Background The need for a concise and objective immune-specific gene set database is crucial in the era of immune checkpoint blockade (ICB) and adoptive cell cancer (ACT) treatments. It is essential for immunologists to understand treatment mechanisms, molecular distinctions between responders and non-responders, and drivers underlying better survival. However, the current lack of such gene sets hampers immunological research, as existing immune pathway databases are limited and carelessly exploited. Objectively constructed and immunologically relevant pathways provide immunologists with unbiased enrichment results and greater clinical interpretability.

Methods We collected 83 Bulk-RNAseq datasets from the Molecular Signature Database C7. These datasets contain samples challenged with infections of different kinds and magnitudes, possessing yet-to-be discovered immune functions that lie beneath the transcriptomic profiles. Using non-negative matrix factorization (NMF), we identified gene sets with coordinated expression, curated robust NMF programs and merged into meta programs based on Jaccard metric. We validated the clinical utilities of these gene sets with Cancer Genome Atlas (TCGA) pan-cancer, a melanoma ICB cohort and a 10X Genomics Visium FFPE Human Breast Cancer spatial slide.

Results 19 lymphoid and 9 myeloid novel gene sets were constructed (table 1), describing diverse range of immune functions. We confirmed their functions with relevant single cell RNA and T cell receptor sequencing data. These gene sets not only recovered the TCGA immune subtypes (figure 1A) but also defined a novel immune-microenvironment subtype (figure 2) with lowest aneuploidy, TCR diversity, and neoantigen loads but significantly preferable survival (figure 1C,D). These gene sets also provided better discriminatory power for ICB response (figure 1E) and alluded that ICB non-response is pre-destined with high activities in these gene sets at baseline, suggesting possible T cell exhaustion that is irreversible by ICB (figure 1F,G). A risk score derived from these gene sets has better prognostic power in TCGA survival data (figure 1H). Lastly, these gene sets accurately delineate the tumor-immune boundaries in the H&E sections in breast cancer spatial data (tables 1 and 2).

Conclusions The translational utilities of these gene sets in diverse cancer contexts are promising, as gene sets were derived mainly from sepsis experiments, suggesting similarities in immune microenvironment between cancerous and sepsis conditions, assuring the wide applicability of these gene sets in cancer research across various domains. Through the study of gene set activities, immunologists can better understand the immune microenvironment, the drivers behind cancer survival, dissect the ICB treatment mechanism and potentially overcome therapeutic resistance.
Abstract 1289 Figure 1  Translational Implications of these gene sets. (A) Single sample gene set enrichment scores calculated for each TCGA sample across cancer types can well cluster the samples into immunologically quiet, inflammatory, and wound healing/IFN-gamma dominant subtypes. (B) Six clusters were identified by performing kmeans clustering algorithm, cluster 2 (yellow) is a combination of a portion of inflammatory samples and all immunologically quiet samples. (C) TCGA signatures stratified by 6 clusters show that cluster 2 has lower aneuploidy score, neoantigen load and intratumor heterogeneity. (D) Kaplan Meier plot with survival curves for different kmeans clusters shows that cluster 2 has significantly better survival in comparison to the rest of the clusters (Log-Rank Test p-value < 0.0001). (E) ROC for ICB response classification accuracy. (F) ICB cohort: comparing gene sets activity levels at different treatment timepoints and for patients with different responding status (PD: Progressive Disease; SD: Stable Disease; CR: Complete Response; PR: Partial Response). (G) Comparing T cell exhaustion signature at baseline between responders (PR+CR) and non-responders (SD+PD). (H) The risk score derived from COX LASSO model separates TCGA patients into 4 percentiles groups with significantly different survival outcomes regardless of cancer types.
Gene sets can well cluster spatial-omics data. The top panel shows a H&E section for breast cancer tumor sample with pathologist annotation (purple: immune spots; green: tumor spots). The bottom panel shows the expression level of three example gene sets (TCR Anchoring, Cell Killing, and Cytokine Signaling Pathway) across the spatial spots.

http://dx.doi.org/10.1136/jitc-2023-SITC2023.1289