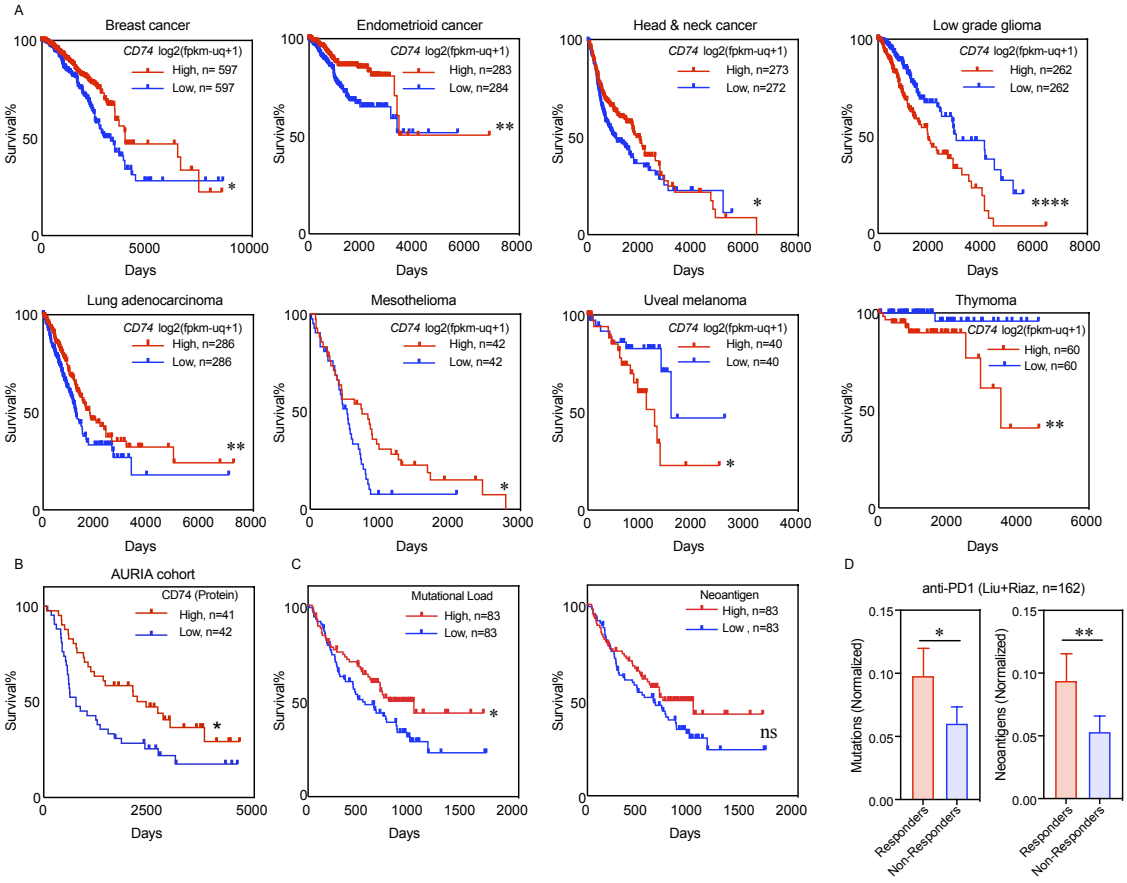
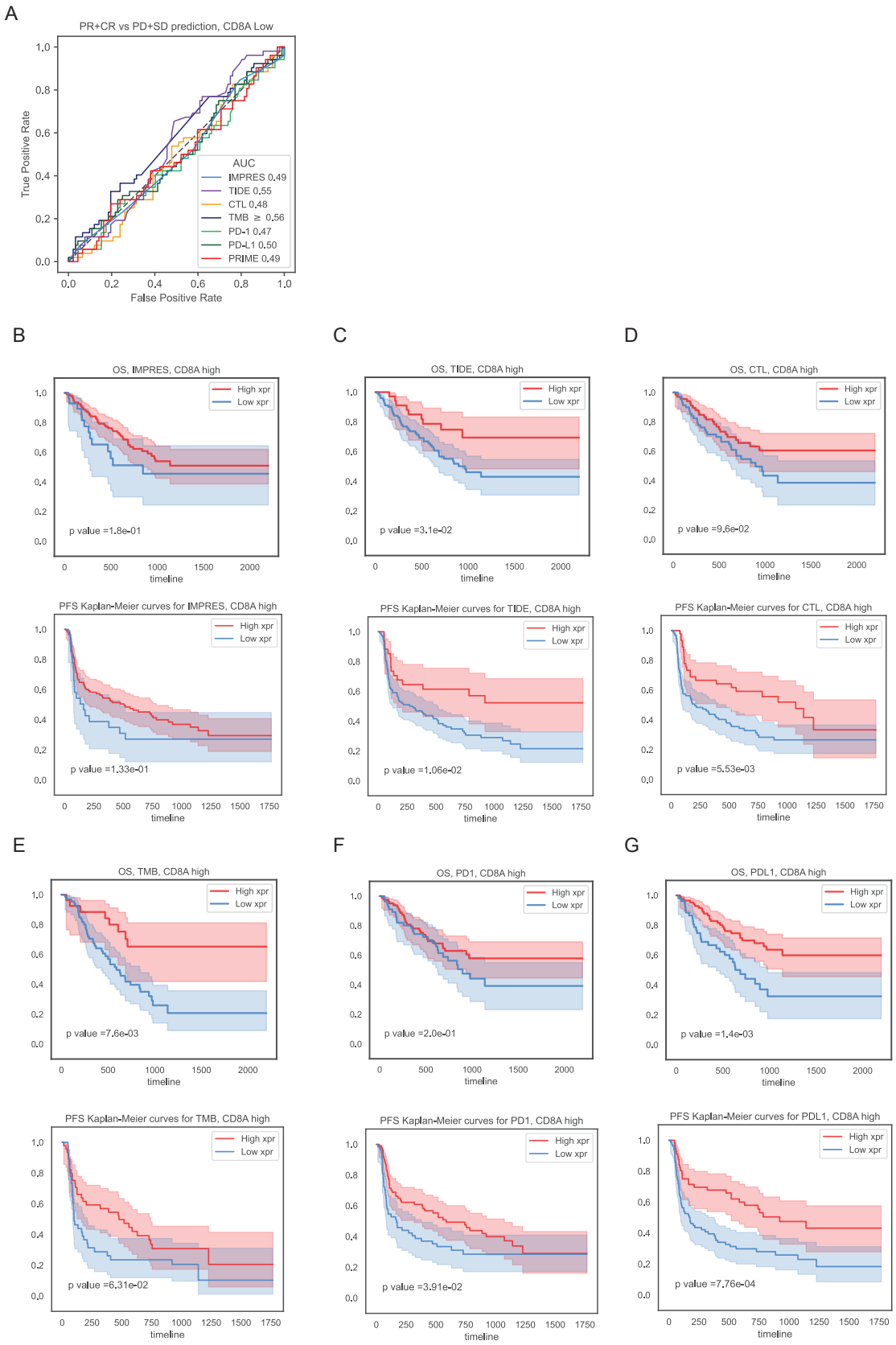
Supplementary Information

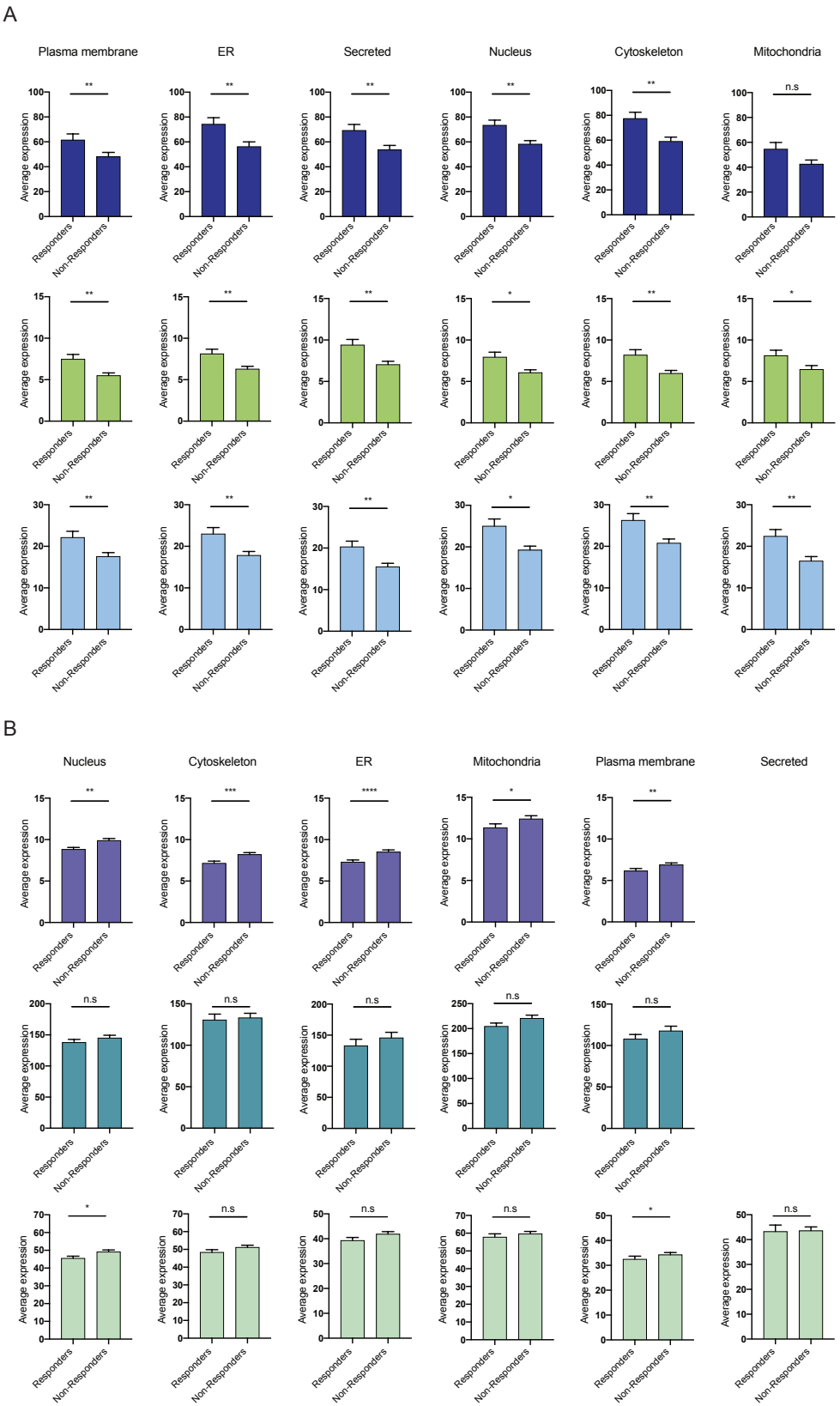
Supplementary Figures



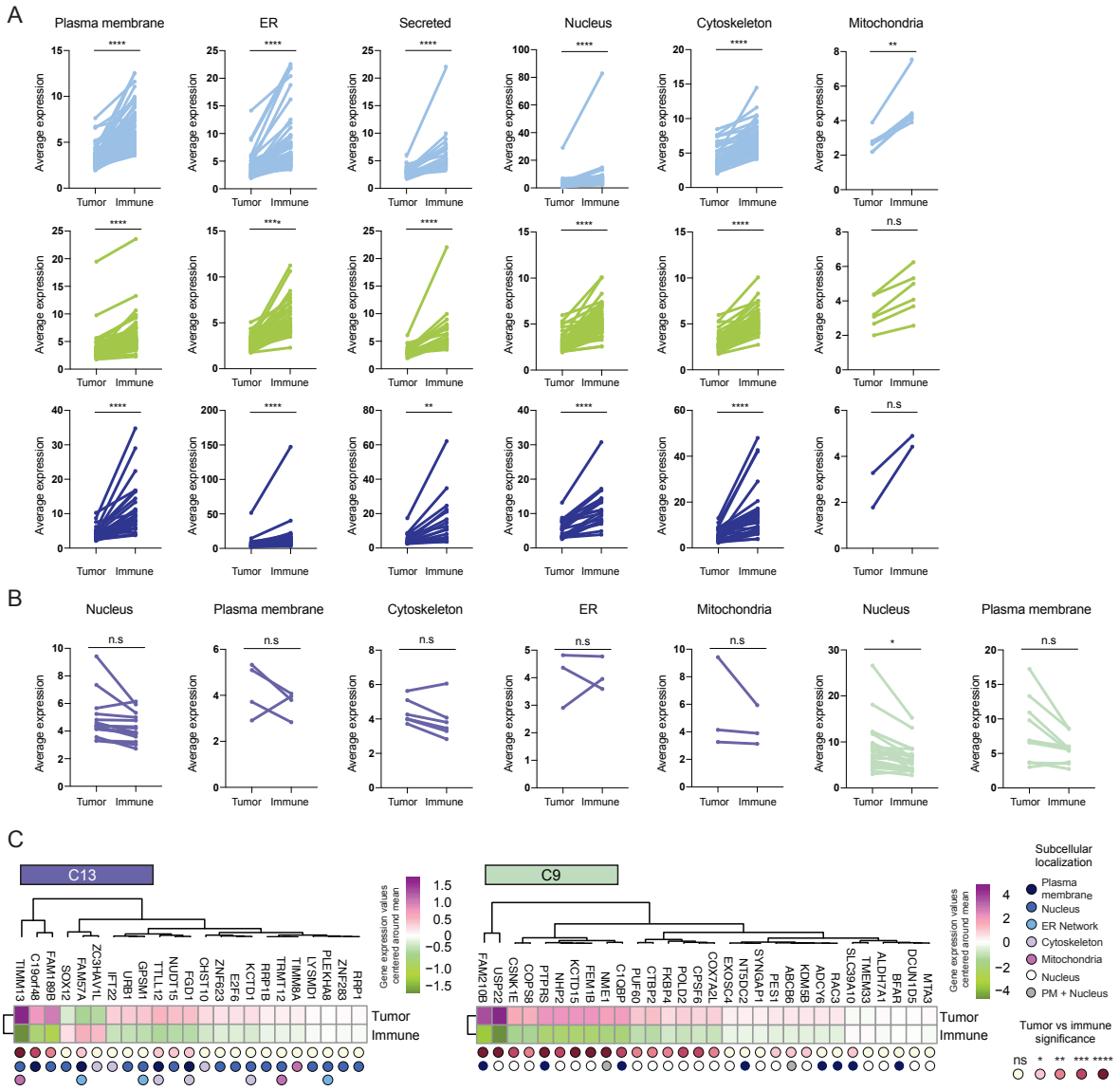
Supplementary Figure 1. *CD74* expression is a predictor of overall survival across different cancers
A) Kaplan Meier survival curves (Log-rank test) of *CD74* in GDC-TCGA cohorts, with cutoff value set at expression median. **B)** Kaplan Meier survival curve (Log-rank test) of CD74 protein measured by immunohistochemistry in AURIA cohort of skin and lymph node primary and metastatic biopsies from metastatic melanoma patients. **C)** Kaplan Meier survival curves of mutation and neoantigen load in anti-PD1 treated Liu [1] and Riaz [2] cohorts (n=166). **D)** Mutation and neoantigen load response prediction power in anti-PD1 Liu [1] and Riaz [2] cohorts (n=171) compared between responders (CR, PR) and non-responders (SD, PD) (Mann-Whitney U test).



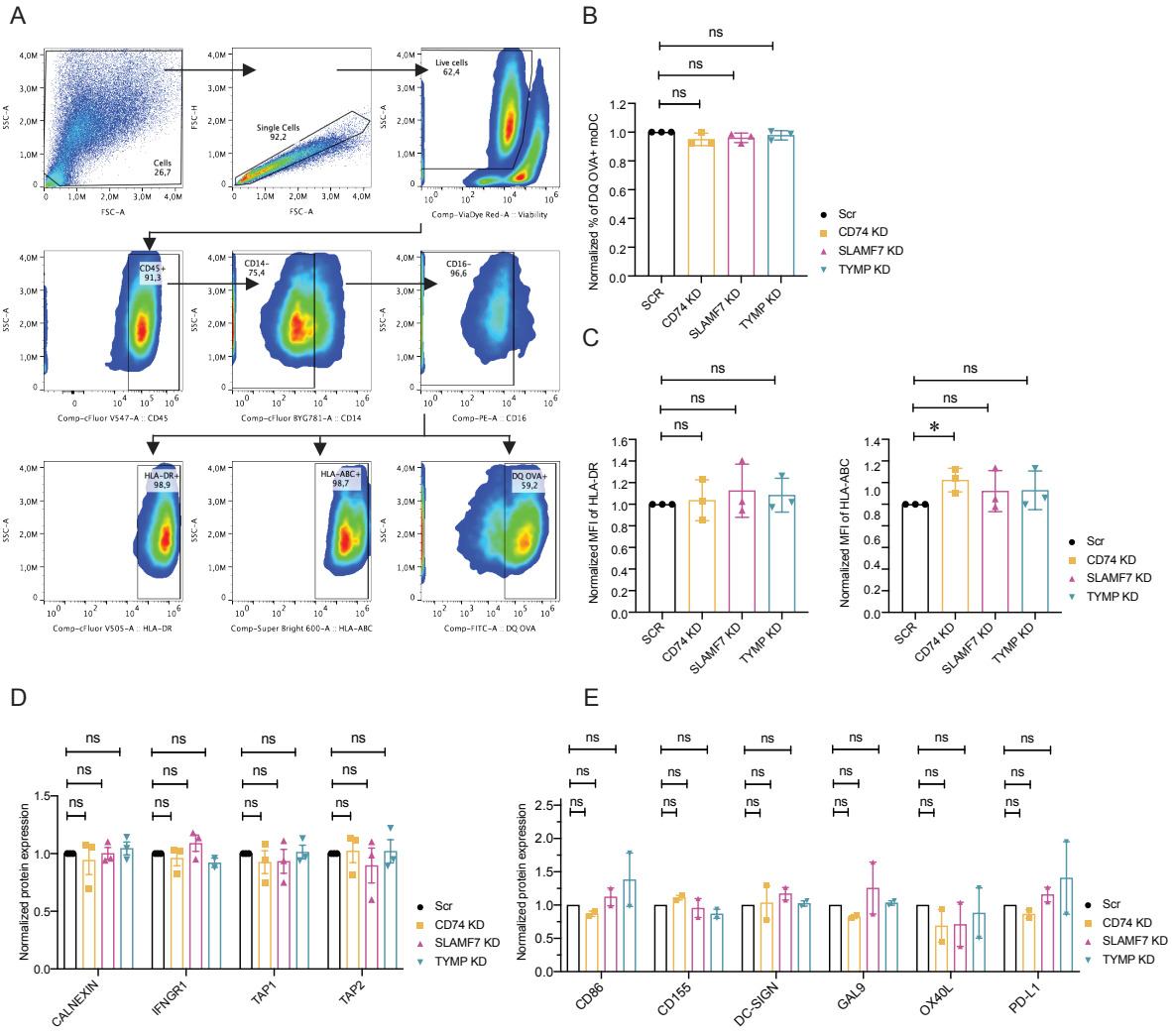
Supplementary Figure 2. *CD8A* stratification reveals predictive limitations in CD8A-low tumors and survival prognostic insights in *CD8A*-high tumors. A) ROC curve assessing predictive performance for RECIST response in melanoma patients with low *CD8A* tumor levels. Kaplan-Meier curves for OS and PFS after *CD8A* stratification for **A)** IMPRES [3] **B)** TIDE [4] **C)** CTL score **D)** TMB **E)** PD-1 and **F)** PD-L1



Supplementary Figure 3. GSP gene expression is enriched in anti-PD1 responders while PSP genes are enriched in non-responders. Average expression of **A)** GSP top 3 cluster genes (C7 dark blue, C2 light green, and C8 light blue) and **B)** PSP top 3 cluster genes (C13 amethyst, C10 teal, and C9 sage) in each subcellular localization compared between responders (CR, PR) and non-responders (SD, PD) in the integrated anti-PD1 dataset of metastatic melanoma patients (n=268) [1, 2, 5, 6]. Mann-Whitney U-test was used after normality testing to test for differences between responders and non-responders for each cluster separately.



Supplementary Figure 4. GSP gene expression is enriched in immune compartments while PSP genes are enriched in tumor compartments of metastatic melanoma skin biopsies. A) Average expression of GSP top 3 cluster genes in each subcellular localization compared between tumor and immune compartments of metastatic melanoma skin biopsies (n=49; Mann-Whitney U-test) **B)** Average expression of PSP top 3 cluster genes in each subcellular localization compared between tumor and immune compartments of metastatic melanoma skin biopsies (n=49; Mann-Whitney U-test). **C)** Heatmap showing mean expression of each PSP gene from response relevant clusters of amethyst and sage clusters in the tumor and immune compartments of metastatic melanoma skin samples and p-value of the difference. T-tests with False Discovery Rate (FDR) corrections were used to assess the significance for each gene between tumor and immune compartments. Heatmap shows gene expression values centered around the mean expression of the cluster where blue indicates values below the mean and red indicated values above the mean. The heatmap was generated using pheatmap package in R.



Supplementary Figure 5. SLAMF7 and TYMP do not regulate the expression of many antigen processing and presentation -associated markers on dendritic cells. A) Gating strategy to identify HLA-ABC, HLA-DR and DQ-Ovalbumin median fluorescence intensities in CD45+CD14-CD16- human monocyte-derived dendritic cells (moDCs) **B)** Percent of DQ Ovalbumin positive moDCs normalized against Scramble (Scr) negative control sample **C)** Normalized median fluorescence intensity of HLA-DR and HLA-ABC in moDCs normalized against Scramble negative control sample **D)** Western blot analysis of calnexin, IFNGR1, TAP1 and TAP2 protein levels in human monocyte-derived dendritic cells with *CD74*, *SLAMF7* or *TYMP* knockdown (KD) compared to Scr. N=3 independent experiments. **E)** Western blot analysis of CD86, CD155, DC-SIGN, Galectin-9, OX40L and PD-L1 protein levels in human monocyte-derived dendritic cells with *CD74*, *SLAMF7* or *TYMP* KD after 24 hours LPS, compared to Scr. N=2 independent experiments. Western blot band intensities were measured using ImageJ and normalized against GAPDH and then against Scr sample for each KD sample. Multiple t-tests with Bonferroni correction were used to adjust for multiple comparisons.

Supplementary Notes

HKPCA batch correction method

The method consists of the following steps:

1. The log2-transformed TPM gene expression values are standardized by subtracting the mean and by dividing by the standard deviation computed over all patients.

2. All data are transformed by PCA with maximum number of components (in this case, the minimum of N(patients)=323 and N(genes)=18505), i.e. with minimal information loss, to a latent space spanned by the principal components, which are ordered by their magnitude of variation in the data as is usual to PCA. This results in a vector $\mathrm{v}(p)$ for each patient p in the latent space. If G denotes the (patients x genes)-matrix of expression values, and P the (patients x genes)-matrix of principal components, then

$$\mathrm{V} = \mathrm{G}P^T$$

is a (patients x patients)-matrix with $\mathrm{v}(p)$ as row vectors. Notice that the Euclidean norm of each row vector of P equals 1.

3. The housekeeping gene data (all patients represented by their housekeeping genes only) are represented in the basis of principal components, i.e. in the latent space, found in step 1. This leads to another representation of each patient p by vector $v_{hk}(p)$ in the latent space. More precisely, let H be the (patients x hk-genes)-matrix with patients represented by the housekeeping genes as row vectors, and let P_{hk} be the restriction of P to housekeeping genes as its column vectors and renormalized in such a way that row vectors have unit Euclidean norm. Then,

$$V_{hk} = \mathrm{H}P_{hk}^T$$

is again a (patients x patients)-matrix with $v_{hk}(p)$ as row vectors representing patients in the latent space by their housekeeping genes.

4. For each batch (cohort) b:

> 4.1 Compute the vector $m_{hk}(b) = \mathrm{mean}\{v_{hk}(p): p \in b\}$ pointing the mean position of the batch represented only by the housekeeping genes in the latent space.
>
> 4.2 The mean vectors $m_{hk}(b)$ are projected in the latent space to a hyperplane spanned by a few (we used 3) first principal components, with most variation in all data. This yields new vectors $m_{hk}^0(b)$ with condensed information about the highest variation in the data.
>
> 4.3 For each patient p in batch b:
>
> > 4.3.1 A Batch-corrected position vector $v_{bc}(p) = \mathrm{v}(p) - \sigma * m_{hk}^0(b)$ is computed.
> >
> > Here $\sigma\text{^}2 = \mathrm{N}(\text{all genes})/\mathrm{N}(\text{housekeeping genes})$ is a scaling factor stemming from the assumption of approximately normally distributed null model points in the latent space.

5. The batch-corrected vectors are transformed back to their original representation by the individual genes by inverse-PCA:

$$G_{bc} = V_{bc}^{T} P ,$$

where V_{bc} is the matrix with rows $v_{bc}(p)$. The matrix G_{bc} is the batch-corrected data with patients as rows and genes as columns.

6. The batch-corrected features can be scaled back to include their original variation by multiplying the batch-corrected vectors by the standard deviation and by adding the mean obtained in step 0.

7. Pseudocounts equal to minimum value in log2(1+batch-corrected TPM) can be added to ensure positive 'counts', if needed.

The HKPCA batch correction was tested to yield similar results with ComBat-seq in terms of correlation of the principal components of batch-corrected data, yet the patients from separate cohorts were mixed better in UMAPs of the batch corrected data. This could be a consequence of ComBat-seq not being designed for TPM data. Our method is also preferred over correction by first few PCA components of the whole data, which is sometimes performed as a simple batch correction, because HKPCA uses only information from housekeeping genes in the latent space and therefore leaves most of the biological variation untouched.

Prognostication framework clustering

To identify clusters of genes with biomarker potential and relevance to antigen presentation and CD8+ T cell generation, genes were classified according to their immune functions using the nCounter immune category list (NanoString Technology) as described previously [59, 60], complemented by literature review. Immune categories were further subcategorized into T cell dysfunction, T cell activation, antigen presentation, immune suppression, immune activation, autoimmunity, epigenetics, metabolism, tumor, motility, and others. The subcellular localization of the corresponding proteins was retrieved from the Human Protein Atlas (https://www.proteinatlas.org), COMPARTMENTS Human knowledge channel (https://compartments.jensenlab.org), or literature review. All classifications are provided in Supplementary Table 4.

For cluster identification, unsupervised clustering was applied using Euclidean distance (dist function) and hierarchical clustering (hclust function) in R, allowing visualization of clusters with shared characteristics. Circos plots were generated using the circlize package in R [100], incorporating survival prediction, gene expression variance (σ^2), immune category, and subcellular localization.

To prioritize clusters with potential biomarker and therapeutic relevance, we developed two scoring metrics:

Priority Cluster Index (PCI) – a metric ranking biomarker clusters based on gene expression levels, expression variance, and cluster robustness.

Functional Priming Index (FPI) – a metric integrating PCI with the functional enrichment of genes related to antigen presentation, immune activation, and T cell priming.

PCI Calculation: PCI ranks biomarker clusters by integrating gene expression properties and cluster robustness, computed as:

$$PCI = (E_{avg} \times \sigma^2)^{(G_{cluster} / G_{total})}$$

where:

E_{avg}: average gene expression in the anti-PD1 dataset
σ^2: expression variance
$G_{cluster}$: number of genes in a given cluster
G_{total}: total number of genes in GSP or PSP

This formula ensures that clusters with higher gene representation and biological coherence receive higher PCI scores.

FPI Calculation: FPI integrates PCI with functional immune category enrichment, calculated as:

$FPI = PCI \times f_{ic}$

where:

f_{ic}: frequency of genes assigned to a functional immune category (ic)

Functional categories were defined based on immune function:
GSP (immune activation categories): antigen presentation, T cell activation, immune activation, and motility.
PSP (immune suppression categories): antigen presentation, T cell dysfunction, immune suppression, and motility.

Response Analysis and Subcellular Localization: To assess the association of subcellular localization with ICT response, we evaluated top PCI-scoring clusters (C8, C2, C7 for GSP; C9, C10, C13 for PSP). The average gene expression for each subcellular category (e.g., ER, plasma membrane, nucleus) was computed, and patients were stratified by clinical response (complete/partial response vs. stable/progressive disease). Normality and significance (p-values) were assessed, with response analysis statistics plotted alongside FPI results.

References

1. Liu, D., et al., *Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma.* Nat Med, 2019. **25**(12): p. 1916-1927.
2. Riaz, N., et al., *Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab.* Cell, 2017. **171**(4): p. 934-949.e16.
3. Auslander, N., et al., *Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma.* Nat Med, 2018. **24**(10): p. 1545-1549.
4. Jiang, P., et al., *Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response.* Nat Med, 2018. **24**(10): p. 1550-1558.
5. Hugo, W., et al., *Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma.* Cell, 2017. **168**(3): p. 542.
6. Gide, T.N., et al., *Distinct Immune Cell Populations Define Response to Anti-PD-1 Monotherapy and Anti-PD-1/Anti-CTLA-4 Combined Therapy.* Cancer Cell, 2019. **35**(2): p. 238-255.e6.