Title

Computational Identification of Migrating T cells in Spatial Transcriptomics Data

Authors

Lin Zhong,[1,2] Bo Li,[3,4] Siyuan Zhang[5], Qiwei Li, [6,] * Guanghua Xiao [1]*

Affiliations

1. Quantitative Biomedical Research Center, University of Texas Southwestern Medical Center, Dallas, TX, USA
2. Department of Statistics and Data Science, Southern Methodist University, Dallas, TX, USA
3. Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
4. Department of Pathology and Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA.
5. Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX, USA
6. Department of Mathematics, University of Texas at Dallas, Richardson, TX, USA.

Abstract

T cells are the central players in antitumor immunity, and effective tumor killing depends on their ability to infiltrate into the tumor microenvironment (TME) while maintaining normal cytotoxicity. However, late-stage tumors develop immunosuppressive mechanisms that impede T cell movement and induce exhaustion. Investigating T cell migration in human tumors in vivo could provide novel insights into tumor immune escape, although it remains a challenging task. In this study, we developed ReMiTT, a computational method that leverages spatial transcriptomics data to track T cell migration patterns within tumor tissue. Applying ReMiTT to ovarian tumor samples, we identified key genes and pathways that enriched on algorithm-identified T cell migration trails, including leukocyte chemotaxis, cell-cell adhesion, and ECM remodeling. We also characterized the phenotypes of T cells on the migrating trails, suggesting that regulatory T cells may accompany cytotoxic T cells during migration. Our findings provide a novel approach to studying T cell migration and interactions within the TME, offering new insights into tumor-immune dynamics.

Introduction

T cell cytotoxicity is essential for the anti-tumor adaptive immune response(*1, 2*). After primed by tumor antigens in the draining lymph nodes(*3*), effector T cells travel through the lymphatic vessels to enter the tumor microenvironment (TME)(*4*). This process is mediated by the T cells expressing necessary cell adhesion molecules and chemokine receptors for migration and infiltration in the tumor(*5, 6*). Efficient tumor killing hinges on 1) the ability of T cells to infiltrate into the tumor core regions(*7*) and 2) T cells maintaining a high level of functionality and cytotoxicity. In late-stage tumors, however, either or both processes were disabled, creating a hostile environment to prevent T cell infiltration(*8*) , or rapidly inducing T cell exhaustion during their migration inside the tumor region(*9*).

Clinical efforts have been prioritized to lift T cell exhaustion by immune checkpoint blockade(*10, 11*), or to break down the tumor barrier to turn immune 'cold' tumors 'hot'(*12*). In addition, reduced motility has been observed among exhausted T cells(*13*).Therefore, investigation on how T cell migrate inside the tumor region and how they become dysfunctional along the way may provide novel therapeutic insights. To date, *in vitro* experiments have been performed to track T cell movement on 2-dimensional (2D) or 3D culture conditions, which improved the quantitative understanding of T cell migration(*14-17*). Non-invasive *in vivo* tracing of cell movement requires the use of radioactive or fluorescent dyes which can be monitored with intravital imaging. This has been done using transgenic mouse models with high single-cell resolution (*18, 19*). However, this approach is usually technically demanding and cannot reach the resolution and scale to study T cell migration in the TME. Consequently, there is still a lack of investigations on T cell migration in the TME.

The recent development of the cutting-edge spatial transcriptomics technologies, such as 10xVisium, MERFISH, Slide-Seq, NanoString CosMx, etc, (*20-23*) provided a new opportunity to address this problem. Spatial transcriptomics data maps the location of gene expression directly within a tissue's spatial arcitecture and allows researchers to study gene activity within the context of tissue structure. In this work, we developed a novel computational method, ReMiTT (Reconstruction of Migration Trails for T cells), to identify T cell migration trails using spatial transcriptomics data. Previous in vitro studies demonstrate that when navigating through a 3-D collagen matrix, T cells prefer to migrate through less dense or channel-like regions, often following each other along the same trajectory (15). We hypothesize that specific areas within the tumor extracellular matrix (ECM), particularly those with less dense collagen fibers, serve as physical highways facilitating T cell migration. Considering that T cells inevitably become exhausted in solid tumors, an increasing trend of exhaustion levels may be observed along some of these migration trails, where streams of T cells follow each other to enter and traverse these regions. ReMiTT is designed to locate these trafficking T cells by searching for closely located T cell loci or spots that exhibit an increasing trend of exhaustion. This might be a snapshot of a stream of migrating T cells on the "physical highway". We implemented this method to study human ovarian tumor samples and identified over 20 candidate T cell migration trails in a tissue slide. Systematic

investigations of the genes overexpressed on the migration trails show enrichment of gene pathways related to T cell chemotaxis, cell-cell adhesion and ECM remodeling on these trails. Furthermore, we observed that both proliferating cytotoxic T cells and regulatory T cells (Tregs) are enriched on the migration trails. Our study might provide a novel approach to unbiasedly explore T cell migration patterns and related genes in human cancer samples.

Results

Characterization of T cell migration trails

We implemented ReMiTT to analyze tumor samples profiled using the 10x Visium spatial transcriptomics platform. This approach is based on a simple observation that migrating T cells in the tumor microenvironment will gradually lose their ability to kill the cancer cells and enter an exhaustion state (*24*). Hence, we hypothesized that the continued movement of T cells can be captured by 'stitching' neighboring spots on a 2-D slide with matched T cell phenotypes, which leads to the development of ReMiTT. In brief, we first identify spots that are most likely to contain newly entry T cells from the draining lymph nodes as the 'start point'. ReMiTT then searches for 1D trails on a 2D slide such that T cells on these trails display an increasing trend of a pre-defined score of T cell exhaustion, using either a minimum spanning path (MSP) searching method or a 3-D line-based approach (Figure 1).

We applied ReMiTT to a human ovarian tumor sample and identified a total of 28 T-cell migration trails (Figure 2A). Trail lengths were between 6 to10 spots, with a mean span of 7.2 spots (~0.7 mm). Of the 28 migration trails, 27 were located in the tumor region, and one was found approaching the tumor region. The pathology slide shows multiple regions of hemorrhagic necrosis (HNR) in this tumor sample. Interestingly, we observed that approximately 1/2 of the trails ran along the edges of these regions (Figure 2B). A subset of trails made their way into the core regions of the tumor (those trails that are not along, or going towards HNR), with a median length of 6.8, indicating that T cells might quickly become exhausted after a short trip inside the immunosuppressive TME.

As experimental tracing of T cell migration *in vivo* is challenging, we sought to validate the trails using a statistical approach. To do this, we first obtained one matching trail for each of the 28 algorithm-identified migration trails. The corresponding 28 matched control trails is one control set. We obtained 10,000 such matched control sets. The systematic analysis in this study is based on comparing algorithm-identified migration trails with their matched control trail sets, rather than simply comparing spots on and off the migration trails. This approach accounts for the spatial distribution of gene expression, as requiring a 'trail' to only consist of adjacent spots influences the expected gene expression pattern. T cell antigen priming and proliferation occur in the draining lymph nodes(*25*) , prior to the migration to the tumor site. Hence, it is expected that T cells of the same clone are more likely to share a migration trajectory. Here, as paired TCR hypervariable regions were not sequenced, we used the TCR α variable (TRAV) and joining (TRAJ) genes and the TCR β variable (TRBV) genes as surrogates to trace clonal sharing. We calculated the mean number of shared TCR V/J genes among spots on each of the 28 algorithm-identified trails. We then calculated the same statistics for each of the 10,000 control sets and got the empirical distribution of the median of the mean shared TCR V/J genes of each trail for each control set. (Fig. S1A). This works as a surrogate for the shared TCR of each trail. The median number of the shared TCR V/J genes among the 28 migration trails is significantly higher (median=12.53) compared to the median in the empirical distribution of these statistics in control sets (median=10.92), with the 1-

sided empirical p-value being 0.033. As different immune cells are reported to migrate together(*6*), we also investigated the clonal sharing of B cells, under the premise that the physical passages for T cell migration might also be utilized by B cells. Consistently, we observed that the median number of shared BCR variable genes among the 28 migration trails is significantly higher (median=2.33) compared to the median in the empirical distribution (median=2.00), with the 1-sided empirical p-value being 0.0012 (Fig. S1B). While these results may not provide direct validation, they supported our conclusion that the predicted trails are statistically enriched for immune cell migration paths.

Gene expression signatures along the migration trails

Next, we sought to investigate the gene expression signatures of cells on the migrating trails. We performed differential gene expression analysis by comparing the mean expression levels of genes on the migration trails to the empirical distribution of gene expression generated with the 10,000 matched control trail sets. We identified 242 significantly upregulated genes (empirical FDR<0.05, Supplementary Table 1). One of the top hits is BMP and activin membrane bound inhibitor (BAMBI). This gene has an established role in promoting the migration of macrophages(*26*), which may company T cells in the TME and cause their exhaustion(*27*). Besides, we observed putative regulators of T cell chemotaxis, including CCL5, CXCL12, and CCL21 (Figure 3A). Specifically, CXCL12 is known to induce T cell migration(*14, 28, 29*), while CCL21 has recently been reported to trigger integrin-mediated adhesion between the T cell and endothelium, a crucial step in the transendothelial migration of T cells(*30*). In addition, genes encoding cell adhesion molecules that shown to promote cell migration and matrix remodeling were also upregulated on the migration trails, such as ITGA11 and ITGBL1(*31-34*).

To understand the gene pathways related to T cell migration, we performed gene-set enrichment analysis (GSEA)(*35*). As expected, the top hits contained many key components of cell migration (Figure 3B). These pathways included the ones involved in the regulation of immune cell chemotaxis, cell adhesion, and cytoskeleton remodeling. Besides, we also noticed pathways involved in extracellular matrix (ECM) remodeling, such as collagen trimmers, which may play a role in remodeling ECM to allow cell migration(*36, 37*).

To validate the above observations, we analyzed another ovarian tumor sample generated with the 10x Visium spatial transcriptomics platform (Method). Implementing the same trail-identification method, we found 15 migration trails and they are all in the tumor region (Supplementary Figure 2A). The length of these trails varied between 5 to 9 nodes with the average being 6.7 spots, similar to the first sample. For this sample, we also obtained 10,000 control trail sets, each containing 15 matched trails corresponding to the 15 algorithm-identified migration trails. We analyzed the upregulated genes on the migration trails and identified a total of 140 genes that are overexpressed along these trails compared with the control trails (Supplementary Figure 2B). They showed similar gene pathway enrichment patterns as the previous one. Specifically, we observed top hits including pathways related to cell adhesion, leukocyte chemotaxis, and T cell/lymphocyte

migration are enriched on the migration trails (Supplementary Figure 2C), providing support for the validity of our method on identifying moving T cells in the TME.

Phenotype of on-trail migrating T cells

We next sought to explore the phenotypes of the T cells on the migrating trails using single-cell RNA-seq data profiled from human T cells. Specifically, we obtained the data from a previous study of 12,346 T cells from 14 non-small cell lung cancer patients(*38*). We used this dataset due to its high data quality and well-accepted T cell subset annotations. We performed unbiased clustering based on the top 30 PCs on these T cells and identified 13 clusters (Figure 4A). We compared the 242 genes upregulated on T cell migration trails to investigate which cluster showed similar gene expression signatures. We observed that cluster 9 and cluster 3 were top enriched (Figure 4B). This analysis was repeated using the second ovarian tumor sample and we observed the same results (Supplementary Figure 2D), suggesting that cluster 9 and 3 present the phenotype of the moving T cells on the migration trails.

T cells of cluster 9 expressed high level of MKI67 and several members of the kinesin family (Figure 4C). These T cells are likely proliferating T cells with massive cytoskeleton remodeling and intracellular protein trafficking. We also identified cytotoxic markers in the top-upregulated genes in this cluster, suggesting that it might contain cytotoxic CD8 T cells. Cluster 3 expressed high levels of Treg markers FOXP3 and IL2RA (Figure 4D). It is highly likely to be a migrating Treg population, which has been previously reported(*39*). It is surprising that this cluster appeared to be top enriched in this analysis, which might indicate that Tregs company the migration of expanding cytotoxic T cells in the tumor microenvironment, hence taking a ‘cis-’ instead of ‘trans-’ role in immune suppression.

Investigation of inter-trail heterogeneity

In the previous results, we noticed that T cell migration trails displayed heterogeneous spatial characteristics. By direction, 6 out of the 28 migration trails were moving towards the HNR (Figure 5A), and 16 trails were moving along or out of the HNR (Figure 5B), with the remaining not showing clear direction regarding these regions, thus denoted as ‘unrelated’ (Figure 5C). There may exist different ‘modes’ of T cell migration in the heterogeneous tumor microenvironment. Here, we did not attempt to reproduce this observation in the second ovarian tumor sample because there was no HNR in this sample.

To better understand this cross-trail heterogeneity, we analyzed the gene expression profiles of each of the 3 trail categories. Specifically, we performed one-way ANOVA to test the equality of the mean expression of each of the 3,000 most variable genes across the 3 groups of trails. We performed adjustment for the p values with Benjamin-Hochberg procedure. We found 21 genes with different expressions passing FDR (Figure 5D). In the group of trails that are ‘unrelated’ to hemorrhagic necrosis, the expression of CTLA4 is significantly higher than in the other groups.

We next investigated the composition of immune cell infiltration across the three types of migration trails. Immune subsets were estimated by CIBERSORT after pooling all the spots on a trail to serve as a 'pseudo-bulk'(*40*). We compared the proportions of different immune cell types (focusing on immune cells specifically, not all cell types) across the three types of migration trails. Interestingly, we observed distinct immune cell compositions among the trails. For example, trails directed along or away from the HNR exhibited significantly higher proportions of immunosuppressive regulatory T cells (Tregs), undifferentiated M0 macrophages, and pro-tumoral M2 macrophages (Figure 5E). In contrast, trails moving towards the HNR showed an increased proportion of CD4 helper T cells, while trails unrelated to the HNR had a significantly larger proportion of CD8 T cells. As expected, these unrelated trails also demonstrated higher levels of overall immune cell infiltration (Figure S6). These findings highlight potential inter-regional tumor heterogeneity related to T cell dysfunction and immune evasion, which has been understudied.

Discussion

Tumor infiltrating lymphocytes (TILs) are the functional carriers of antitumor immunity. As T cells need to make physical contact with their target cells to perform the cytotoxic function, their successful infiltration into the tumor core is the basis of an effective immunotherapy of adoptive T cells(*41-45*). However, how T cells migrate within TME and how to improve their chance of reaching their targets are still understudied. Although a few *in vitro* studies provide valuable insights on how T cells migrate in 3D matrix structure, it is very difficult to track T cells within human tumor tissues(*15, 16*). Compared to normal tissues or any 3D matrix structures created in an in vitro experiment, ECM throughout tumors is highly dysregulated with substantial chemical and structural differences. Hence, *in vitro* experiments on T cells migration may not accurately reflect this crucial part of antitumor immune response in human TME. With the spatial transcriptomics data, we developed a novel method to identify migration trails of T cells *in vivo*, under the assumption that T cells in the TME are gradually increasing their exhaustion levels. ReMiTT could offer an unbiased approach to investigate T cell migration in the tumor. When applied to two human ovarian cancer samples, we made 3 interesting discoveries: 1) genes involved in chemotaxis, actin cytoskeleton remodeling and ECM remodeling are overexpressed in the on-trail spots; 2) regulatory CD4 T cells are migrating with cytotoxic CD8 T cells within the trails; 3) T cell migration trails exhibited spatial heterogeneity, with regard to the locations of the hemorrhagic necrosis regions.

Migration of CD4 regulatory T cells in the tumor has not been fully understood. It is commonly believed that Tregs exert their immunosuppressive functions mainly through the secretion of inhibitory cytokines, such as IL10(*46*). There are also studies suggesting that Tregs can directly kill cytotoxic CD8 T cells through antigen-specific recognition through TCR binding(*47*). Our finding favors the second mechanism, as we provided evidence to show that the migrating T cells are enriched for both the proliferating cytotoxic T cells as well as the regulatory CD4 T cells. Future emphasis should be given to investigating if the inhibition of Tregs in ovarian tumors is controlled by their TCRs or the expression of other regulatory surface molecules.

Macrophage migration marker, BAMBI, appeared to be the top upregulated gene on the trails. We speculated that T cell migration is also accompanied by migrating macrophages or other myeloid cells. This is a reasonable hypothesis given the fact that tumor-associated macrophages (TAM) can be highly immunosuppressive, by expressing diverse checkpoint ligands(*48*). However, it should be noted that further evidence will be needed to prove the physical interactions between TAMs and T cells on the migration trails.

Why is it possible to observe 3-D migration trails on a thin-layer 2-D slide? Most of the trails were found along the edges of the hemorrhagic necrosis regions, which are expected to be spherical, instead of flat. Hence, it is possible that the ‘trails’ are derived from the intersection between a 2-D manifold and the cutting surface of the slide. The 2-D manifold might be the thin gaps between blocks of tissues in the 3-D space. If this is

true, the picture of real T cell migration in the tumor might be much more complicated than previously thought, where T cells can choose different directions to move through.

There are several limitations to this work. First, the Visium spatial RNA-seq data is not at the single-cell level. With each spot having a diameter of 55 microns, it can cover 0-10 cells, depending on the spot's location. Therefore, our analysis is based on the assumption that if multiple T cells are present in one spot, their exhaustion levels are similar. It will be interesting to implement this approach to data produced using the next-generation spatial transcriptomic technologies, such as Visium HD. Second, the sensitivity of this method hinges on high-level of T cell infiltration and large-span of T cell exhaustion markers, which may not be observed for all the tumor spatial transcriptomics samples. Finally, the scope of this work remains exploratory, and most conclusions in this work remain descriptive. Although it might not be easy to trace T cell migration in human tumors, that some of our conclusions can be validated with *in vivo* experiments. For example, it will be interesting to explore which cell types expressed some of the upregulated genes on the trails, such as BAMBI. Despite these limitations, findings in this study might provide novel insights into T cell migration and exhaustion in the tumor microenvironment and facilitate the improvement of cancer immunotherapies.

Materials and Methods

T cell infiltrated spots and exhaustion score

We only include spots with T cells for the analysis in this study. We identify T cell infiltrated spots using T cell surface markers CD8A, CD8B, CD4, CD3D, CD3E, and CD3G. To confirm the presence of T cells in our selected spots, we only include the spots with a sum of these T cell markers>=5. For the T cell spots defined previously, we assigned a T cell exhaustion score for each of them. T cell exhaustion markers used in this analysis are PDCD1, LAG3, HAVCR2, TIGIT, CTLA4, and CD37. These genes have been shown to have increased expression among exhausted T cells in numerous studies (*10, 46, 49-55*) and we use their sum expression as the level of exhaustion of T cells in this study. This sum could be correlated with the number of T cells in each spot. Thus, we performed linear regression with the dependent variable being the sum of gene expression of all these exhaustion markers in a spot and the independent variable being the sum of gene expression of all the T cell markers in the same spot. We assume that spots with more T cells would have higher expression of T cell surface markers. Then, T cell exhaustion score of each T cell spot is defined as the residuals of this linear model, reflecting how exhausted the T cells are in this spot. For convenience, we shift all the residuals to the right until the smallest exhaustion score is 0.

Potential start and end nodes of trails

We clustered adjacent spots with very similar exhaustion scores together to reduce the impact of data noise during trail elongation. Specifically, spatially adjacent spots with similar exhaustion scores are clustered as one single node. The spatial coordinates of these nodes are the centroids of all the spots within a node. In this study, we applied very stringent criteria for 'similar exhaustion scores' so most nodes only have 1 spot.

Assuming that activated T cells experience gradual exhaustion as they navigate through TME, the starting point of a migration trail would be a T cell node with a low exhaustion score, while the endpoint would be a node with a high exhaustion score. We ranked all T cell nodes based on their exhaustion scores and selected the first 33% of nodes with the lowest exhaustion scores as the potential start points and the last 33% with the highest exhaustion scores as the potential endpoints. With these potential start and endpoints, we identified potential migration trails with two methods.

Minimum-spanning path

The first method of trail identification attempts to connect start points to endpoints using Dijkstra's algorithm(*56*). In this method, we defined a directed, weighted graph G(V,E) with all the T cell nodes and their spatial locations. In this graph, V stands for vertices, in our case, all the T cell nodes. E stands for edges. The 10x Visium spatial gene expression data are organized in a hexagonal grid. Each spot (except those at the boundary of the slide) has 6 neighboring spots, with the distance between adjacent spots being

approximately 110 microns. On graph G, edges only exist between neighboring nodes, with 2D distances smaller than 130 microns. In other words, we do not allow for spatial gaps along consecutive pairs along the migration trail. The direction of edges is set to point from the node with the smaller exhaustion score to the node with the larger exhaustion score. Then, for a pair of potential start and end node, if they can be connected on graph G, by a few intermediate spots, then we have a potential migration trail with T cells gradually getting more and more exhausted. It is possible that two nodes can be connected by multiple routes. Among these routes, we chose the one with gradual changes in T cell exhaustion stage rather than having big jumps in exhaustion stages between consecutive pairs. Thus, we assign the weight of edges to be the difference in exhaustion score between the two spots and use Dijkstra's algorithm to find the shortest spanning path based on the weights of edges(*57, 58*). For each of the pairs of potential start and end nodes, we apply this method and got potential T cell migration trails. Then, we exclude trails featuring short distances between start and end nodes (< 500 microns) and trails comprising fewer than 5 nodes.

Line in 3-D space

The minimum spanning path method, while helping to identify trails on which T cells have gradually changing exhaustion stages, requires exhaustion score to be strictly increasing along the trail. The exhaustion score in this study is a rough measure on T cell exhaustion stage. To require it to be strictly increasing along the way might exclude some of the real trails. Thus, we developed the 3-D line method to release this limitation and allow small fluctuations on exhaustion scores along the route while making sure the trend is still an increasing one.

For each T cell node, we use X and Y to represent its 2-dimensional coordinates on the tissue slide, and Z to represent its exhaustion score. Consequently, the 3-dimensional coordinates formed by X, Y, and Z of the T cell nodes along a migration trail would approximate a segment of a straight line in this 3-D space for those T cells migrating roughly in a linear pattern. To get this line segment, we first rescaled the XY coordinates of T cell nodes based on the minimum and maximum value of the exhaustion score (Z coordinate) in this data sample. After rescaling, the range of both X and Y coordinates becomes the same as the range of Z coordinates. Then, for each pair of potential start and end node, we construct a straight line in the 3D space connecting these two nodes. We then calculate the 3-D distance between all other T cell nodes and this 3-D line and keep the ones that have a very small distance to this line as a potential migration trail. The cutoff point is set arbitrarily but we advise users to use a small value. We then determine the indexing of nodes along the trail based on projections of each node to this straight line, beginning from the end with the smaller exhaustion scores.

Similar to the minimum spanning path method, we only include trails with at least 5 nodes as potential migration trails and exclude trails with start and end node too close to each other. Furthermore, we only kept trails with all the T cell nodes adjacent to each other (<=130 microns in 2D (XY) distance). Finally, while a rough line segment ensures that the trend of exhaustion score is increasing, small fluctuations is permissible.

However, significant drawbacks between consecutive pairs of nodes in exhaustion score along the trail should be avoided so we removed such trails.

Trail selection

With potential migration trails identified using these two strategies, we select the final set of trails based on the assumption that as a stream of T cells become more and more exhausted while migrating along some collagen fiber loosen area, the correlation of expression of T-cell stage markers of these T cells on a migration trail would be higher compared to the correlation of the expression of these genes of T cells in neighboring areas. To make this comparison, we first obtained 5-15 alternative routes for each potential migration trail (Figure 1). Specifically, for a migration trail with n nodes, we first get n circular areas with a radius being 1000 microns, and the center being each node. Then, within the oval region surrounding the trail, formed by these overlapped circular areas, we obtained alternative routes of T cell nodes connecting the same pair of start and end nodes. For the alternative trails, there is no requirement for an increasing exhaustion score along the way, but all the other criteria for a potential migration trail apply, such as consecutive pairs of nodes along the trail must be neighbors and a trail must have >=5 spots.

Among markers indicating T cell stages(*38*), we selected the ones that significantly positively correlated with the level of T cell markers (with correlation coefficient>0.2) for this filtering. For each migration trail, we calculated the mean pairwise correlation of the normalized expression level of the selected T-cell stage-related genes(*59*). We then calculated the mean pairwise correlation for each of its alternative routes. Final set of migration trails all met two criteria: 1) mean pairwise correlation larger than 75th percentile of all alternative routes and 2) mean pairwise correlation larger than all of its own alternative routes.

Systematic analysis of algorithm-identified migration trails

To test whether T cells on migration trails are more likely to descend from the same clone compared to T cells not on these trails, we generated 10,000 control trail sets. This process involved clustering our spatial data and dividing the spots into six clusters based on dimensionality reduction methods(*60*) according to the expression level of 3000 variable genes. (Supplementary Figure 7) Then, for each migration trail, a matched control trail of the same length within the same cluster was generated. Additional criteria for migration trail identification were also applied when generating the control trail, including requiring consecutive pairs along the trail to be adjacent and ensuring the spatial distance between the start and end spots of a trail to be ≥ 0.5 mm. One control set includes one matched control trail for each migration trail. For example, in the first human ovarian cancer sample, we identified 28 migration trails, and each control set consists of 28 control trails, with each one matched to a specific migration trail. (Supplementary Figure 7, right panel). Matching gene-expression clusters and trail lengths between algorithm-identified migration trails and control trails can significantly

reduce potential bias from spatial gene expression patterns when comparing features of T cells on and off the migration trails. We generated 10,000 such control sets.

For each of the 28 algorithm-identified migration trails, we calculated its mean number of shared TCR variable genes. Specifically, for each trail, we examined the number of shared TCR variable genes between neighboring pairs of spots along the trail. The mean number of shared TCR variable genes for each trail is then calculated by averaging these values. The median of the mean number of shared TCR variable genes of the 28 migration trails are then compared to the empirical distribution calculated by the 10,000 control sets and we reported the one-sided p value as our hypothesis is that the median number of shared TCR variable genes is higher among migration trails. We performed the same analysis for BCR variable genes.

We used transcripts per million (TPM) for gene expression in each spot to systematically analyze observed trails. To identify upregulated genes and gene pathways on these identified migration trails, we calculated the mean expression for each of the 3000 variable genes among migration trails and control sets. We then calculated one-sided empirical p values for each gene using the empirical distribution generated by the 10,000 control sets. Benjamini-Hochberg adjustment was performed on these p-values to account for multiple comparisons. With genes significantly overexpressed on migration trails compared to the control sets, we performed gene set enrichment analysis (GESA)(*35*). We used C5 collection GO term of Human Molecular Signature Dataset (MSigDB) for this analysis (*61*). Go term contains 10402 gene pathway sets that are related to oncology.

We then characterized the phenotype of the moving T cells on our migration trails using deep single-cell RNA-seq data from 12,346 T cells from 14 non-small cell lung cancer patients(*38*). Initially, we performed clustering analysis and divided these T cells into 13 clusters(*60*). Next, we calculated the mean expression for each of the significantly upregulated genes on our migration trails within each cluster. For each gene, we ranked the 13 means from smallest to largest, with rank 13 corresponding to the largest mean among the clusters. The T cell clusters with more top ranks indicate a higher enrichment of markers from our migration trails, suggesting that these clusters represent the phenotype of the migrating T cells on the identified trails.

Lastly, we overlaid the trails onto the pathology slide. This tumor sample contains many HNRs, and we observed that many migration trails are either along these areas or very close to them. We manually categorized the trails into three groups based on their location and direction relative to HNRs. The first category is "away from hemorrhage," referring to trails where T cells appear to be moving along or away from the hemorrhage regions. The second category is "towards the hemorrhage," where T cells seem to be moving towards the hemorrhage areas. The third category is "unrelated," indicating trails that do not intersect with or attach to any HNRs. We performed ANOVA for each of the 3000 variables to compare their mean expression among the 3 categories of trails and performed p value adjustment. We also performed the GSEA and phenotype analysis in each of these groups. To get the immune cell composition of the migration trails, we summed gene expression for each of the 547 LM22 genes that our data contains for each

trail as a pseudo-RNA-seq bulk. LM22 is a validated leukocyte gene signature matrix for the deconvolution of 22 immune cell types(*62*). We then perform immune cell type deconvolution with CIBERSORT, one of the most commonly used computational tools for the deconvolution of tumor-infiltrating immune cells (*62, 63*). Of the 22 immune cell types, we excluded those with an average fraction of less than 1% among the algorithm-identified migration trails. Additionally, we combined the different stages of CD4 helper T cells into one category 'CD4 helper T cells'. Similarly, we combined the different stages of CD4 helper T cells into one category 'NK cells'. As a result, the final seven distinct immune cell types are shown in Figure 5D.

References

1. H. Raskov, A. Orhan, J. P. Christensen, I. Gogenur, Cytotoxic CD8(+) T cells in cancer and cancer immunotherapy. *Br J Cancer* **124**, 359-367 (2021).
2. A. D. Waldman, J. M. Fritz, M. J. Lenardo, A guide to cancer immunotherapy: from T cell basic science to clinical practice. *Nat Rev Immunol* **20**, 651-668 (2020).
3. H. du Bois, T. A. Heim, A. W. Lund, Tumor-draining lymph nodes: At the crossroads of metastasis and immunity. *Sci Immunol* **6**, eabg3551 (2021).
4. M. M. Steele *et al.*, T cell egress via lymphatic vessels is tuned by antigen encounter and limits tumor control. *Nat Immunol* **24**, 664-675 (2023).
5. I. Melero, A. Rouzaut, G. T. Motz, G. Coukos, T-cell and NK-cell infiltration into solid tumors: a key limiting factor for efficacious cancer immunotherapy. *Cancer Discov* **4**, 522-526 (2014).
6. K. Kohli, V. G. Pillarisetty, T. S. Kim, Key chemokines direct migration of immune cells in solid tumors. *Cancer Gene Ther* **29**, 10-21 (2022).
7. J. Zhang, S. Endres, S. Kobold, Enhancing tumor T cell infiltration to enable cancer immunotherapy. *Immunotherapy* **11**, 201-213 (2019).
8. M. M. Melssen, N. D. Sheybani, K. M. Leick, C. L. Slingluff, Jr., Barriers to immune cell infiltration in tumors. *J Immunother Cancer* **11**, (2023).
9. J. Crespo, H. Sun, T. H. Welling, Z. Tian, W. Zou, T cell anergy, exhaustion, senescence, and stemness in the tumor microenvironment. *Curr Opin Immunol* **25**, 214-221 (2013).
10. S. Agarwal *et al.*, Deletion of the inhibitory co-receptor CTLA-4 enhances and invigorates chimeric antigen receptor T cells. *Immunity* **56**, 2388-2407 e2389 (2023).
11. L. J. Rupp *et al.*, CRISPR/Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells. *Sci Rep* **7**, 737 (2017).
12. G. Xu *et al.*, Hydrogel-mediated tumor T cell infiltration and immune evasion to reinforce cancer immunotherapy. *Nanoscale Horiz* **9**, 295-304 (2024).
13. B. H. Zinselmeyer *et al.*, PD-1 promotes immune exhaustion by inducing antiviral T cell motility paralysis. *J Exp Med* **210**, 757-774 (2013).
14. C. E. Day, C. Guillen, G. B. Willars, A. J. Wardlaw, Characterization of the migration of lung and blood T cells in response CXCL12 in a three-dimensional matrix. *Immunology* **130**, 564-571 (2010).
15. P. Friedl, E. B. Brocker, T cell migration in three-dimensional extracellular matrix: guidance by polarity and sensations. *Dev Immunol* **7**, 249-266 (2000).
16. Z. Sadjadi, R. Zhao, M. Hoth, B. Qu, H. Rieger, Migration of Cytotoxic T Lymphocytes in 3D Collagen Matrices. *Biophys J* **119**, 2141-2152 (2020).
17. K. Wolf, R. Muller, S. Borgmann, E. B. Brocker, P. Friedl, Amoeboid shape change and contact guidance: T-lymphocyte crawling through fibrillar collagen is independent of matrix remodeling by MMPs and other proteases. *Blood* **102**, 3262-3269 (2003).
18. H. M. Cheng, A primer on in vivo cell tracking using MRI. *Front Med (Lausanne)* **10**, 1193459 (2023).
19. M. J. Miller, S. H. Wei, M. D. Cahalan, I. Parker, Autonomous T cell trafficking examined in vivo with intravital two-photon microscopy. *Proc Natl Acad Sci U S A* **100**, 2604-2609 (2003).
20. P. L. Ståhl *et al.*, Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78-82 (2016).

21. S. G. Rodriques *et al.*, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463-1467 (2019).
22. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
23. G. K. Geiss *et al.*, Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol* **26**, 317-325 (2008).
24. C. U. Blank *et al.*, Defining 'T cell exhaustion'. *Nat Rev Immunol* **19**, 665-674 (2019).
25. G. Maass *et al.*, Priming of tumor-specific T cells in the draining lymph nodes after immunization with interleukin 2-secreting tumor cells: three consecutive stages may be required for successful tumor vaccination. *Proc Natl Acad Sci U S A* **92**, 5540-5544 (1995).
26. D. Wang, X. Chen, R. Zhang, BAMBI promotes macrophage proliferation and differentiation in gliomas. *Mol Med Rep* **17**, 3960-3966 (2018).
27. N. B. Hao *et al.*, Macrophages in tumor microenvironments and the progression of tumors. *Clin Dev Immunol* **2012**, 948098 (2012).
28. R. Janssens, S. Struyf, P. Proost, The unique structural and functional features of CXCL12. *Cell Mol Immunol* **15**, 299-311 (2018).
29. J. T. Siveke, A. Hamann, T helper 1 and T helper 2 cells respond differentially to chemokines. *J Immunol* **160**, 550-554 (1998).
30. L. L. de Boer *et al.*, T cell migration requires ion and water influx to regulate actin polymerization. *Nat Commun* **14**, 7844 (2023).
31. P. Erusappan, J. Alam, N. Lu, C. Zeltz, D. Gullberg, Integrin alpha11 cytoplasmic tail is required for FAK activation to initiate 3D cell invasion and ERK-mediated cell proliferation. *Sci Rep* **9**, 15283 (2019).
32. W. Huang *et al.*, ITGBL1 promotes cell migration and invasion through stimulating the TGF-beta signalling pathway in hepatocellular carcinoma. *Cell Prolif* **53**, e12836 (2020).
33. X. Qiu *et al.*, ITGBL1 promotes migration, invasion and predicts a poor prognosis in colorectal cancer. *Biomed Pharmacother* **104**, 172-180 (2018).
34. C. Zeltz *et al.*, Integrin alpha11beta1 in tumor fibrosis: more than just another cancer-associated fibroblast biomarker? *J Cell Commun Signal* **16**, 649-660 (2022).
35. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
36. J. Ren, S. Zhao, J. Lai, Role and mechanism of COL3A1 in regulating the growth, metastasis, and drug sensitivity in cisplatin-resistant non-small cell lung cancer cells. *Cancer Biol Ther* **25**, 2328382 (2024).
37. F. Yang, L. Lin, X. Li, R. Wen, X. Zhang, Silencing of COL3A1 represses proliferation, migration, invasion, and immune escape of triple negative breast cancer cells via down-regulating PD-L1 expression. *Cell Biol Int* **46**, 1959-1969 (2022).
38. X. Guo *et al.*, Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat Med* **24**, 978-985 (2018).
39. A. W. Mailloux, M. R. Young, Regulatory T-cell trafficking: from thymic development to tumor-induced immune suppression. *Crit Rev Immunol* **30**, 435-447 (2010).
40. A. M. Newman *et al.*, Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457 (2015).

41. Y. Tian *et al.*, CXCL9-modified CAR T cells improve immune cell infiltration and antitumor efficacy. *Cancer Immunol Immunother* **71**, 2663-2675 (2022).
42. I. Scarfo, M. V. Maus, Current approaches to increase CAR T cell potency in solid tumors: targeting the tumor microenvironment. *J Immunother Cancer* **5**, 28 (2017).
43. S. J. Priceman *et al.*, Regional Delivery of Chimeric Antigen Receptor-Engineered T Cells Effectively Targets HER2(+) Breast Cancer Metastasis to the Brain. *Clin Cancer Res* **24**, 95-105 (2018).
44. C. E. Brown *et al.*, Optimization of IL13Ralpha2-Targeted Chimeric Antigen Receptor T Cells for Improved Anti-tumor Efficacy against Glioblastoma. *Mol Ther* **26**, 31-44 (2018).
45. P. S. Adusumilli *et al.*, Regional delivery of mesothelin-targeted CAR T cell therapy generates potent and long-lasting CD4-dependent tumor immunity. *Sci Transl Med* **6**, 261ra151 (2014).
46. E. R. O. Allan *et al.*, A role for cathepsin Z in neuroinflammation provides mechanistic support for an epigenetic risk factor in multiple sclerosis. *J Neuroinflammation* **14**, 103 (2017).
47. A. M. Schmidt *et al.*, Regulatory T cells require TCR signaling for their suppressive function. *J Immunol* **194**, 4362-4370 (2015).
48. J. Wang, D. Li, H. Cang, B. Guo, Crosstalk between cancer and immune cells: Role of tumor-associated macrophages in the tumor microenvironment. *Cancer Med* **8**, 4709-4721 (2019).
49. M. Ahmadzadeh *et al.*, Tumor antigen-specific CD8 T cells infiltrating the tumor express high levels of PD-1 and are functionally impaired. *Blood* **114**, 1537-1544 (2009).
50. J. Duraiswamy, G. J. Freeman, G. Coukos, Dual blockade of PD-1 and CTLA-4 combined with tumor vaccine effectively restores T-cell rejection function in tumors--response. *Cancer Res* **74**, 633-634; discussion 635 (2014).
51. J. Fourcade *et al.*, Upregulation of Tim-3 and PD-1 expression is associated with tumor antigen-specific CD8+ T cell dysfunction in melanoma patients. *J Exp Med* **207**, 2175-2186 (2010).
52. E. Ruffo, R. C. Wu, T. C. Bruno, C. J. Workman, D. A. A. Vignali, Lymphocyte-activation gene 3 (LAG3): The next immune checkpoint receptor. *Semin Immunol* **42**, 101305 (2019).
53. R. Yamamoto *et al.*, PD-1-PD-1 ligand interaction contributes to immunosuppressive microenvironment of Hodgkin lymphoma. *Blood* **111**, 3220-3224 (2008).
54. Z. Z. Yang *et al.*, TIGIT Expression Is Associated with T-cell Suppression and Exhaustion and Predicts Clinical Outcome and Anti-PD-1 Response in Follicular Lymphoma. *Clin Cancer Res* **26**, 5217-5231 (2020).
55. A. B. van Spriel *et al.*, A regulatory role for CD37 in T cell proliferation. *J Immunol* **172**, 2953-2961 (2004).
56. R. C. Prim, Shortest connection networks and some generalizations. *The Bell System Technical Journal* **36**, 1389-1401 (1957).
57. N. T. Csárdi G, Traag V, Horvát S, Zanini F, Noom D, Müller K, igraph: Network Analysis and Visualization in R. (2024).
58. G. C. a. T. Nepusz, The igraph software package for complex network research. *InterJournal, Complex Systems*, **Complex Systems**, 1695 (2006).
59. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019).

60. T. Stuart *et al.*, Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
61. A. Liberzon *et al.*, The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
62. B. Chen, M. S. Khodadoust, C. L. Liu, A. M. Newman, A. A. Alizadeh, Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol* **1711**, 243-259 (2018).
63. C. B. Steen, C. L. Liu, A. A. Alizadeh, A. M. Newman, Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods Mol Biol* **2117**, 135-157 (2020).

Acknowledgments

Funding:

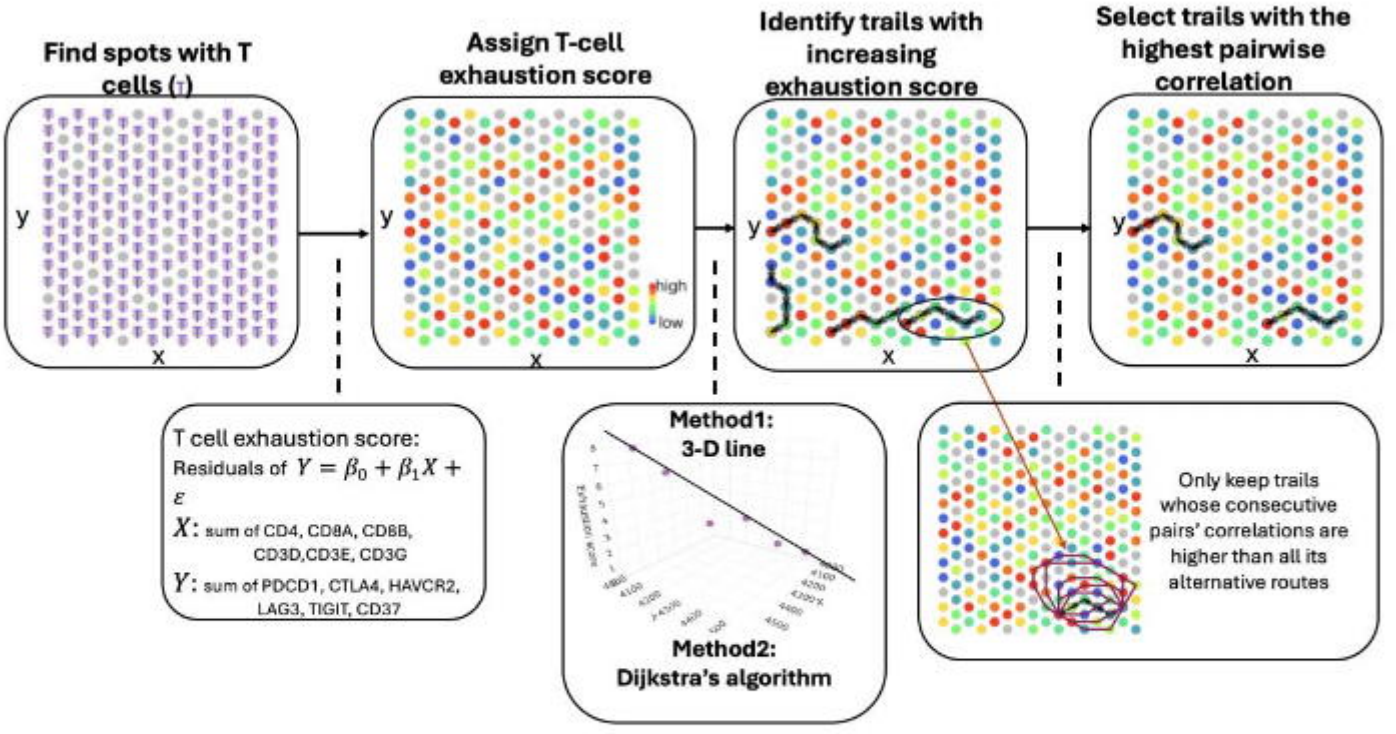
This work was supported by the following funding: the National Science Foundation [2210912, 2113674] and the National Institutes of Health [1R01GM141519] (to Q. L.); the National Institutes of Health [R01GM140012, R01GM141519, U01CA249245], and the Cancer Prevention and Research Institute of Texas [CPRIT RP230330] (to G. X.), NCI R01 grants CA258524 (to B.L.) and CA245318 (to B.L.)

Data and materials availability:

The authors analyzed two publicly available SRT data. Raw count matrices, images, and spatial information for two SRT data from 10x Visium are accessible on the 10x Genomics website at https://support.10xgenomics.com/spatial-gene-expression/datasets. The first sample can be downloaded at https://www.10xgenomics.com/datasets/human-ovarian-cancer-1-standard and the second sample can be downloaded at https://www.10xgenomics.com/datasets/human-ovarian-cancer-11-mm-capture-area-ffpe-2-standard . The deep sequencing single cell data for T cells can be downloaded from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE99254

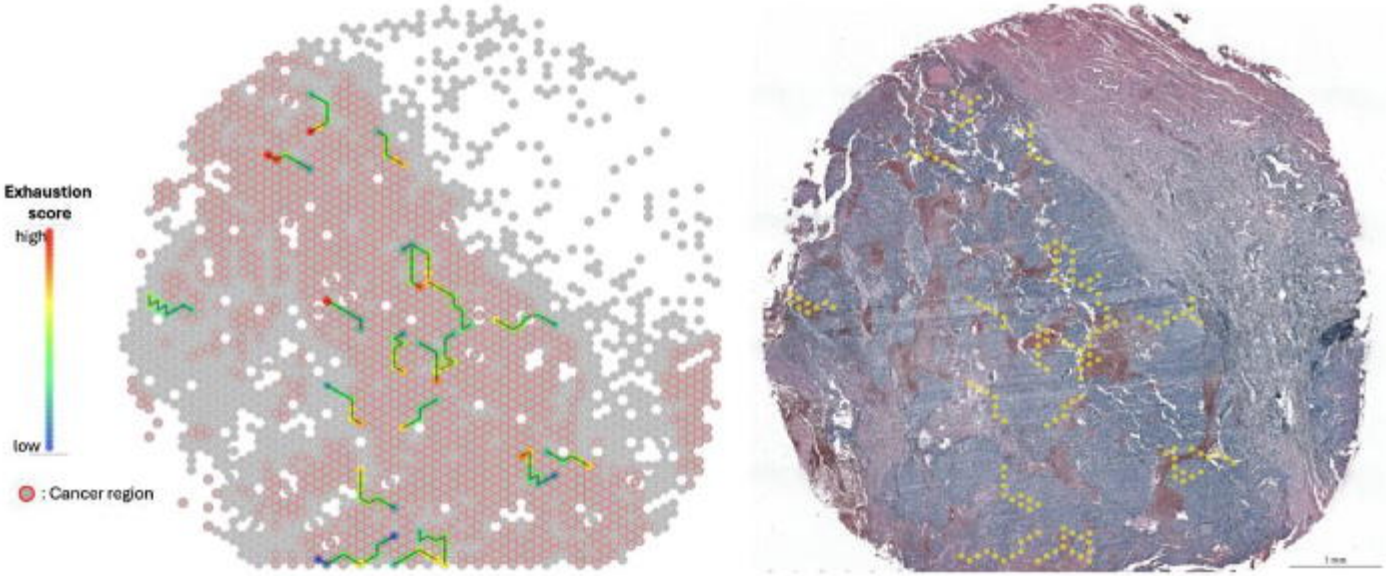
An open-source implementation of the algorithm in R is available at: https://github.com/zhonglin1234/ReMiTT

Fig. 1. **Schematic illustration of the T cell migration trail identification method.**



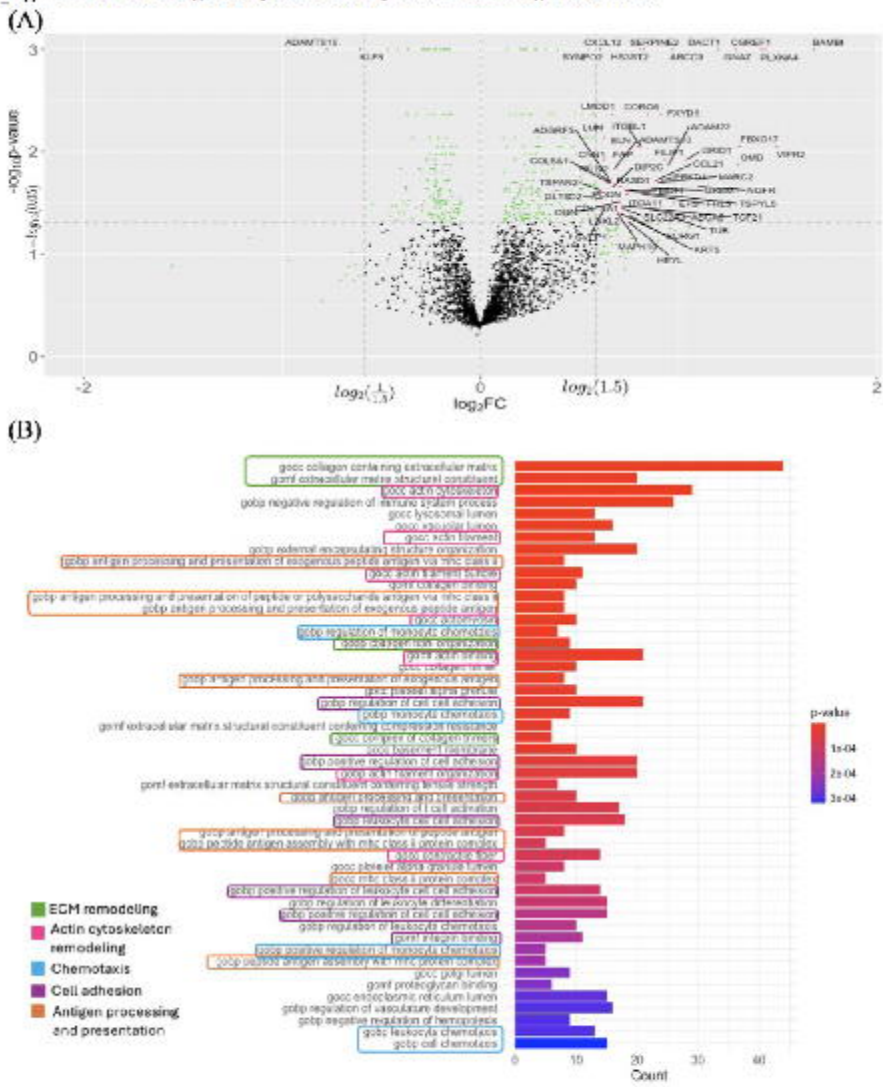
We first identified spots with expression of T cell surface markers (CD4, CD8 and CD3) to select those with T cell infiltration. A T cell exhaustion score was then assigned for each T cell spot using putative exhaustion markers. Candidate migration trails were identified by searching for increasing exhaustion scores among adjacent spots. To filter for natural migration trails, we only keep the ones whose mean pairwise correlation among genes associated with T cell stage is higher compared to all its alternative routes.

Fig. 2. **Migration trails identified in the human ovarian cancer sample and their projections on the pathology slide.**



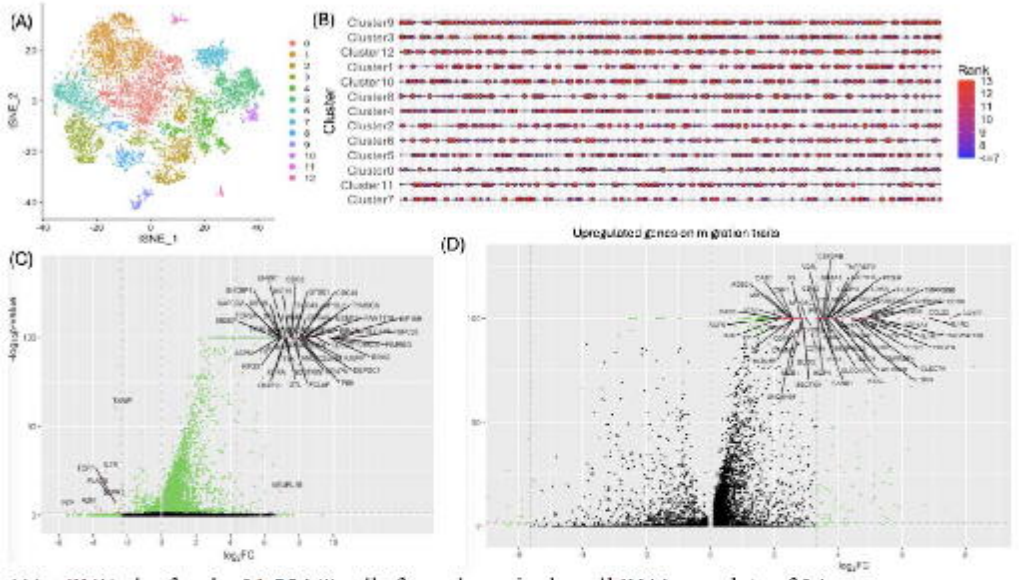
Left: Algorithm-identified migration trails labeled as splines on the spatial coordinates, with the starting location marked as blue. Exhaustion scores along the trail were indicated as in the legend. Grey spots are spots with T cell infiltration and red circles indicating spots with high levels of expression of ovarian cancer markers. **Right:** same trails overlaid on the pathology (H&E) slide of the tumor.

Fig. 3. Differential gene expression analysis for T cell migration trails.



(A) Volcano plot of gene fold change (FC) and adjusted empirical p values between those on migration trails vs. 10000 control sets. Genes with FDR<0.05 and the absolute values of FC ≥1.5 were labeled on the plot. **(B)** Barplot showing the top 50 enriched Gene Ontology biological process (GOBP) gene pathways computed from the 242 significantly upregulated genes on migration trails by GSEA. Colors of the textbox borders indicated the type of biological processes that may involved in T cell migration. Colors of the bars indicate the adjusted p values. Statistical significance was evaluated using Fisher's exact test, with FDR corrected by the Benjamini-Hochberg approach.

Fig. 4. Phenotypic investigation of T cells on the migration trails.



(A) t-SNE plot for the 11,236 T cells from deep single-cell RNA-seq data of 14 non-small lung cancer patients, with the nearest-neighboring clusters color-labeled. **(B)** Bubble plot showing the enrichment of overexpressed genes on T cell migration trails in each of the T cell clusters in (A). The large bubble sizes and red colors indicate higher ranks of mean expression of each of the 242 upregulated genes on migration trails among all the 13 clusters of T cells. Clusters (y-axis) were ordered by their median ranks of the 242 genes. **(C)** Volcano plots of gene FC and adjusted p values for Cluster 9 compared with other clusters. Statistical significance was evaluated using the Wilcoxon rank sum test with FDR corrected using the Benjamini-Hochberg procedure. **(D)** Volcano plots of gene FC and adjusted p values for Cluster 3 compared with other clusters. Statistical significance was evaluated using the Wilcoxon rank sum test with FDR corrected using the Benjamini-Hochberg procedure.

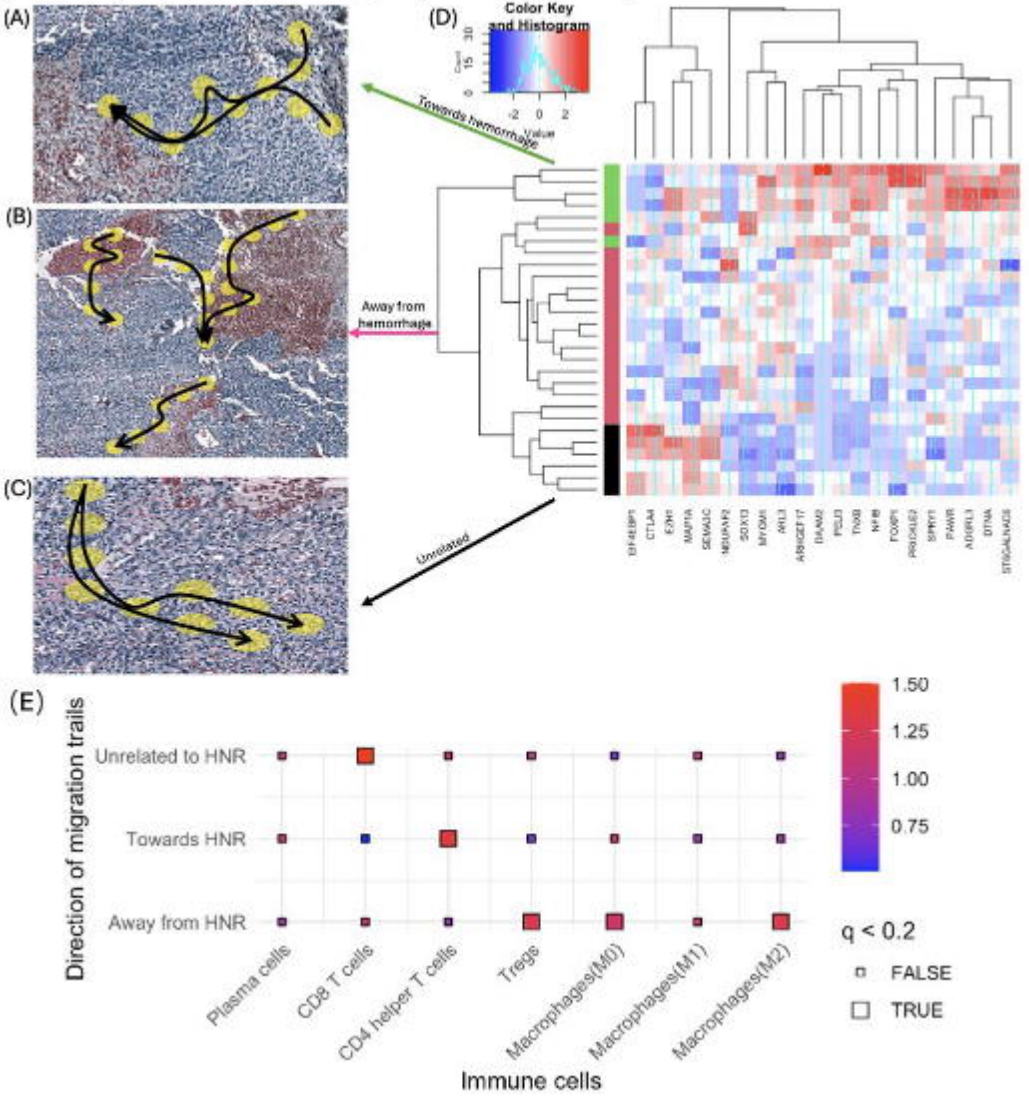



Fig. 5. Investigation of directional heterogeneity of different T cell migration trails