Title:
A Gene-expression Module Identifies Circulating Immune Cells with Enhanced Recruitment to Sites of Inflammation

Debajyoti Sinha*[1], Thomas Laurent*[1], Alexis Broquet[1], Cynthia Fourgeux[1], Thibault Letellier[2], Gaelle Tilly[1], Sarah Bruneau[1], Simon Ville[1,2], Laurence Bouchet-Delbos[1], Julien Brancherau[1,2], Clarisse Kerleau[2], Sophie Brouard[1], Gilles Blancho[1,2], Magali Giral[1,2], Regis Josien[1,2], Richard Danger[1] , Antoine Roquilly[1,3], Nicolas Degauque[1], Jeremie Poschmann[1,4]

[1]Nantes Université, CHU Nantes, INSERM, Center for Research in Transplantation and Translational Immunology, UMR 1064, ITUN; F-44000, Nantes, France.
[2] CHU Nantes, Nantes Université, Institut de Transplantation Urologie Néphrologie (ITUN), Service de Néphrologie et Immunologie Clinique, 44000 Nantes, France.
[3] Nantes Université, CHU Nantes, INSERM, Anesthesie Réanimation, CIC 1413, F-44000 Nantes, France
[4] Corresponding author (orcid : 0000-0002-9613-5297), 30 Boulevard Jean-Monnet, CR2TI, 44093 Nantes, France, jeremie.poschmann@univ-nantes.fr
[*] contributed equally

Abstract:
Circulating immune cells are critical mediators of inflammation upon recruitment to tissues, yet how their gene expression state influences this recruitment is not well understood. Here, we report longitudinal single-cell transcriptome profiling of peripheral blood mononuclear cells in patients undergoing kidney transplantation rejection. We identify a novel gene expression module, termed ALARM (early activation transcription factor module), associated with transcriptional regulation, homing, and immune activation across multiple immune cell types. Circulating cells expressing this module are significantly reduced in patients experiencing graft rejection, a finding confirmed in a pig model of acute kidney transplantation rejection. Correspondingly, module expression is markedly increased in kidney grafts undergoing rejection, indicating preferential recruitment of ALARM-expressing cells to the inflamed tissue.
Within this module, we identify the receptor CXCR4 and its ligand CXCL12, expressed in the graft, as a likely mechanism for recruitment. In vitro transwell assays combined with scRNA-seq reveal that this CXCR4-CXCL12 interaction is critical for T cell migration and upregulation of CD69, an early activation marker, and is accompanied by a metabolic switch towards glycolysis. Further exploration of publicly available transcriptomic data demonstrates that this module is generally expressed in healthy individuals and is strongly associated with responses to infection, including SARS-CoV-2 infection. This finding is further supported by experiments in a pneumonia mouse model, which confirm the recruitment of CXCR4-expressing T cells during lung infection. Moreover, we find that module expression is predictive of immune-mediated diseases.
In summary, we have identified a key gene expression module in circulating immune cells that orchestrates their preferential recruitment to inflamed tissues, metabolic reprogramming, promoting tissue residency and effector functions. These insights advance our understanding of immune cell recruitment and activation mechanisms in transplant rejection and infectious diseases, with potential implications for therapeutic interventions.

Introduction:

Circulating immune cells are critical to be recruited to the site of inflammation, infection, and cancer. This compartment and particularly the peripheral blood mononuclear cells (PBMCs) thus offer an attractive resource for precision medicine as in a single experiment, diverse cell types, including the $CD4^+$ T cells, $CD8^+$ T cells, B cells, NK cells and monocytes are profiled. Especially since the introduction of single cell RNA-seq (scRNA-seq), profiling PBMCs has been highly successful in identifying gene expression signatures and cell-types associated with immune-related diseases[1–4]. For example, a monocyte signature associated with sepsis was discovered in circulating cells [1]. Recently, a study on systemic lupus erythematosus revealed gene expression changes with disease state and genetic variation [2]. In addition, distinct Covid-19 studies revealed signatures associated to infection and disease severity in blood [5,6].

Transcriptomics profiling enables the characterization of genes groups (i.e., modules) which perform critical cellular functions such as maintaining a cell identity, homeostasis & metabolism and respond to external signals. Notably, in circulating immune cells we have previously shown that monocytes express a gene module associated with Herpes Simplex virus reactivation after traumatic brain injury [7]. While gene expression programs of circulating immune cells are likely to be distinct from the same cells which migrated into the tissue, identifying modules in circulating cells may reveal early immune activation programs or modules associated with homing and migration. For example, in a previous study we identified large gene regulatory and gene expression alterations in circulating monocytes during active Tuberculosis which improved pathogen clearance for these cells [8].

Leveraging on scRNA-seq and the availability of recent module identification approaches tailored for single cell transcriptomics [9] we aimed to identify gene expression programs associated with kidney transplantation rejection. Currently, rejection status is monitored in clinical practice by analyzing metabolites in blood and urine, such as creatinine, to assess renal function. The diagnosis is then confirmed through pathologic examination of kidney biopsies[10]. However, metabolite monitoring is not specific to rejection and can be approximative. Pathologic analysis of kidney biopsies remains the primary diagnostic tool, offering reliable results. However, despite minimal risk to patients, it is an invasive procedure that cannot be performed regularly. Thus identifying gene expression signatures in circulating cells may improve precision medicine diagnostics of kidney rejection [11]. In addition, acute and chronic rejection are characterized by the infiltration of immune cells into the graft, which then mediate an inflammatory response in the tissue ultimately leading to the rejection of the graft. Blood thus constitutes an easily accessible compartment to identify gene expression modules associated with homing and early activation. Two archetypes of rejection are prominent according to the Banff classification[10], the antibody mediated rejection (ABMR) and T-cell mediated rejection (TCMR), which can also arise in a mixed form. In both cases, immune cell infiltration into the graft occurs via the bloodstream through either donor specific antibodies binding to the graft endothelium in ABMR, or cytokine and homing signals in TCMR[12].

In this study, our aim was to identify putative modules in circulating cells which may be associated to kidney transplantation rejection. We profiled a longitudinal patient cohort consisting of 3 stable patients, 3 TCMR and 3 ABMR patients at 0-month, 3 month and 12 months after transplantation or when rejection occurred. The PBMCs were collected at the same time of graft biopsy, allowing us to characterize relationship of graft rejection status with gene expression modules of peripheral immune cell-types. We identified a module associated with transcriptional regulation and early activation in the blood and used a pig-transplantation model to validate its association with rejection status. Further characterization of this module was carried out in transcriptomics data from over 1500 kidney biopsies revealing a cytokine-receptor interaction between the graft and circulating cells, respectively. Finally, we demonstrate that this module is not specific to graft rejection but is implicated and predictive of a variety of immuno-pathologies.

104 **Results:**

105 **Single cell transcriptome analysis of circulating immune cells in a longitudinal kidney**
106 **transplantation cohort**

107 In order to identify gene expression changes in circulating immune cells during kidney transplantation
108 rejection, we used scRNA-seq on PBMC isolated from 3 patients with stable allograft function (STA) for
109 which no sign of rejection was observed clinically after more than a year of follow-up, 3 antibody
110 mediated rejection (ABMR) patients and 3 T-cell mediated rejection (TCMR) patients (Figure 1A, Table
111 1). The patients were selected based on their treatment, age, sex, and collection time. (See Table 1).
112

Table 1: Clinical summary of the cohort composition.

	Name	Patients characteristics					
		Age	Sex	Treatment *	Rejection time (in days)	HLA ** mismatches	Induction Therapy
Batch 1	STA1	35	*Male*	*Tacrolimus, MMF, Corticoids*	–	5	*Depleting*
	TCMR1	70	*Male*	*Tacrolimus, MPA, Corticoids*	244	5	*Depleting*
	ABMR1	51	*Female*	*Tacrolimus, MMF, Corticoids*	374	2	*Depleting*
	TOL1	75	*Male*	–	–	4	*Depleting*
Batch 2	STA2	69	*Male*	*Tacrolimus, MMF, Corticoids*	–	5	*Non – depleting*
	TCMR2	35	*Male*	*Tacrolimus, MPA, Corticoids*	369	3	*Depleting*
	ABMR2	56	*Male*	*Tacrolimus, MPA, Corticoids*	427	4	*Depleting*
	TOL2	45	*Male*	–	–	4	*NA*
Batch 3	STA3	25	*Female*	*Tacrolimus, MPA, Corticoids*	–	5	*Depleting*
	TCMR3	24	*Male*	*Tacrolimus, MMF, Corticoids*	213	3	*Non – depleting*
	ABMR3	61	*Female*	*Tacrolimus, MPA, Corticoids*	365	2	*Depleting*
	TOL3	39	*Female*	–	–	0	*NA*
		Mean Age 48.75	Sex Ratio 0.66				

* Antiproliferative treatments: MMF=Mycophenolate mofetil, MPA=Mycophenolic Acid
** HLA mismatches: number of mismatches on the A, B and DR loci (0-6)

113
114
115
116

117 For each patient, three time points were profiled; T0, at the kidney transplantation, T1 at 90-150 days
118 after the transplantation and T2, which was sampled at the time of rejection for ABMR and TCMR or
119 at 1 year for STA patients after transplantation (Figure 1A). In addition, PBMCs from 3 kidney
120 transplantation patients who maintained allograft function in the absence of immunosuppression (i.e.,
121 operational tolerant (TOL)) were included in the cohort (Table 1).

122 Rejection status was defined by clinical pathology assessment of kidney biopsies performed at T1 and
123 T2 for all patients. ABMR and TCMR were defined by pathology biopsy assessment at time T2 whereas
124 patients were determined as STA when they which no sign of rejection in biopsies at T1 and T2. To
125 minimize scRNA-seq-related experimental variation, we performed CITE-seq (Cell Hashing)[13] using
126 hashtag oligo-conjugated antibodies (HTO) to label each patient and time point separately and then
127 pooled 10 samples (1 ABMR, 1 TCMR and 1 STA patient across T0, T1 & T2 and 1 TOL patient) into a
128 single experiment. We thus generated the complete transcriptomic data in three balanced batches
129 (Figure 1B). After removing doublets, data cleaning, normalization, and batch correction, we obtained

50,507 cells across the three batches (see Methods). Cells were automatically annotated using Azimuth[14,15] and manually validated for cell-type specific markers (Figure 1C and Supplemental Figure S1A). Cell-type proportions varied minimally between conditions when compared to PBMCs from two separate cohorts of healthy volunteers (HV) from publicly available scRNA-seq data [1,5] (Supplemental Figure S1B). For example, NK cells were significantly decreased in the stable and rejection conditions as compared to HV. However, we note that there were also significant differences in the HV (e.g., CD14 and CD16 monocytes) from the two distinct control cohorts indicating that this may be due to individual variation. We then inspected the distribution of cells across patients and by time points (Supplemental Figure S1C) and across disease states (Figure 1D). Neither of these variables formed unique clusters suggesting that the clusters were driven by cell-type specific expression rather than by condition or cellular states as was also observed in other PBMC studies in patient cohorts [1,2,5]. We then explored whether the clusters were affected by merging the three experimental batches (Figure 1E). As no batch effects were apparent through visual inspection, we used the KBET metric to quantitatively assess potential batch effects. KBET evaluates whether cells from different batches are clustering together in shared neighborhoods (i.e., clusters) (Supplemental Figure S1D)[16]. The acceptance rate of the KBET for the complete data set was 0.969, indicating that batch integration was successful. In summary, the pooling strategy and subsequent bioinformatics analysis resulted in a robust dataset of 50,507 cells to be analyzed for time and disease state specific gene expression.

Gene co-expression analysis identifies a module associated to rejection state

To identify modules, ie. co-expressed groups of genes, we used gene co-expression analysis across all three batches independently (Figure 2A). This approach was chosen to avoid potential signal alterations induced during the batch correction step. We applied consensus non-negative matrix factorization (cNMF)[9] to identify gene expression programs which may either be associated to cell-type specific gene expression programs or to cellular activity (see methods). Nine overlapping modules were identified which were evenly distributed across the three batches (Figure 2B, Supplemental Figure 2A). These modules mostly revealed cell-type specific expressions, notably three of these modules were associated to monocytes (Mod_1-Mod_3), and two modules were mostly expressed in a specific cell type such as B cells and pDC cells (Mod_4 and Mod_5 respectively) (Figure 2C). Three modules were enriched for the CD4 & CD8 T lymphocytes (Mod_6) and/or NK cells (Mod_7, Mod_8, Figure 2C). Interestingly, the Mod_9 was expressed in all cell types, but with notable higher expression in B cells, T cells, pDCs and NK cells as revealed by its module score (Figure 2C, Figure 2D).

Next, we aimed to independently validate the co-expression of the 9 modules (Supplemental file 1). To achieve this, we quantified the gene co-expression using the pair-wise Pearson correlation coefficient R. As the module was identified using cNMF, Pearson correlation thus acts as an independent evaluation of gene co-expression. We note that this approach is inherent to weighted gene co-expression analysis (WGCNA), a prominent method to identify modules in bulk and single cell transcriptomics[17]. We compared the average Pearson correlation per cell for each module to the same number of randomly picked genes in the same cells. The eight cell type specific modules showed robust and significant correlation between module genes, and this correlation was strongest in the cell-types to which they were associated (Supplemental Figure S2B). The cell-activity module Mod_9 was also significantly correlated in all cell types (Supplemental Figure S2C). Of note, randomly chosen genes picked 1000 times revealed a Pearson correlation of 0 in all cell-types indicating that unrelated genes typically do not correlate with each other (Supplemental Figure S2B, S2C). Therefore, the modules identified above are robustly co-expressed, as inferred by the two most prominent module detection methods.

To test whether this cell-activity module or any of the cell-specific modules were associated to disease state, (i.e., rejection or stable) and if it would vary throughout time, we estimated the module score for each cell type and calculated the trend of the module score longitudinally in each cell-type (Figure 2E). A positive or negative slope thus indicates whether a module changes over time across the stable, humoral & cellular rejection. Indeed, Mod1 and Mod2 showed a significant positive trend in both

181 rejection conditions (ABMR and TCMR) but not in the STA condition. Interestingly, the activity module
182 (Mod_9) showed a significant negative trend (regression β value) in the rejection conditions but not in
183 the stable condition in multiple cell types (B cells, CD4 and CD8 T cells and NK cells). Further inspection
184 of this negative trend was carried out by displaying the module score of each patient separately in the
185 form of a combined violin plot (Figure 2F). These module scores show that there was indeed a
186 reduction in ABMR patients, while in TCMR patients the module score followed a U-shape, reduced
187 more during T1 and increased again at T2. Interestingly, the 3 stable patient's module scores remained
188 consistent across time in NK, CD4 and CD8 and B cells. These results signify that Mod_9 expression is
189 associated to rejection state in a time dependent manner.
190
191 **Discovery of the early activation, transcription factor module (ALARM)**
192 To explore the function of the 61 genes found in Mod_9 we first investigated whether it was enriched
193 for ribosomal, proliferation or cell cycle genes using SEURAT-based list [18]. We did not observe any
194 genes involved in these cellular processes (data not shown). We thus explored the genes within this
195 module by performing gene ontology analysis. Enrichment of the module genes was quantified for
196 molecular function (MF) and biological processes (BP) compared to the combined set of 4000 most
197 variable genes from the three batches (Figure 2G). MF could be associated to 24 genes and was
198 significantly enriched (FDR <0.05) for helicase activity, ubiquitin-like protein binding, RNA catalytic
199 activity, transcription activator activity and ribonucleoprotein binding. The BP (20 genes) was
200 associated to NF-κB signaling, response to peptide and oxidative stress, regulation of RNA metabolic
201 processes and viral gene expression. Of the 61 genes in the module, 56 were annotated in the GSEA
202 database and 30 of these genes were linked to gene ontology enrichment. This indicates that Mod_9
203 is likely to be involved in multiple molecular functions associated with transcription, mRNA process
204 and ubiquitination. The BP suggested involvement in response to immune conditions (i.e., viral gene
205 expression, NF-κB signaling, oxidative stress). This result was further supported by the 5-fold
206 enrichment of Transcription factor genes in this module (OR: 4.9; Fisher-Test P-value $2.5e^{-5}$), such as
207 the AP-1 complex (*JUN, JUND, FOS*), *REL* (NF-κB subunit), *MAFF* and *NR4A2* (see methods,
208 Supplemental file 1). Further manual examination exposed the early activation marker CD69, a cell
209 surface type II lectin. This receptor was described to be rapidly expressed at the membrane in T cells
210 upon TCR activation[19]. In addition, CD69 has been described as a marker of tissue retention of T cells
211 [20–22]. Interestingly, *CD69* gene promoter is controlled by AP-1 TF complex and NF-κB, both of which are
212 also members of this module[23,24]. We also found the cell surface marker CXCR4 in this module, a
213 chemokine receptor known to play a role in recruiting CXCR4 positive cells to the kidney after an
214 ischemic injury via the chemokine CXCL12[25]. The role of the CXCR4/CXCL12 axis in kidney rejection is
215 still unclear[26] but an elevated expression of CXCL12 has been described in chronic kidney rejection[27]
216 suggesting it may act as a chemotactic signal to recruit immune cells in the inflamed tissues. In
217 summary, the Mod_9 module comprises genes implicated in the response to stress, mRNA processing,
218 early activation, and tissue-homing. For clarity, we named this module ALARM, which stands for for
219 eArLy activation trAnscription factoR Module.
220
221 **Circulating ALARM cells are recruited to the graft during acute graft rejection in a pig kidney**
222 **transplantation model**
223 We found that circulating cells expressing ALARM are altered in ABMR and TCMR in a timely fashion
224 (Figure 2). To independently validate this observation, i.e., whether ALARM high expressing cells are
225 depleted in the circulation during kidney transplantation rejection, we took advantage of an acute
226 rejection pig kidney transplantation model (Figure 3A). We decided to use this pig kidney
227 transplantation model as they share anatomical, physiological and genetic similarities to human and
228 offer the advantage to have well defined swine leukocyte antigen (SLA) genotypes [28]. Two SLA-
229 mismatched pigs were subjected to kidney transplantation, keeping one of their own kidneys (see
230 methods). This model typically results in an acute TCMR within a few days after transplantation, as no
231 immunosuppressive treatment is given (Figure 3A). Kidney biopsies and PBMCs were collected daily
232 before and after the transplantation. Microscopical analysis of the biopsies at D2, D4 and D6 indicated

233 a time-dependent infiltration of immune cells, culminating at day 6 (Figure 3B). This infiltration was
234 quantified in three areas (excluding glomeruli) of each biopsy time point (see methods, Supplemental
235 Figure S3A). Cell counts drastically increased from day 2 to D6, indicating a continuous immune cell
236 accumulation over time (Figure 3C). We noted that the second pig did not display any signs of rejection,
237 possibly due to early arterial ischemia of the transplanted kidney and it was thus discarded from the
238 subsequent analysis. PBMCs collected at D0, D2, D4 and D6, were pooled in a single scRNA-seq
239 experiment (see methods), resulting in a total of 4,411 annotated cells across cell-types and time-
240 points (Figure 3D and Supplemental Figure S3B). We found that cellular proportions within the PBMC
241 compartment drastically changed from D0 to D2, characterized by a drastic increase in monocyte
242 proportions concomitant with a reduction of B, CD4+ and CD8+ T cells (Figure 3E). This suggests that
243 the lymphocytes are rapidly recruited to the kidney graft and accumulate there, as demonstrated by
244 the cellular invasion observed in the biopsies at the same time (Figure 3C). To test whether the decline
245 of blood lymphocytes is accompanied by a reduction of ALARM high expressing cells, we quantified
246 ALARM expression across the time-points in CD4+, CD8+, B cells and NK cells (Figure 3F). Interestingly,
247 as soon as D2 the levels of ALARM expression drastically decreased and remained low in the blood
248 until sacrifice of the animal (D6). Taken together, this controlled time-course experiment reveals a
249 drastic immune cell infiltration in the graft associated with the depletion of ALARM high expressing
250 cells in the circulation. This result mirrors the reduction of circulating ALARM cells observed during the
251 kidney transplantation rejection in the human cohort (Figure 2). It is thus probable that ALARM high
252 expressing cells are readily circulating in healthy condition. Upon the kidney graft transplantation
253 ALARM high expressing cells are then preferentially recruited to the organ to mediate the immune
254 response.
255
256 **ALARM gene expression increases in kidney grafts undergoing rejection**
257 To further support the hypothesis that ALARM high-expressing cells are recruited to the graft during
258 rejection from the bloodstream, we evaluated the expression of ALARM genes in graft biopsies from
259 kidney transplantation patients. For this, we used a previously published Canadian transcriptomics
260 analysis of 569 transplant biopsies collected from 13 clinical sites and with a patient classification of
261 STA, TCMR, ABMR and mixed rejection (TCMR and ABMR)[29]. In parallel, we exploited a second similar
262 Belgian transcriptomics study performed on kidney biopsies in 224 patients who were either stable
263 (168 patients) or undergoing ABMR[30]. After precleaning and QC controls of the available microarray
264 data (Methods), we quantified the ALARM gene expression in each cohort separately (Figure 4A). The
265 ALARM score was consistently and significantly elevated in all three rejection cases compared to stable
266 biopsies. Similarly, in the second study, ABMR samples showed a significant increase in ALARM
267 expression, regardless of the presence of donor-specific antibodies. Analysis of the ALARM genes
268 ranked by expression, further revealed that this score is driven by the upregulation of a large fraction
269 of the ALARM genes, including *CD69*, *CXCR4*, *JUN* and *IRF1* in both cohorts (Figure 4B). Quantification
270 of the ALARM module across rejection and stable revealed a significant upregulation of ALARM
271 expression in both data sets (Figure 4C). These results indicate that ALARM genes are significantly
272 increased during graft rejection across over 793 biopsies in two distinct studies. Given that graft
273 rejection is defined by immune cell infiltration and that circulating ALARM high cells are depleted at
274 the same time, it is possible that ALARM expressing cells are preferentially recruited to the graft to
275 mediate the rejection. To investigate how circulating cells could be preferentially recruited to the graft,
276 we explored the cytokine expression in the graft and receptor expression in the circulating cell subsets.
277 We first identified all possible ligand-receptor pairs and then tested whether these pairs were
278 differentially expressed between rejection and stable status. We found 10 differentially expressed
279 cytokines in the graft pairing with 7 receptors upregulated in circulating immune cells in both cohorts
280 independently (Figure 4D). The most prominent receptor was CXCR4, expressed in CD4, NK, CD8 and
281 B cells. As mentioned above, CXCR4 is a member of ALARM genes, indicating a likely mechanism of
282 signaling from the graft via CXCL12 and recruitment of ALARM cells expressing *CXCR4*. This cytokine
283 receptor pair has been previously described as a homing mechanism in various distinct tissues, such as
284 bone marrow[31,32] and in cancer[33]. To test the relationship of this interaction, we quantified *CXCL12*

285 expression in the graft biopsies stratified by ALARM expression (low (<25%), mid (25-75%), high
286 (>75%)) and found that there was a significant increase in CXCL12 expression in the ALARM high group
287 in both cohorts (Figure 4E). In summary, these results show that ALARM is increased during rejection
288 in the graft, which strengthens the notion that circulating cells with high ALARM expression are
289 preferentially recruited to the kidney graft via the homing signaling axis of CXCL12 and CXCR4.
290
291 **Analysis of ALARM Module Expression in an In Vitro Transwell Assay**
292 To experimentally verify the observed CXCL12-mediated recruitment, we employed an *in vitro*
293 transwell assay using a cytokine gradient of CXCL12. This assay consists of a membrane covered with
294 primary microvascular endothelial cells (HDMEC), allowing for the comparison of unstimulated cells,
295 cells in direct contact with CXCL12, and those migrating based on the CXCL12 gradient (Figure 5A).
296 Flow cytometry quantifications of T cells indicate that CXCL12 significantly induces the recruitment and
297 migration of T cells, observed 4 hours post-deposition (Figure 5A; p=0.0286). The composition of naïve,
298 central memory (CM), effector memory expressing CD45RA (TEMRA), and effector memory (EM) T cells
299 remained similar after migration and in response to CXCL12 alone (Figure 5B, Supplemental Figure
300 S4A). This suggests that all T cell subsets are equally attracted and migrate in response to a CXCL12
301 gradient. We further analyzed the expression of several T cell surface markers, focusing on CD69 as a
302 component of the ALARM module, CD25 (an activation marker), CD49a (a tissue residency marker),
303 and CD95 (an apoptosis marker) (Figure 5C). Interestingly, CD69 increased slightly with CXCL12
304 addition but was even more highly upregulated in migrated cells, while CD25 expression did not
305 change, decoupling the role of CD69 in early activation from its role in tissue residency[34]. CD49a and
306 CD95 levels remained unchanged, indicating that CXCL12 in combination with migration specifically
307 induced the extracellular display of CD69. It is noteworthy that CD69 expression depends on both
308 CXCL12 and the direct contact with HDMEC cells (Supplemental Figure S4B).
309 To assess whether CXCL12-induced migration altered the T cells' response to immune stimuli, CXCL12
310 exposed or transmigrated cells were purified and restimulated polyclonally for an additional 24 hours
311 (Figure 5D). Migrated cells showed a significantly increased expression of IFN-γ in both CD4+ and CD8+
312 T cells in contrast to TNF-α, granzyme B (GZMB) or granulysin (GNLY). The increased expression of IFN-
313 γ suggests that migration via the CXCL12 gradient may enhance the effector functions of T cells but
314 that other signals are needed to induce cytotoxic mechanisms in this model. The observed increase in
315 CD69 membrane display indicates a functional role for the ALARM module, which enables T cells to
316 acquire "new functions" in the tissue i.e. to establish residency via CD69 expression and increased IFN-
317 γ expression.
318 To further investigate the impact of CXCL12 signaling on T cell migration and the role of ALARM module
319 expression, we performed scRNA-seq using the transwell assay under three distinct conditions (Figure
320 6A). In the first condition (CXCL12-), T cells were placed below the transwell membrane without any
321 exposure to CXCL12. In the second condition (CXCL12+), T cells were placed below the transwell
322 membrane in direct contact with CXCL12. The third condition, (transmigrated), involved placing T cells
323 above the membrane, which were then collected from below the membrane after migrating in the
324 presence of CXCL12. This setup allowed us to assess the transcriptional changes associated with T cell
325 migration in response to CXCL12 (Figure 6A). Notably, all T cell subtypes identified by gene expression
326 were found in the three conditions (Supplemental Figure S4C).
327 To more accurately quantify gene expression changes and minimize the impact of subtype differences,
328 we focused on the most abundant subsets, namely naïve CD4 and CD8 T cells. These naïve T cells
329 clustered into four distinct groups, primarily due to differences in ribosomal gene expression and the
330 annotation of CD4 versus CD8 cells (Figure 6B). Additionally, there was a clear separation between
331 CXCL12- T cells and those exposed to CXCL12 or that had transmigrated, indicating significant
332 underlying gene expression differences. This separation was further evidenced by distinct changes in
333 the expression of ALARM module genes, with the CXCL12- condition showing markedly lower
334 expression patterns compared to the CXCL12+ and transmigrated groups (Figure 6B). Next, we
335 investigated specific gene expression alterations which would increase or decrease from CXCL12-,
336 CXCL12+ to the transmigrated condition (see methods).

337 In addition to ALARM module genes (Supplemental Figure S4D), several others exhibited notable
338 changes in expression across the different conditions (Figure 6C, Supplemental file 1). HSP90AB1, a
339 member of the HSP90 family of chaperone proteins, which is crucial for stabilizing proteins involved in
340 cell survival and esponses[35] was significantly upregulated in both CD4 and CD8 T cells. This suggests
341 that HSP90AB1 may play an important role in enhancing the functional stability of proteins required
342 for T cell migration and adaptation during CXCL12 stimulation. Interestingly, two genes with roles in
343 cell migration, VIM[36] and STK38[37] (serine/threonine kinase 38), also showed differential expressions in
344 both cell types. VIM, a key regulator of cytoskeletal organization that promotes cell motility[36], was
345 upregulated in migrated cells, aligning with its role in facilitating the cytoskeletal rearrangements
346 necessary for migration. On the other hand, STK38 was upregulated predominantly in the CXCL12-
347 condition. To better understand the roles of all differentially expressed genes, we grouped them by
348 function (Supplemental file S1, Figure 6D). We found gene sets involved in immune response,
349 migration, cytoskeleton, adhesion, stress response and metabolism were gradually upregulated in the
350 CXCL12+ and migrated conditions in both CD4 and CD8 T cells. Conversely, gene sets associated with
351 structural organization, gene regulation, and membrane transport were downregulated. These
352 findings suggest that CXCL12 signaling, and migration induce profound metabolic and functional
353 changes in T cells, preparing them for new roles that require increased energy and biosynthetic
354 demands. Given the prominent upregulation of metabolic pathways, we further investigated the
355 metabolic reprogramming that accompanies T cell migration and activation in response to CXCL12. To
356 achieve this, we performed a comprehensive metabolic pathway analysis on T cells using Compass, an
357 algorithm designed to characterize the metabolic state of cells by integrating single-cell RNA-Seq data
358 with flux balance analysis[38]. This in silico approach allows us to infer the metabolic status of individual
359 cells based solely on transcriptomic data, providing insights at single-cell resolution. The analysis
360 revealed significant upregulation in several metabolic pathways, notably glycolysis/gluconeogenesis,
361 phosphatidylinositol signaling, and amino acid metabolism in response to CXCL12 and migration
362 (Figure 7A). To visualize the overall metabolic differences between the conditions, we performed PCA
363 on the Compass score matrix, which quantifies the metabolic state in each cell (see methods)[38]. The
364 PCA results showed clear clustering of samples according to their condition, with non-migrated T cells
365 forming a distinct cluster separate from CXCL12+ and migrated groups (Figure 7B). This separation
366 underscores the significant impact of CXCL12-induced migration on the metabolic state of T cells.
367 Because the metabolic shift towards glycolysis is an essential hallmark of T cell activation[39] we focused
368 on the glycolysis and gluconeogenesis pathways to investigate the gene expression changes involved
369 in these processes (Figure 7C). Key glycolytic enzymes, such as glucose-6-phosphate isomerase and
370 pyruvate dehydrogenase, were significantly upregulated in migrated T cells compared to controls. To
371 validate the transcriptomic findings experimentally, we performed flow cytometry analyses to assess
372 glucose uptake and the expression of glycolytic enzymes. We measured protein levels of glucose
373 transporter 1 (GLUT1) and lactate dehydrogenase A (LDHA), a key glycolytic enzyme that converts
374 pyruvate to lactate for rapid ATP production under anaerobic conditions. Although the increase in
375 GLUT1-positive cells post-transmigration was not statistically significant ($p = 0.68$) (Figure 7D), LDHA
376 expression showed a significant increase in migrated T cells compared to controls ($p = 0.019$) (Figure
377 7D). Additionally, uptake of the fluorescent glucose analog 2-NBDG was significantly elevated in
378 migrated cells ($p = 0.031$) (Figure 7D), indicating enhanced glucose metabolism. These results confirm
379 an increased glycolytic activity observed in migrated T cells, consistent with the metabolic
380 reprogramming identified in our pathway analysis. In summary, our results demonstrate that T cells
381 migrating in response to CXCL12 undergo functional reorganization, enabling their transition from
382 circulating cells to active immune responders at sites of inflammation or tissue injury.
383
384 **ALARM is expressed in healthy individuals and variation is associated with infectious disease**

385 Up to this point, we have examined the role of ALARM cells primarily in the context of kidney rejection
386 and stable kidney transplantation patients. However, the findings from the Transwell assay, using
387 healthy donor cells in an in vitro model, suggest that ALARM cells may play a broader role beyond
388 kidney-specific contexts. Therefore, we investigated whether the ALARM module is expressed across
389 a healthy population to assess its broader function. For this, we explored a publicly available scRNA-
390 seq data set of ~25,000 PBMC from 45 healthy volunteers (HV) for ALARM expression [40]. The data was
391 generated from the *LifeLines DEEP* cohort in Netherland, ranges in age from 20 to 79 and contained
392 46,6% female individuals and was described to be healthy time of collection as estimated by two
393 general practitioner visits [41]. The ALARM gene expression was prominent in all cell types as shown by
394 its module score suggesting that ALARM is generally expressed in HV (Supplemental Figure S5A). We
395 note that age and sex did not result a significantly different expression of the ALARM (Supplemental
396 Figure S5B). Since this module was generally expressed in HV, we next asked whether it may be
397 involved in other disease conditions than transplantation rejection, but which implicate the
398 recruitment of circulating immune cells to specific tissues. To test this hypothesis, we exploited a
399 publicly available scRNA-seq data on PBMCs in which healthy individuals were intravenously injected
400 with the endotoxin lipopolysaccharide (LPS), a component of the cell wall of Gram-negative bacteria [5].
401 LPS in the bloodstream causes an immediate systemic release of a variety of inflammatory mediators,
402 a fever and a rapid but transient leukopenia [42,43]. This experiment is thought to mimic an acute systemic
403 inflammatory response (SIRS) [5], and thus provides an ideal proxy of how ALARM expressing cells are
404 responding to LPS-induced SIRS. We used the preprocessed available scRNA-seq data which contained
405 the 0-time point (10 HV), 90 minutes (6 HV) and 10h after the LPS injection (6 HV) and first evaluated
406 how the cellular proportions changed over time. We measured the ALARM score in the three
407 conditions across the CD4, CD8, NK and B cells (Figure 8A). There was rapid and significant reduction
408 of ALARM high expressing cells as soon as 90 minutes and which further decreased until 10h after the
409 LPS injection (Figure 8A). This decrease was continuous within three individuals for which both
410 timepoints were available, indicating that ALARM cells are reduced in a time dependent manner
411 (Figure 8B). This drastic change of ALARM expression in such a short time frame suggests that cells
412 which highly express this module rapidly egress from circulation, directly contributing to the transient
413 leukopenia observed upon LPS injection. This is also consistent with the transient leukopenia
414 associated with LPS i.v. injection. Furthermore, this response to LPS which is thought to be initiated
415 via TLR4 receptor signaling expressed on circulating monocytes [44,45] signifies that ALARM is not solely
416 implicated in transplantation rejection or kidney immune cell invasion but appears to also be involved
417 in the inflammatory response to endotoxin.
418 Next, we investigated how ALARM may regulate when the site of inflammation is localized to a single
419 organ as in urinary tract infection. For this we explored a publicly available PBMC scRNA-seq dataset
420 which contained patients with leukocyte infiltrating urinary tract infection (UTI). We chose this
421 condition as the data was generated on patients which presented a localized infection with infiltrating
422 leukocytes. The study also provided results of HV and sepsis patients, notably patients with
423 bacteremia, i.e. bacterial presence in the blood. [1] (Figure 8C). The bacteremia patients were thus also
424 used for comparison since this condition reflects a generalized or systemic infection which is distinct
425 from a localized infection such as UTI (Figure 8C). Interestingly, we found that ALARM expression was
426 reduced in circulating cells in UTI patients (except B cells), consistent with the recruitment of
427 leukocytes to the tissue. In contrast, in patients with bacteremia, the ALARM cells accumulated in the
428 circulation, indicating that under this condition ALARM expressing cells may not be recruited to a
429 specific tissue. This contrasted with the response to LPS iv which is also thought to engender a systemic
430 response (see discussion).
431 To further explore the role of ALARM in response to localized infection, we analyzed two scRNA-seq
432 datasets generated from Covid-19 patients [5,46]. The aim was to explore the dynamics of ALARM cells in
433 the circulation in comparison to the lung. For this, we compared ALARM expression in PBMCs with

434 ALARM expression in broncho-alveolar lavage fluid (BALF). The two separate original studies stratified
435 the patients by healthy, moderate, and severe Covid-19 disease and we used this stratification to
436 compare the ALARM module expression in the blood (PBMC) and in the lung (BALF) (upper panel figure
437 8D). The ALARM cells diminished according to disease severity (T, B, Mono, but not NK) in the blood
438 stream. This reduction was concomitant with an increase of ALARM high cells in the lung suggesting
439 that ALARM cells are migrating to the site of infection. Interestingly, these changes were cell-type
440 specific, notably while CD4+ T cells increased it was not the case for CD8+T cells. We note that the cell-
441 type specific modules did not change between disease state and between blood and lung (lower panel,
442 figure 8D) indicating that the similar cell types were analyzed and that the cell-type specific modules
443 were not related to disease state.
444 Collectively these results show that ALARM displays a normal distribution of expression across healthy
445 individuals and changes in response to distinct disease states (bacterial, viral, kidney rejection). It is
446 noteworthy that expression alterations of this module are apparent in distinct cell types depending on
447 the disease conditions.
448
449 **Recruitment of CXCR4+ T Cells During Lung Infection In Vivo**
450 To independently validate the rapid recruitment of immune cells during lung infection, as exemplified
451 in the COVID-19 results above, we used a well-established pneumonia mouse model[47,48] to study T cell
452 recruitment. The infection was induced with *E. coli* and resolved after 7 days, with the peak of infection
453 occurring between day 1 and day 3. To specifically evaluate the recruitment of T cells from the blood,
454 we employed CD45-PE mediated immune staining of blood immune cells before and during infection[48]
455 (Figure 9A). This method allowed us to precisely quantify the recruitment of cells from the blood during
456 the infection. We observed that T cells were recruited from the blood as early as day 1, with a
457 significant peak at day 3, indicating a rapid response to the lung infection (Figure 9B). Upon analyzing
458 the membrane expression of CXCR4 on these cells, we found that 15% to 20% were CXCR4+ T cells, a
459 key component of the ALARM module, suggesting that a diverse set of T cells is recruited to the lung
460 (Figure 9B, Pie Charts). Next, we investigated the membrane display of CD69 and found that most
461 CXCR4+ CD69+ T cells were predominantly CD45+ cells (60-80% at day 1 and day 3), indicating their
462 blood-derived origin (Figure 9C). This phenomenon was particularly evident at day 1 and day 3,
463 corresponding with the peak of infection. Additionally, stratification of the CXCR4+ T cells by CD4+ and
464 CD8+ subsets showed that CD4+ T cell recruitment was much more abundant than CD8+ T cells,
465 mirroring observations from COVID-19 lung infection studies (Figure 9D).
466 These findings demonstrate that CXCR4+ T cells, particularly CD4+ T cells, are rapidly recruited from
467 the blood to the lung in response to infection, with peak recruitment occurring between day 1 and day
468 3. The presence of CD69+ cells among the recruited CXCR4+ T cells suggests that these cells are not
469 only migrating but are also likely being activated or display a residency program[34]. The observation of
470 the concomitant *protein* expression and membrane display of CXCR4 and CD69 is characteristic of the
471 ALARM module, which is defined by the *gene* expression of both CXCR4 and CD69. The results
472 underscore the crucial role of the ALARM module in orchestrating an effective immune response
473 during the peak of lung infection, highlighting its broader relevance beyond kidney-specific contexts.
474
475
476 **Disease classification of ALARM in immune mediated diseases**
477
478 The above results suggest that ALARM expression changes are associated to immune diseases, likely
479 via the recruitment of circulating immune cells to the site of inflammation. It may thus be possible that
480 the ALARM genes could be used as predictors for immune disease classification. The rationale is that
481 if ALARM genes are relevant for a precise disease state, they would be strong predictors to classify
482 healthy from disease [49]. To investigate this hypothesis, we analyzed 10 immune-mediated diseases
483 with distinct tissue tropisms (Supplemental Figure S8A). Out of these 10 diseases, systemic lupus
484 erythematosus (SLE), Sjogren syndrome and Anti-neutrophil cytoplasmic antibody-associated
485 vasculitis (ANCA) are known to favor the kidney among other organs. In contrast, mixed connective

tissue disease and systemic sclerosis rather favor the connective tissues and rheumatoid arthritis targets the joints. Thus, we aimed to test whether ALARM genes are good classifiers of disease vs healthy condition and whether the classification of disease state varies depending on the tissue tropism. This was possible thanks to a large bulk RNA-seq dataset comprising 28 circulating immune cell-types which included 337 patients across 10 immune mediated diseases and 79 healthy controls [50] (Supplemental Figure S8A). We evaluated total gene expression using UMAP of this data set (Figure 10A) and found that ALARM module expression was present in all major cell types (Figure 10B). To test the discriminative ability of the ALARM module, we then devised a classification pipeline comparing disease state (all 10 immune diseases) vs healthy using logistic regression (Figure 10C). As in the data set there were transcriptomics data from 28 cell types available, we focused on B cells, T cells, NK and Monocytes by regrouping their respective sub cell types together (see Figure 10A).

We then generated an ALARM gene classification model compared it to two other models. The first one using the most variable genes (coefficient of variation (CV^2), see methods) and the second model was based on genes in cell type identity modules (see figure 2). The rationale for using CV^2 gene selection was to use an independent gene selection process which is more predictive than random gene selection. We compared the prediction performance of the three models using the F-1 score as it gives equal weight to precision and recall (Figure 10D, F1 score). Interestingly, the ALARM genes were the best predictors for CD8, Monocytes and NK cells and were similar in performance to the cell type specific modules of CD4 and B cells. To account for possible imbalances in the numbers of disease and control samples we also computed the Cohen's Kappa score (Supplemental Figure S8B). The results were consistent with the F-scores. Of note, the CV^2 gene selection approach was less predictive in all cell types (Figure 8D and Supplemental S8B).

To evaluate whether the ALARM module was prominent for any specific cell sub types within the major cell types (e.g., CD4 T helper cells vs CD4 naïve cells) we estimated disease classification performance of the ALARM module separately for each sub type. Predictability as estimated by F1 score remained robust when each subtype was analyzed separately in comparison to the CV^2 method (Figure 10E). We note that in certain subtypes there were too few samples to compute an accurate Kappa score (Supplemental Figure S8C). Nevertheless, this result indicates that ALARM was found to be relevant in all the subtypes analyzed.

Next, we evaluated whether each of the 10 immune mediated diseases could be individually classified from healthy (Figure 10F). The best classification ability of ALARM was found for Mixed connective tissue disease, rheumatoid arthritis, Sjogren's syndrome, SLE and Takayasu's Arteritis. This indicates that ALARM genes are not specifically predictive for tissue tropism but appear to be relevant independently of the targeted tissue. In most cases ALARM outperformed the CV^2 feature selection suggesting that ALARM genes are likely to be implicated in their disease etiology. In summary, this comprehensive classification analysis indicates that ALARM genes are strong predictors of disease state across the majority of circulating immune cells and within the 10 immune related diseases.

ALARM is enriched for genetic disease associations

Since the ALARM genes are strong predictors of immune mediated disease and its general role within multiple cell-types and across multiple infections, immune related and autoimmune diseases, it is likely that ALARM is enriched for genes known to be associated to diseases. To test this, we exploited the DisGeNET database[51] which provides a comprehensive compilation of published and curated human gene disease associations (GDAs) from repositories including Mendelian, complex and environmental diseases and enables enrichment analysis of such GDAs. Notably, we found that 24 out of the 61 ALARM genes (39%) were associated with a disease. To estimate the probability of this occurring by chance, we compared this to random sampling of 61 genes and their quantification of GDAs (Supplemental Figure S8D permutation). The probability of reaching 40 % of genes or more was below > 0.001, indicating that ALARM genes are highly enriched for GDAs.

Next, we tested whether the ALARM genes were enriched for diseases associations (Figure 10G). The top three disease categories enriched for GDAs were Juvenile arthritis, Hodgkin disease and polyarthritis comprising by themselves distinct 13 genes. This enrichment analysis indicates that

538 ALARM is also genetically connected to disease state. In summary, ALARM is a strong classifier and
539 genetically linked to immune disease and thus is likely to play a general role in multiple immune
540 mediated diseases.
541
542 **Discussion:**
543 In this study, we gathered a cohort of matching ABMR, TCMR and stable patients and generated a
544 comprehensive scRNA-seq atlas of circulating immune cells across time and conditions. We then
545 identified multiple gene co-expression modules. Five out of nine modules were related to a single cell-
546 type while three were expressed in closely related cells (CD4+, CD8+ and NK cells) and only the ALARM
547 module was prominent in multiple cell-types. The observation that single cell transcriptomes mostly
548 reveal cell-type specific modules was also described by Kotliar et al., [9] in which they distinguish
549 between identity (i.e., cell-type specific) gene expression programs (GEP) and activity GEPs. It is
550 possible that cell-type specific co-expressed genes are better detected as they show a greater
551 coherence within a well-defined group of cells. This is also notable in the presented data as the gene
552 expression scores of cell-type specific modules show less variance than the ALARM module.
553 Nevertheless, cell-type specific gene expression may not necessarily imply that it remains constant
554 across conditions. For example, we noted that some cell type specific modules were associated to
555 disease state (Figure 2E), notably the NK cell and monocyte specific modules were increased during
556 rejection. It is also possible that because of the relatively low number of genes per cell detected when
557 compared to bulk transcriptomics, cell identity programs are preferentially detected, and more subtle
558 condition specific modules are not robustly detected. Indeed, while cNMF revealed additional modules
559 in the separate batches, only the ALARM module was consistently identified across the three batches.
560 The ALARM genes were found to be highly enriched for transcription factors and gene ontology
561 pathways associated with the gene expression machinery including transcription, mRNA processing
562 and ubiquitination. Prominent transcription factors included the AP-1 complex and the NFKB subunit
563 REL which are both associated with stress responses and inflammation. The membership of CD69 in
564 the ALARM module also suggests a role of stress response. CD69 is a classical early activation marker
565 of lymphocytes, as demonstrated by its rapid display on the surface of T cells after TCR stimulation[34].
566 In addition, CD69 is also known to be a tissue retention marker as it is expressed on resident memory
567 T cells in distinct tissues. In blood, this gene has been associated with chronic inflammation in various
568 diseases including rheumatoid arthritis [52] and systemic lupus erythematosus [53]. Concomitant with this,
569 CD69 protein expression is increased on infiltrated immune cells at the site of inflammation in immune
570 mediated diseases including systemic sclerosis, rheumatoid arthritis, systemic lupus erythematosus[34].
571 The membership of CD69 in this module thus indicates that this module role could be to prepare
572 circulating cells T cells for TCR stimulation and for tissue retention once moved into a tissue, i.e., to
573 become T resident effector cells. This notion is also consistent with the increased expression of the
574 ALARM module in the kidney biopsy transcriptomics data (Figure 4).
575 The cytokine receptor CXCR4 was also identified in this module. CXCR4 is predominantly expressed by
576 lymphocytes as well as monocytes and through which the CXCL12 ligand promotes chemotaxis to
577 tissues via a concentration gradient [54]. CXCL12 is expressed in multiple tissues including the kidney and
578 is altered during pathophysiological responses including immune diseases. Indeed, an alteration of
579 CXCL12 expression was observed in the kidney transplantation biopsies and this increase was
580 associated to elevated ALARM gene expression in the tissue (Figure 4E). It is possible that the CXCR4-
581 CXCL12 axis contributes to the recruitment of ALARM expressing cells in the case of kidney
582 transplantation rejection and other immune diseases. This is also consistent with the observation that
583 cells expressing the ALARM module decrease in the circulation during kidney graft rejection. This
584 observation was confirmed by both transcriptomics and histological studies of pigs as well as
585 transcriptomics in human kidney biopsies, thus, via CXCR4-CXCL12 leading ALARM cells to infiltrate the
586 tissues during rejection. This mechanism was further supported by an in vitro trans well assay, where
587 CXCL12 was shown to induce T cell migration. More importantly, it was found that both CXCL12
588 presence and migration significantly increased CD69 protein expression at the cell membrane.
589 Specifically, the combination of HMEC contact and CXCL12 presence was necessary for the highest

590 expression of CD69, similar to that observed during transmigration. Single cell transcriptomic analysis
591 of migrated and non-migrated cells further revealed that certain gene groups, particularly those
592 involved in cytoskeleton organization, migration, and immune response, were upregulated, while
593 others were downregulated, indicating a shift in cellular state to adapt to new functions after
594 migration. Furthermore,we found that migration of T cells in response to CXCL12 is accompanied by
595 significant metabolic reprogramming. We observed upregulation of glycolytic pathways, increased
596 expression of LDHA at protein levels, and enhanced glucose uptake in migrated T cells. Metabolic
597 reprogramming towards increased glycolysis is a hallmark of activated T cells and is essential for their
598 effector functions during immune responses [39]. These findings suggest that ALARM module expression
599 not only primes T cells for migration and tissue retention but also prepares them metabolically for the
600 demands of their new functional roles at sites of inflammation.
601 This CXCR4-CXCL12 axis also highlights the notion that the ALARM module is not necessarily specific to
602 transplantation rejection or the kidney. Indeed, reanalysis of circulating immune cells from publicly
603 available scRNA-seq data showed that the ALARM module was expressed in 45 unrelated healthy
604 individuals [40] and showed significant alteration between distinct pathological conditions (Figure 5).
605 First, the ALARM response to LPS iv injection in healthy individuals revealed that it is time-dependent,
606 illustrated by a gradual decrease of ALARM cells within the same individuals over time. Second, there
607 was a significant difference between ALARM cells depending on the location of the pathological
608 condition. ALARM cells were shown to be decreased in response to kidney rejection, in response to
609 leukocyte infiltrating UTI and Covid-19 infection of the lung highlighting the role of ALARM in the
610 recruitment of cells to the site of inflammation and infection. This was further supported by
611 bacteremia sepsis, a state of systemic inflammation in which ALARM cells were increased in the blood.
612 While LPS iv injection and bacteremia induce both a systemic immune response, the former is
613 associated is essentially an endotoxemia response associated to a transient leukopenia[42,43] while the
614 latter is a complex and heterogeneous condition that involves multiple factors beyond LPS, such as
615 pathogen virulence, host susceptibility, and coexisting medical conditions. It thus makes sense that
616 leukopenia is associated to the decrease of ALARM, while in bacteremia ALARM expression is
617 increased. Third, ALARM cells showed a gradually measurable response to disease severity. This notion
618 was observed by combining two distinct and complementary Covid-19 datasets one of which was
619 collected on BALF, and which had stratified their patients according to disease severity. ALARM cells
620 decreased in response to severity in the blood with a corresponding increase in the lung.
621 Fourth, we collected several lines of evidence suggesting that ALARM cells are indeed recruited to the
622 site of inflammation and/or infection. During acute rejection induced in the pig model there was a
623 rapid infiltration of leukocytes concomitant with the reduction of ALARM cells in the blood. The
624 analysis of kidney biopsies revealed an increase of ALARM gene expression during kidney transplant
625 rejection. Similarly, the recruitment of ALARM cells to the lung was observed during Covid-19 lung
626 infection. Finally, an in vivo mouse model of *E. coli* pneumonia demonstrated that CXCR4+ and CD69+
627 T cells are rapidly recruited from the blood to the lung during the peak of infection further supporting
628 the role of the ALARM module in mediating immune cell recruitment to sites of inflammation.
629 We thus propose a model in which ALARM expression priorities the infiltration capacity of each
630 circulating cell (see figure 11 model). This model has wide ranging consequences in precision medicine
631 as blocking of ALARM cells to the site of inflammation in the case of kidney rejection may prevent
632 further organ damage or attenuate the immune response in the case of Covid-19 lung infection. It may
633 also be useful to predict disease state as we have shown in figure 8. ALARM was a strong classifier of
634 immune disease when compared to healthy individuals. The importance of ALARM was independently
635 demonstrated by its enrichment for genes mutated in notably juvenile and poly- arthritis. Further
636 investigation is however required to test the specificity of disease detection or whether ALARM is
637 merely a response to inflammatory state in the circulation.
638 There are several limitations of this study, first it is based on gene transcription and thus remains to
639 be explored for protein expression, however this is difficult to achieve at single cell resolution and we
640 are not aware of any gene co-expression modules estimated at the protein level. Nevertheless, the
641 transwell assay and in vivo mouse model experiments indicate a connection between the CXCL12-

642 CXCR4 axis and CD69 and their display during and after migration. While direct evidence of the
643 recruitment of these cells to the tissue has been experimentally confirmed the module expression was
644 shown to be low in the blood and high in the kidney tissue during rejection which may also be caused
645 be lack of the source of these cells, or a slowing in cellular maturation before expressing ALARM genes
646 an avenue that should be explored in subsequent studies. A further limitation is that as of now we do
647 not have a protein surface marker panel that could be associated to cells with high or low ALARM
648 expression. Such markers would enable the purification of ALARM cells enabling further molecular and
649 cellular characterization. Markers would also allow the targeted modulation of the recruitment of
650 ALARM cells to the graft or during Covid19 as well as other immune mediated diseases may thus impact
651 disease severity. Nevertheless, our study remains important in terms of precision medicine,
652 highlighting the discovery of ALARM, which expression enables cells to be preferentially recruited to
653 the inflamed tissue. This notion is likely to open novel strategies of disease monitoring and disease
654 intervention.
655
656
657
658
659 **Methods:**
660
661
662 Study of kidney transplant patients
663 Kidney transplantation patients
664 The PBMC samples used in this study (see table 1) were obtained from the DIVAT biocollection (CNIL
665 agreement n°891735, Réseau DIVAT: 10.16.618). Every patient included in the study was enrolled in
666 the DIVAT biocollection following their informed consent. The PBMC from patients were isolated from
667 kidney transplantation biopsies, frozen with DMSO 10% and stored in liquid nitrogen at the Centre de
668 Ressources Biologiques (CRB, CHU Nantes, France).
669 Cell preparation
670 Frozen PBMC samples were rapidly thawed and resuspended in complete Roswell Park Memorial
671 Institute (RPMI) 1640 media (Invitrogen, Carlsbad, CA) with 5% FBS, pre-heated at 37°C.
672 Following washing steps in PBS+0.04%BSA, cell pellets were resuspended in 200µL FACS buffer (1X PBS
673 supplemented with 2mM EDTA, 2% FBS) in which dead cells were labelled by adding 0.1 µg/mL DAPI
674 (Invitrogen, Carlsbad, CA). Cells were filtered on 70 µm cell strainer and living cells were then sorted
675 using a Fluorescence-activated cell sorting (FACS) Aria II cell sorter (BD Biosciences, Mountain View,
676 CA).We used the same method as previously described for single cell RNAseq[55,56]. One million cells
677 were kept for each sample and resuspended in 100µL of staining buffer (PBS,2%BSA,0.01% Tween)
678 according to the cell hashing protocol[13] recommendations. Cells were incubated for 10 min with 10µL
679 of human Fc blocking reagent. Each sample was then mixed with 1uL of a specific TotalSeq-A hashtag
680 antibody (BioLegend, San Diego, CA) and incubated on ice for 30 min. Following 3 washing steps with
681 the staining buffer, cells were counted and their viability measured using an ADAM-MC automatic cell
682 counter (NanoEntek, Seoul,South Korea) to ensure a viability above the recommended 70%. All the
683 samples were pooled at an equal cell concentration in a single vial, centrifugated and resuspended in
684 PBS to obtain a concentration of 700 cells/µL, to match the targeted cell recovery of 32,000 cells.
685 Encapsulation of single cells was performed on a 10XChromium (10X Genomics, San Francisco, CA)
686 with the Chromium Single Cell 3' Library and Next GEM reagent kit v3. The libraries were sequenced
687 twice for each of the three experiments on the NovaSeq 6000 (Illumina, San Diego, CA) with S1 flow
688 cells. The sequenced libraries were aligned to the GRCh38-2020-A reference genome with CellRanger
689 v5.0.0 (10X Genomics, San Francisco, CA). The scRNA-seq was performed in 3 different experiments
690 following the same protocol. Each experiment included longitudinal samples from three patients (one
691 stable patient, one humoral rejection and one cellular rejection) as well as one late sample of a tolerant
692 patient.
693 Method demultiplexing and Seurat analysis

694 The count matrices were analyzed in R 4.0.3 using the Seurat R package (v4.0.2, Satija Lab[14]). Each
695 experiment was first processed separately, with the same workflow. First, following the standard
696 workflow recommendations, cells with less than 200 unique feature counts were removed (potential
697 empty droplets). Cells with a percentage of mitochondrial genes greater than 15% were excluded as it
698 results from mitochondria degradation from dead or dying cells. The hashing antibody sequences were
699 then collected to demultiplex and assign each cell to its sample using the MULTIseqDemux function.
700 Cells with too little labels information were called "Negative" while cells with a high count of two or
701 more different oligo-conjugated antibody sequences were called "Doublets". Only cells with a unique
702 HTO were kept for downstream analysis. Singlet cells were annotated automatically with the Azimuth
703 workflow within Seurat, by mapping the query cells on an annotated reference of 162,000 PBMC
704 measured with 228 antibodies[15].
705 All the runs were then merged in a single Seurat object. Doublets and contaminant cells to exclude
706 were selected by identifying cells co-expressing marker genes from distinct cell types. After
707 normalization of the global object, the 2,000 most variable genes in the data were selected to compute
708 the correction using the reciprocal principal component analysis (RPCA). The final annotated and
709 corrected object gathering the 12 patients was composed of 50,507 cells.
710 Gene Module Identification
711 Consensus Non-Negative Matrix Factorization (cNMF)[9] was used to decompose the cell vs gene
712 expression matrix into cell vs module and usage vs gene low-rank matrices. Non-Negative Matrix
713 Factorization is a stochastic method and therefore it was run with 200 NMF replicates to find a
714 consensus robust factorization. For each batch, the top 2000 over dispersed genes were selected as
715 input to the cNMF run. Different K values (7 to 14) were explored to determine the optimal number of
716 modules. For each K, the *stability* and *error* metrics were examined. The best K was chosen such that
717 the error was minimum, and the stability was maximum. Each batch was independently processed to
718 mitigate batch effects from the three different runs. For the three batches, the optimal number of
719 modules (K) were 14, 11 and 13.
720 Module filtering: All modules from the three batches were collated and hierarchical clustering was
721 performed to identify matching modules. Jaccard similarity score was used to define the similarity
722 between two modules. Finally, only those clusters were retained that could represent all three
723 batches. In this way, nine consistent modules were identified. A unique geneset was determined for
724 each consistent cluster by intersecting the top-ranking 200 genes from the modules. The threshold
725 was achieved by observing a scree plot of input number of genes vs the number of genes after
726 intersection.
727 Module association: The genes in the module were examined for a module to be associated with a
728 known cell type. The modules containing marker genes were associated with their respective cell type;
729 for example, the module with MS4A1 and CD79A genes was associated with B cells. However, one of
730 the modules (Module 9, later named as the ALARM module) could not be associated with a known cell
731 type as it did not contain cell type specific markers and was well expressed in multiple cell types.
732 UCell module score
733 The enrichment of a particular set of genes in an individual cell was measured with the UCell[58] module
734 score. The score is calculated using the Mann-Whitney U statistic, which compares the expression
735 levels of the module genes relative to the total gene expression of the cell. The U-statistic outcome is
736 then normalized between 0 to 1 to produce the UCell score.
737 Regression analysis
738 Change in the module score (μ) along time (τ) for each condition was modelled with linear regression
739 method. The 'lm()' function from R *stats* package was used to fit a distinct model $\mu \sim \tau$, per celltype
740 within the group of Stable, ABMR and TCMR individuals. A positive slope indicated that the module
741 score increased with time.
742 Gene Ontology analysis
743 The gene ontology (GO) functional enrichment of the ALARM module gene list was performed using
744 the R package WebGestaltR (v0.4.4) for Biological Process (BP) and Molecular function (MF)
745 annotation. P-value are obtained with the hypergeometric test for ORA (Over-Representation

746 Analysis). As background dataset for enrichment the top 2000 variable genes from each run were
747 merged and used.
748 Transcription factor enrichment analysis
749 To test for enrichment of transcription factors in the ALARM module, GSEA molecular signature
750 database was used to count the transcription factors gene family in the module and in the background
751 dataset used for gene ontology analysis. Fisher exact test was used to calculate the P-value and
752 enrichment of transcription factors in the ALARM module.
753
754 Allogeneic kidney rejection model in pig
755 Animal model
756 The study protocol was approved by the French Ministry of Higher Education, Research and Innovation
757 (APAFiS #30136). The experiment was performed on 60 to 80kg male pigs (*Sus scrofa*). Test card with
758 pre-applied antibodies from Serafol (Berlin, Germany) were used to identify the pigs' ABO blood
759 groups. The alloreactivity was performed by mixed lymphocyte reaction assay between donor and
760 recipient. The donor pig was selected from a different breed as inbred pigs might escape rejection.
761 Donor and recipients blood groups were ensured to be compatible to avoid hyper-acute rejection, and
762 mixed lymphocyte reaction assays were positive thus proving their alloreactivity.
763 Allogeneic transplantation
764 Unilateral nephrectomies were performed on two recipient pigs under general anesthesia with a
765 premedication by Zolazepam/Tiletamine (Zoletil ® Virbac, Carros, France) 15 mg/kg IM, before
766 intubation and a maintained ventilation with a mixture of 49% oxygen, 49% nitrous oxide and 2%
767 isoflurane. The two kidneys from a third donor pig were harvested in the same operating time. The
768 two recipient pigs received one collected kidney each for an orthotopic transplantation. During
769 surgery, a central venous catheter (CVC) was inserted into the internal jugular vein for hydration and
770 medication. Post-operative analgesia was performed every day with intravenous injections of
771 Nalbuphine (Nubain ®, Mylan, Canonsburg, Pennsylvania) and Paracetamol at a dose of 25 mg/kg.
772 Prophylactic antimicrobial therapy was conducted with Cefazolin 1 g (Cefovet ®, Dopharma, Ancenis,
773 France).
774 PBMC collection in Pig model
775 Blood samples were collected daily through the CVC and frozen in a CoolCell® container (Corning ®,
776 Corning, NY, USA) at -80°C following the PBMC isolation. Kidney transplant biopsies were collected
777 daily using automated biopsy needles of 16 gauges under ultrasound guidance while pigs were sedated
778 by Zolazepam/Tiletamine and locally anesthetized with Lidocaine. Kidney samples were then placed in
779 cryovials with 1mL fetal bovine serum (FBS) and 10% dimethylsulfoxyde (DMSO) for gradual cooling in
780 a CoolCell chamber.
781 Single cell preparation for pig model
782 Blood samples were processed as described previously (see Methods 1.2). After filtering, cells were
783 centrifugated at 300g for 10 minutes at 4°C and resuspended in staining buffer for the HTO antibody
784 labeling (see Methods 1.3). The sequenced libraries were aligned to the Sscrofa 11.1 (February 2017
785 release) reference genome with CellRanger v5.0.0.
786
787 Pig biopsy immunostaining analysis
788 Kidney biopsies fixed Carnoy's solution for 30 minutes followed by a fixation in formaldehyde for 24h
789 for optical microscopy purpose. A second batch of kidney biopsies was prepared for immunostaining
790 purpose: biopsies were placed in cryomold, covered with optimal cutting temperature (OCT)
791 compound and immersed in cold isopentane. Following their solidification, cryomolds were stored in
792 liquid nitrogen. Cryosectioning was performed and the resulting slides of kidney biopsies were stained
793 with periodic acid-Schiff (PAS) and Masson's trichrome stains (TM).
794 The cellular infiltration was counted using ImageJ[59] software on the PAS-stained kidney biopsies. Areas
795 of interest were selected to exclude areas with glomeruli. Pictures were first converted to 8-bit
796 grayscale, and the threshold of detection was set to capture only the stained cells.

797
798 Transmigration Model
799 HDMECs (10 × 10^4 cells) activated with TNF-α (100 U/ml) for 24 hours were seeded O/N onto 1%
800 gelatin-coated Transwell membrane inserts (24-well, 3-μm-pore polycarbonate membrane; Corning
801 Life Science) in endothelial cell growth medium at 37°C. On the day of the assay, purified CD3 T-cell
802 subsets (4×10^5) were added to the upper transwell migration chamber, and the chemokine CXCL12 (50
803 ng/ml) was added to the lower transwell migration chamber. Migration was assessed after 4 hours by
804 quantifying the number and phenotype of migrated cells in the lower chamber using 123count eBeads
805 counting beads and a Cytek AURORA flow cytometer (5 lasers). Migrated CD3 were surface stained
806 with specific antibodies to characterize phenotype CD3, CD8, CD4, CD45RA, CCR7 and activation
807 molecules CD69, CD25, CD127, CD95, CD103 and CD49. The antibodies used for the cytometric
808 analyses are listed belowExpression of Cytotoxic Molecules by Human transmigrated CD3 T Cell
809 Subsets
810 To define the expression of cytotoxic molecules, transmigrated CD3 were restimulated with PMA
811 (50ng/mL), ionomycin (500ng/mL) and BFA (5ug/mL). Transmigrated CD3 were surface stained with
812 specific antibodies for phenotypic characterization of CD3, CD8, CD4, CD45RA, CCR7, and after fixation
813 and permeabilization (BD Cytofix/Cytoperm), intracellular staining was performed using antibodies
814 against granzyme B (GZMB) and perforin-1 (PERF-1), granulysin, and TNFa. The antibodies used for
815 cytometric analyses are listed below.
816 **Metabolic characterization of Human transmigrated CD3 T Cell Subsets**
817 Transmigrated T cells were stimulated over-night with plate bound anti-CD3 (1ug/mL) and anti-CD28
818 (2ug/mL) mAb. Cells were washed, surface stained with anti-CD3, CD4 and CD8 mAbs and cultured for
819 30' at 37°C 5% CO2 in glucose-free medium containing 50 µM 2-NBDG. Alternatively, cells were surface
820 stained with anti-CD3, CD4, CD8 and GLUT1 mAbs and after fixation and permeabilization (BD
821 Cytofix/Cytoperm), intracellular stained with anti-LDHA mAb. Data were acquired using a 5 lasers Cytek
822 AURORA flow cytometer and analyzed using OMIQ.
823 Antibodies used:

Target	Clone	Fluorochrome	Provider
CD3	OKT3	Purified	In house production
CD28	CD28.2	Purified	In house production
CD3	SK7	PE Cy7	BD
CD4	SK3 R7	CF568	Cytek
CD8	SK1	Sparkblue 550	Biolegend
CD69	FN50	APC Cy7	Biolegend
GLUT1	202915	FITC	R&D Systems
LDHA	APC	ab310866	Abcam
CD3	UCHT1	PE-Cy7	Biolegend
CD8	SK1	Spark Blue 550	Biolegend
CD4	SK3	cL-Fluor YG584	Cytek
CD45RA	HI100	BUV496, BV786	BD
CCR7	G043H7	APC Fire810	Biolegend
CD69	FN50	APC-Cy7	Biolegend
CD25	M-A251	PE-Fire700	Biolegend
CD127	HIL-7R-M21	APC R700	BD Biosciences
CD95	DX2	PE-Cy5	BD Biosciences
CD103	Ber-Act8	BV605	BD Biosciences

CD49a	SR84	BUV661	BD Biosciences
GZMb	QA16A02	PE-Cy5	Biolegend
IFNg	B27	APC	BD Biosciences
GNLY	RB1	AF488	BD Biosciences
TNFa	Mab11	BUV737	ThermoFisher

824
825 scRNA-seq of Transmigration model
826 T cells were collected as indicated in figure 5A and then processed for chromium loading. The three
827 conditions were processed at the same time using CITE-seq approach (see above).
828 Primary Analysis: Fastq files were generated from Illumina bcl files using Bcl2fastq version 2.2.
829 Cellranger v7.2 was employed to create a filtered scRNA gene expression matrix from the fastq files,
830 utilizing default parameters and the human genome reference version GRCh38-2020.
831 Seurat version 5.1 was used for subsequent quality control and preprocessing. HTODemux, with a
832 positive quality threshold of 0.95, was applied to demultiplex cells, identifying singlets and associating
833 each singlet with the corresponding condition. Cells with an RNA count exceeding 30,000 or exhibiting
834 mitochondrial gene expression above 20% were excluded. Cell type annotation was performed using
835 Celltypist annotation tool[60].
836 Secondary Analysis: For each CD8 and CD4 cell type, genes exhibiting significant gradients across the
837 three conditions were identified using a linear model. The dependent variable in the model
838 represented the condition, with control CXCL12-, control CXCL12+, and migrated groups assigned
839 values of 0, 1, and 2, respectively. To simulate multiple individuals, cells were randomly grouped into
840 10 groups. These groups served as pseudo-individuals, each containing at least 50 cells, created using
841 the `createfolds` function from the caret package. The final linear model was formulated as ~
842 condition + (1 | individual). Top genes were identified based on significant p-values (< 0.05), and the
843 direction of their gradients across the conditions was noted.
844 Among the top gene modules, several were found to participate in known pathways. The average gene
845 expression for these modules was calculated using the `AddModuleScore` function of Seurat.
846 Metabolome Pathway Analysis: The Compass algorithm[38] was used to characterize the metabolic
847 states of CD4+ T cells across three different conditions: Migrated, CXCL12-positive (CXCL12+), and
848 CXCL12-negative (CXCL12-). The algorithm designed to infer the metabolic state of cells from scRNA-
849 Seq data through flux balance analysis. It addresses the limitations of traditional metabolic assays in
850 assessing metabolic states at the single-cell level, leveraging transcriptome data to predict metabolic
851 activities. For reference to metabolic reactions and pathways, the RECON2 database was utilized for
852 this analysis. The scRNA data were first micropooled, resulting in 20 pseudobulk samples for each
853 condition in order to compare the same number of samples between conditions. The reaction
854 penalties were estimated for various metabolic pathways based on gene expression levels in each
855 pseudobulk sample. Reaction penalties were then converted to negative log scores, with higher values
856 indicating greater predicted activity. Significant active reactions were identified using the Wilcoxon
857 rank-sum test, comparing the Migrated samples to the CXCL12- samples. Reactions with an adjusted
858 p-value of less than 0.1 were considered significant.
859
860 Analysis of the ALARM module in kidney biopsies
861 Two separate studies were analyzed here. The Reeve *et al.* Affymetrix Microarray data in RAW CEL
862 format was downloaded from Gene Expression Omnibus (GEO) website with accession number
863 GSE98320. The samples were pre-processed using robust multi-chip averaging (RMA) implemented in
864 Bioconductor. The patient condition was obtained from the 'd96' metadata column as designated in
865 the corresponding Series Matrix file. The patients strictly defined as either TCMR (n=76), ABMR
866 (n=197), Mixed (n=39) or no major abnormality (STA, n=257) as a stable condition, were retained for

867 downstream analysis. The Callemeyn *et al.* dataset was downloaded from GEO with accession number
868 GSE147089. The CEL files were similarly pre-processed using the RMA method. The labels for each
869 sample were obtained from the Series Matrix file and the phenotypes are defined as biopsies without
870 ABMR (n=168), DSA negative ABMR (n=26) and DSA positive ABMR (n=30).
871 The score used to stratify patients was computed by averaging the z-scores of the ALARM module
872 genes.

873 ALARM mean z-score distribution across conditions
874 The Kolmogorov-Smirnov (KS) test was then used to compare the distribution of the ALARM module
875 score between stable and rejection conditions. This non-parametric statistical test compares the
876 cumulative distribution functions of the mean ALARM z-scores in both groups. The KS test statistic (D)
877 is the maximum vertical distance between the two distributions. The p-value of the test is the
878 probability of obtaining a test statistic as extreme as D or more extreme, assuming that the null
879 hypothesis is true. The null hypothesis is that the two samples are drawn from the same distribution.

880 Receptor-Ligand analysis
881 The 'iTalk' R package was used for the receptor-ligand (RL) analysis. The receptors were gene
882 candidates in single-cell kidney transplant stable and rejection patients. The *rawParse()* function with
883 `stats='mean'` was used to identify the candidate genes. For each celltype, the genes were ordered
884 by their average count expression. Only the top 50 percent of these genes were selected for the
885 subsequent RL analysis. The same criteria were adopted for selecting the ligand candidates from the
886 bulk RNAseq biopsy data. The significant interaction pairs were discovered from the iTalk database
887 restricted on the cytokine interactions only.
888

889 Analysis of publicly available scRNA-seq datasets

890 PBMCs of 45 healthy Volunteers (Van Der Wijst MG et al.):
891 Processed (de-anonymized) single-cell RNA-seq data and its relevant meta data was obtained from the
892 European Genome-phenome Archive (EGA) accession number EGAS00001002560. The available data
893 was merged to build a single Seurat object for downstream analysis. Azimuth reference for PBMCs was
894 used to annotate the cells.

895 LPS and Covid PBMC dataset (Stephenson et al.):
896 The processed data was downloaded from Array Express under accession number E-MTAB-10026. Only
897 individuals from the same batch containing the LPS-treated volunteers were selected to mitigate batch
898 effects. Fig 5B shows only those individuals where the major cell types were available. The Covid
899 patients originally annotated as Mild/Moderate was included as 'Moderate' in Fig 5D.

900 Sepsis-PBMC Dataset (Reyes et al.):
901 The pre-processed scRNA-seq data was downloaded through the Broad Institute Single Cell Portal
902 (SCP548). The data was further analyzed with Seurat to obtain the Module score based on UCell.

903 Covid – BALF Dataset (Liao et al.):
904 The data was accessed from GSE145926. The filtered cell matrix was processed with Seurat with the
905 code as provided by the authors of the original article.

906 In vivo Mouse model experiments
907 For induction of pneumonia, E. coli (DH5α strain), OVA-coated E. coli, grown for 18 h in Luria broth
908 medium at 37 °C, were washed twice (1,000g, 10 min, 37 °C), diluted in sterile isotonic saline and
909 calibrated by nephelometry. Bacteria (75 µl, OD600 = 0.6–0.7) were injected i.t. in anesthetized mice
910 to induce nonlethal acute pneumonia. Infected mice were intravenously (i.v.) injected with 10 µg of

CD45-PE on days 0, 1, 3 and 5 to evaluate T cell trafficking towards the lung. Five minutes before sample collection on each of these days, 10 µg of CD45-BV480 was i.v. injected to evaluate blood contamination during lung excision. To specifically assess the expression of membrane markers CD3, CD4, CD8 CXCR4 and CD69 conjugated monoclonal antibodies were used on cell suspensions: CD3-bv711 (145-2C11, 7311597, BD Biosciences, 1:200 dilution); CD4-buv395 (GK1.5, 1097734, BD Biosciences, 1:200 dilution); CD8-AlexaFluor700 (RPA-T8, 9025745, BD Biosciences, 1:200 dilution); anti-CD69-APC (H1.2F3, 9204727, BD Biosciences). Two independent experiments with each 3-4 mice were carried out. Anova test were used to evaluate for statistical significance across time points.

Analysis of publicly available data set of 10 immune mediated diseases

Bulk RNAseq of 28 pure immune cell types from 339 individuals divided into 10 immune-mediated diseases and 92 healthy controls was obtained from the National Bioscience Database Center (NBDC) Human Database with the accession number E-GEAD-397. The dataset was assembled as a large matrix with genes listed in rows and columns are individuals with homogeneous cell types. Functions from the Seurat pipeline were used to compute the module scores and generate the population's UMAP embedding.

Disease Classification

Combined diseases vs healthy approach was deployed for each major cell type and by cell subtype at the primary level. The Logistic Regression classifier from the R package 'caret' and the repeated cross-validation strategy for model evaluation were used.

In the next phase, for each major cell type, classification was evaluated for Healthy vs One Disease. An identical classification model was built in this phase as well. The coefficient of variation (CV^2)[61] method produced an unsupervised set of highly variable genes as a control for the ALARM module genes. The mean vs (variance/$means^2$) was modelled with *glmgam.fit* from the *statmod* R package for the variance estimate of every gene. The genes were ranked by the significance of deviation from the fit. The same number of variable genes was then used in the modules.

Gene association to disease terms

The association between the ALARM genes and immune disease terms was performed using the disgenet2r (v0.99.2) R package [51]. The ratio of genes associated with immune diseases to the total number of genes within the modules was compared to the corresponding ratio obtained from 1000 randomly selected sets of 61 genes.

Methodological Clarifications and Editing

Parts of this manuscript were edited and refined with the assistance of ChatGPT, the AI language model developed by OpenAI. This tool was used to improve the clarity and coherence of the text without altering the scientific content.

Acknowledgements:

We are most grateful to the Genomics Core Facility **GenoA**, member of Biogenouest and France Genomique and to the Bioinformatics Core Facility **BiRD**, member of Biogenouest and Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013) for the use of their resources and their technical support. DS was supported by Progreffe foundation. The project was supported by Region Pays de Loire "Etoile Montantes" grant for JP. We would like to thank the members of the DIVAT consortium for their involvement in the study, the physicians who helped recruit patients, and all patients who participated in this study. We also thank the clinical research associates who participated in the data collection. Data were collected from the French DIVAT multicentric prospective cohort of kidney and/or pancreatic transplant recipients (www.divat.fr, N° CNIL 914184, ClinicalTrials.gov recording: NCT02900040). We also thank the biological resource centre for biobanking (CHU Nantes, Nantes

Université, Centre de ressources biologiques (BB-0033-00040), F-44000 Nantes, France). The authors have no conflicting financial interests.

Contributions:

TL, DS & JP analyzed the data and wrote the manuscript. CF and TL performed scRNA-seq experiments. TL and SV, JB and GB performed pig transplantation model experiments. LBD performed cell sorting experiments. CK, RD, MG and SB collected and provided patient data and kidney transplantation PBMC samples. GT & ND performed and analyzed all transwell experiments. AB & AR performed and analyzed mouse pneumonia experiments and critically revised the manuscript. RJ, SB provided samples, contributed to study design, result interpretation and critically revised the manuscript. All authors contributed to the draft of the manuscript. JP conceived the study and oversaw experiments and analyses.

Data availability

The data underlying Figures 1 and 2 will be openly available in GEO and embargo will be lifted upon acceptance.

Figure legends:

Figure 1. Comprehensive longitudinal single-cell RNA-sequencing of circulating immune cells in a cohort of kidney allograft recipient.

A) Timeline of the blood sampling points post-transplantation for the patients followed longitudinally. STA=Stable patient (n=3), TCMR=T cell mediated rejection (n=3), ABMR=Antibody-mediated rejection (n=3). Tolerant patients are not shown.

B) Schematic diagram of the scRNA-seq preparation workflow using cell hashing. Peripheral blood mononuclear cells (PBMCs) were collected from stable, ABMR and TCMR patients and then stained with one different oligo-conjugated antibody before being pooled and processed using microfluidic encapsulation.

C) UMAP dimensional-reduction embedding of the integrated samples (n=30). Each colour represents a different cell subpopulation, adapted and manually curated from the automatic Azimuth annotation.

D) UMAP projection showing the disease state distribution, TOL=Tolerant patients.

E) UMAP projection coloured according to experimental batch of origin

Figure 2. ALARM Module identification

A) Schematic workflow of the gene co-expression analysis using consensus nonnegative matrix factorisation (cNMF) module detection separate in each batch. The module selection was then refined based on overlapping genes between the three batches.

B) Clustering of 9 modules across the 3 batches, using the Jaccard distance.

C) Heatmap showing the combined gene expression of each module, i.e. module scores of the 9 modules summarized for each distinct cell-type.

D) UMAP projection of the expression of Module 9 across all cells.

E) Regression analysis of the 9 modules by cell types. The outcome variable (y) was time (T0, T1 and T2) and the independent variable (x) was the module score. Heatmap shows the beta values (trend) for each cell-type and disease state. Negative values correspond to a decrease of the module score across the three time points, positive values to an increase.

F) Super violin plots[62] showing the longitudinal trend of ALARM module in NK, CD4 T cells (top), CD8 T cells and B cells (bottom), stratified by individuals from each batch.

G) Gene Ontology analysis of biological process (BP) and molecular function (MF) identified by comparing the ALARM gene enrichment using the 4000 most variable genes as background (See Supplemental file 1).

1010
1011 **Figure 3. Pig model scRNA-seq analysis**
1012 A) Schematic diagram of the acute kidney allograft rejection in pig model. Recipient pig with a
1013 unilateral nephrectomy received a kidney graft from a second healthy pig. Kidney biopsies and PBMCs
1014 were collected daily and observed in immunohistochemistry. PBMCs were prepared for a scRNA-seq
1015 analysis.
1016 B) Immunohistochemistry of pig biopsies stained with Periodic Aid Schiff (PAS), to stain
1017 polysaccharides, glycoproteins and glycolipids at day 0, day 2, day 4 and day 6
1018 C) Quantification of cell populations in the kidney graft at given days using Image J cell counting
1019 software (See supplemental figure 3 for details)
1020 D) UMAP dimensional-reduction projection of the circulating immune cell types (PBMCs) scRNA-seq
1021 after PBMC isolation at D0, D2, D4 and D6.
1022 E) Proportions of circulating immune cells (PBMC) across the different time points (D0-D6) in the
1023 recipient transplanted pig.
1024 F) Violin plot of the ALARM module score in B cells, monocytes, CD4 T cells, CD8 T cells and NK cells
1025 throughout acute rejection in the kidney tissues (Day 0, Day2, Day 4 Day6). *$P < 0.05$, **$P < 0.01$, ***P
1026 < 0.001, ****$P < 0.0001$.
1027
1028 **Figure 4. ALARM gene expression increases in kidney tissues during rejection**
1029 Left panels from Reeve et al., 2017 and right panels from Callemeyn et al 2020. STA =Stable patient,
1030 ABMR=Antibody-mediated rejection, TCMR=T cell mediated rejection, Mixed=Graft undergoing ABMR
1031 and TCMR, DSA- = Donor specific antibody negative and DSA+= Donor specific antibodies positive.
1032 A) Violin plot of the sum of z-scores of ALARM genes across conditions. Wilcoxon P-values are shown
1033 in panel comparing Stable (STA) to rejection status (ABMR, Mixed or TCMR and DSA- and DSA+ ABMR).
1034 The mean comparison p-values were computed using the Wilcoxon rank-sum test. B) Heatmap
1035 showing z-scores of ALARM genes (one gene per row) in all graft biopsies. Phenotype denotes the
1036 disease states. Patients are sorted on the mean of the module gene z-scores.
1037 C) Cumulative distribution of the mean of z-scores of ALARM genes comparing stable vs the combined
1038 rejection conditions. K-S = Kolmogorov -Smirnov P-values and distance.
1039 D) Ligand-Receptor analysis between receptors genes identified in circulating immune cells and
1040 cytokines genes expressed in the allograft kidney tissue under no rejection and rejection condition.
1041 Width of the arrow line is proportional to the expression of the ligands and the width of the arrowhead
1042 is proportional to the receptor. Only the top 5 associations from each cell type of differentially
1043 expressed receptors (PBMC scRNA-seq) and cytokines (Biopsy microarray) are shown.
1044 E) Boxplots showing CXCL12 expression in biopsies from patients with high ALARM module expression
1045 (>75%), medium (25-75%) and low (<25%). Wilcoxon rank-sum test was used to calculate P-values
1046 shown above.
1047
1048 **Figure 5: Analysis of ALARM Module Expression and T Cell Behavior in an In Vitro Transwell Assay**
1049 A) Schematic representation of the in vitro transwell assay used to study CXCL12-mediated recruitment
1050 of T cells. The assay involved a membrane covered with human dermal microvascular endothelial cells
1051 (HDMEC), allowing for the comparison between unstimulated cells, cells in direct contact with CXCL12,
1052 and those migrating across the membrane in response to a CXCL12 gradient.
1053 B) Quantification of T cell subsets following migration. The composition of naïve, central memory (CM),
1054 effector memory RA (TEMRA), and effector memory (EM) T cells remained similar post-migration and
1055 in response to CXCL12 alone, indicating that all T cell subsets are attracted to and migrate in response
1056 to the CXCL12 gradient.
1057 C) Analysis of T cell surface marker expression. CD69, a component of the ALARM module, showed a
1058 slight increase in response to CXCL12, with a more significant upregulation in migrated cells. CD25,
1059 CD49A, and CD95 levels remained unchanged.
1060 D) Functional analysis of migrated T cells. CXCL12-exposed and transmigrated cells were stimulated
1061 with PMA/ionomycin after 24 hours. Migrated cells showed significantly increased expression of IFN-

γ in both CD4+ and CD8+ T cells, while TNF-α, granzyme B (GZMB), and granulysin (GNLY) expression remained constant.
Statistical significance was determined using one-way ANOVA, followed by post-hoc tests where appropriate. Significance levels are indicated as follows: *< 0.05, ** < 0.01, and *** < 0.001.

Figure 6: Single Cell Characterization of T Cell Behavior in an In Vitro Transwell Assay
A) Schematic representation of the transwell assay used to investigate CXCL12-mediated signaling and its effect on T cell migration. Three distinct conditions were assessed: (1) CXCL12-: T cells were placed below the transwell membrane without exposure to CXCL12, (2) CXCL12+: T cells were placed below the transwell membrane in direct contact with CXCL12, and (3) Migrated: T cells were placed above the membrane and collected from below after migrating in the presence of CXCL12.
B) UMAP visualizations displaying the clustering of naïve CD4 and CD8 T cells based on gene expression. Clusters were identified based on differences in ribosomal gene expression and CD4 versus CD8 annotation (top left & right). Annotation according to condition (bottom left) shows a notable separation between CXCL12- T cells and those exposed to CXCL12 or that had transmigrated. The ALARM module shows lower expression in the CXCL12- condition compared to the CXCL12+ and Migrated groups (bottom right).
C) Dot plot showing the expression levels of specific genes across the different conditions (Migrated, CXCL12+, CXCL12-).
D) Violin plots depicting the expression distribution of gene groups in pooled CD4 and CD8 T cells across the three conditions. The plots demonstrate changes in gene expression associated with immune response, migration, cytoskeleton organization, and stress response, as well as the downregulation of genes related to structural organization and gene regulation.

Figure 7. ALARM Module associates with a shift of immune cells metabolic fonctions
A) Metabolome pathway analysis of CD4+ T cells, using 20 pseudobulk sample for each experimental condition (Migrated, CXCL12-positive, and CXCL12-negative). The differential activity of metabolic reactions was evaluated by comparing the mean of Migrated samples and CXCL12- samples
B) PCA visualization of the pseudobulk samples by condition based on the metabolome pathway analysis scores
C) Volcano plot illustrating glycolytic enzyme changes in migrated T cells compared to controls. The x-axis shows Cohen's d effect sizes, and the y-axis indicates the statistical significance (−log10 p-values). Key enzymes in the glycolysis/gluconeogenesis pathway are labeled.
D) Flow cytometry analysis and quantification of GLUT1 and 2-NBDG uptake in T cells before and after transmigration. Left panels show representative flow cytometry plots, and right panels show paired comparisons for each parameter. LDHA expression levels (measured by MFI) are also shown before and after transmigration. Statistical significance is indicated, with p-values provided for each comparison.

Figure 8. ALARM gene expression is altered in distinct immune conditions
A) Outline of study on Lipopolysaccharide (LPS) intravenous injection (iv) scRNAseq experiment performed on healthy volunteers obtained from Stephenson et al. The annotated expression data was used to compute the ALARM module score across time points after LPS injection in healthy volunteers. Violin plots show the module score across cell types and condition. Total number of cells are mentioned below the plot. P-values are shown above the violin plots and were calculated using Wilcoxon rank sum test comparing each cell type between healthy and LPS conditions.
B) Violin plots of ALARM module score for three individuals which had matching timepoints across all cell types in the condition LPS 90 min and LPS 10h. Lines connect the median ALARM module score across time points for each individual separately. P-values are shown above the violin plots and were calculated using Wilcoxon rank sum test comparing the two distinct time points.
C) Outline of patients with sepsis and urinary tract infection obtained from Reyes et al,. Annotated data was used to compute ALARM module score across the different cell-types in urinary tract infection

1114 (UTI) patients and bacteraemia sepsis patients. Number of cells and individuals used are shown below
1115 the violin plots. P-values are shown above the violin plots and were calculated using Wilcoxon rank
1116 sum test comparing each cell type between healthy and LPS conditions.
1117 D) Outline of PBMC scRNA-seq generated on healthy individuals, moderate covid-19 and severe covid-
1118 19 patients (Stephenson et al.,) and Bronchoalveolar lavage fluid (BALF) from a distinct cohort (Liao et
1119 al.). Upper panel shows median ALARM module score for each cell-type for PBMC (blue) and BALF
1120 (red). Lower panel shows for each cell-type the module score for each corresponding cell-type specific
1121 module.
1122
1123 **Figure 9 Recruitment and Characterization of CXCR4+ T Cells During Lung Infection in Vivo**
1124 A) Schematic representation of the experimental setup used to study T cell recruitment during *E. coli*
1125 pneumonia in a mouse model. The infection was induced on day 0, with peak infection occurring
1126 between days 1 and 3, and resolution by day 7. CD45-PE was intravenously injected at different time
1127 points (days 0, 1, 3, and 5) to stain and track blood-derived immune cells migrating into the lungs
1128 during infection.
1129 B) Quantification of CD45+ T cells in the lungs at various time points post-infection. The left plot shows
1130 the number of CD45+ T cells at each time point. The accompanying pie charts depict the proportion of
1131 CXCR4+ and CXCR4- T cells over the course of the infection.
1132 C) Percentage of CD45+ cells within the CXCR4+ CD69+ T cell population over time.
1133 D) Comparative analysis of CD45+ CXCR4+ CD4+ and CD8+ T cell counts at different time points.
1134 Independent experiments with 3-4 mice per condition were carried out, and statistical significance was
1135 determined using one-way ANOVA, followed by post-hoc tests where appropriate. Significance levels
1136 are indicated as follows: *< 0.05, ** < 0.01, and *** < 0.001.
1137
1138 **Figure 10 ALARM genes are implicated in and predictive of immune mediated diseases**
1139 A) UMAP of bulk transcriptomic study (Ota et al., 2020). Colours represent major cell types and each
1140 single point represents a bulk transcriptomic dataset. Cell sub types are annotated in the plot.
1141 B) The same UMAP showing expression of ALARM module expression across samples.
1142 C) Outline of classification approach used to test disease prediction
1143 D) Barchart showing the F-1 score of disease classification for ALARM, cell-type specific modules and
1144 coefficient of variation (CV^2) selected genes in each major cell type separately.
1145 E) Barchart of F1-scores computed for disease classification in each sub cell type separately for ALARM
1146 and CV^2 genes
1147 F) Radar chart showing the F1-score in each cell type for classification between each disease and
1148 healthy separately using ALARM genes in each major cell type.
1149 G) Barchart showing the top three enriched diseases in the ALARM genes using the DisGeNET curated
1150 database.
1151
1152 **Figure S1 ScRNA-seq integration of a cohort of kidney allograft recipients**
1153 A) Expression profiles of cell-specific markers distinguish the PBMC populations. Average expression is
1154 the log-normalized expression average of the cells by cell type, size of the dots is associated to the
1155 fraction of cells of the cluster in which the gene is detected.
1156 B) Cell type proportion by cluster in the overall PBMC population. Comparison include Stable patients,
1157 Rejection (ABMR+TCMR) and two public datasets: healthy volunteers from Stephenson et al, 2021, and
1158 healthy volunteers from Reyes et al, 2020. One-way ANOVA with Tukey's multiple comparisons post
1159 hoc test was performed, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.
1160 C) UMAP projection showing the sample distribution, ABMR=Humoral rejection, TCMR=Cellular
1161 rejection, STA=Stable, TOL=Tolerant. Second UMAP shows the timepoint distribution, T0=Time point
1162 0 (Graft), T1=Time point 1, T2=Time point 2.
1163 D) K-bet acceptance rate by cell types following the CCA batch integration. Complete k-bet acceptance
1164 rate was computed on the overall PBMC population.
1165

Figure S2 Assessing module distribution

A) Alluvial plot showing the shared origin of each module across batches. Each cell is assigned to the most enriched module they express.

B) Boxplots showing the distribution of Pearson correlation coefficient *R* of pairwise gene expression correlation by cell type across modules. A module of random genes was generated to compare with the nine other modules. Multiple gene pairs from each module are selected to compute the rho values using a subset of common cells within the cell type. 1000 repetitions of 50 pairs each yielded 50000 *R* values, as represented in every box. The subset of common cells changes with every repetition.

C) Boxplots showing the distribution of Pearson correlation coefficient *R* of pairwise gene expression correlation by cell type for Module 9.

Figure S3 Cellular characterization of pig PBMC by scRNA-seq analysis

A) Image J analysis of immuno-stained slices to quantify cell infiltration. Red boxes show counting areas which excludes glomeruli. Right panel shows the cells which were then counted using Image J software.

B) Pig data were annotated using the Sscrofa reference genome and associated to their human homologue. The expression profiles of cell type-specific markers is shown in the dotplot of relative expression by cell type. Size of the dots are associated to the fraction of cells of the cluster in which the gene is detected. The UMAP embeds the 4 samples (D0, D2, D4, D6) with their corresponding cell type annotation.

Fig S4 Transwell assay experiments

A) Gating strategy used in the transwell assay for cell type annotation. Lymphocytes were gated based on forward scatter (FSC) and side scatter (SSC), followed by singlet selection and the identification of CD3+ T cells. CD4+ and CD8+ T cells were further delineated based on CD45RA and CCR7 expression to distinguish between naïve (Tn), central memory (Tcm), and effector memory RA (TemRA) subsets. Additionally, CD4 T regulatory (Treg) cells were identified by CD25 and CD127 expression to differentiate between activated (aTreg) and non-Treg populations.

B) Percentage of CD4+ and CD8+ T cells expressing CD69 under five different experimental conditions: before the transwell assay (baseline), CXCL12 only, HDMEC alone, CXCL12 with HDMEC, and after migration through HDMEC with CXCL12.

C) UMAP visualization of scRNA-seq data from the transwell experiment showing all cell populations. The top plot displays T cell subtypes identified by their transcriptional profiles, including CD4+ and CD8+ T naïve cells, CD8+ T effector memory RA (TEMRA) cells, mucosal-associated invariant T (MAIT) cells, and CD4+ and CD8+ T effector memory (TEM) cells annotated using Celltypist. The bottom plot shows the distribution of cells based on their experimental condition: control which is CXCL12-, CXCL12+ and transmigrated. The clustering of cells in the UMAP space indicates differences in transcriptional states based on both cell type and condition.

D) ALARM gene expression by condition. Dot plot showing the relative expression levels of ALARM-associated genes in CD4+ and CD8+ T cells across the different conditions of the transwell assay (Migrated, CXCL12+, and CXCL12-). The size of the dots represents the percentage of cells expressing the gene, and the color intensity represents the level of expression, with darker shades indicating higher expression.

Fig 5S. Overview of the cohort distribution

A) ALARM module score distribution by cell type based on scRNA-seq analysis of PBMC from 45 unrelated healthy individuals[40].

B) Left: ALARM module score distribution across sex. No significance was found for the sex parameter with a t-test. Right: ALARM module score distribution across age categories. No significance was found for the age parameter with a one-way ANOVA test.

Fig 6S. ALARM genes can classify immune-mediated diseases

1217 A) Distribution of the bulk RNA-seq samples across the 28 circulating immune cell types for the 337
1218 patients distributed across 10 immune mediated diseases and the 79 healthy controls.
1219 B) Area Under Curve (AUC) and Cohen's Kappa score across cell types in disease and healthy patients.
1220 C) AUC and Cohen's Kappa score across cell subtypes in disease and healthy patients
1221 D) Distribution of the ratio of genes associated to Gene-Disease Association (GDA) terms in 1000
1222 modules of 61 randomly selected genes. The ALARM module has a ratio of 0.39 genes associated to
1223 GDA terms (red line).
1224
1225
1226 **References:**
1227
1228

1229 1. Reyes, M. *et al.* An immune cell signature of bacterial sepsis. *Nat. Med.* **26**, 333–340 (2020).

1230 2. Perez, R. K. *et al.* Single-cell RNA-seq reveals cell type-specific molecular and genetic associations
1231 to lupus. *Science* **376**, eabf1970 (2022).

1232 3. Kazer, S. W. *et al.* Integrated single-cell analysis of multicellular immune dynamics during
1233 hyperacute HIV-1 infection. *Nat. Med.* **26**, 511–518 (2020).

1234 4. Zhang, F. *et al.* Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by
1235 integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* **20**, 928–942 (2019).

1236 5. Stephenson, E. *et al.* Single-cell multi-omics analysis of the immune response in COVID-19. *Nat.*
1237 *Med.* **27**, 904–916 (2021).

1238 6. Bernardes, J. P. *et al.* Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes,
1239 Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. *Immunity* **53**, 1296-1314.e9
1240 (2020).

1241 7. Chaumette, T. *et al.* Monocyte Signature Associated with Herpes Simplex Virus Reactivation and
1242 Neurological Recovery After Brain Injury. *Am. J. Respir. Crit. Care Med.* (2022)
1243 doi:10.1164/rccm.202110-2324OC.

1244 8. del Rosario, R. C. H. *et al.* Histone acetylome-wide associations in immune cells from individuals
1245 with active Mycobacterium tuberculosis infection. *Nat. Microbiol.* **7**, 312–326 (2022).

1246 9. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity
1247 with single-cell RNA-Seq. *eLife* **8**, e43803 (2019).

1248 10. Loupy, A., Mengel, M. & Haas, M. Thirty years of the International Banff Classification for
1249 Allograft Pathology: the past, present, and future of kidney transplant diagnostics. *Kidney Int.* **101**,
1250 678–691 (2022).
1251 11. Sirota, M. & Sarwal, M. M. Transplantomics: Toward Precision Medicine in Transplantation
1252 Research. *Transplantation* **101**, 1777–1782 (2017).
1253 12. Moreau, A., Varey, E., Anegon, I. & Cuturi, M.-C. Effector Mechanisms of Rejection. *Cold*
1254 *Spring Harb. Perspect. Med.* **3**, a015461 (2013).
1255 13. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet
1256 detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
1257 14. Stuart, T. *et al.* Comprehensive integration of single-cell data. *Cell* **177**, 1888-1902.e21
1258 (2019).
1259 15. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29
1260 (2021).
1261 16. Büttner, M., Miao, Z., Wolf, F. A., Teichmann, S. A. & Theis, F. J. A test metric for assessing
1262 single-cell RNA-seq batch correction. *Nat. Methods* **16**, 43–49 (2019).
1263 17. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis.
1264 *BMC Bioinformatics* **9**, 559 (2008).
1265 18. Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and
1266 progenitor cell differentiation. *Blood* **128**, e20–e31 (2016).
1267 19. Yamashita, I., Nagata, T., Tada, T. & Nakayama, T. CD69 cell surface expression identifies
1268 developing thymocytes which audition for T cell antigen receptor-mediated positive selection. *Int.*
1269 *Immunol.* **5**, 1139–1150 (1993).
1270 20. Ziegler, S. F., Ramsdell, F. & Alderson, M. R. The activation antigen CD69. *Stem Cells* **12**, 456–
1271 465 (1994).

1272 21. Werfel, T., Boeker, M. & Kapp, A. Rapid expression of the CD69 antigen on T cells and natural
1273 killer cells upon antigenic stimulation of peripheral blood mononuclear cell suspensions. *Allergy*
1274 **52**, 465–469 (1997).
1275 22. Turner, J.-E., Becker, M., Mittrücker, H.-W. & Panzer, U. Tissue-Resident Lymphocytes in the
1276 Kidney. *J. Am. Soc. Nephrol. JASN* **29**, 389–399 (2018).
1277 23. López-Cabrera, M. *et al.* Transcriptional Regulation of the Gene Encoding the Human C-type
1278 Lectin Leukocyte Receptor AIM/CD69 and Functional Characterization of Its Tumor Necrosis
1279 Factor-α-responsive Elements (∗). *J. Biol. Chem.* **270**, 21545–21551 (1995).
1280 24. Castellanos, M. C. *et al.* Expression of the leukocyte early activation antigen CD69 is
1281 regulated by the transcription factor AP-1. *J. Immunol. Baltim. Md 1950* **159**, 5463–5473 (1997).
1282 25. Tögel, F., Isaac, J., Hu, Z., Weiss, K. & Westenfelder, C. Renal SDF-1 signals mobilization and
1283 homing of CXCR4-positive cells to the kidney after ischemic injury. *Kidney Int.* **67**, 1772–1784
1284 (2005).
1285 26. Song, A., Jiang, A., Xiong, W. & Zhang, C. The Role of CXCL12 in Kidney Diseases: A Friend or
1286 Foe? *Kidney Dis.* **7**, 167–176 (2021).
1287 27. Hoffmann, U. *et al.* SDF-1 expression is elevated in chronic human renal allograft rejection.
1288 *Clin. Transplant.* **20**, 712–718 (2006).
1289 28. Pabst, R. The pig as a model for immunology research. *Cell Tissue Res.* **380**, 287–304 (2020).
1290 29. Reeve, J. *et al.* Assessing rejection-related disease in kidney transplant biopsies based on
1291 archetypal analysis of molecular phenotypes. *JCI Insight* **2**, 94197 (2017).
1292 30. Callemeyn, J. *et al.* Transcriptional Changes in Kidney Allografts with Histology of Antibody-
1293 Mediated Rejection without Anti-HLA Donor-Specific Antibodies. *J. Am. Soc. Nephrol.* **31**, 2168–
1294 2183 (2020).
1295 31. Moll, N. M. & Ransohoff, R. M. CXCL12 and CXCR4 in bone marrow physiology. *Expert Rev.*
1296 *Hematol.* **3**, 315–322 (2010).

1297 32. Yellowley, C. CXCL12/CXCR4 signaling and other recruitment and homing pathways in
1298 fracture repair. *BoneKEy Rep.* **2**, 300 (2013).
1299 33. Burger, J. A. & Kipps, T. J. CXCR4: a key receptor in the crosstalk between tumor cells and
1300 their microenvironment. *Blood* **107**, 1761–1767 (2006).
1301 34. Cibrián, D. & Sánchez-Madrid, F. CD69: from activation marker to metabolic gatekeeper. *Eur.*
1302 *J. Immunol.* **47**, 946–953 (2017).
1303 35. Scarneo, S. A. *et al.* Expression of membrane Hsp90 is a molecular signature of T cell
1304 activation. *Sci. Rep.* **12**, 18091 (2022).
1305 36. Nieminen, M. *et al.* Vimentin function in lymphocyte adhesion and transcellular migration.
1306 *Nat. Cell Biol.* **8**, 156–162 (2006).
1307 37. Liu, Y. *et al.* Prognostic and Immunological Role of STK38 across Cancers: Friend or Foe? *Int. J.*
1308 *Mol. Sci.* **23**, 11590 (2022).
1309 38. Wagner, A. *et al.* Metabolic modeling of single Th17 cells reveals regulators of autoimmunity.
1310 *Cell* **184**, 4168-4185.e21 (2021).
1311 39. MacIver, N. J., Michalek, R. D. & Rathmell, J. C. Metabolic regulation of T lymphocytes. *Annu.*
1312 *Rev. Immunol.* **31**, 259–283 (2013).
1313 40. van der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs
1314 and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).
1315 41. Tigchelaar, E. F. *et al.* Cohort profile: LifeLines DEEP, a prospective, general population cohort
1316 study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* **5**,
1317 e006772 (2015).
1318 42. Vos, A. F. de *et al.* In Vivo Lipopolysaccharide Exposure of Human Blood Leukocytes Induces
1319 Cross-Tolerance to Multiple TLR Ligands. *J. Immunol.* **183**, 533–542 (2009).
1320 43. Andonegui, G., Goyert, S. M. & Kubes, P. Lipopolysaccharide-Induced Leukocyte-Endothelial
1321 Cell Interactions: A Role for CD14 Versus Toll-Like Receptor 4 Within Microvessels1. *J. Immunol.*
1322 **169**, 2111–2119 (2002).

1323 44. Beutler, B. TLR4 as the mammalian endotoxin sensor. *Curr. Top. Microbiol. Immunol.* **270**,
1324 109–120 (2002).
1325 45. Beutler, B., Du, X. & Poltorak, A. Identification of Toll-like receptor 4 (Tlr4) as the sole conduit
1326 for LPS signal transduction: genetic and evolutionary studies. *J. Endotoxin Res.* **7**, 277–280 (2001).
1327 46. Liao, M. *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with COVID-
1328 19. *Nat. Med.* **26**, 842–844 (2020).
1329 47. Roquilly, A. *et al.* Alveolar macrophages are epigenetically altered after inflammation, leading
1330 to long-term lung immunoparalysis. *Nat. Immunol.* **21**, 636–648 (2020).
1331 48. Broquet, A. *et al.* Sepsis-trained macrophages promote antitumoral tissue-resident T cells.
1332 *Nat. Immunol.* 1–18 (2024) doi:10.1038/s41590-024-01819-8.
1333 49. Mi, Z., Guo, B., Yin, Z., Li, J. & Zheng, Z. Disease classification via gene network integrating
1334 modules and pathways. *R. Soc. Open Sci.* **6**, 190214 (2019).
1335 50. Ota, M. *et al.* Dynamic landscape of immune cell-specific gene regulation in immune-
1336 mediated diseases. *Cell* **184**, 3006-3021.e17 (2021).
1337 51. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update.
1338 *Nucleic Acids Res.* **48**, D845–D855 (2020).
1339 52. Afeltra, A. *et al.* Expression of CD69 antigen on synovial fluid T cells in patients with
1340 rheumatoid arthritis and other chronic synovitis. *Ann. Rheum. Dis.* **52**, 457–460 (1993).
1341 53. Su, C. C., Shau, W. Y., Wang, C. R., Chuang, C. Y. & Chen, C. Y. CD69 to CD3 ratio of peripheral
1342 blood mononuclear cells as a marker to monitor systemic lupus erythematosus disease activity.
1343 *Lupus* **6**, 449–454 (1997).
1344 54. Polyfunctionality of the CXCR4/CXCL12 axis in health and disease: Implications for
1345 therapeutic interventions in cancer and immune-mediated diseases. doi:10.1096/fj.202001273R.
1346 55. Abidi, A. *et al.* Characterization of Rat ILCs Reveals ILC2 as the Dominant Intestinal Subset.
1347 *Front. Immunol.* **11**, (2020).

1348 56. Drouin, M. *et al.* CLEC-1 is a death sensor that limits antigen cross-presentation by dendritic
1349 cells and represents a target for cancer immunotherapy. *Sci. Adv.* **8**, eabo7621 (2022).
1350 57. Feregrino, C. & Tschopp, P. Assessing evolutionary and developmental transcriptome
1351 dynamics in homologous cell types. *Dev. Dyn.* **251**, 1472–1489 (2022).
1352 58. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring.
1353 *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
1354 59. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image
1355 analysis. *Nat. Methods* **9**, 671–675 (2012).
1356 60. Xu, C. *et al.* Automatic cell-type harmonization and integration across Human Cell Atlas
1357 datasets. *Cell* **186**, 5876-5891.e20 (2023).
1358 61. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat.*
1359 *Methods* **10**, 1093–1095 (2013).
1360 62. Kenny, M. & Schoen, I. Violin SuperPlots: visualizing replicate heterogeneity in large data
1361 sets. *Mol. Biol. Cell* **32**, 1333–1334 (2021).
1362

Figure 1

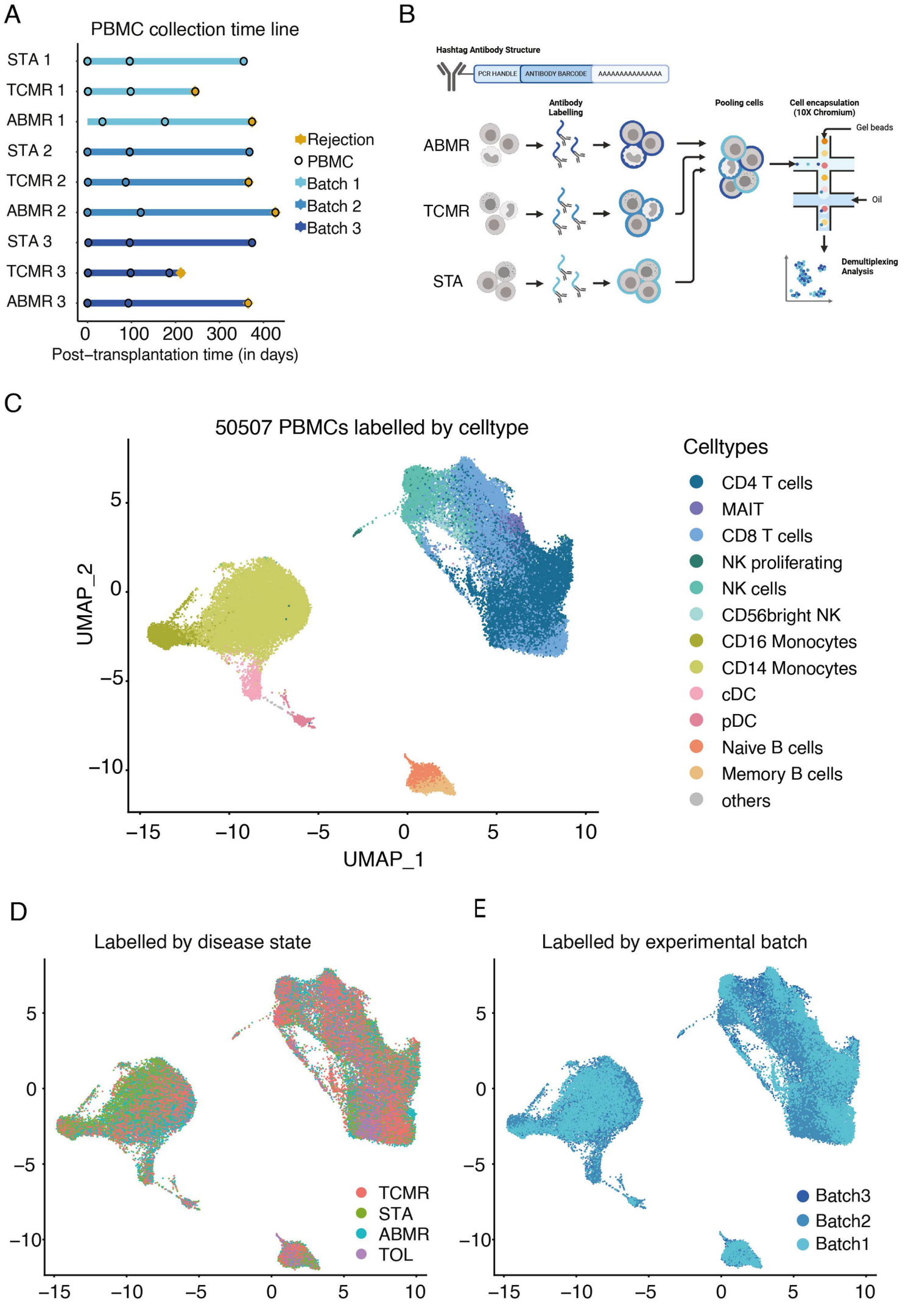

A
PBMC collection time line
STA 1
TCMR 1
ABMR 1
STA 2
TCMR 2
ABMR 2
STA 3
TCMR 3
ABMR 3
0
100
200
300
400
Post–transplantation time (in days)
Rejection
PBMC
Batch 1
Batch 2
Batch 3
B
Hashtag Antibody Structure
PCR HANDLE
ANTIBODY BARCODE
AAAAAAAAAAAAAA
Antibody Labelling
Pooling cells
Cell encapsulation (10X Chromium)
Gel beads
Oil
Demultiplexing Analysis
ABMR
TCMR
STA
C
50507 PBMCs labelled by celltype
UMAP_2
UMAP_1
Celltypes
CD4 T cells
MAIT
CD8 T cells
NK proliferating
NK cells
CD56bright NK
CD16 Monocytes
CD14 Monocytes
cDC
pDC
Naive B cells
Memory B cells
others
D
Labelled by disease state
TCMR
STA
ABMR
TOL
E
Labelled by experimental batch
Batch3
Batch2
Batch1

Figure 2

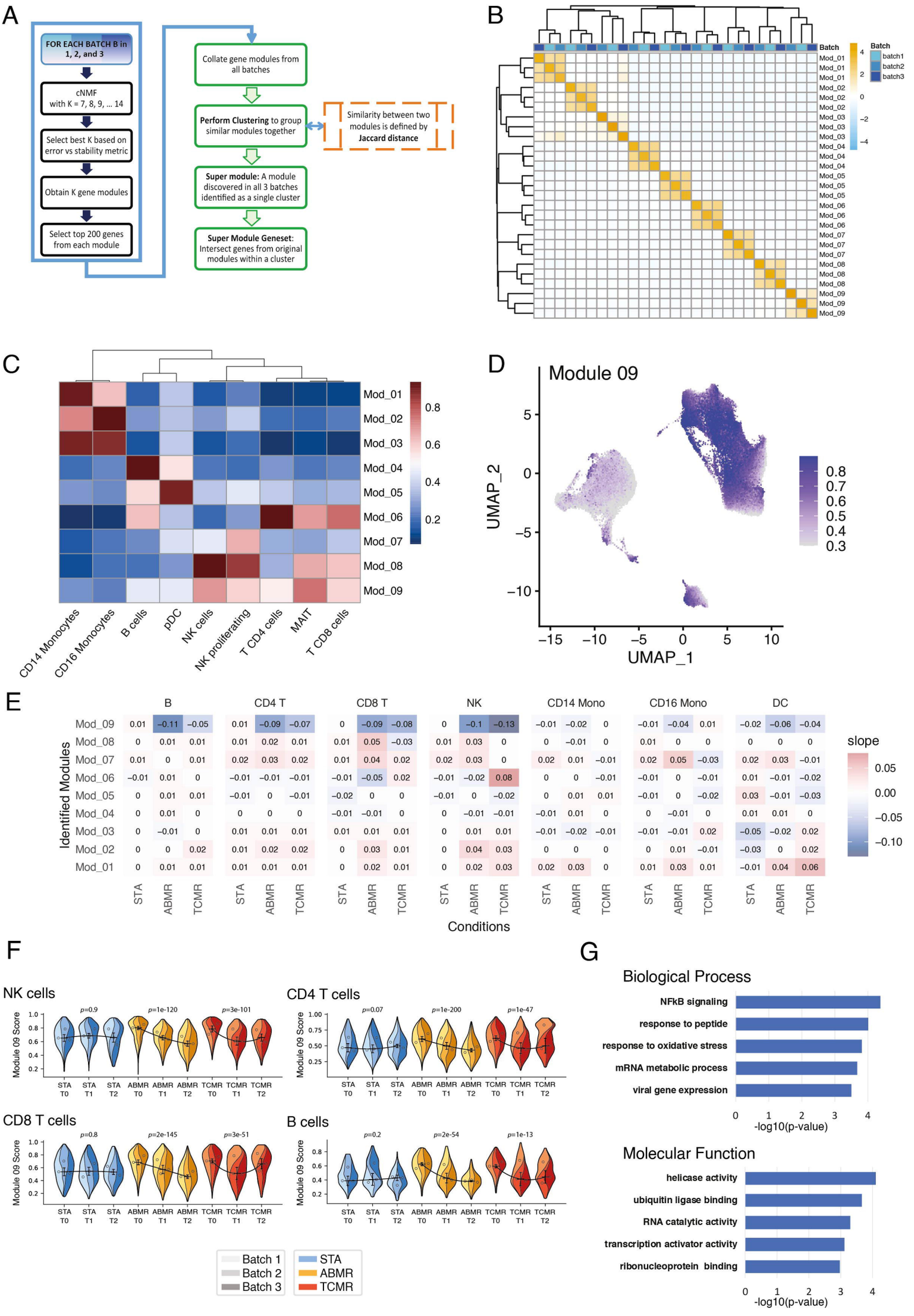

A
FOR EACH BATCH B in 1, 2, and 3
cNMF with K = 7, 8, 9, ... 14
Select best K based on error vs stability metric
Obtain K gene modules
Select top 200 genes from each module
Collate gene modules from all batches
Perform Clustering to group similar modules together
Similarity between two modules is defined by Jaccard distance
Super module: A module discovered in all 3 batches identified as a single cluster
Super Module Geneset: Intersect genes from original modules within a cluster
B
Batch
batch1
batch2
batch3
C
Mod_01
Mod_02
Mod_03
Mod_04
Mod_05
Mod_06
Mod_07
Mod_08
Mod_09
CD14 Monocytes
CD16 Monocytes
B cells
pDC
NK cells
NK proliferating
T CD4 cells
MAIT
T CD8 cells
D
Module 09
UMAP_2
UMAP_1
E
B
CD4 T
CD8 T
NK
CD14 Mono
CD16 Mono
DC
Identified Modules
slope
STA
ABMR
TCMR
Conditions
F
NK cells
CD4 T cells
CD8 T cells
B cells
Module 09 Score
Batch 1
Batch 2
Batch 3
STA
ABMR
TCMR
G
Biological Process
NFkB signaling
response to peptide
response to oxidative stress
mRNA metabolic process
viral gene expression
-log10(p-value)
Molecular Function
helicase activity
ubiquitin ligase binding
RNA catalytic activity
transcription activator activity
ribonucleoprotein binding
-log10(p-value)

Figure 3

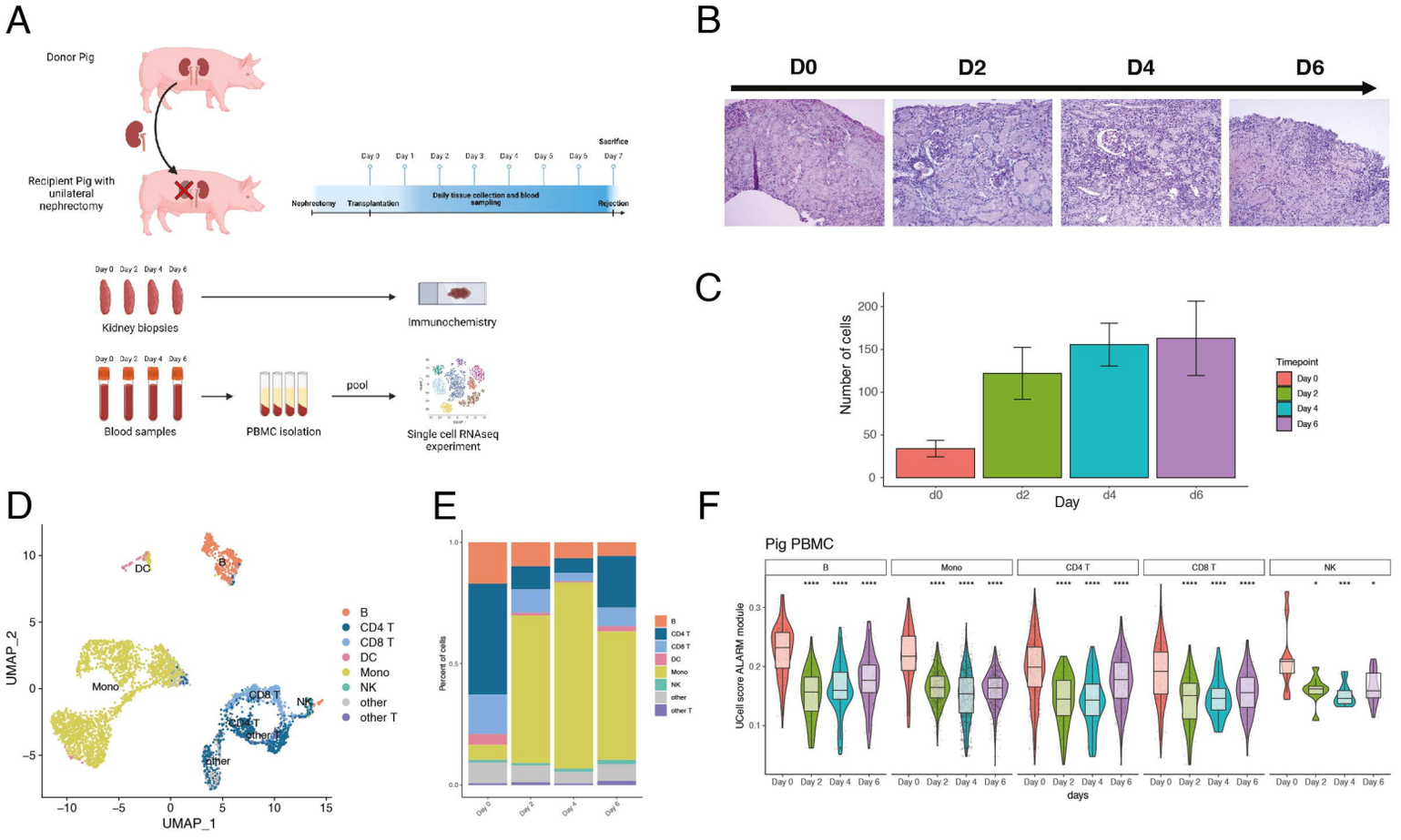

A
Donor Pig
Recipient Pig with unilateral nephrectomy
Day 0
Day 1
Day 2
Day 3
Day 4
Day 5
Day 6
Sacrifice
Day 7
Nephrectomy
Transplantation
Daily tissue collection and blood sampling
Rejection
Kidney biopsies
Immunochemistry
Blood samples
PBMC isolation
pool
Single cell RNAseq experiment
B
D0
D2
D4
D6
C
Number of cells
Day
d0
d2
d4
d6
Timepoint
Day 0
Day 2
Day 4
Day 6
D
UMAP_1
UMAP_2
DC
B
Mono
CD8 T
NK
CD4 T
other T
other
CD4 T
CD8 T
DC
Mono
NK
other
other T
E
Percent of cells
F
Pig PBMC
B
Mono
CD4 T
CD8 T
NK
UCell score ALARM module
days

Figure 4

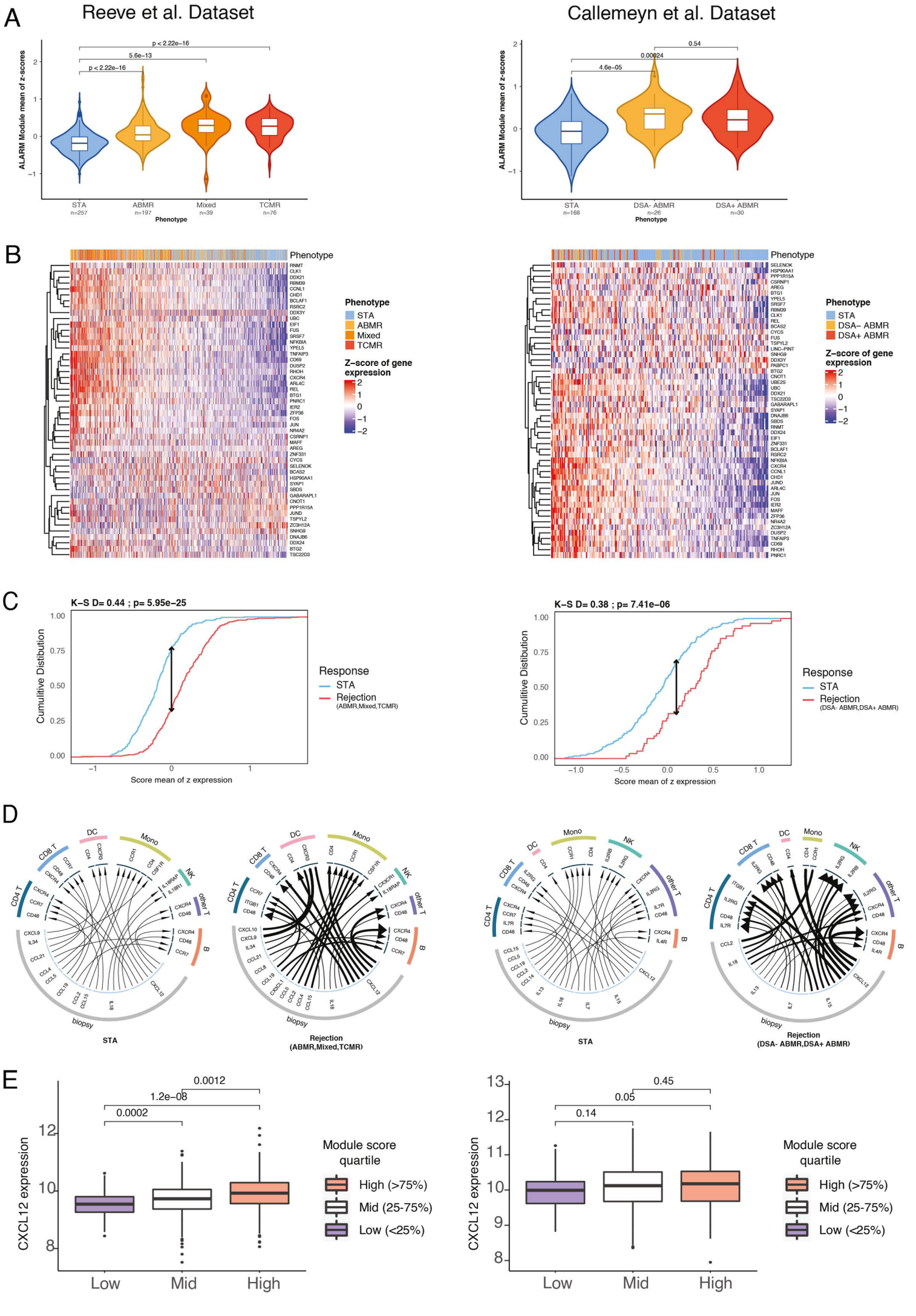

A
Reeve et al. Dataset
Callemeyn et al. Dataset
ALARM Module mean of z-scores
p < 2.22e−16
5.6e−13
4.6e−05
0.00024
0.54
STA n=257
ABMR n=197
Mixed n=39
TCMR n=76
STA n=168
DSA- ABMR n=26
DSA+ ABMR n=30
Phenotype
B
Phenotype
STA
ABMR
Mixed
TCMR
DSA− ABMR
DSA+ ABMR
Z−score of gene expression
C
K−S D= 0.44 ; p= 5.95e−25
K−S D= 0.38 ; p= 7.41e−06
Cumulitive Distibution
Score mean of z expression
Response
STA
Rejection (ABMR,Mixed,TCMR)
Rejection (DSA- ABMR,DSA+ ABMR)
D
DC
Mono
CD8 T
CD4 T
NK
other T
B
biopsy
STA
Rejection (ABMR,Mixed,TCMR)
Rejection (DSA- ABMR,DSA+ ABMR)
E
CXCL12 expression
0.0012
1.2e−08
0.0002
0.45
0.05
0.14
Low
Mid
High
Module score quartile
High (>75%)
Mid (25-75%)
Low (<25%)

Figure 5

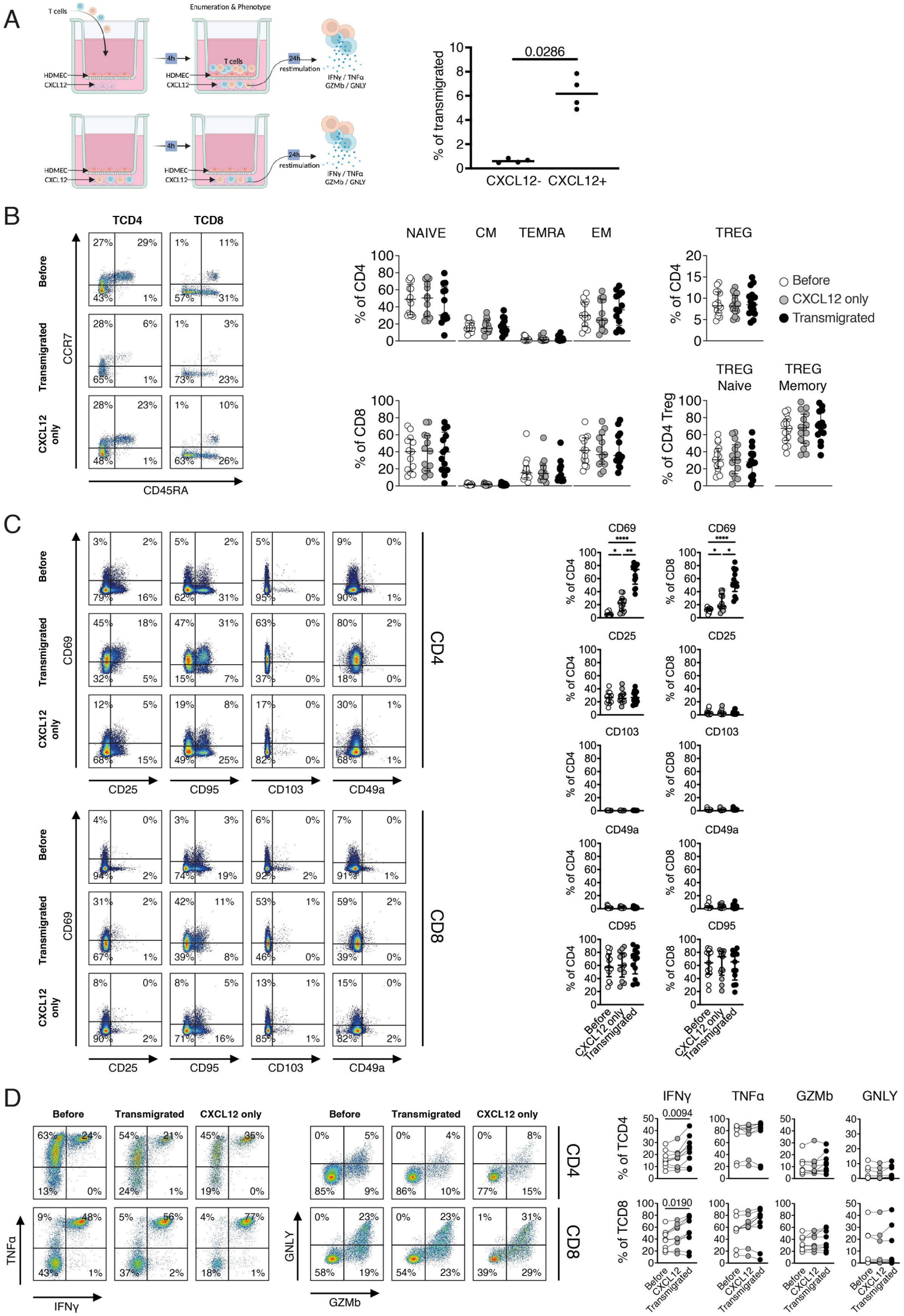
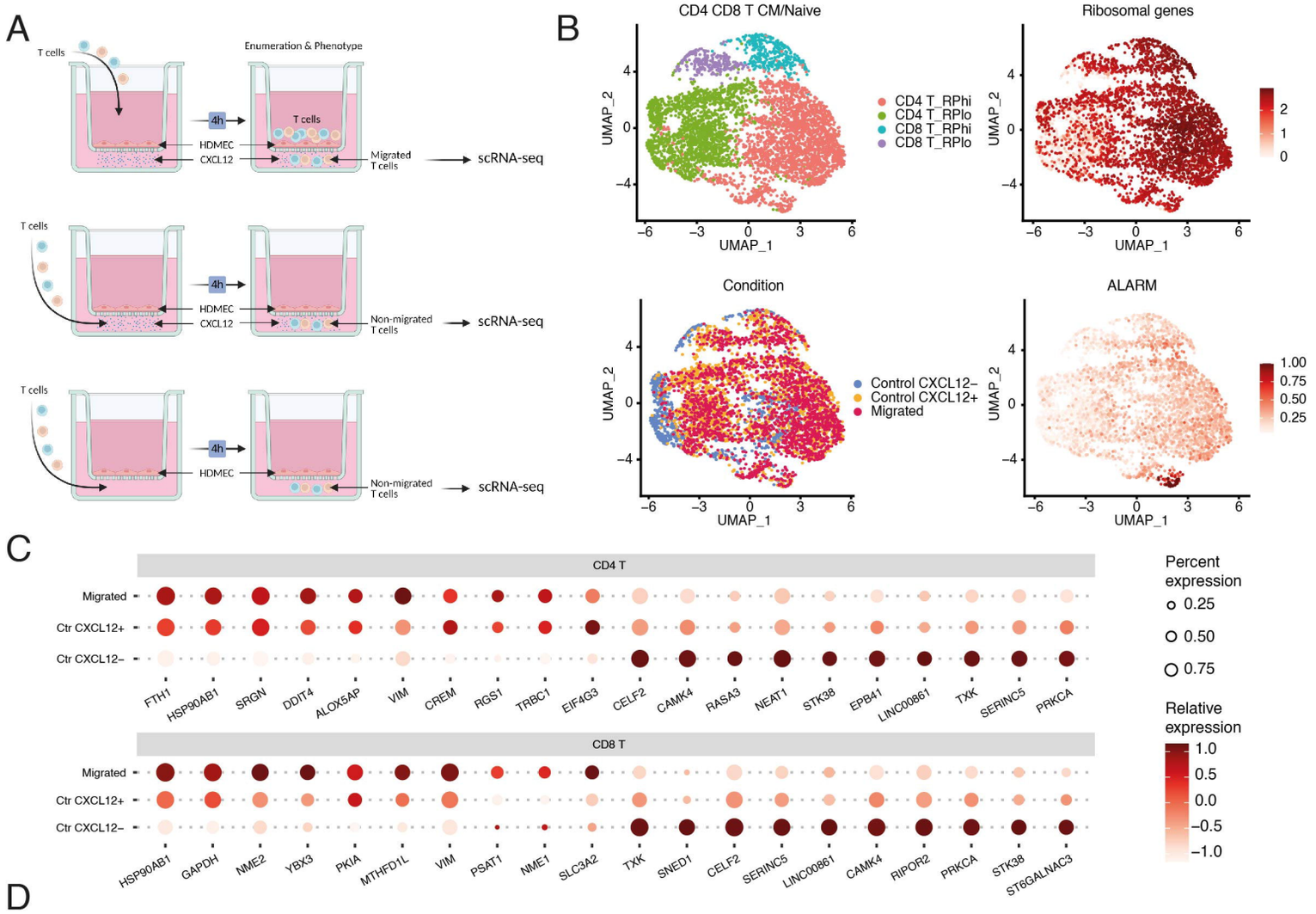

A
T cells
Enumeration & Phenotype
HDMEC
CXCL12
4h
24h restimulation
IFNγ / TNFα
GZMb / GNLY
0.0286
% of transmigrated
CXCL12- CXCL12+
B
TCD4
TCD8
Before
Transmigrated
CXCL12 only
CCR7
CD45RA
NAIVE
CM
TEMRA
EM
TREG
% of CD4
% of CD8
% of CD4 Treg
TREG Naive
TREG Memory
Before
CXCL12 only
Transmigrated
C
CD69
CD25
CD95
CD103
CD49a
CD4
CD8
D
Before
Transmigrated
CXCL12 only
TNFα
IFNγ
GNLY
GZMb
IFNγ
TNFα
GZMb
GNLY
% of TCD4
% of TCD8
0.0094
0.0190

Figure 6


A
T cells
Enumeration & Phenotype
4h
T cells
HDMEC
CXCL12
Migrated T cells
scRNA-seq
Non-migrated T cells
B
CD4 CD8 T CM/Naive
Ribosomal genes
Condition
ALARM
UMAP_1
UMAP_2
CD4 T_RPhi
CD4 T_RPlo
CD8 T_RPhi
CD8 T_RPlo
Control CXCL12−
Control CXCL12+
Migrated
C
CD4 T
CD8 T
Migrated
Ctr CXCL12+
Ctr CXCL12−
FTH1
HSP90AB1
SRGN
DDIT4
ALOX5AP
VIM
CREM
RGS1
TRBC1
EIF4G3
CELF2
CAMK4
RASA3
NEAT1
STK38
EPB41
LINC00861
TXK
SERINC5
PRKCA
GAPDH
NME2
YBX3
PKIA
MTHFD1L
PSAT1
NME1
SLC3A2
SNED1
RIPOR2
ST6GALNAC3
Percent expression
Relative expression

D

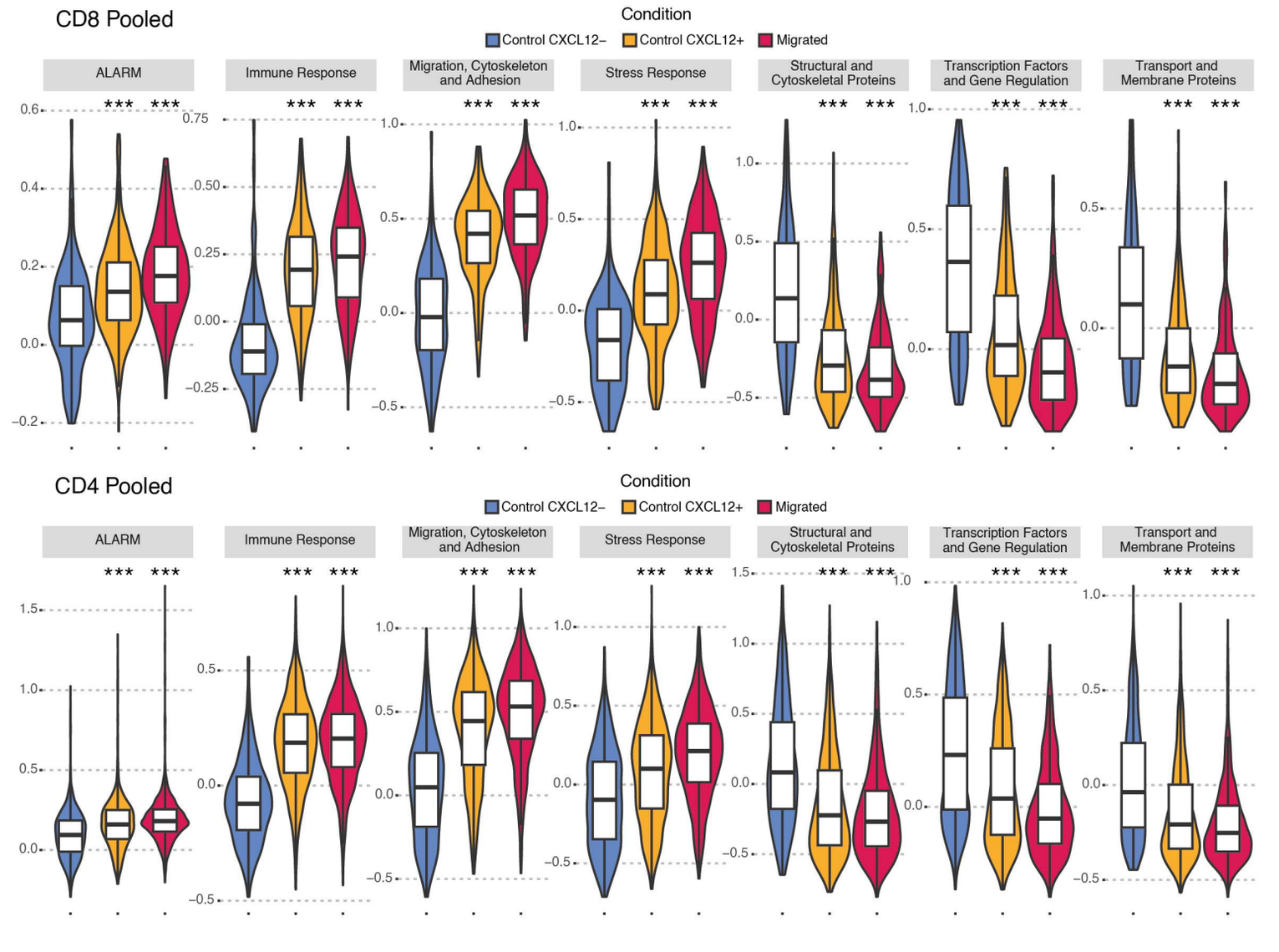
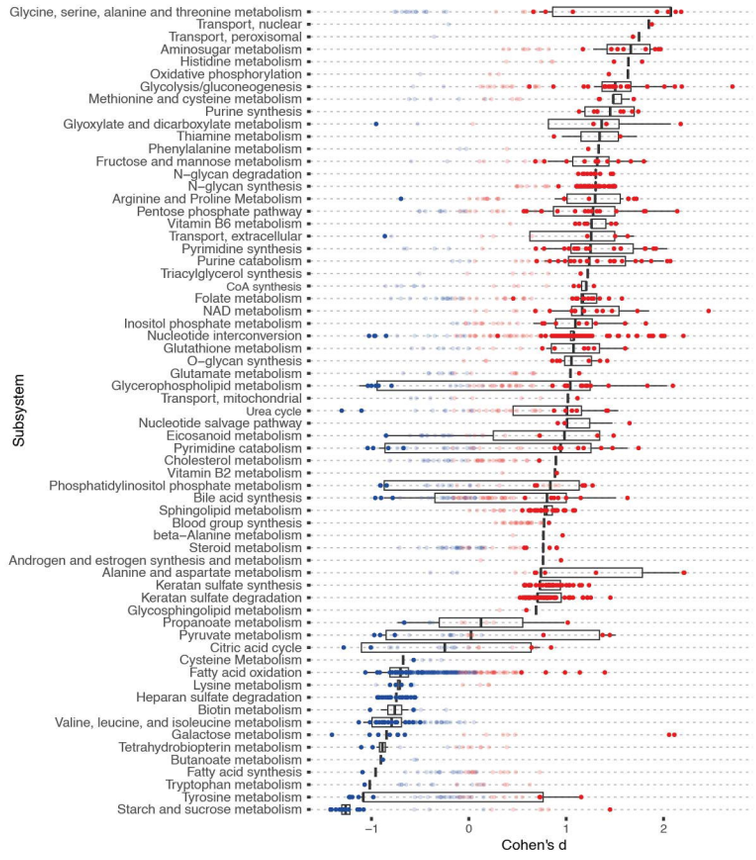
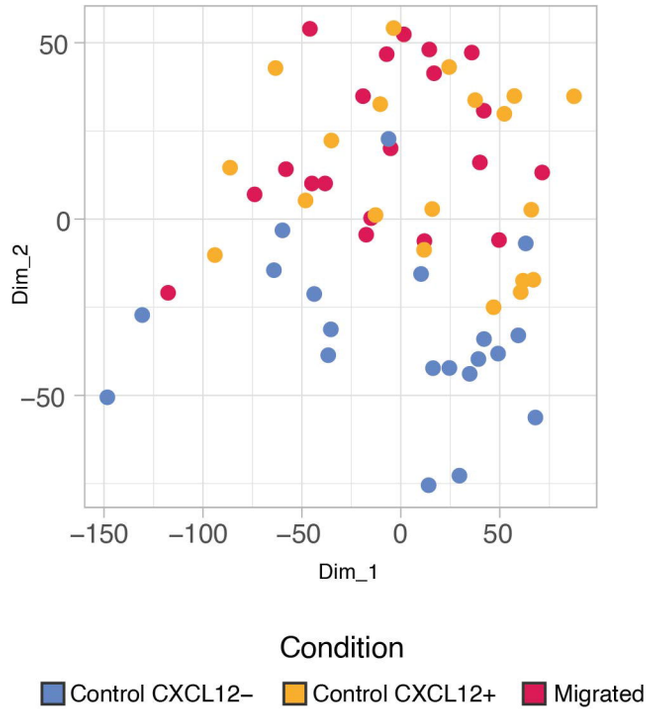

CD8 Pooled
CD4 Pooled
Condition
Control CXCL12−
Control CXCL12+
Migrated
ALARM
Immune Response
Migration, Cytoskeleton and Adhesion
Stress Response
Structural and Cytoskeletal Proteins
Transcription Factors and Gene Regulation
Transport and Membrane Proteins
***

Figure 7

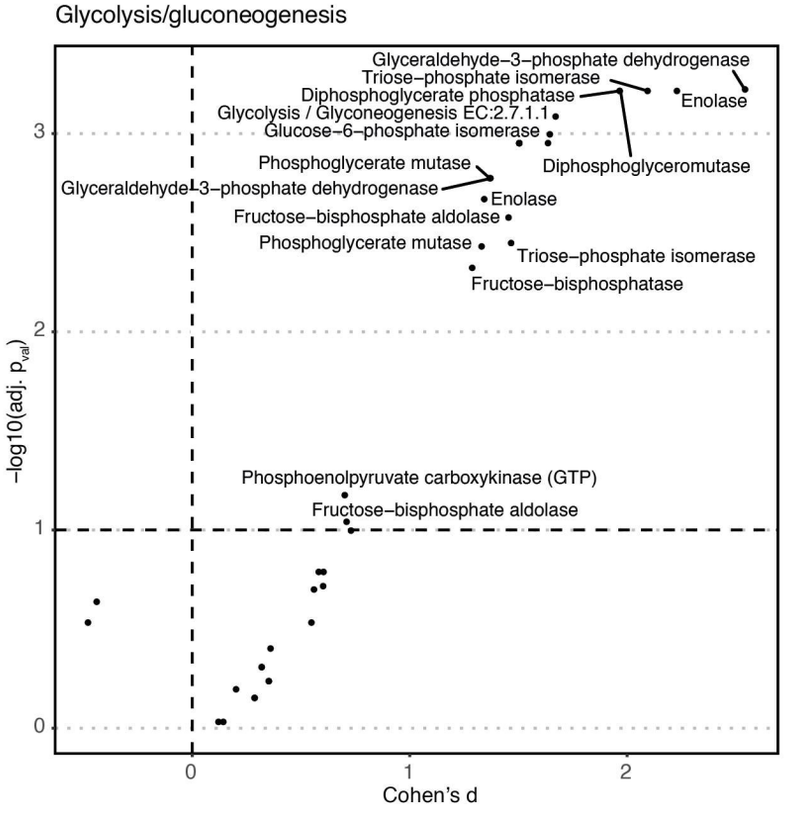
A



B



C



D

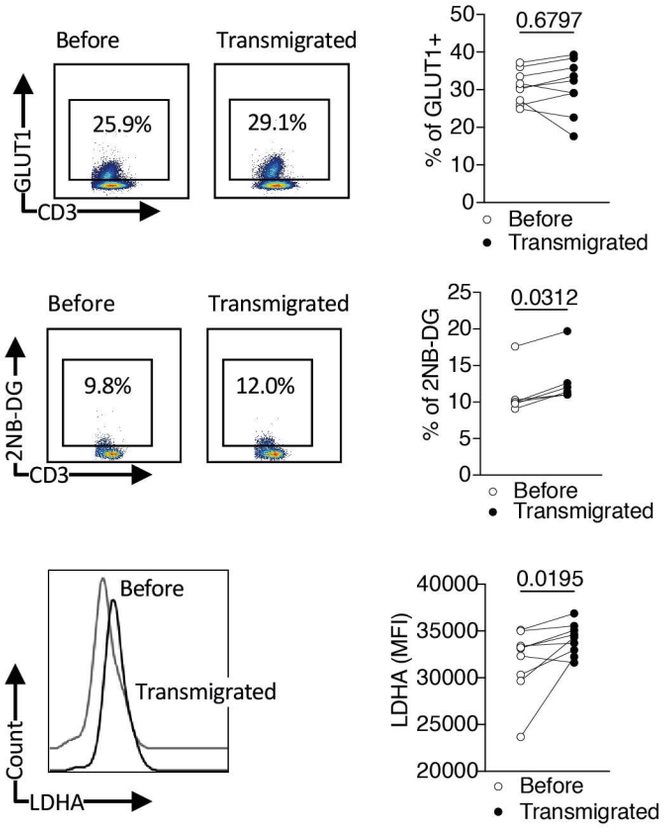
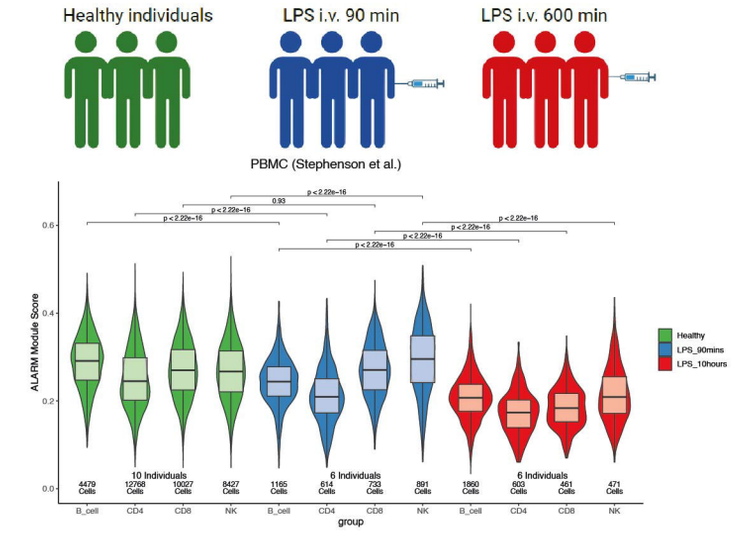
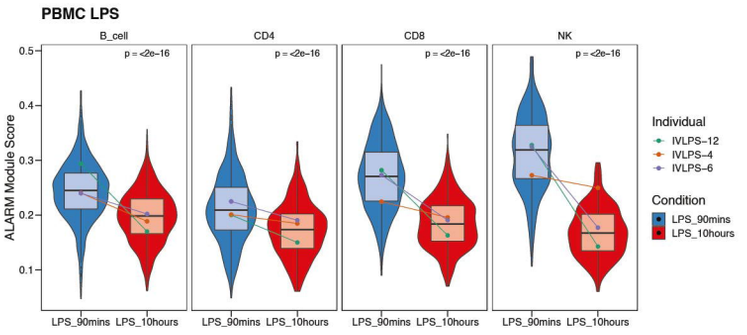


Figure 8

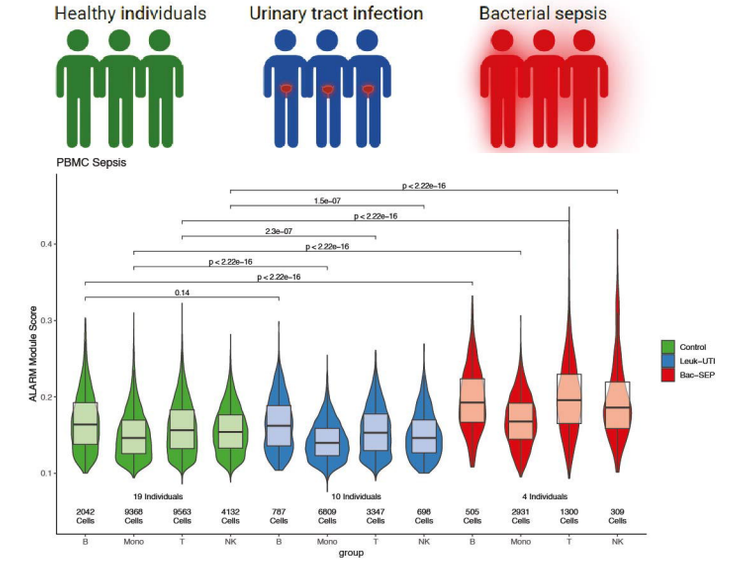
A



B



C



D

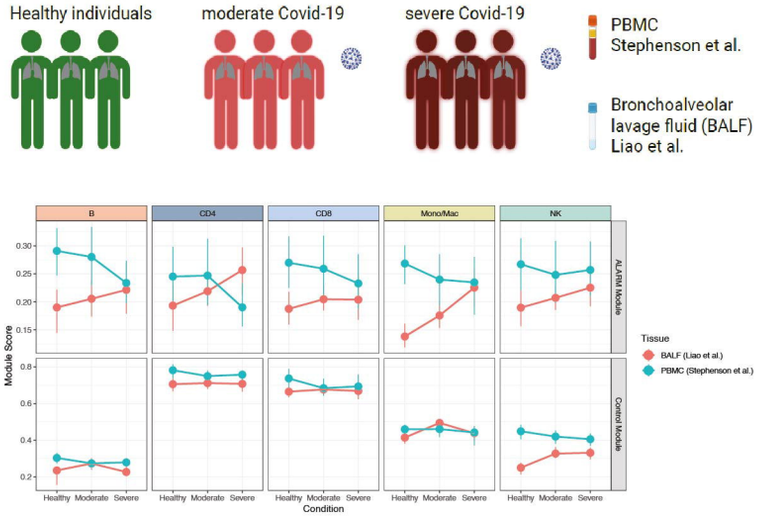
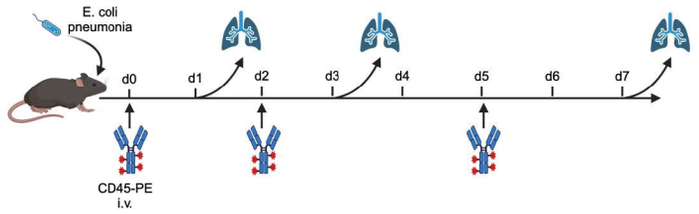
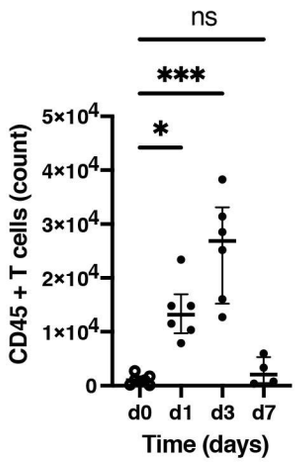


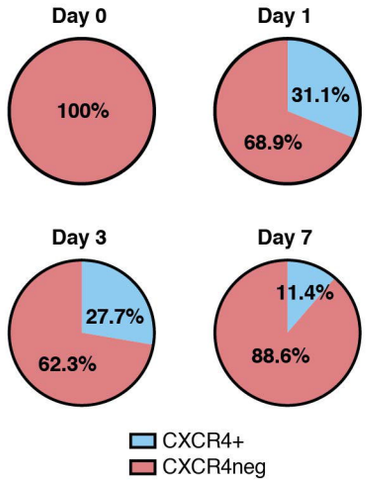
Figure 9

A

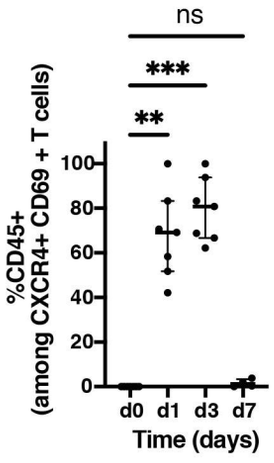


B





C



D

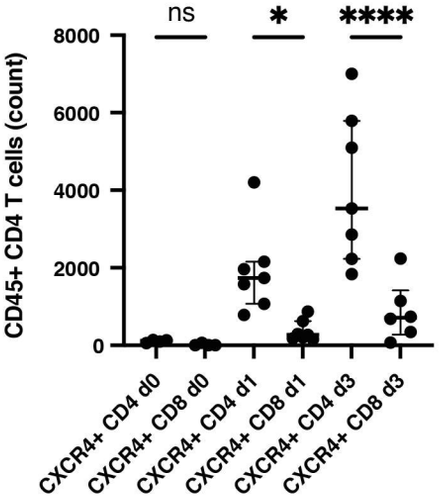


Figure 10

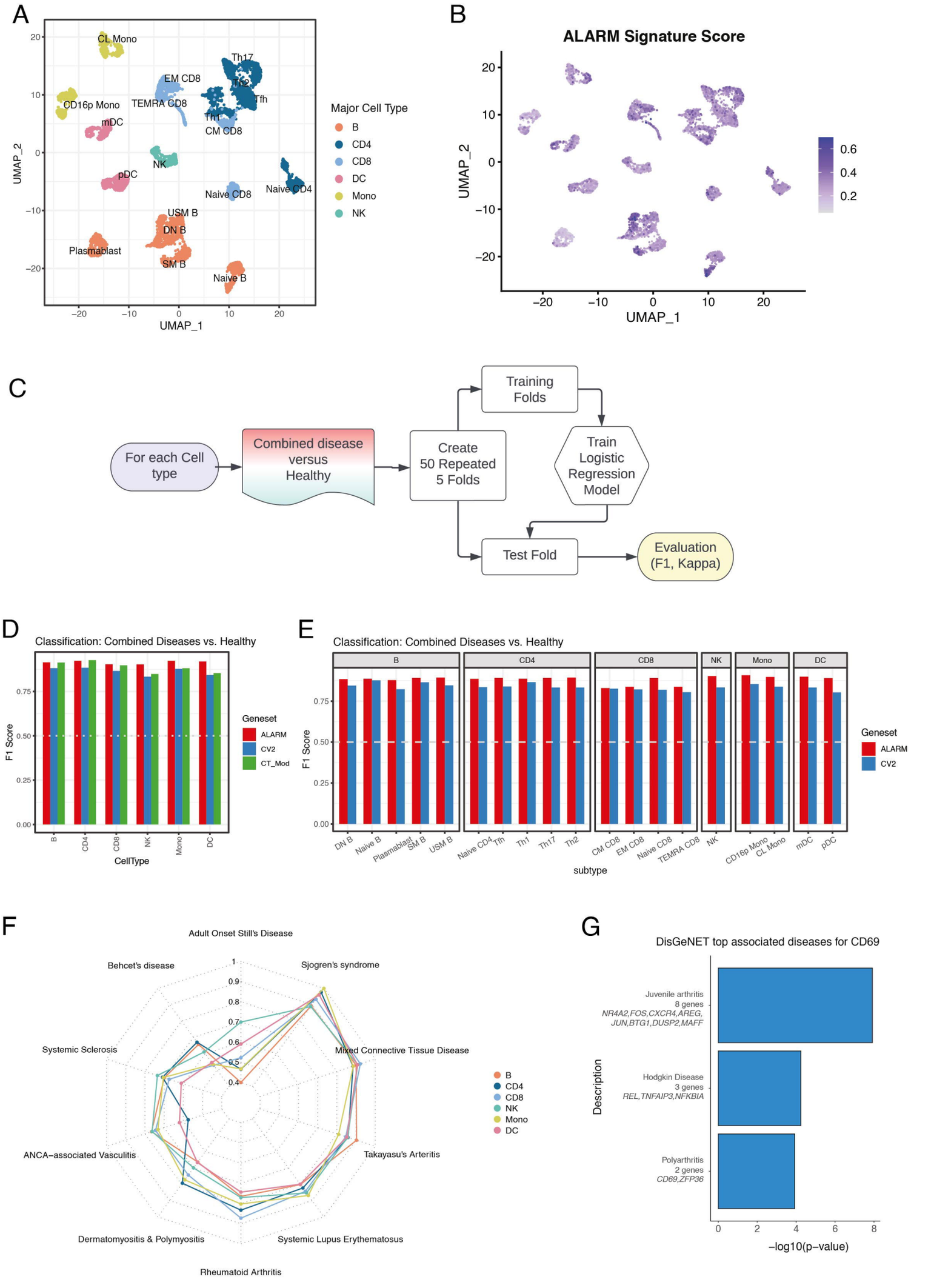
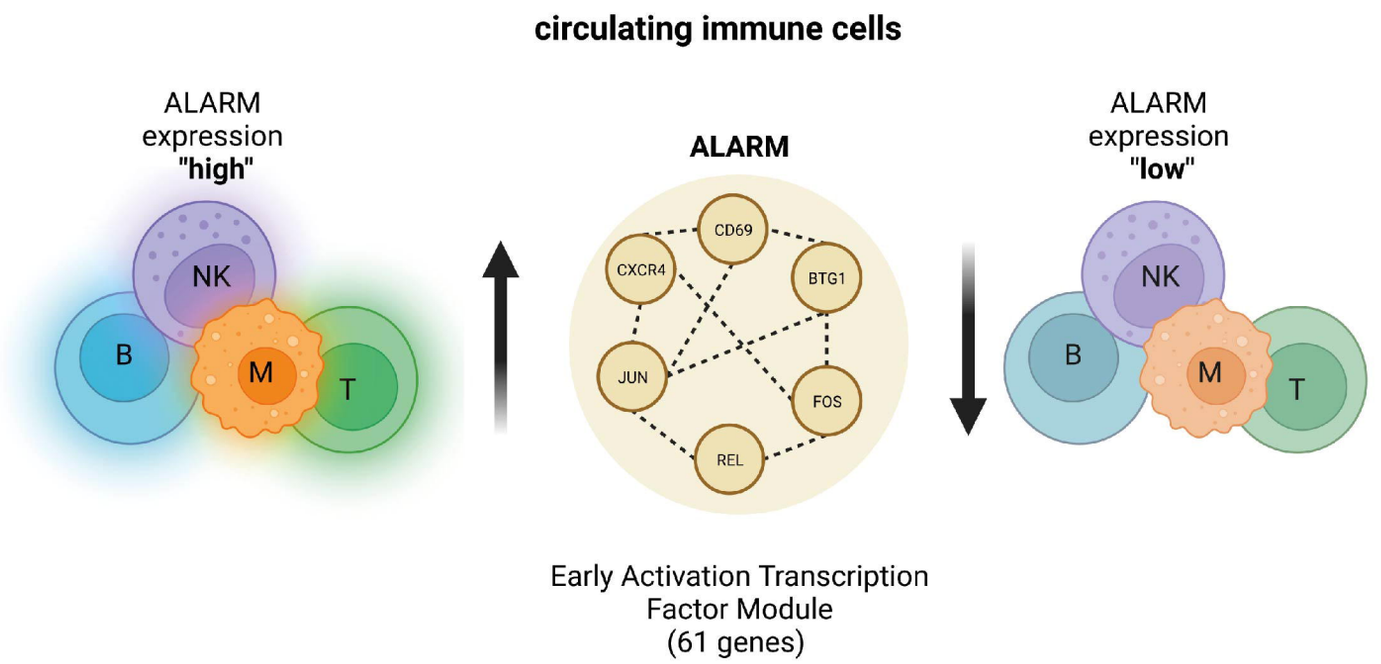
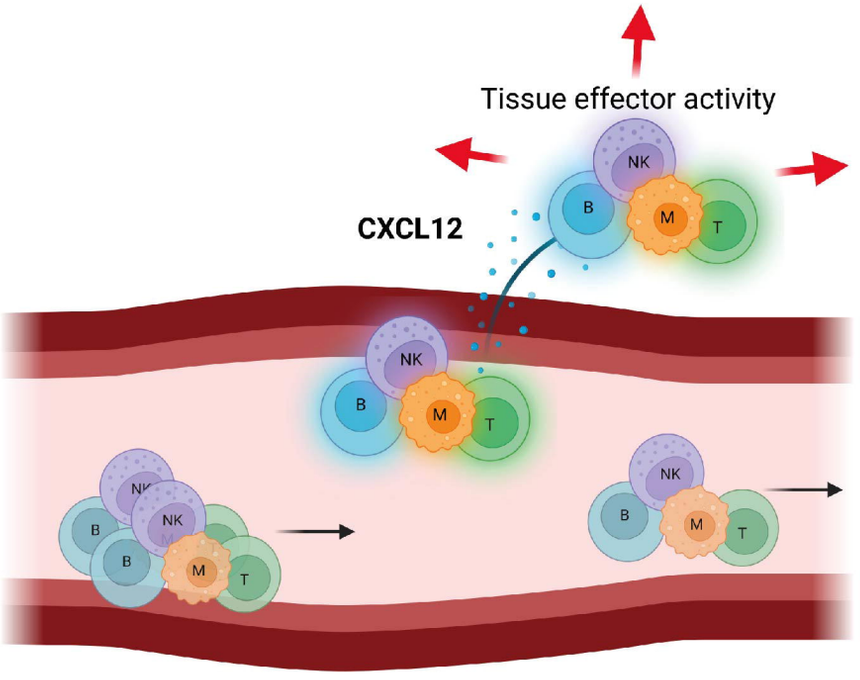

A
CL Mono
Th17
EM CD8
Th2
CD16p Mono
TEMRA CD8
Tfh
mDC
Th1
CM CD8
NK
pDC
Naive CD8
Naive CD4
USM B
DN B
Plasmablast
SM B
Naive B
UMAP_1
UMAP_2
Major Cell Type
B
CD4
CD8
DC
Mono
NK
B
ALARM Signature Score
UMAP_1
UMAP_2
0.6
0.4
0.2
C
For each Cell type
Combined disease versus Healthy
Create 50 Repeated 5 Folds
Training Folds
Train Logistic Regression Model
Test Fold
Evaluation (F1, Kappa)
D
Classification: Combined Diseases vs. Healthy
F1 Score
CellType
Geneset
ALARM
CV2
CT_Mod
E
Classification: Combined Diseases vs. Healthy
F1 Score
subtype
Geneset
ALARM
CV2
F
Adult Onset Still's Disease
Sjogren's syndrome
Mixed Connective Tissue Disease
Takayasu's Arteritis
Systemic Lupus Erythematosus
Rheumatoid Arthritis
Dermatomyositis & Polymyositis
ANCA–associated Vasculitis
Systemic Sclerosis
Behcet's disease
B
CD4
CD8
NK
Mono
DC
G
DisGeNET top associated diseases for CD69
Juvenile arthritis
8 genes
NR4A2,FOS,CXCR4,AREG,
JUN,BTG1,DUSP2,MAFF
Hodgkin Disease
3 genes
REL,TNFAIP3,NFKBIA
Polyarthritis
2 genes
CD69,ZFP36
Description
−log10(p−value)

Figure 11

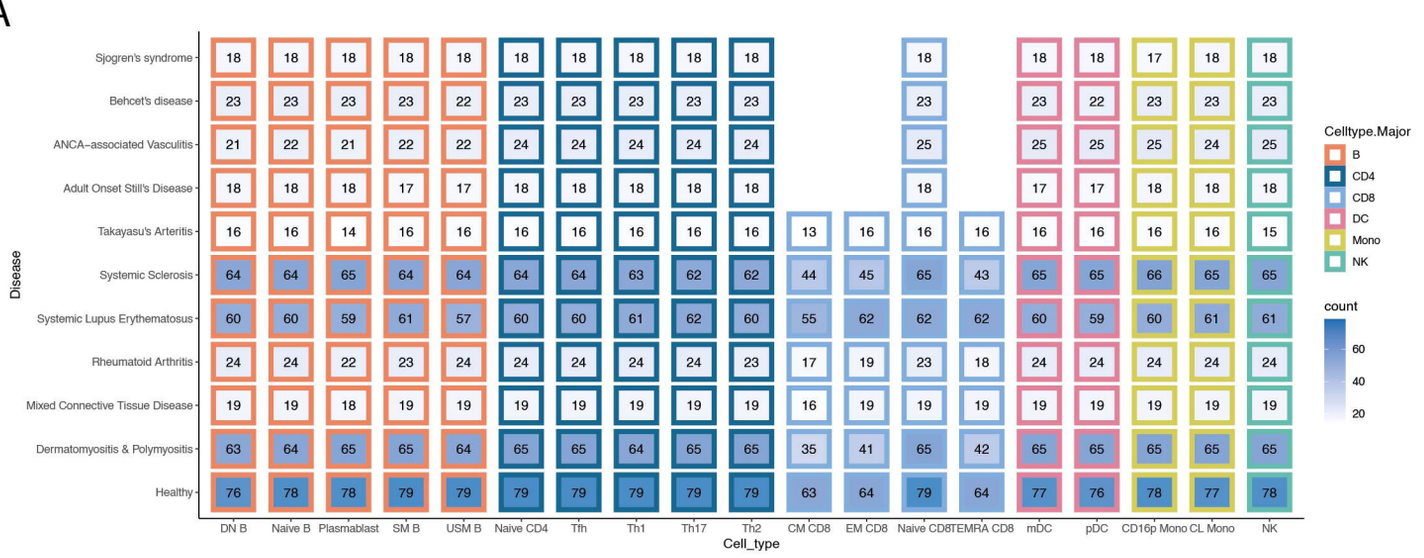


- Peripheral blood mononuclear cells express ALARM at different levels
- ALARM cells are depleted in the blood in response to tissue infection & inflammation
- ALARM cells increase at site of infection and inflammation
- Recruitment of ALARM is associated to CXCL12 expression in the tissue
- ALARM cells are implicated across a wide variety of immune diseases
- ALARM genes are strong predictors of immune mediated diseases

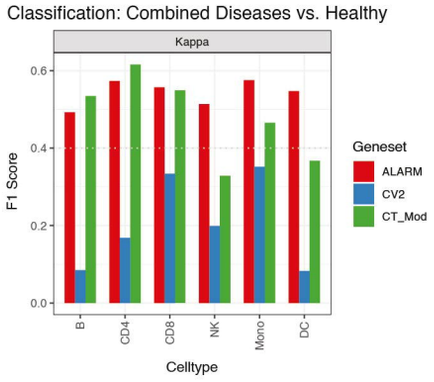


Supp. Figure 6

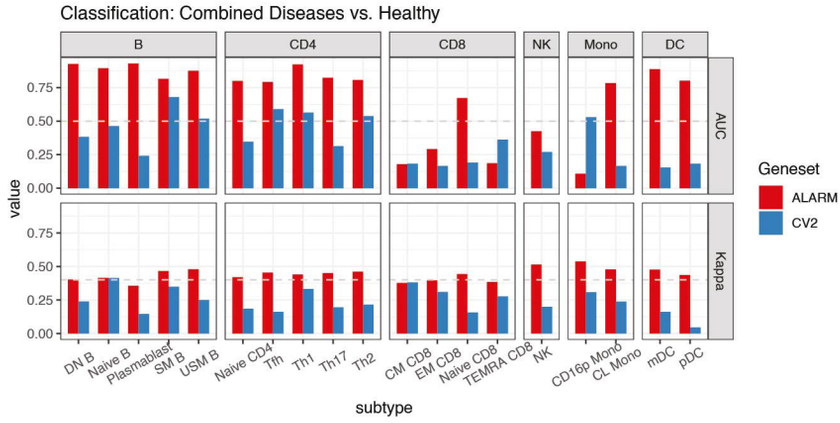
A



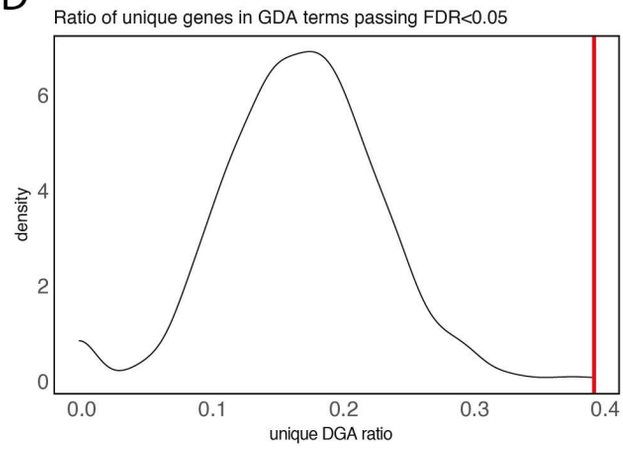
B



C

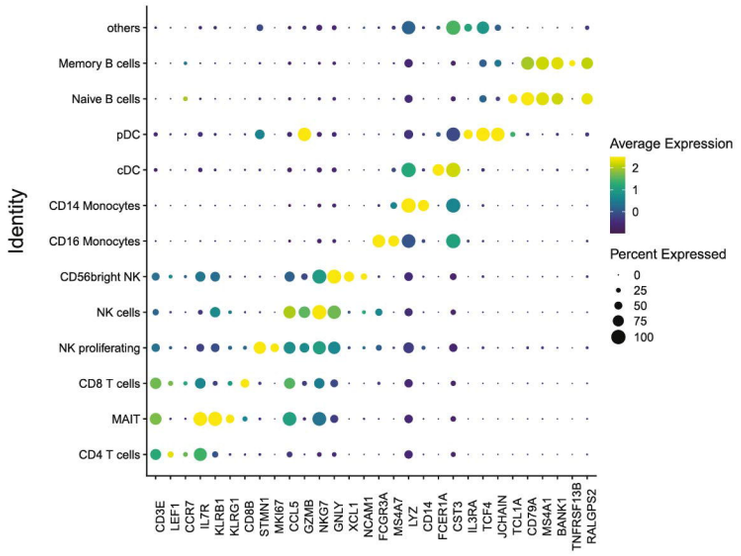


D

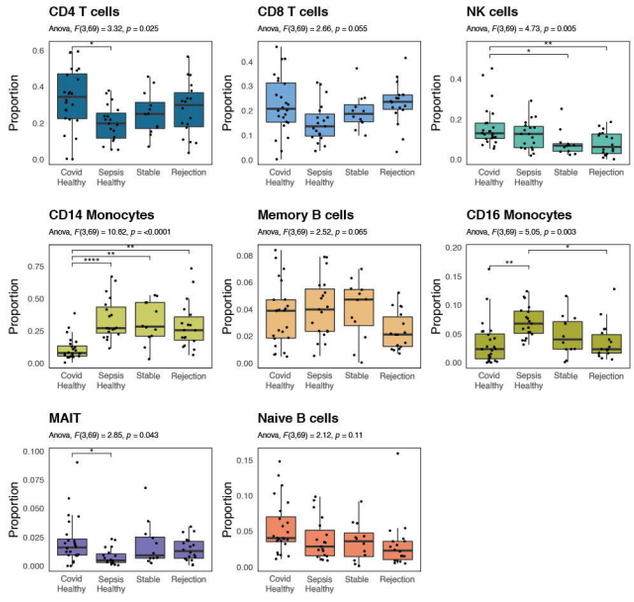


Supp. Figure 1

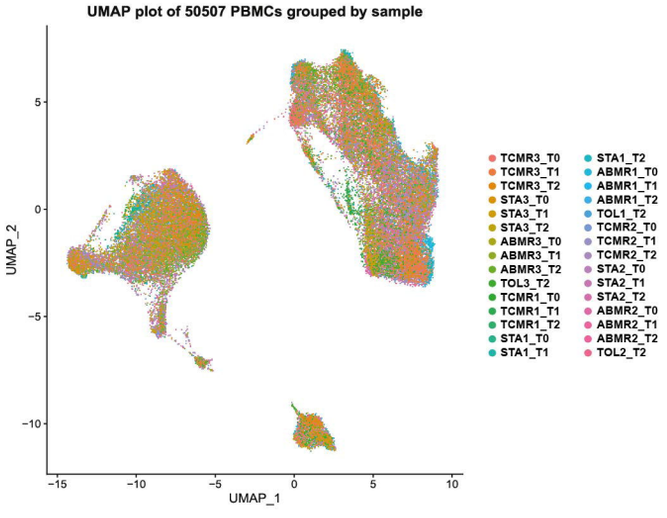
A

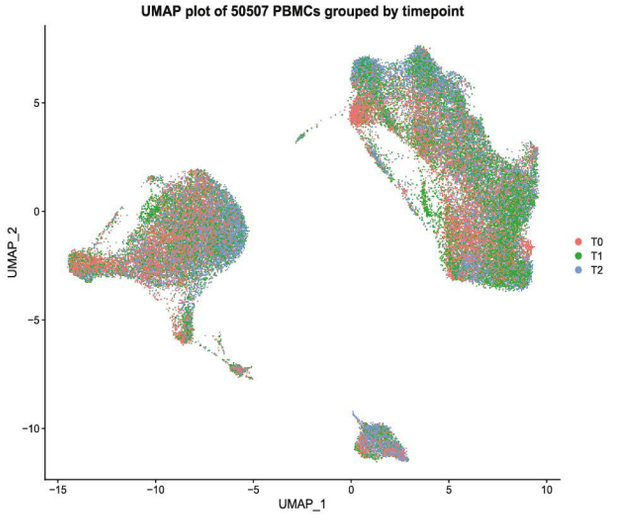


B

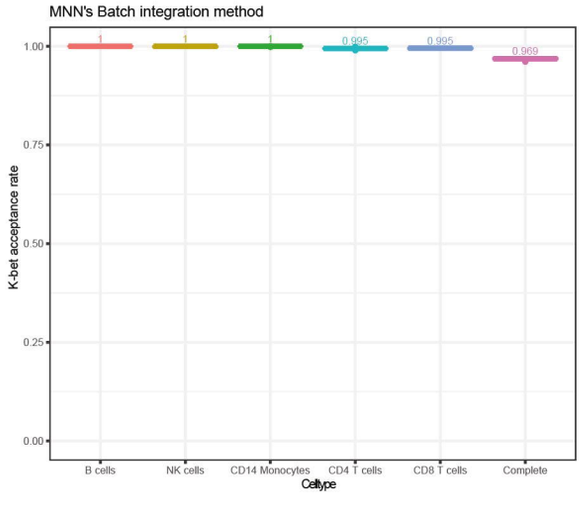


C

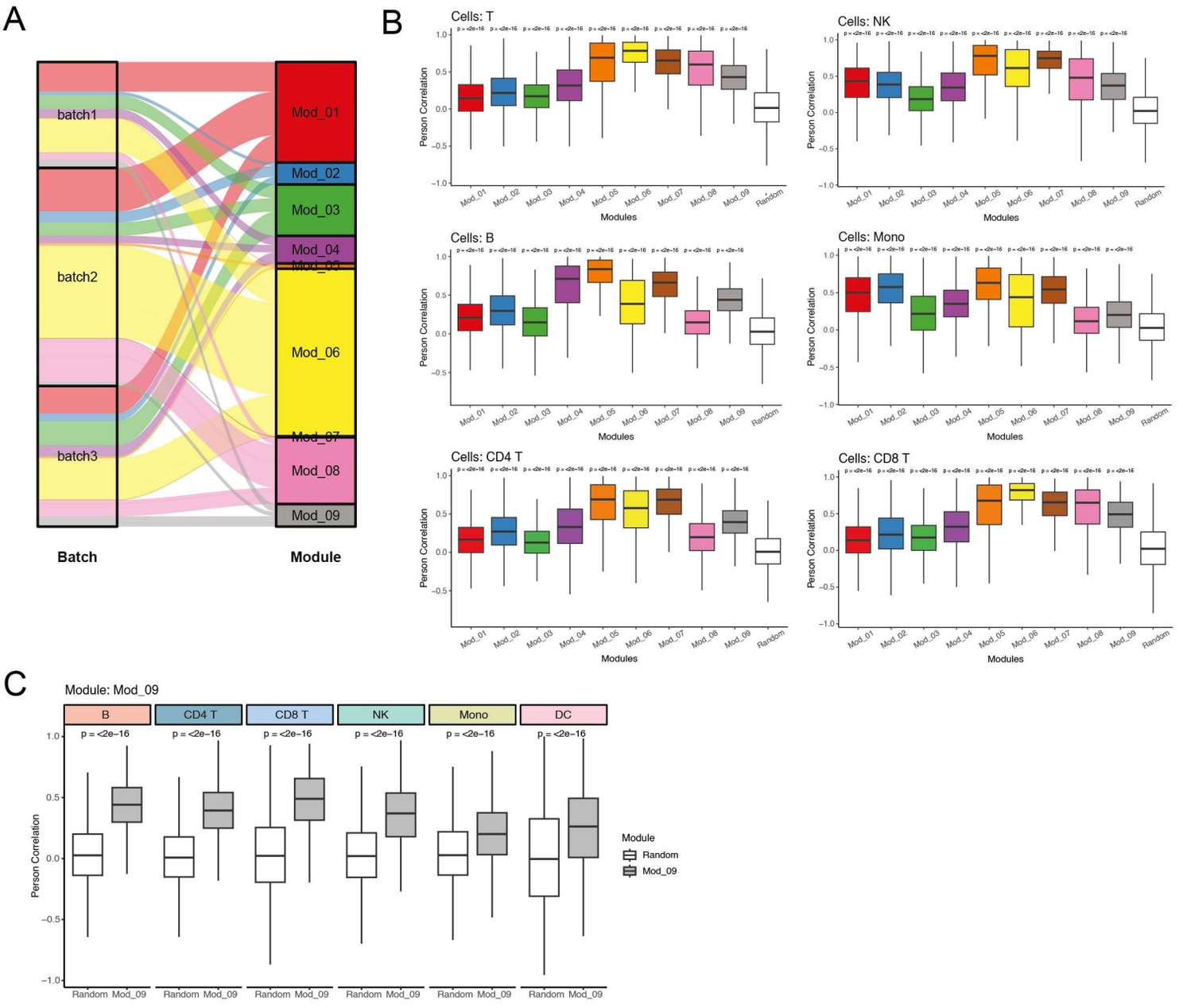




D

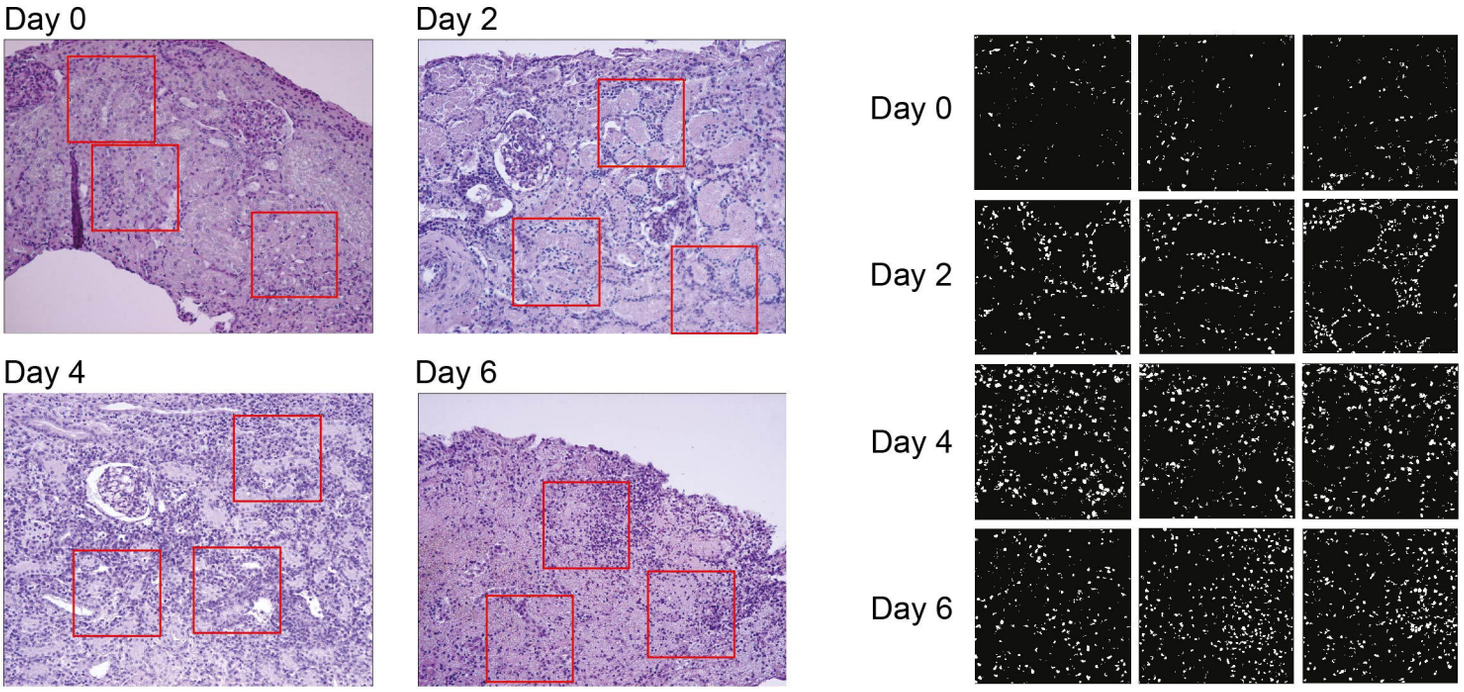


Supp. Figure 2

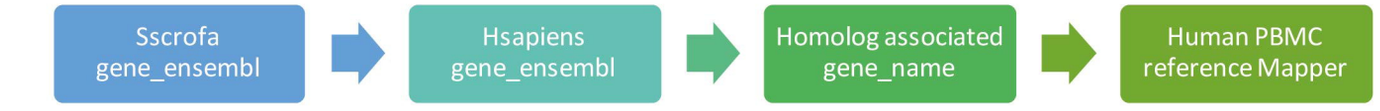


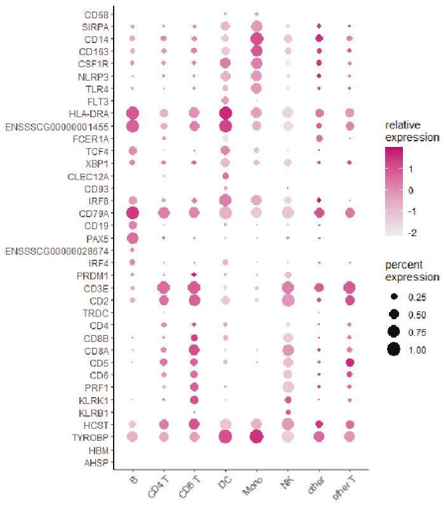
Supp. Figure 3

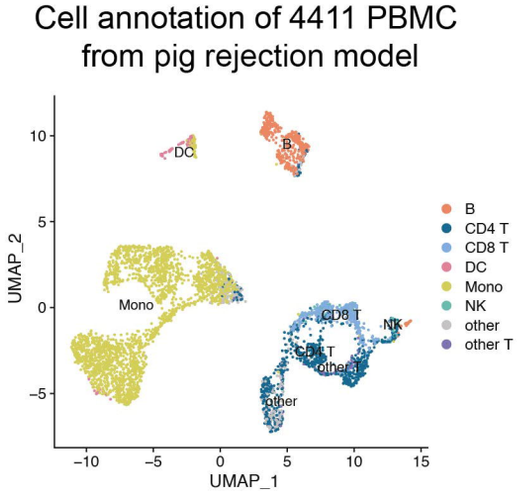
A



B

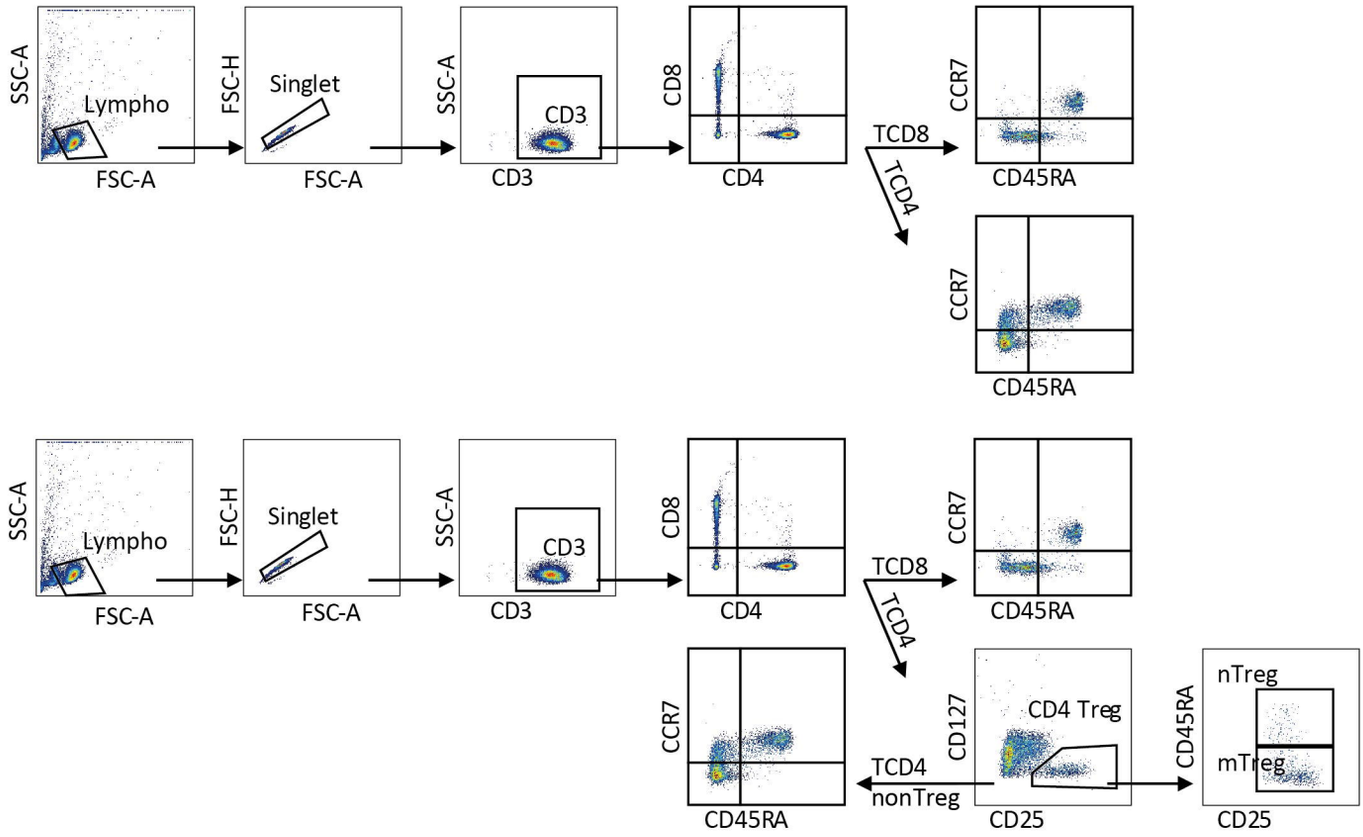




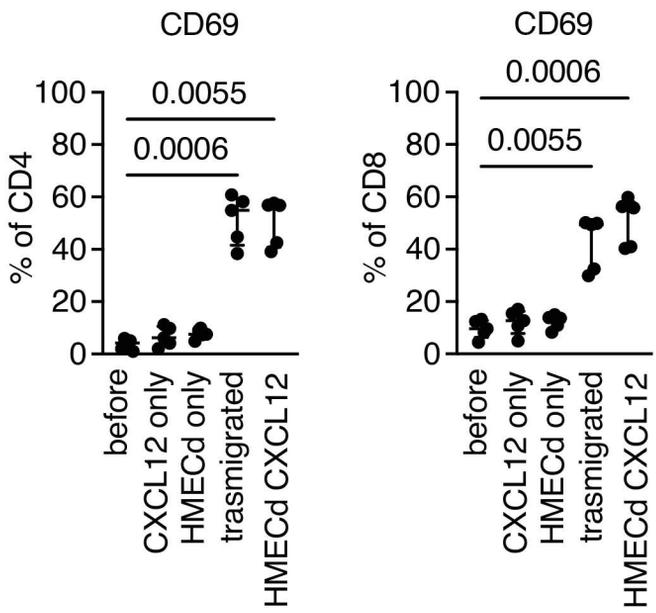


Supp. Figure 4

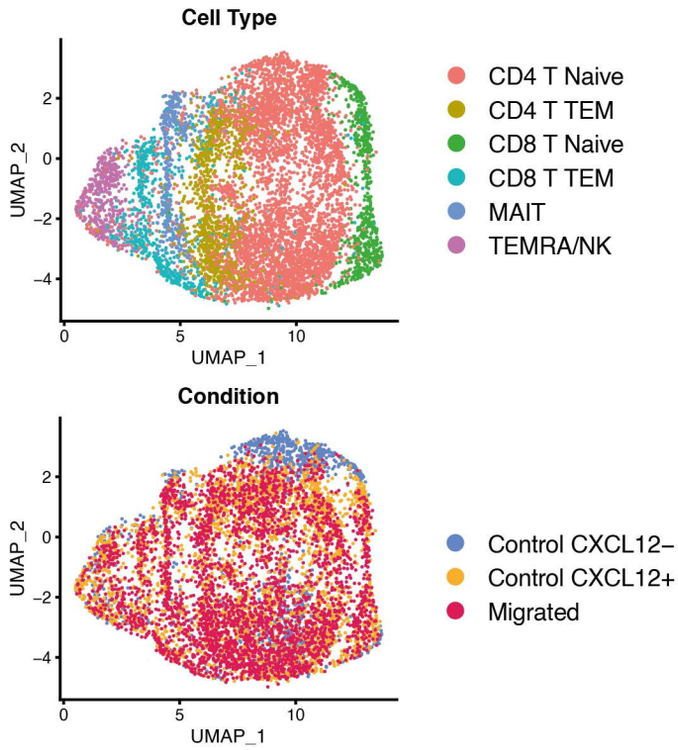
A



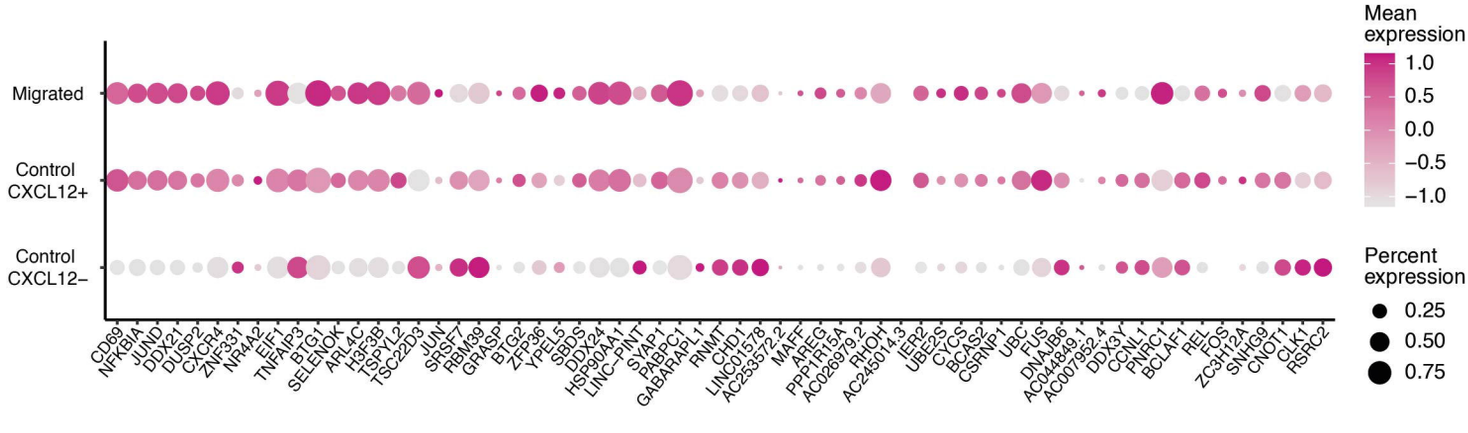
B



C

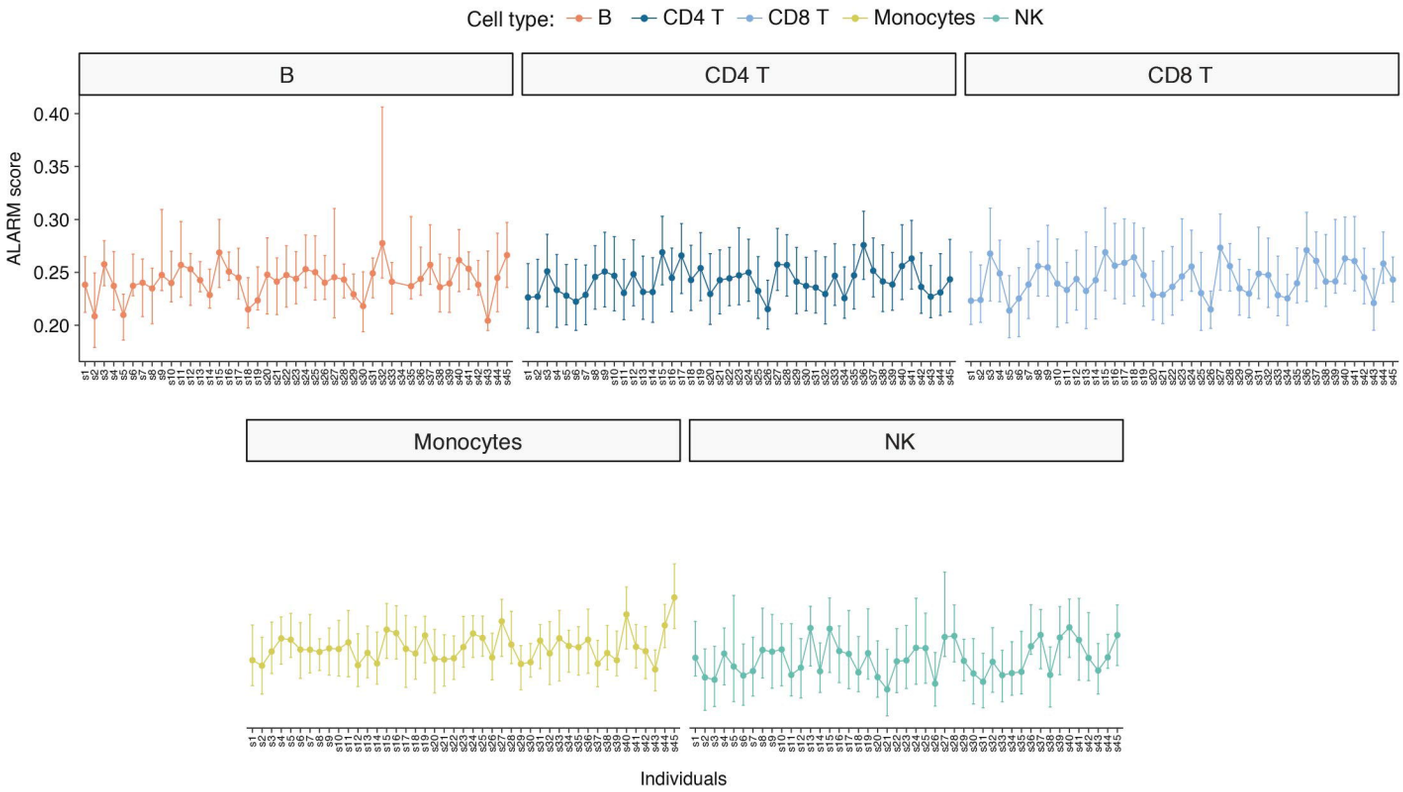
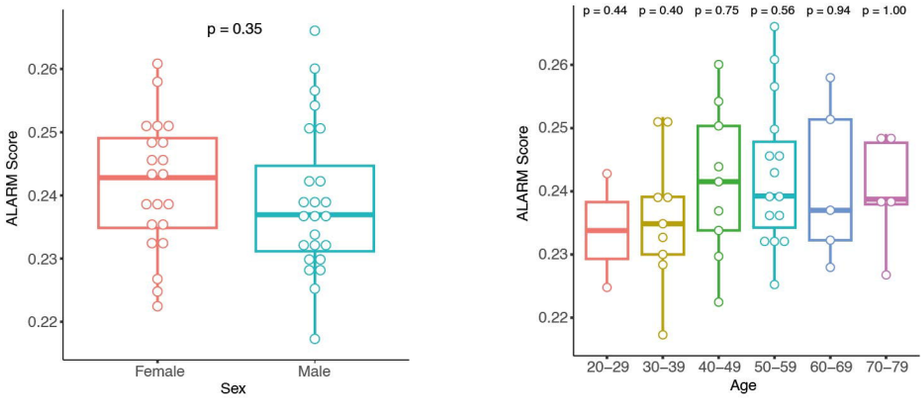


D



Supp. Figure 5

A Healthy Individuals (Van der Wijst et al., dataset)



B