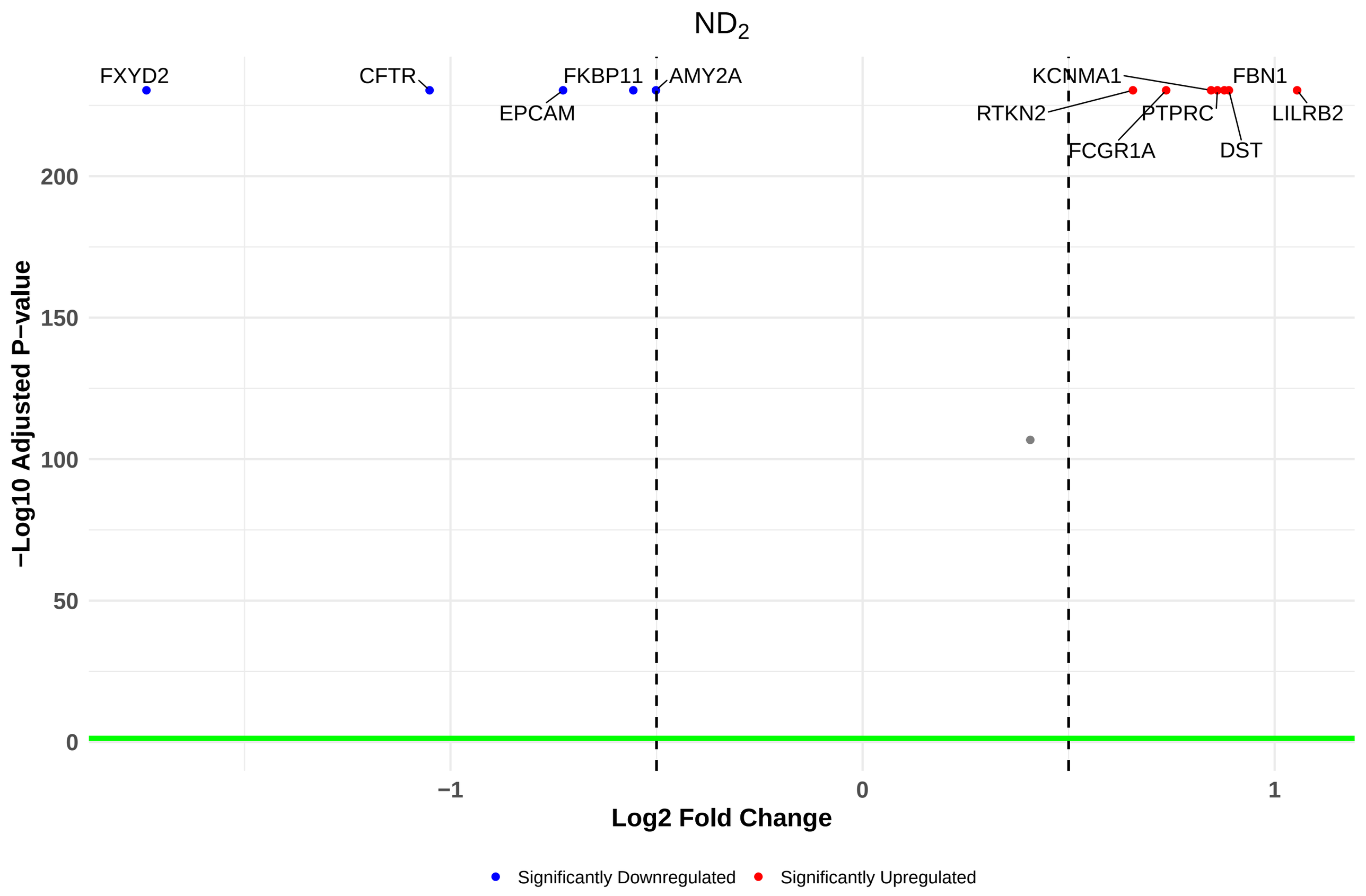
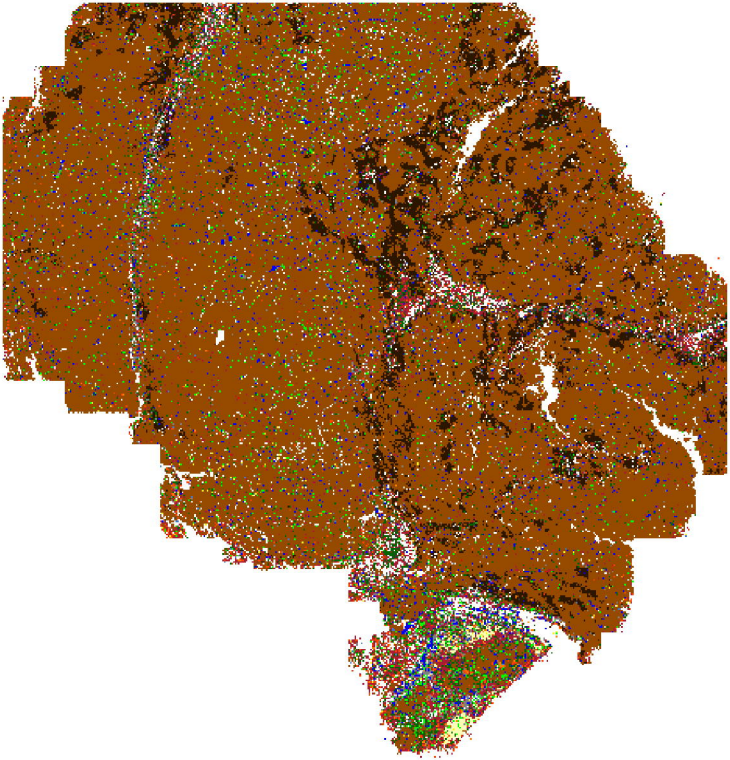


Figure 1: The DEGAS workflow, showing the key inputs and outputs, is overall similar to that of [9]. **B)** Sliding windows progressively shift to capture the entire tissue, enabling localized spatial smoothing of the disease association scores for SWiS and FoVs algorithms. **C, D, E, F)** Spatial visualization of T2D association scores generated by the DEGAS+SWiS model for Xenium tissue samples: ND_1 (panel C), $T2D_1$ (D), ND_2 (E), and $T2D_2$ (F), respectively. Legends display the scaled T2D disease association scores. **G)** Boxplot showing a significant difference in T2D association scores across samples, with higher scores observed in T2D samples compared to non-diabetic samples ($***$ indicates significant p-values). **H)** Combining disease association scores from non-diabetic and T2D samples separately reveals higher overall T2D association scores in diabetic samples. **I, J)** Spatial visualization of LIHC association scores generated by the DEGAS+FoVS model for CosMx tissue samples: Normal (panel I) and LIHC Carcinoma (panel J), respectively. **K)** LIHC association scores for the Carcinoma sample are significantly higher than those of the Non-carcinoma sample. **L)** ROC curve showing patient-level survival prediction by the DEGAS model for classifying 3-year overall survival status in Liver CosMx data. **M)** LIHC association scores by various cell types across the Non-carcinoma and Carcinoma samples, respectively. **N)** Spatial visualization of SKCM association scores generated by the DEGAS+SWiS model for a Xenium tissue sample. **O)** The SKCM association scores for cells within the tumor region are significantly higher than those in the stromal and other non-tumor regions.

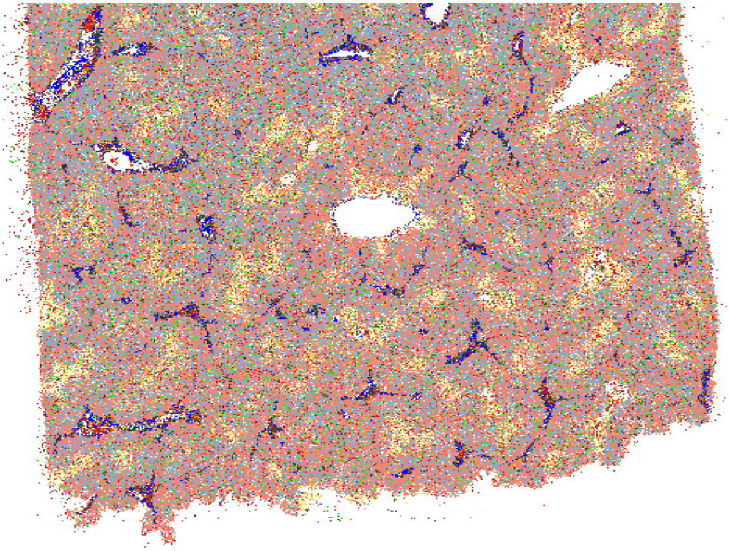

A
DEGAS Workflow
Processing Data
Single cell expression matrix
Patient expression matrix
scSRT data
Patient data
Train DEGAS Model
MMD
Bootstrap Aggregated Densenet Model
Predict with DEGAS
Predict "impressions" of patient outcomes in single cells
Cell-level predictions
Predict single cell characteristics in patients
Patient-level predictions
B
a
b
c
c
Y coordinates
X coordinates
C
ND1
Y coordinates
X coordinates
D
T2D1
T2D Association
0.5
0.0
−0.5
E
ND2
F
T2D2
T2D Association
0.5
0.0
−0.5
G
***
***
T2D Association
ND1
T2D1
ND2
T2D2
Sample
H
***
T2D Association
ND
T2D
ND
T2D
I
Non−carcinoma
LIHC Association
J
Carcinoma
LIHC Association
K
***
LIHC Association
Non−carcinoma
Carcinoma
Sample
L
ROC Curve
Sensitivity
1 − Specificity
AUC = 0.7686
M
Non−carcinoma
Stellate.cells
Non.inflammatory.macrophages
NotDet
Hep.6
Mature.B.cells
Hep.5
Hep.3
Erthyroid.cells
Cholangiocytes
CD3+.alpha.beta.T.cells
NK.like.cells
Inflammatory.macrophages
Hep.4
Hep.1
Central.venous.LSECs
Antibody.secreting.B.cells
gamma.delta.T.cells.1
Periportal.LSECs
Portal.endothelial.cells
Carcinoma
tumor_2
tumor_1
Hep
Stellate.cells
Non.inflammatory.macrophages
NotDet
Mature.B.cells
Erthyroid.cells
Cholangiocytes
CD3+.alpha.beta.T.cells
NK.like.cells
Inflammatory.macrophages
Central.venous.LSECs
Antibody.secreting.B.cells
gamma.delta.T.cells.1
Periportal.LSECs
Portal.endothelial.cells
Average LIHC Association Scores
N
Y coordinate
X coordinate
SKCM Association
0.25
0.00
−0.25
−0.50
−0.75
O
***
SKCM Association
Normal
Tumor


ND1
AMY2A
INS
FXYD2
CFTR
ANPEP
GATM
PROX1
FKBP11
CHGA
MYC
EPCAM
ADAM28
FHL2
PDPN
IGF1
CD163
EDNRB
IL1RL1
PDGFRA
PTPRC
CXCR4
IL7R
DMBT1
AIF1
LYVE1
PPP1R1B
TFPI
MS4A6A
CD14
FBN1
300
200
100
0
−Log10 Adjusted P−value
−2
−1
0
1
2
3
Log2 Fold Change
Significantly Downregulated
Significantly Upregulated

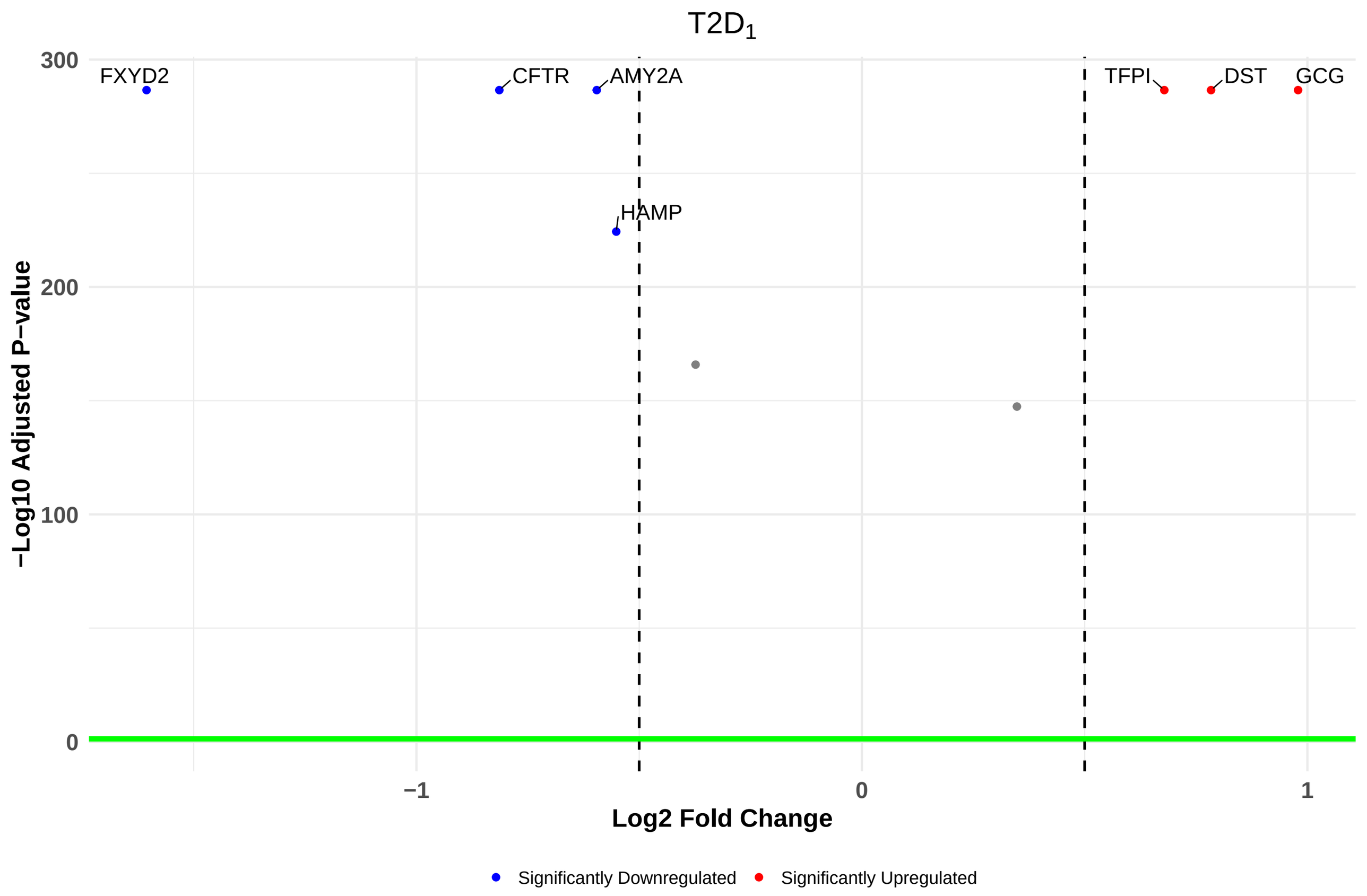

ND2
FXYD2
CFTR
FKBP11
AMY2A
EPCAM
KCNMA1
FBN1
RTKN2
PTPRC
LILRB2
FCGR1A
DST
200
150
100
50
0
−Log10 Adjusted P−value
−1
0
1
Log2 Fold Change
Significantly Downregulated
Significantly Upregulated

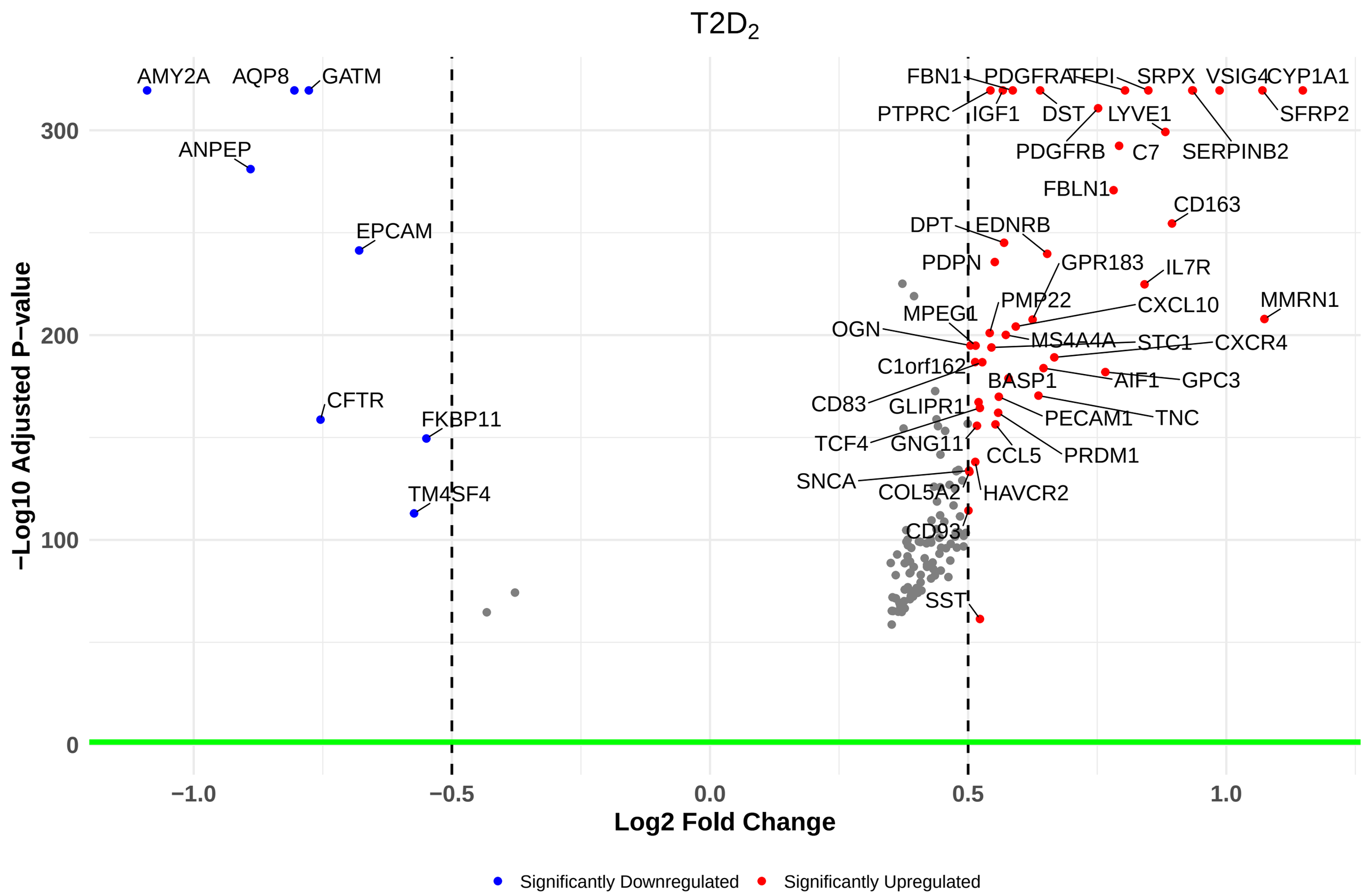


- tumor_2
- tumor_1
- Stellate.cells
- Central.venous.LSECs
- CD3+.alpha.beta.T.cells
- Inflammatory.macrophages
- Portal.endothelial.cells
- Periportal.LSECs
- NK.like.cells
- Non.inflammatory.macrophages
- Mature.B.cells
- Cholangiocytes
- Antibody.secreting.B.cells
- gamma.delta.T.cells.1
- Erthyroid.cells
- NotDet
- Hep
- tumor NA



Hep.3
Hep.5
Hep.4
CD3+.alpha.beta.T.cells
Hep.1
Stellate.cells
Inflammatory.macrophages
Hep.6
Non.inflammatory.macrophages
Mature.B.cells
NK.like.cells
gamma.delta.T.cells.1
Central.venous.LSECs
Antibody.secreting.B.cells
Erthyroid.cells
Cholangiocytes
Portal.endothelial.cells
Periportal.LSECs
NotDet
tumor NA


T2D1
FXYD2
CFTR
AMY2A
HAMP
TFPI
DST
GCG
300
200
100
0
−Log10 Adjusted P−value
−1
0
1
Log2 Fold Change
Significantly Downregulated
Significantly Upregulated


T2D2
AMY2A
AQP8
GATM
ANPEP
EPCAM
CFTR
FKBP11
TM4SF4
FBN1
PDGFRA
TFPI
SRPX
VSIG4
CYP1A1
PTPRC
IGF1
DST
LYVE1
SFRP2
PDGFRB
C7
SERPINB2
FBLN1
CD163
DPT
EDNRB
PDPN
GPR183
IL7R
PMP22
CXCL10
MMRN1
MPEG1
OGN
MS4A4A
STC1
CXCR4
C1orf162
BASP1
AIF1
GPC3
CD83
GLIPR1
PECAM1
TNC
TCF4
GNG11
CCL5
PRDM1
SNCA
COL5A2
HAVCR2
CD93
SST
300
200
100
0
−Log10 Adjusted P−value
−1.0
−0.5
0.0
0.5
1.0
Log2 Fold Change
Significantly Downregulated
Significantly Upregulated