# Immune System Organization is Encoded in Transcription

H. Jabran Zahid*
Microsoft Research - 14820 NE 36th St, Redmond, WA 98052
*Correspondence should be addressed to hzahid@microsoft.com

## Abstract

**The immune system consists of diverse cell types that act in coordination, yet it remains unclear at what biological scale coherent immune organization emerges. Here we analyze single-cell transcriptomics from a large longitudinal cohort of healthy adults and demonstrate that transcription is coordinated across immune cells. Although baseline expression differs by cell type, transcriptional deviations spanning thousands of genes align across immune cells within donors. This alignment resolves into axes of transcriptional variation that organize both inter-donor differences and longitudinal change. Leading axes capture the largest fraction of variance and reflect dynamic shifts shared across donors, whereas higher-order axes encode more stable donor-specific structure. We refer to this shared multi-dimensional variance structure as the immune transcriptional landscape (ITL). An individual's position within the ITL reflects their transcriptional configuration and is recapitulated in matched serum proteomics, linking immune cell transcription to the circulating proteome. Together, these findings demonstrate that coherent immune organization is encoded in coordinated gene expression across immune cells,**

**revealing a multi-dimensional, temporally stratified transcriptional landscape shared across individuals and over time.**

# Introduction

The immune system coordinates diverse cell types and compartments to sustain host defense while preserving tolerance [1, 2, 3]. This coordination reflects integrated regulation across tissues and immune populations, a defining feature of immune homeostasis [4]. Perturbations alter multiple immune compartments, indicating that immune responses operate as a system rather than as isolated cellular processes [5, 6, 7]. Such system-level organization implies that natural variation may be constrained across immune compartments rather than arising independently within each lineage. Large cohort studies further show substantial inter-individual variation in immune composition that remains relatively stable within individuals over time [8, 9, 10, 11]. Together, these observations suggest that immune organization constrains both within-individual dynamics and between-individual differences. Transcription integrates diverse upstream signals and captures cellular state, making it a natural scale at which system-level immune organization may be observed. However, it remains unclear whether donor-level immune organization is encoded in transcription and what form it takes.

Technological advances now enable direct interrogation of immune organization at single-cell resolution. High-dimensional profiling approaches, including transcriptomics, proteomics and mass cytometry, permit systematic characterization of immune cell populations at scale [12, 13, 14]. In particular, single-cell RNA sequencing quantifies gene expression across thousands of genes in individual cells, providing a direct readout of cellular state [15, 16, 17]. While these approaches have revealed extensive cellular heterogeneity [18], they also make it possible to test whether coordinated donor-level organization is encoded in transcription.

Despite advances in single-cell technologies, coherent organization in immune cell tran-

scription has not been systematically characterized. Detecting shared donor-level structure across heterogeneous immune populations requires repeated sampling within individuals and careful integration across cell types. However, many single-cell immune studies are cross-sectional, and longitudinal multi-donor cohorts have been limited by throughput, cost and analytical complexity [19, 20, 21]. In addition, the high dimensionality of single-cell data and the challenge of integrating diverse immune populations complicate detection of shared patterns that may reflect system-level organization. Recent large-scale longitudinal immune monitoring cohorts [22, 23] now provide the depth and repeated sampling necessary to directly test whether immune organization is encoded in transcription.

We analyze single-cell transcriptomes from a large longitudinal cohort of healthy adults sampled repeatedly over time [23]. Gong et al. used this cohort to characterize age-related immune dynamics and vaccine responses through integrated multi-omic profiling. Here we examine whether gene expression across immune lineages encodes coherent donor-level organization. We quantify gene expression deviations relative to lineage-specific baseline profiles and show that these deviations are aligned across immune cell types within donors. This shared structure spans major immune lineages and resolves into a multi-dimensional set of transcriptional axes that organize both inter-donor variation and change over time. We refer to this organizational structure as the immune transcriptional landscape (ITL).

# Results

The Sound Life cohort comprises peripheral blood samples from 96 healthy adults (49 young, 25–35 years; 47 older, 55–65 years) collected at multiple visits across two annual cycles (Fig. 1), enabling longitudinal analyses of immune transcription within donors. Immune cells are annotated at high resolution into 71 immune cell subsets, spanning T cells, B cells, NK cells, monocytes and additional peripheral immune populations [23]. This structure enables direct

examination of whether transcriptional variation across immune lineages is coordinated within individuals.

To examine donor-level immune organization, our analysis proceeds in three biologically motivated steps. First, we summarize gene expression within each immune subtype for every donor and visit by averaging expression profiles across all cells assigned to that subtype. This defines subtype-specific transcriptional profiles for each donor and visit. Second, we remove subtype-specific baseline expression by centering profiles within each subtype relative to the cohort mean profile for that subtype. This step isolates donor-level transcriptional variation by removing the component associated with subtype identity. Third, we examine whether these deviations are coordinated across immune lineages and whether they organize into common axes of transcriptional variation across donors.

For each donor, visit and immune subtype, we compute a mean expression profile by averaging log1p-normalized expression vectors across all cells assigned to that subtype. Groups containing fewer than 100 cells are excluded to ensure stable estimates. Analyses are restricted to a shared set of 3,530 genes defined by prevalence-based filtering across immune subtypes after exclusion of mitochondrial and ribosomal genes, enabling gene-wise comparison of transcriptional patterns across lineages. To focus on homeostatic variation, cells with strong transient activation, interferon-associated or cell-cycle signatures were removed before aggregation (see Methods).

Averaging within each donor–visit–subtype group reduces cell-to-cell heterogeneity while preserving lineage identity and longitudinal structure. These mean profiles therefore serve as the fundamental units of donor-level analysis. For each immune subtype, we define a baseline transcriptional profile as the average donor–visit–subtype mean profile across all donors and visits. Subtracting this baseline from each donor–visit–subtype profile yields a deviation profile that captures gene-wise variation relative to the characteristic transcriptional state of that

subtype. These deviation profiles remove subtype-specific baseline expression while retaining donor- and visit-specific variation and form the basis for all subsequent analyses.

## Shared transcriptional deviations across immune lineages

Immune cell lineages exhibit distinct transcriptional programs that dominate baseline gene expression differences across subtypes (Fig. 2A). After removing these lineage-specific baselines we examine whether deviations from these baselines are coordinated across immune lineages within individual donors. For an illustrative pair of immune subtypes, deviation profiles from the same donor are positively correlated across genes (Fig. 2B), indicating coordinated up- or down-regulation across lineages despite their distinct baseline transcriptional states. This alignment reflects donor-level coordination of transcriptional shifts across immune lineages. When subtypes are randomly paired across donors at the same visit the correlation is lost (Fig. 2C). Extending this analysis across T cells, B cells, NK cells and monocytes reveals consistently positive gene-wise correlations (Fig. 2D), demonstrating that coordinated transcriptional variation generalizes across immune lineages and donors. Although Fig. 2 illustrates a single donor–visit example, significant positive correlations of deviation profiles are observed across all donor–visit pairs in the cohort.

To quantify how much of this coordinated variation reflects stable inter-donor differences versus temporal change within donors, we partition total subtype-resolved variance in deviation profiles into between-donor and within-donor components. Across subtypes, within-donor temporal variation accounts for approximately 55–65% of total variance. Conversely, stable between-donor differences account for approximately 35–45%.

These results demonstrate coordinated gene-wise shifts across immune lineages within donors and establish that temporally varying within-donor changes modestly exceed stable between-donor differences in magnitude. We next ask whether this coordinated variation reflects a shared

donor-level structure.

## Shared donor-level axes of transcriptional variation

If donor-level transcriptional variation is coordinated across immune lineages, this shared structure should be representable using common transcriptional axes. Because principal components are ordered by variance explained rather than biological importance, biologically meaningful structure may be distributed across multiple components. We perform principal component analysis (PCA) independently within immune compartments using deviation profiles (see Methods). Within each compartment, donor ordering along the first principal component is highly concordant across subtypes, indicating that dominant donor-level variation is recovered despite differences in subtype identity. Shared donor-level structure extends beyond this single component to additional principal components. Because the first principal component is consistent across subtypes, subtype-level scores represent repeated measurements of a common compartment-level donor signal. We therefore aggregate these scores within each donor and visit to obtain a single compartment-level donor score.

Donor ordering along this compartment-averaged first principal component is strongly correlated across T cells, B cells, NK cells and monocytes at a fixed visit (Fig. 3A). This concordance indicates that the first principal component reflects a transcriptional axis shared across immune compartments. Projecting deviation profiles into a reference PCA basis (CD4 naive) further shows that this shared structure extends beyond the first principal component (Fig. 3B), supporting a multi-dimensional organization of transcriptional variation shared across compartments. These results demonstrate that deviations from subtype-specific baselines align across immune compartments at the donor level, providing a basis for consolidating compartment-level principal component scores into a single donor representation.

To evaluate the biological coherence of this consolidated donor representation, we examine

the temporal behavior of the first principal component score derived by averaging across high-prevalence L3 subtypes. Donor ordering is strongly preserved between Year 1 Day 0 and Day 7 (Fig. 3C) but is weaker between Day 0 and Day 90 (Fig. 3D), with the same pattern observed in Year 2. This indicates that donor-level transcriptional organization is stable over short intervals yet evolves over longer timescales, a pattern that extends across additional principal components. This structured temporal pattern suggests coordinated biological organization.

Together, these analyses show that the coordinated transcriptional shifts observed across immune lineages in Fig. 2 reflect a shared donor-level structure that is consistent across immune compartments. This shared structure is captured by a multi-dimensional set of transcriptional axes, motivating a unified framework for representing donor-level transcriptional variation across immune lineages.

## Defining the immune transcriptional landscape

Given the shared donor-level structure recovered independently across compartments, we construct a joint representation of transcriptional variation across immune lineages by applying PCA to deviation profiles pooled across immune subtypes and donor visits. The resulting principal components define a coordinate system in which donor-level transcriptional organization can be represented and quantified. For each principal component, we aggregate subtype-level scores across well-represented subtypes to obtain a single score for each donor–visit pair. Each axis corresponds to a distinct pattern of correlated transcriptional variation across many genes. We refer to this unified representation as the immune transcriptional landscape (ITL).

The ITL reveals that donor-level transcriptional variation can be summarized by a relatively small number of shared axes, with the leading components capturing a substantial fraction of total variance (Fig. 4A). These axes account for a substantial fraction of variance within each lineage (Fig. 4B), recapitulating that donor-level transcriptional variation across immune com-

partments is organized along shared directions. However, the fraction of variance explained by individual components differs modestly across lineages, indicating that while donor-level structure is broadly shared, its expression remains partially lineage-dependent. The first principal component of the ITL preserves the donor ordering observed in compartment-specific analyses (Fig. 4C), confirming that the ITL consolidates previously identified donor-level structure within a unified coordinate system. Importantly, similar cross-compartment structure extends beyond this first component, consistent with the ITL representing a multi-dimensional organization of transcriptional variation.

Genes contributing to the first principal component of the ITL (Fig. 4D) span diverse functional categories rather than being dominated by a single transcriptional program, indicating that this axis reflects broad transcriptional organization. Similar distributed gene contributions are observed across ITL axes more generally. We observe that temporal coherence varies with component index (Fig. 4E); leading components capture short-term transcriptional fluctuations, whereas higher-order components capture transcriptional variation that is stable within donors over time but differs across donors. Intraclass correlation analysis further shows that stable donor-specific structure is distributed across multiple principal components (Fig. 4F). Although the ITL is constructed without temporal information, this stratified pattern emerges naturally.

Although principal components are ordered by the amount of variance they explain, biologically meaningful signals can be distributed across many components and do not necessarily coincide with the largest sources of variation. Consistent with this pattern, higher-order ITL components retain highly stable donor-associated signal. In particular, PCs 100–150 show strong donor stability. Under a strict annual holdout used for this nearest-neighbor analysis, all Flu Year 2 samples are excluded during PCA fitting and subsequently projected into the learned basis. In this setting, nearest-neighbor matching in the space defined by PCs 100–150 recovers donor identity across years with near-perfect accuracy, compared to 7% using the leading

principal components (PCs 1–10). This finding is consistent with the temporal coherence and intraclass correlation patterns observed across ITL components. Permutation of donor labels reduces recovery to chance levels (mean accuracy 1%, empirical $p < 0.001$).

To assess whether these regimes differ in the transcriptional programs they capture, we examined gene loadings associated with the leading dynamic components (PCs 1–10) and the donor-identity components (PCs 100–150). The dynamic regime shows strong contributions from canonical stimulus-response genes, including immediate early transcription factors such as FOS, JUN, DUSP1 and IER2, as well as genes involved in antigen presentation and immune activation. The prominence of these genes indicates that this structure remains detectable after removal of transiently activated cells. In contrast, these activation-associated genes contribute less strongly to the donor-identity regime, whose loadings are distributed across genes associated with baseline immune signaling, metabolism and cellular regulation. This contrast is consistent with the temporal structure of the ITL. Early components reflect rapid transcriptional responses that vary within individuals, whereas higher-order components capture more stable regulatory configurations that distinguish individuals.

Together, these analyses show that the coordinated transcriptional shifts observed across immune lineages in Fig. 2 reflect a shared donor-level structure that is consistent across immune compartments. This structure is captured by a multi-dimensional set of transcriptional axes, with leading components reflecting dynamic transcriptional variation within donors and higher-order components capturing stable donor-specific variance.

## External replication and robustness of the ITL

To determine whether the donor-level transcriptional organization identified in the primary cohort reflects a general property of human immune variation, we analyze an independent cross-sectional PBMC cohort of 166 healthy donors generated using a distinct study design and pro-

cessing pipeline [24]. Applying the same analytical framework, we observe that deviation profiles correlate across immune lineages within donors (Fig. 5A), PCA recovers a shared donor ordering across compartments (Fig. 5B) and cross-compartment coherence extends across multiple principal components (Fig. 5C). Because the external cohort is cross-sectional, these analyses assess replication of cross-lineage deviation alignment and shared donor-level structure but do not test the temporal stratification observed in the longitudinal Sound Life cohort.

We further evaluate robustness to preprocessing choices and technical variation through targeted sensitivity analyses. The overall geometry and donor-level temporal coherence of the ITL remain stable when PCA is fit after excluding Flu Year 2 Day 0 and Day 7 samples, recomputing subtype centering using training data only and projecting the held-out data into the learned basis, as well as when the number of cells contributing to each donor–visit–subtype profile is downsampled prior to aggregation. The ITL is similarly stable under alternative gene prevalence thresholds and exclusion of activation-associated or high-loading genes. Although PC1 correlates with mean detected genes and total UMI counts, consistent with known relationships between global transcriptional output and sequencing complexity in single-cell RNA-seq data [25], this association is largely confined to the first principal component. The donor-level immune organization described here is distributed across many axes and remains evident beyond the first principal component, indicating that the ITL structure is not sensitive to sequencing complexity effects captured in PC1. Together, these analyses demonstrate that the ITL structure is robust to preprocessing choices, sampling variability and sequencing-related effects.

## The immune transcriptional landscape across molecular and physiological modalities

To determine whether the ITL reflects system-level immune organization beyond transcriptional measurements, we evaluate its correspondence with independent external modalities. We test

whether donor positions within the ITL are reflected in matched circulating measurements, including routine hematologic indices and serum protein profiles derived from a predefined panel of immune-related plasma proteins. Donor position within the ITL is highly constrained. Thus, for these analyses we summarize donor positions in ITL space by applying PCA to the first 150 ITL components. Whereas the ITL defines axes of coordinated transcriptional variation across genes, this reduced representation captures the dominant axes of variation in donor immune state defined by donor position within the ITL. For proteomic measurements, we derive an analogous low-dimensional representation, enabling direct comparison of donor positions across transcriptional and proteomic spaces (see Methods).

Correlation of components of the reduced ITL representation with routine hematologic indices reveals a structured pattern. The first few components of the reduced representation, which capture the dominant donor-level and longitudinal variation, show minimal association with complete blood count measurements. Modest correlations with hematologic variables emerge in intermediate components, while stronger associations appear in later components, most notably with hemoglobin, hematocrit and platelet count. Thus, classical blood physiology is reflected in identifiable directions of the ITL but does not account for its dominant transcriptional structure.

Extending this cross-modal evaluation to molecular measurements, we next assess correspondence between donor-level structure in the ITL and circulating proteomic space. Using the same reduced ITL representation, we compare donor positions across transcriptional and proteomic manifolds. For a single ITL axis, donor ordering defined by transcriptional variation is accurately recovered from proteomic features in held-out data (Fig. 6A), demonstrating alignment between modalities. To evaluate correspondence beyond a single axis, we compare donor positions across the first ten principal components of both representations under visit-level and donor-level holdout. In each case, cross-modal alignment exceeds permutation expectations

(Fig. 6B–C). These results indicate that donor-level immune organization captured by the ITL is coherently reflected in circulating proteomic measurements.

We observe coordinated transcriptional variation across immune lineages that reflects donor-level structure across immune compartments. This organizational structure is concordant with hematological indices and proteomic measurements. These results show that donor-level immune organization is directly observable in the transcriptional states of immune cells.

# Discussion

Immune cell transcription varies across individuals and over time. These changes are not random but follow consistent patterns across immune lineages. The dominant component of variance reflects longitudinal change rather than fixed differences between donors. Beneath this dynamic layer lies a more stable component of expression that persists across annual visits and enables recovery of donor identity from expression profiles. Together, these observations indicate that immune transcription exhibits a layered structure in which coordinated, shared dynamics account for most variation, while stable donor-specific expression set points persist but account for less variance. This layered organization reflects a simple mode of immune regulation: transcription shifts coherently across immune lineages, moving the transcriptional state of the system as a whole.

The immune system operates as an integrated network of diverse cell types [26], yet the scale at which coherent system-level organization becomes directly detectable has remained uncertain. We demonstrate that this organization is observable in cellular transcription and formalize it as the immune transcriptional landscape (ITL). The ITL provides a coordinate system for quantifying shared donor-level transcriptional variation. Donor positions within this space correspond to independent circulating measurements, including routine hematologic indices and matched circulating proteomic profiles [11, 27], linking cellular transcription to systemic

immune state. Because transcription integrates upstream regulatory inputs and shapes downstream cellular behavior [28, 29], the ITL situates donor-level immune organization within cellular state rather than limiting it to higher-order phenotypes such as immune cell frequencies or circulating cytokine levels [13]. The ITL does not replace lineage identity or specify underlying mechanisms; instead, it provides a formal framework for representing and quantifying donor-level transcriptional organization.

The structure of the ITL constrains plausible mechanisms of immune organization. Coordinated transcriptional variation across immune compartments indicates that the dominant drivers are unlikely to be lineage-intrinsic and instead act in parallel across cell types [30, 31]. The signal spans thousands of genes and remains evident in sensitivity analyses excluding canonical activation-associated programs, implying distributed, transcriptome-wide modulation rather than a limited pathway effect [32, 33]. Moreover, the relative dynamism of leading components and the longitudinal stability of higher-order components suggest regulatory processes operating across multiple timescales [34, 32, 10]. Stable donor-associated axes may reflect multiple influences, including genetic variation, sex, environmental exposures and persistent immune regulatory set points that shape baseline immune configuration across individuals. Together, these features are consistent with systemic regulators that couple immune compartments, including physiological and environmental influences [34, 32, 33], homeostatic cytokine and circulating protein signaling [11, 27] and stable donor-specific regulatory architecture shaped by genetic and non-heritable factors [8, 35, 30].

Prior work has shown that the human immune system exhibits substantial inter-donor diversity while remaining relatively stable within donors over time [8, 9, 10, 11]. Persistent immune signatures and donor-specific set points have been described across multiple cohorts [36, 37, 14]. Our results resolve these phenomena at the level of transcription, revealing that stability and diversity are embedded within a shared, multi-dimensional variance structure. Rather than

reflecting a single dominant axis of immune state, this structure arises from coordinated transcriptional dynamics layered over stable donor-specific set points. The ITL formalizes shared donor-level structure consistent with prior observations across immune compartments [13, 26, 3] and provides a quantitative foundation for analyzing immune organization directly in transcriptional space.

The present analysis characterizes donor-level immune organization in circulating immune cells from healthy adults. Because the longitudinal design includes seasonal influenza vaccination, an important next step will be to determine how defined immune perturbations such as vaccination reshape donor trajectories in the ITL. We aggregate expression across many cells within a subtype to isolate donor-level structure and therefore do not resolve fine-grained cell-to-cell heterogeneity layered upon this donor-level signal. Analyzing single cells could clarify how donor-level set points are instantiated at cellular resolution. Whether comparable multi-dimensional structure is preserved across tissue-resident immune populations, diverse disease states or defined perturbations remains to be determined. In addition, while the ITL captures coordinated transcriptional variation, it does not specify the upstream regulatory processes that give rise to these axes. Applying this framework to longitudinal perturbation studies, tissue-resolved profiling and integrative multi-omic analyses will be essential for delineating the causal drivers of donor-level immune organization and for determining how this organization is altered in pathological contexts.

We establish the immune transcriptional landscape as a quantitative framework for interrogating donor-level immune organization directly in gene expression space. By defining a shared coordinate system for immune transcription, the ITL enables systematic analysis of how immune states vary across individuals, evolve over time and relate to systemic phenotypes such as circulating proteomic profiles. By revealing that system-level immune structure is encoded in coordinated transcriptional dynamics across cell types, the ITL links organism-level immune

organization to cellular state. This framework provides a foundation for investigating how stable donor-specific set points are maintained, how shared transcriptional programs are regulated and how these layered structures are reshaped across aging, environmental exposure, physiological perturbation and disease.

# Materials & Methods

## Cohorts

**Sound Life longitudinal cohort.** Primary analyses use data from the Sound Life longitudinal healthy adult cohort, a prospective multi-omic study with repeated immune profiling of the same individuals over time [23]. Participants are sampled across annual cycles with a standardized visit structure including baseline (Day 0), short-term (Day 7) and intermediate-term (Day 90) follow-up visits, with seasonal influenza vaccination administered at Day 0 of each cycle. In addition to these core visits, a subset of donors contributes additional intermediate visits. The same individuals are followed across multiple years, enabling within-subject analysis of immune transcriptional organization over weeks, months and years.

**Independent healthy PBMC atlas.** To assess generality of the transcriptional structure identified in the Sound Life cohort, we analyze an independent large-scale single-cell atlas of healthy human peripheral blood mononuclear cells described by Terekhova et al. [24]. This cohort comprises single-cell transcriptomic profiles from 166 healthy donors and represents a primarily cross-sectional design. Raw UMI count matrices and subtype annotations are obtained and processed using the same normalization, filtering and aggregation workflow applied to the Sound Life data, as described below. Analyses in the external cohort are restricted to the Sound Life ITL gene universe, retaining the intersection of genes present in both datasets (all but one gene). The external cohort is used exclusively to test whether cross-lineage deviation alignment and

shared donor-level structure are present in an independent dataset; no projection of external samples into the Sound Life ITL is performed.

## Starting data and preprocessing state

For both cohorts, analyses begin from post-quality-control, annotated single-cell datasets with raw UMI count matrices retained. Cell-level quality control, doublet removal and initial subtype annotation are performed by the original study authors [23, 24] and are not repeated here. Downstream analyses operate on these post-QC data with additional analysis-specific filtering and aggregation steps described below.

## Cell filtering, normalization and aggregation

All primary analyses start from raw UMI count matrices stored in backed AnnData objects. For each cell, raw counts are library-size normalized to a common target sum $T = 4524$ and transformed using $\log(1 + x)$. Specifically, if $x_{ig}$ denotes the raw UMI count for gene $g$ in cell $i$, and $L_i = \sum_g x_{ig}$ is the library size of cell $i$, we compute

$$\tilde{x}_{ig} = \log\left(1 + \frac{T}{L_i} x_{ig}\right),$$

for cells with $L_i > 0$.

To reduce the influence of transient non-homeostatic states on donor-level baseline structure, cells exhibiting strong transient activation, cell-cycle or interferon-associated signatures are filtered using subtype-specific thresholds computed within L3 labels. For each cell, module scores for three predefined gene sets (cell-cycle S, cell-cycle G2M, and a core interferon-stimulated gene set) are computed as the mean of $\tilde{x}_{ig}$ over genes in the set. For each L3 subtype, a cutoff is defined as the 99th percentile of the corresponding score distribution across retained cells, and cells are excluded if they exceed the subtype-specific cutoff for any of the three scores. Results are insensitive to the exact cutoff used.

For each donor–visit–L3 group, subtype-resolved mean expression profiles are computed by taking the arithmetic mean of $\tilde{x}_{ig}$ across all retained cells in the group for each gene. Only groups containing at least 100 cells are retained for downstream analysis. The resulting group-by-gene matrix of mean log-normalized expression values forms the basis for gene filtering, baseline centering and dimensionality reduction.

## Gene universe definition

To define a consistent gene universe for dimensionality reduction and cross-lineage comparison, genes are filtered using a prevalence-based criterion applied to retained cells. Gene prevalence is defined as the fraction of cells with nonzero raw UMI counts for a given gene. Prevalence is computed within coarse immune label groups derived from L3 annotations, corresponding to the major T cell compartments represented in each dataset and consistent with the file structure of the data. Genes are retained if their prevalence exceeds 10% within each coarse label group. The resulting gene sets are intersected across groups to define a shared gene universe.

Mitochondrial and ribosomal genes are excluded from the retained gene universe prior to principal component analysis. No additional exclusion of activation-associated genes was applied in the primary gene universe. Instead, transient non-homeostatic states were controlled at the cell level through the filtering procedure described above. Exclusion of selected activation-related gene programs was performed only in dedicated sensitivity analyses. Genes with names beginning with the prefix `MT-` are classified as mitochondrial genes, and genes with prefixes `RPL` or `RPS` are classified as ribosomal genes. Although excluded from the PCA gene set, aggregate mitochondrial and ribosomal expression levels are computed separately for each donor–visit–subtype group and retained for diagnostic and alignment analyses.

## Baseline centering and deviation profiles

Subtype-resolved mean expression profiles are centered to define lineage-specific baseline expression. For each L3 subtype, the baseline expression vector is defined as the mean of subtype-resolved group-level expression profiles across all donors and visits retained for analysis. Deviation profiles are computed by subtracting this subtype baseline from each donor–visit–subtype mean expression profile. These centered deviation profiles remove systematic baseline differences between immune subtypes while preserving coordinated variation across donors and visits, and serve as input to deviation alignment analyses and principal component analysis. Except where noted for illustrative purposes (Fig. 2A), all downstream analyses operate on these donor–visit–subtype deviation profiles or linear transformations derived from them.

## Deviation alignment analyses

To assess whether transcriptional deviations from lineage-specific baselines are shared across immune lineages within individuals, we analyze centered deviation profiles defined as described above.

**Illustrative single donor–visit analysis.** For visualization, a representative donor–visit with high cell counts in paired immune subtypes is selected. Gene-wise Spearman correlation across deviation profiles is used to quantify alignment between lineages, illustrating the effect of subtype baseline centering on cross-lineage correspondence.

**Population-level deviation alignment.** For population-level analysis, gene-wise Spearman correlations are computed between deviation profiles for all donor–visit combinations in which the relevant lineage pairs are present. Correlations are calculated across genes independently for each donor–visit, yielding a distribution of correlation coefficients across the cohort. This

procedure is applied uniformly to T cell subsets, B cells, NK cells and monocytes using the same centering strategy.

**Null model construction.** Statistical significance of deviation alignment is evaluated using a visit-preserving donor-shuffled null model. For each visit independently, donor identities are permuted with the constraint that no donor is paired with itself. Deviation profiles from one lineage are paired with deviation profiles from a different donor at the same visit, preserving visit-specific context while disrupting within-individual correspondence. The resulting distribution of gene-wise Spearman correlations defines the null expectation.

## Compartment-specific PCA and cross-compartment coherence

To characterize donor-level transcriptional structure within immune compartments, principal component analysis (PCA) is performed independently within each major compartment using deviation profiles as input. PCA is fit using randomized singular value decomposition with a fixed random seed. Principal component scores and gene loading vectors are retained for downstream analyses.

**Donor-level aggregation.** For cross-compartment comparisons, subtype-level principal component scores are aggregated to donor–visit level by averaging across L3 subtypes and then averaging across coarse compartment labels (L2) within each donor at a fixed visit (L3 $\rightarrow$ L2 $\rightarrow$ donor). This produces a single donor-level score per principal component for each compartment and visit.

**Cross-compartment coherence of the first principal component.** To assess whether dominant donor-level structure is shared across immune compartments, donor-level scores for the first principal component are compared across compartments at a fixed visit using Spearman

correlation, yielding a cross-compartment correlation matrix.

**Cross-compartment coherence across principal components.** To evaluate whether shared donor-level structure extends beyond the first principal component, deviation profiles from each target compartment are projected into a fixed PCA basis defined in a reference compartment. Let $X^{(t)} \in \mathbb{R}^{n_t \times G}$ denote the centered group-by-gene matrix for target compartment $t$, and let $L_k^{(r)} \in \mathbb{R}^{k \times G}$ denote the loading vectors corresponding to the first $k$ principal components from the reference compartment, with all compartments restricted to a shared ordered gene list. Projected scores are computed as

$$Z_k^{(t \to r)} = X^{(t)} {L_k^{(r)}}^{\mathsf{T}}.$$

Projected scores are aggregated to donor–visit level as described above. For each principal component index $j \in \{1, \ldots, k\}$, donor-level scores from the reference compartment (native PCA space) are compared to donor-level projected scores from each target compartment using Spearman correlation at a fixed visit. Correlation curves as a function of principal component index summarize cross-compartment coherence across leading and higher-order principal components.

## Construction of the immune transcriptional landscape

The immune transcriptional landscape (ITL) is constructed by pooling donor–visit–subtype deviation profiles across immune compartments and performing principal component analysis (PCA) on the resulting joint matrix. PCA is fit with `n_components=None`, retaining all principal components of the joint matrix. For each input pipeline (cd8_naive, cd8_memory, cd4_naive, cd4_memory, b, nk, mono), a shared gene list is defined as the intersection of `genes_kept` across all pipelines and each pipeline's group-level mean expression matrix

(`group_means`) is subset and reindexed to this shared ordered gene list. The resulting matrices are concatenated row-wise to form a joint group-by-gene matrix of deviation profiles. In total, this matrix contains 14,979 deviation profiles across 3,530 genes.

Principal component analysis is then performed on this joint matrix using randomized singular value decomposition with a fixed random seed. Principal component scores, gene loading vectors and variance explained are retained. The resulting principal components define the ITL axes.

## Variance captured within lineages by the ITL subspace

To quantify how much within-lineage variance is captured by the ITL subspace (Fig. 4B), we use the joint centered matrix and the ITL gene-space basis. Let $X_L$ denote the subset of rows of the joint centered matrix corresponding to lineage $L$ (mapping sources to five lineages: CD8 includes cd8_naive and cd8_memory; CD4 includes cd4_naive and cd4_memory; B, NK, and Mono correspond to b, nk, and mono). Let $U_k$ denote the first $k$ ITL axes in gene space (rows of `pca_model.components_`). For each lineage, the fraction of variance captured by the first $k$ ITL axes is computed as

$$\mathrm{FracVar}_L(k) = \frac{\|X_L U_k^{\mathsf{T}}\|_F^2}{\|X_L\|_F^2},$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Curves $\mathrm{FracVar}_L(k)$ are evaluated over component index $k$ using a fixed maximum $K_{\max}$.

## Temporal coherence across ITL axes

Temporal coherence of donor-level ITL structure across axes (Fig. 4E) is evaluated using donor–visit ITL coordinates constructed by robust aggregation across high-prevalence L3 subtypes. An L3 subtype is considered present for a donor if it appears in the ITL `pca_scores` for that donor at any visit. Subtypes present in at least 95% of donors are retained, and analyses are

restricted to the six canonical visits (Flu Year 1 Day 0, Day 7, Day 90; Flu Year 2 Day 0, Day 7, Day 90). Donor–visit coordinates are computed by averaging ITL principal component scores across the retained high-prevalence L3 subtypes (direct L3 $\rightarrow$ donor–visit aggregation without intermediate L2 averaging).

For each pair of visits $(v_1, v_2)$, temporal coherence is quantified for each ITL axis by computing the Spearman correlation of donor-level scores between $v_1$ and $v_2$ across donors with measurements at both visits. Short-term coherence is computed for Day 0 vs Day 7 within each year and averaged across Year 1 and Year 2; intermediate-term coherence is computed for Day 0 vs Day 90 within each year and averaged across Year 1 and Year 2; and long-term coherence is computed using matched cross-year comparisons (Year 1 Day 0 vs Year 2 Day 0 and Year 1 Day 7 vs Year 2 Day 7), averaged across the two comparisons. For visualization, correlation curves across axis index are smoothed using a running median with a fixed window size (window size = 11).

## Intraclass correlation across ITL axes

To quantify the extent to which each ITL axis reflects stable between-donor differences versus within-donor fluctuations (Fig. 4F), we compute an intraclass correlation-like statistic on the same donor–visit ITL coordinates used for temporal coherence. Using the robust L3 subtype set (95% donor prevalence) and the six canonical visits, donor–visit scores are obtained by averaging ITL principal component scores across retained L3 subtypes.

For each ITL axis $i$ (restricted to the first 150 axes), between-donor variance is computed as the variance across donors of the donor mean score averaged across visits, $\mathrm{Var}_d(\mathbb{E}_v[\mathrm{PC}_i])$. Within-donor variance is computed as the mean across donors of the variance across visits within each donor, $\mathbb{E}_d[\mathrm{Var}_v(\mathrm{PC}_i)]$. The intraclass correlation statistic is defined as

$$\mathrm{ICC}_i = \frac{\mathrm{Var}_d(\mathbb{E}_v[\mathrm{PC}_i])}{\mathrm{Var}_d(\mathbb{E}_v[\mathrm{PC}_i]) + \mathbb{E}_d[\mathrm{Var}_v(\mathrm{PC}_i)]}.$$

High values indicate axes dominated by stable donor-specific differences, whereas low values indicate axes dominated by within-donor temporal variation.

## External cohort analyses

Analyses in the independent healthy PBMC atlas [24] test whether the empirical conditions motivating construction of the immune transcriptional landscape in the Sound Life cohort are present in an independent dataset. Raw UMI count matrices from the external cohort are processed using the same preprocessing and analysis pipeline described above, including normalization, cell filtering, aggregation to subtype-resolved mean expression profiles, gene universe restriction and baseline centering.

All analyses therefore operate on deviation profiles constructed using the identical workflow applied to the Sound Life cohort. External samples are not projected into the Sound Life ITL, and no model parameters or representations are transferred between cohorts.

To evaluate cross-lineage transcriptional alignment in the external cohort, gene-wise Spearman correlations are computed between deviation profiles from representative lineage pairs within individual donors, mirroring the deviation alignment analyses performed in the Sound Life cohort.

To assess donor-level transcriptional structure, principal component analysis is performed on centered subtype-resolved mean expression profiles using the same procedure applied in the Sound Life analyses. Donor-level scores for the first principal component are aggregated within compartments by averaging across L3 subtypes and compared across compartments at a fixed visit using Spearman correlation.

Finally, correlations across donors are evaluated as a function of principal component index to determine whether multi-dimensional donor-level transcriptional organization, as observed in the Sound Life cohort, is also present in the independent dataset.

## Robustness and Technical Sensitivity Analyses

To evaluate the robustness of the Immune Transcriptional Landscape (ITL) to technical and preprocessing choices, we performed a series of targeted sensitivity analyses designed to test whether the observed donor-level transcriptional structure depends on specific modeling assumptions or data characteristics.

**Season-level holdout.** To test generalization across annual vaccination cycles, we excluded Flu Year 2 Day 0 and Day 7 samples from principal component analysis (PCA) fitting. Subtype-specific centering was recomputed using training data only to avoid leakage of held-out visits into subtype baseline estimates. The ITL was then refit using the remaining data, and the held-out season was projected into the learned space. Donor-level temporal coherence between Day 0 and Day 7 was quantified using Spearman rank correlation across donors for each principal component (PC) and summarized as the median correlation across PCs 1–20. Temporal coherence in the held-out season was comparable to that obtained from the full model, indicating that the ITL geometry does not depend on inclusion of a specific annual vaccination cycle.

**Gene program exclusion.** As a gene-level sensitivity analysis, we excluded gene sets corresponding to interferon-stimulated genes, immediate early response genes and heat shock genes from the shared gene universe and refit the PCA. This exclusion was not part of the primary analysis, in which such programs were controlled at the cell level through filtering of transient non-homeostatic cells. Multi-dimensional donor-level temporal coherence across PCs 1–20 remained unchanged following exclusion of each gene set, indicating that the ITL is not reducible to any single activation-associated transcriptional program.

**Sequencing depth and complexity.** We evaluated the association between ITL coordinates and sequencing complexity metrics computed at the donor-by-visit level, including mean detected genes per cell, mean unique molecular identifiers (UMIs) and mean mitochondrial read fraction. The first principal component exhibited a strong correlation with mean detected genes and UMIs, consistent with a global transcriptional complexity axis. This association persisted within sequencing batch, pool, chip and calendar year strata, arguing against a simple between-batch artifact. Importantly, leading components beyond the first principal component (PCs 2–20) showed minimal correlation with sequencing complexity metrics while retaining strong donor-level temporal coherence.

**Gene universe threshold sensitivity.** To test sensitivity to gene inclusion criteria, we reconstructed the T cell ITL using a relaxed gene prevalence threshold (2% versus 10% in the primary analysis), expanding the shared gene universe from 3,530 to 8,171 genes. Multi-dimensional donor-level temporal coherence across PCs 1–20 remained unchanged, demonstrating that the ITL structure is not sensitive to the specific gene prevalence threshold used for gene universe definition.

**Cell count sensitivity.** To assess the influence of variable group sizes, we downsampled each donor–visit–subtype group to 100 cells prior to aggregation and PCA fitting. Downsampling was performed at the cell level before computation of subtype-resolved mean expression profiles. The resulting ITL retained comparable multi-PC donor-level temporal coherence, indicating that the shared transcriptional structure is not driven by unequal sampling depth across groups.

## Association of the ITL with routine hematologic indices

To assess whether dimensions of the immune transcriptional landscape (ITL) correspond to classical whole-blood physiological measurements, we analyzed their relationship with routine hematologic indices obtained from complete blood count (CBC) testing.

CBC-derived variables were obtained from the Sound Life laboratory metadata file (`sound_life_labs_me` Variables analyzed included hemoglobin concentration, hematocrit, platelet count, total white blood cell count (WBC), absolute neutrophil count, absolute lymphocyte count, absolute monocyte count, and corresponding percentage-based measures. Donor identifiers were matched using `subject.subjectGuid`. Visit labels were harmonized by extracting study day (e.g., Day 0, Day 7, Day 90) from visit names. Only donor–visit pairs with both ITL coordinates and hematologic measurements available were retained for analysis ($n = 776$).

Associations were evaluated using the reduced ITL manifold representation used in the proteomic correspondence analyses. Donor–visit ITL coordinates were constructed by averaging subtype-specific principal component scores across immune subtypes present in $\geq 95\%$ of donors. Principal component analysis was then applied to the first 150 ITL principal components to obtain a reduced manifold representation. Associations between each ITL manifold coordinate and individual hematologic variables were quantified using Spearman rank correlation.

## Proteomics manifold construction and ITL–proteomics correspondence analyses

Serum proteomic measurements from the Sound Life cohort were used to assess whether donor-level position within the immune transcriptional landscape corresponds to circulating protein profiles. A fixed panel of immune-related plasma proteins measured by Olink was defined a priori based on functional annotation and included cytokines, chemokines, soluble receptors,

TNF-family ligands and receptors, complement components, lectins, Siglecs and selected trafficking markers.

Proteomic features were normalized as described in the Sound Life study [23]. Donor–visit protein values were obtained by averaging replicate measurements where applicable. For each holdout analysis, proteins with high number of missing values were excluded using the training data only, and remaining missing values were imputed using training-set medians prior to dimensionality reduction. Principal component analysis was applied to the normalized proteomic matrix to construct a low-dimensional proteomic manifold representation, and the first ten proteomic components were retained for analysis.

To enable cross-modal correspondence analyses, donor–visit ITL coordinates were first constructed by aggregating subtype-level principal component scores across immune subtypes present in $\geq 95\%$ of donors. Because the ITL representation is high dimensional, principal component analysis was applied to the first 150 ITL dimensions to define a reduced ITL manifold. The first ten ITL manifold dimensions were used for correspondence analyses. For each holdout analysis, the reduced proteomic and ITL manifold representations were fit using training data only, and held-out samples were projected into the resulting spaces before regression evaluation.

We evaluated the correspondence between proteomic and transcriptional manifolds using ridge regression models mapping proteomic manifold coordinates to ITL manifold coordinates. Model generalization was assessed using two strict holdout strategies: donor holdout (all visits from a subset of donors withheld during training) and temporal holdout (all samples from specific visits withheld). For each analysis, dimensionality reduction and regression models were fit using training data only, and held-out samples were projected into the resulting spaces for evaluation. Model performance was quantified using Spearman correlation between predicted and observed ITL manifold coordinates in the held-out data.

Statistical significance was assessed using permutation testing by randomly permuting ITL targets in the training data, refitting the regression model and evaluating performance on the fixed test set. Empirical $p$-values were computed from the resulting null distribution.

## Statistical analysis

Spearman rank correlation was used throughout to quantify relationships between gene-wise deviation profiles, donor-level transcriptional coordinates and cross-modal features. This choice avoids assumptions of linearity or normality and focuses on preservation of relative ordering across donors.

Where statistical significance was evaluated, null distributions were generated using permutation testing. For deviation alignment analyses, donor identities were permuted within visits while forbidding self-matches. For proteomics correspondence analyses, ITL targets were permuted in the training data prior to model fitting and evaluated on fixed test sets. Empirical $p$-values were computed from the resulting null distributions.

Bootstrap resampling of test samples was used to visualize sampling variability of selected summary statistics. Bootstrap results were used for visualization only and were not used for formal hypothesis testing.

All analyses were performed in Python using NumPy, SciPy, pandas and scikit-learn. Random seeds were fixed where applicable to ensure reproducibility of dimensionality reduction and regression procedures.

**Data Availability:** All data used in this study is publicly available. Details of the Sound Life cohort data may be found in Gong et al. [23] and data may be accessed at `https://apps.allenimmunology.org/aifi/insights/dynamics-imm-health-age/cohorts/`. The Terekhova et al. [24] validation dataset is available via Synapse (syn49637038).

**Acknowledgments:** We gratefully acknowledge the Sound Life investigators for generating and making publicly available the longitudinal immune profiling dataset used in this study. The extensive experimental design, longitudinal sampling and multi-omic integration underlying this resource made the present analyses possible. We thank Miah Wander and Claire Gustafson for providing feedback that improved the manuscript.

**Funding Statement:** The work was funded by Microsoft Corporation.

**Competing Interests:** HJ Zahid has employment and equity ownership with Microsoft. The author declares no other competing interests.

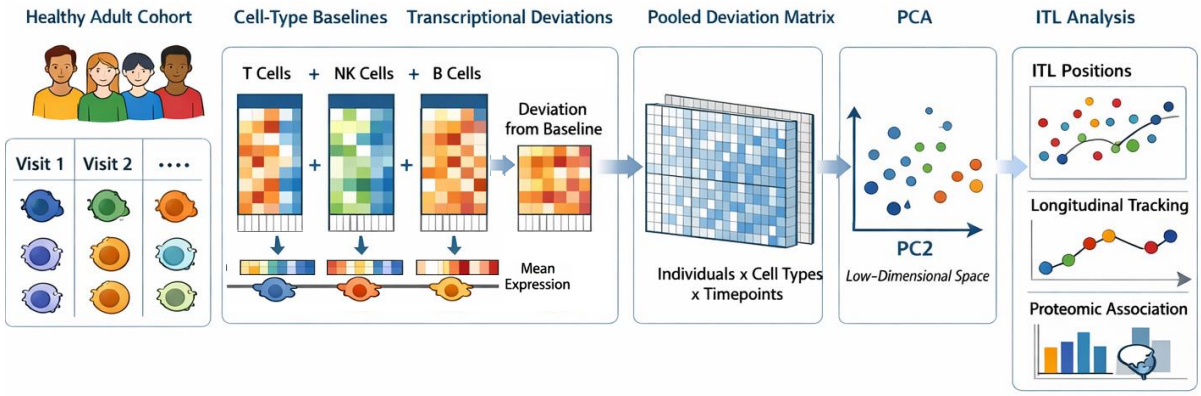


Figure 1: **Overview of cohort structure and analytical workflow.** Schematic illustrating the longitudinal cohort design and the construction of analytical objects used throughout the Results. Healthy adult donors are sampled repeatedly across multiple visits organized into annual cycles, with core visits at day 0, day 7, and day 90 and additional visits for a subset of donors. At each donor and visit, single-cell transcriptomic profiles are annotated at fine-grained subtype resolution across major immune lineages. Gene expression is summarized as mean expression vectors for each donor–visit–subtype group, requiring a minimum of 100 cells per group. Subtype-level mean profiles are mapped to broader immune compartments for interpretation. Deviations from subtype-specific baseline expression are computed and pooled across donors, visits, and subtypes to form a deviation matrix used for dimensionality reduction. The resulting axes define positions within the immune transcriptional landscape (ITL), which are used for longitudinal analysis, external cohort replication, and comparison with circulating serum proteomic measurements.

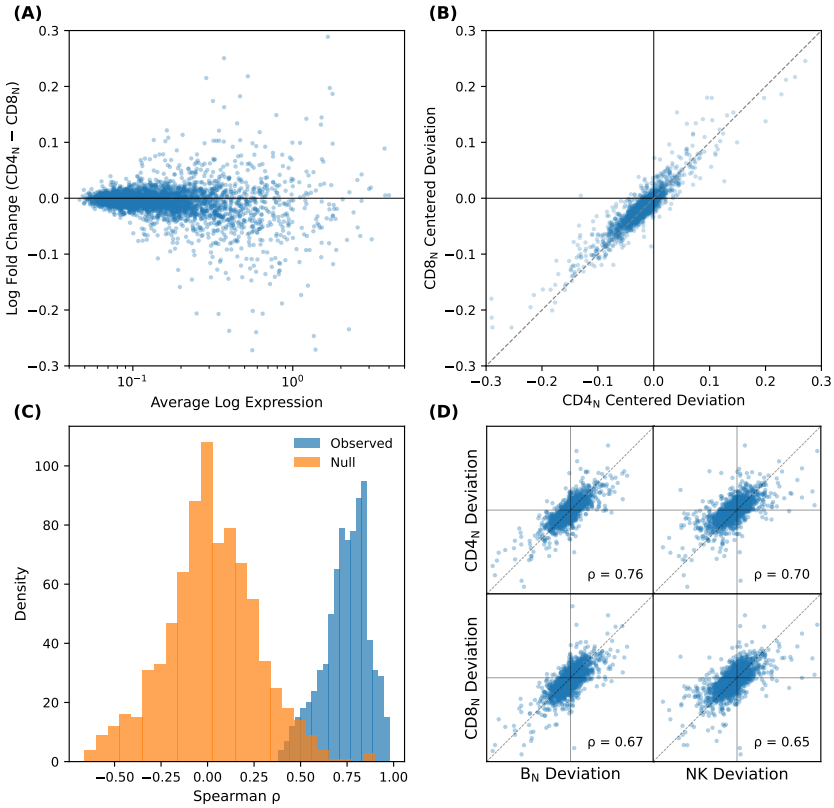


Figure 2: **Correlated transcriptional deviations across immune lineages within donors.** (A) Mean gene expression profiles differ systematically across immune lineages, illustrated here for a representative pair of immune subtypes from a single, well-sampled donor–visit, reflecting lineage-specific baseline transcriptional programs. (B) After centering expression relative to global lineage-specific baselines, gene-wise deviations from baseline are positively correlated across genes within a single donor–visit for a representative lineage pair, indicating coordinated up- or down-regulation across lineages. (C) Distribution of gene-wise correlation coefficients across all donor–visit pairs compared with a null generated by shuffling donor identities within each visit (preserving visit context and forbidding self-matches), demonstrating that the observed alignment depends on preserving within-donor correspondence. (D) Extending this analysis across additional immune lineages at a fixed visit reveals consistent positive correlations of gene-wise deviations across donors between T cells, B cells, NK cells and monocytes, with varying magnitudes across lineage pairs. Together, these analyses show that after accounting for lineage-specific baselines, transcriptional deviations are systematically aligned across immune lineages within donors, indicating structured and coordinated variation rather than independent fluctuations.

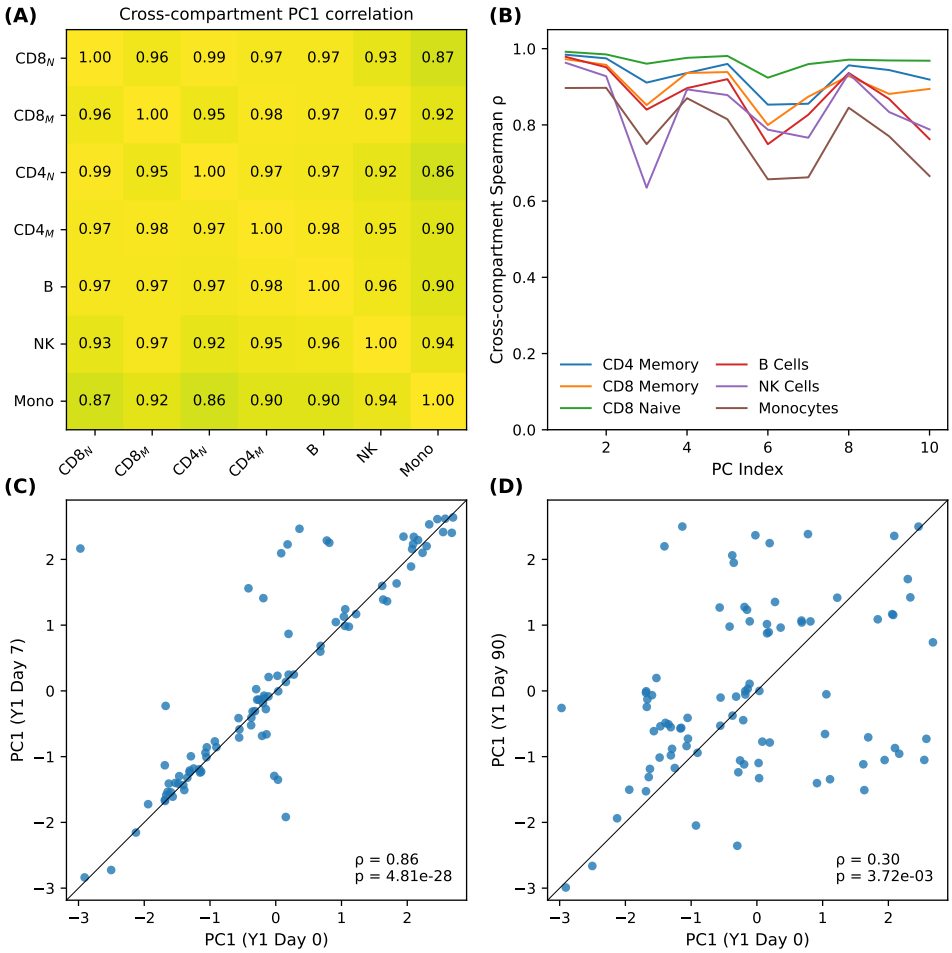


Figure 3: **Independent recovery of shared donor-level transcriptional structure and its temporal coherence across immune compartments.** (A) Principal component analysis (PCA) performed independently within immune compartments reveals that donor-level scores along the first principal component (PC1) are strongly correlated across compartments at a fixed visit, demonstrating that shared donor ordering emerges prior to pooling. (B) Cross-compartment Spearman correlations as a function of principal component index, obtained by projecting compartment-specific deviation profiles into a fixed reference PCA basis (defined in CD4 naive T cells), show that shared donor-level structure extends beyond the first principal component to additional principal components, supporting a multi-dimensional representation of donor-level variation. (C) Comparison of donor-level PC1 scores between Year 1 Day 0 and Day 7 demonstrates strong preservation of donor ordering over short timescales ($\rho = 0.86$), indicating structured temporal coherence. (D) Comparison of PC1 scores between Day 0 and Day 90 shows substantially weaker correspondence ($\rho = 0.30$), indicating that donor-level transcriptional organization evolves over longer intervals. Together, these analyses establish that shared donor-level transcriptional structure is independently recoverable within compartments, extends across multiple principal components and exhibits temporally structured coherence.

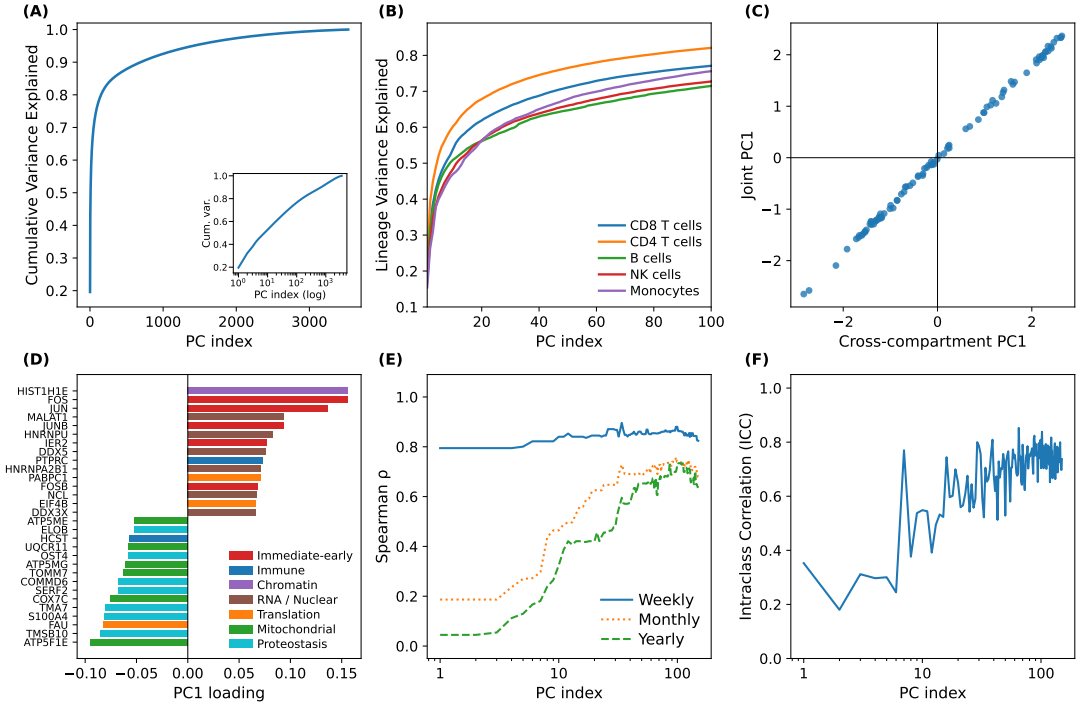


Figure 4: **Shared and temporally structured donor-level immune organization.** (A) Cumulative variance explained by principal component analysis (PCA) of pooled donor-level transcriptional deviation profiles, showing a compressed but high-dimensional structure. Inset shows the same relationship on a logarithmic principal component index. (B) Fraction of within-lineage donor-level variance captured by the ITL axes, shown separately for CD8 T cells, CD4 T cells, B cells, NK cells and monocytes. A substantial portion of lineage-specific variance lies within the same shared multi-dimensional subspace. (C) Comparison of donor ordering along the first principal component of the ITL with the previously defined cross-compartment PC1 demonstrates that the joint representation recovers the same shared donor-level structure previously identified. (D) Gene loadings for the first principal component reveal coordinated contributions from diverse functional categories, consistent with broad transcriptional organization rather than a single gene program. (E) Temporal coherence of donor ordering across ITL principal components, quantified as Spearman correlation between matched visits at weekly, monthly and yearly intervals. Principal components differ in temporal persistence, with early components exhibiting stronger short-term coherence and later components retaining stability over longer timescales. Curves are averaged across annual cycles and a median smoothing is applied for clarity. (F) Donor-level stability of ITL principal components, quantified as intraclass correlation (ICC). Early components show greater within-donor variability, whereas many higher-order components exhibit stable donor-specific structure. Together, panels (E) and (F) demonstrate that the ITL resolves donor-level immune organization into shared principal components with distinct temporal and inter-donor stability profiles.

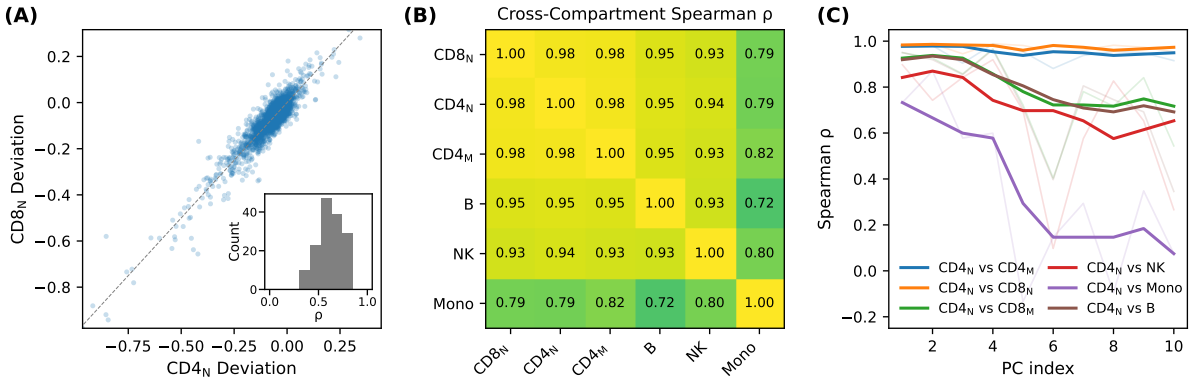


Figure 5: **Independent replication of shared donor-level transcriptional structure in an external cohort.** (A) Direct comparison of gene-wise transcriptional deviations between immune lineages in an independent healthy PBMC cohort shows positive correlation after centering relative to lineage-specific baselines, indicating coordinated deviations across lineages without dimensionality reduction. (B) Spearman correlation matrix of donor ordering along the first principal component of a joint PCA fit to the external cohort, with PC1 aggregated within immune compartments, demonstrates strong cross-compartment agreement of the primary donor-level component. (C) Cross-lineage correspondence of donor ordering as a function of principal component index, evaluated using representative anchor subtypes for each lineage, shows that shared donor-level structure extends beyond the first principal component to additional principal components. Together, these analyses demonstrate that the shared donor-level transcriptional structure motivating construction of the immune transcriptional landscape is reproducible in an independent cohort, supporting the robustness and generality of the ITL framework.

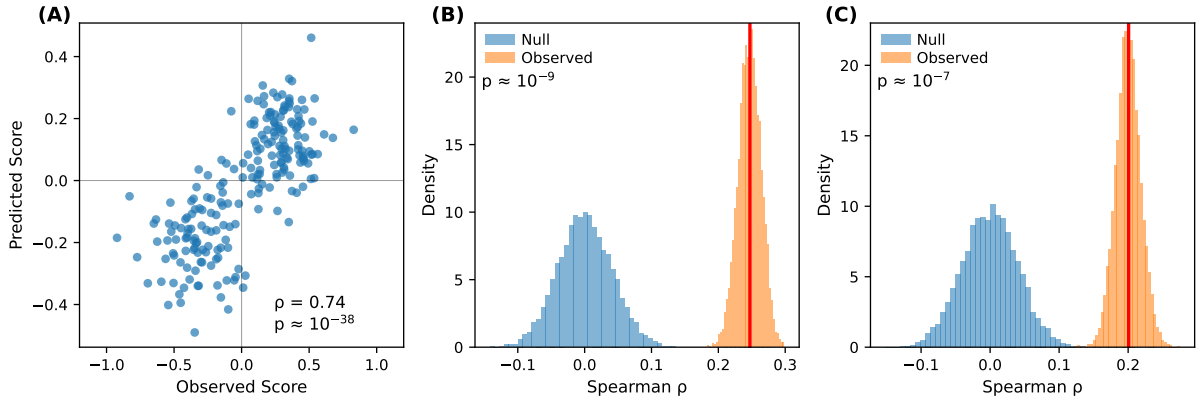


Figure 6: **Correspondence between immune transcriptional landscape position and circulating serum proteomics.** (A) Out-of-sample correspondence between observed donor-level score along one principal component (PC7) and values predicted from serum proteomic features. Proteomic features are mapped to ITL scores using ridge regression with donor holdout, and correspondence is quantified by Spearman correlation. (B) Multivariate assessment of proteomics–ITL correspondence under temporal holdout, simultaneously modeling the first ten principal components of the reduced ITL representation. Proteomic features are mapped to these axes using multi-output ridge regression, and correspondence is quantified as the mean Spearman correlation across axes on held-out visits. The observed distribution (bootstrap resampling of test samples) is compared to a permutation null that disrupts the training-set correspondence between proteomic features and ITL scores. (C) Corresponding robustness analysis under donor holdout, in which entire donors are withheld during training. The observed correspondence again exceeds the permutation null, demonstrating that the proteomics–ITL relationship generalizes to unseen donors. Together, these analyses show that donor-level position within the immune transcriptional landscape is reflected in independent circulating protein profiles when evaluated on held-out data.

# References


1. Sakaguchi S, Yamaguchi T, Nomura T, and Ono M. Regulatory T cells and immune tolerance. cell 2008; 133:775–87
2. Kotas ME and Medzhitov R. Homeostasis, inflammation, and disease susceptibility. Cell 2015; 160:816–27
3. Belkaid Y and Harrison OJ. Homeostatic immunity and the microbiota. Immunity 2017; 46:562–76
4. Rankin LC and Artis D. Beyond host defense: emerging functions of the immune system in regulating complex tissue physiology. Cell 2018; 173:554–67
5. Dudley ME, Wunderlich JR, Robbins PF, Yang JC, Hwu P, Schwartzentruber DJ, et al. Cancer regression and autoimmunity in patients after clonal repopulation with antitumor lymphocytes. Science 2002; 298:850–4
6. Sylwester AW, Mitchell BL, Edgar JB, Taormina C, Pelte C, Ruchti F, et al. Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. The Journal of experimental medicine 2005; 202:673–85
7. Liston A, Humblet-Baron S, Duffy D, and Goris A. Human immune diversity: from evolution to modernity. Nature immunology 2021; 22:1479–89
8. Brodin P, Jojic V, Gao T, Bhattacharya S, Angel CJL, Furman D, et al. Variation in the human immune system is largely driven by non-heritable influences. Cell 2015; 160:37–47
9. Carr EJ, Dooley J, Garcia-Perez JE, Lagou V, Lee JC, Wouters C, et al. The cellular composition of the human immune system is shaped by age and cohabitation. Nature immunology 2016; 17:461–8
10. Alpert A, Pickman Y, Leipold M, Rosenberg-Hasson Y, Ji X, Gaujoux R, et al. A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. Nature medicine 2019; 25:487–95
11. Lakshmikanth T, Muhammad SA, Olin A, Chen Y, Mikes J, Fagerberg L, et al. Human immune system variation during 1 year. Cell reports 2020; 32
12. Querec TD, Akondy RS, Lee EK, Cao W, Nakaya HI, Teuwen D, et al. Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nature immunology 2009; 10:116–25
13. Tsang JS, Schwartzberg PL, Kotliarov Y, Biancotto A, Xie Z, Germain RN, et al. Global analyses of human immune variation reveal baseline predictors of postvaccination responses. Cell 2014; 157:499–513

14. Smithmyer ME, Hu A, Dufort MJ, Hocking AM, Wiedeman AE, Fasano KJ, et al. Longitudinally stable T cell function and innate immune activation distinguish healthy adult immunotypes. Science Translational Medicine 2025; 17:eadt4605

15. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 2014; 343:776–9

16. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 2017; 356:eaah4573

17. Papalexi E and Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nature Reviews Immunology 2018; 18:35–45

18. Ginhoux F, Yalin A, Dutertre CA, and Amit I. Single-cell immunology: Past, present, and future. Immunity 2022; 55:393–404

19. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. Genome biology 2020; 21:31

20. Perez RK, Gordon MG, Subramaniam M, Kim MC, Hartoularos GC, Targ S, et al. Single-cell RNA-seq reveals cell type–specific molecular and genetic associations to lupus. Science 2022; 376:eabf1970

21. Wynn EA, Mould KJ, Vestal BE, and Moore CM. Simulating paired and longitudinal single-cell RNA sequencing data with rescueSim. Bioinformatics 2025; 41:btaf442

22. Wimmers F, Donato M, Kuo A, Ashuach T, Gupta S, Li C, et al. The single-cell epigenomic and transcriptional landscape of immunity to influenza vaccination. Cell 2021; 184:3915–35

23. Gong Q, Sharma M, Glass MC, Kuan EL, Chander A, Singh M, et al. Multi-omic profiling reveals age-related immune dynamics in healthy adults. Nature 2025 :1–11

24. Terekhova M, Swain A, Bohacova P, Aladyeva E, Arthur L, Laha A, et al. Single-cell atlas of healthy human blood unveils age-related loss of NKG2C+ GZMB- CD8+ memory T cells and accumulation of type 2 memory T cells. Immunity 2023; 56:2836–54

25. Luecken MD and Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular systems biology 2019; 15:MSB188746

26. Brodin P and Davis MM. Human immune system variation. Nature reviews immunology 2017; 17:21–9

27. Sparks R, Rachmaninoff N, Lau WW, Hirsch DC, Bansal N, Martins AJ, et al. A unified metric of human immune health. Nature medicine 2024; 30:2461–72

28. Medzhitov R and Horng T. Transcriptional control of the inflammatory response. Nature Reviews Immunology 2009; 9:692–703

29. Smale ST. Selective transcription in response to an inflammatory stimulus. Cell 2010; 140:833–44

30. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, et al. Single-cell eQTL mapping identifies cell type–specific genetic control of autoimmune disease. Science 2022; 376:eabf3041

31. Van Der Wijst MG, Brugge H, De Vries DH, Deelen P, Swertz MA, Study LC, et al. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. Nature genetics 2018; 50:493–7

32. Dopico XC, Evangelou M, Ferreira RC, Guo H, Pekalski ML, Smyth DJ, et al. Widespread seasonal gene expression reveals annual differences in human immunity and physiology. Nature communications 2015; 6:7000

33. Temba GS, Kullaya V, Pecht T, Mmbaga BT, Aschenbrenner AC, Ulas T, et al. Urban living in healthy Tanzanians is associated with an inflammatory status driven by dietary and metabolic changes. Nature immunology 2021; 22:287–300

34. Möller-Levet CS, Archer SN, Bucca G, Laing EE, Slak A, Kabiljo R, et al. Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. Proceedings of the National Academy of Sciences 2013; 110:E1132–E1141

35. Piasecka B, Duffy D, Urrutia A, Quach H, Patin E, Posseme C, et al. Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. Proceedings of the National Academy of Sciences 2018; 115:E488–E497

36. Kaczorowski KJ, Shekhar K, Nkulikiyimfura D, Dekker CL, Maecker H, Davis MM, et al. Continuous immunotypes describe human immune variation and predict diverse responses. Proceedings of the National Academy of Sciences 2017; 114:E6097–E6106

37. Kotliarov Y, Sparks R, Martins AJ, Mulè MP, Lu Y, Goswami M, et al. Broad immune activation underlies shared set point signatures for vaccine responsiveness in healthy individuals and disease activity in patients with lupus. Nature medicine 2020; 26:618–29