Molecular imprint of epithelial origin identifies immuno-subtype of thymic epithelial tumors

Zhongwei Xin  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Mingjie Lin  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Zhixing Hao  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Di Chen  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Yongyuan Chen  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Xiaoke Chen  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Xia Xu  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Jinfan Li  
Zhejiang University

Dang Wu  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Ying Chai  
The Second Affiliated Hospital, Zhejiang University School of Medicine

Pin Wu  (pinwu@zju.edu.cn)  
The Second Affiliated Hospital, Zhejiang University School of Medicine  https://orcid.org/0000-0002-0472-1425

Article

Keywords: Thymic epithelial tumors, Immune landscape, T cell development, Epithelial origin, Histological classification, Prognostic prediction

Posted Date: September 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-863052/v1
Abstract

Thymic epithelial tumors (TETs) are common tumors in human anterior mediastinum with limited biological understanding. Through decoding the immune landscape of tumors, we reclassify TETs into three types based on T cell developmental patterns. We uncover the developmental dysfunction and TCR repertoire of tumor-infiltrating T cells by cell atlas. Moreover, we identify the unique subset of tumor cells with distinct epithelial origin in each TETs type. Furthermore, we demonstrate that KRT14/GNB3+ mTEC-like cell accumulation inhibits the T cell positive selection in type 1 TETs, while CCL25+ cTEC-like cell promotes T cell positive selection in type 2. Interestingly, although CHI3L1+ mTEC-like cell in type 3 TETs loses the function of supporting T cell development, it acquires the capacity to induce CD8+TRMs-mediated response. Finally, we propose a new molecular classification of human TETs using GNB3 and CHI3L1 to distinguish the epithelial origin of tumor cells, which is promising in prognostic prediction.

Introduction

Thymic epithelial tumors (TETs), including thymomas and thymic carcinomas, are common primary tumors in human anterior mediastinum. Owing to lack of cell lines and animal models, the cognition regarding the etiology and biology of TETs is far from adequate. All types of TETs are considered potentially malignant in the latest clinical guideline because previously defined benign tumors also exhibit aggressive malignant behavior clinically. These evidence suggests that human heterogeneous TETs needs to be reclassified to accurately reflect the biological characteristics of tumors.

Generally, TETs are considered to be transformed from thymic epithelial cells (TECs) which play a crucial role in the T cell development of mammal. There are two major subtypes of TECs that may be the epithelial origin of human TET, cortical TECs (cTECs) and medullary TECs (mTECs), which mainly located in cortex or medulla structure of thymus respectively. It is well demonstrated that cTECs are essential for the positive selection of T cells, while mTECs are mainly involved in the negative selection during T cell development in thymus to establish central self-tolerance. Clinical studies showed that patients with type B thymoma, which may be derived from mTECs, have higher incidence of autoimmune diseases than patients with type A and type C thymoma. These evidence suggests that the tumor cells of human TETs originated from different subset of TECs may impact the T cell development in tumor differently.

WHO classification is the most common used histological classification of human TETs, however its prognostic predictive value and clinical repeatability remain unsatisfactory. Another clinically used Masaoka staging system is also limited in the prognostic prediction, especially in early TETs. In addition, the WHO classification and Masaoka staging assessment of TETs mainly depends on the experience of pathologists and surgeons to examine the tumor morphology, which unavoidably lead to certain subjectivity. Recently, emerging molecular classifications provided insight in stratification of prognosis and treatment for patients with TETs at the molecular level. We proposed a hypothesis...
that the cellular heterogeneity of human TETs which derived from distinct epithelial origin, for instance cTECs or mTECs, resulted in the discrepancy in the formation of immune landscape within each tumor type. Therefore, we conducted a comprehensive study using mass cytometry, single cell sequencing, TCR repertoire, histological analysis, FCM detection and immunofluorescence testing to decode the epithelial origin and its impact on immune composition of human TETs. We successfully reclassified human TETs into three subtypes according to the immune landscape of tumor, pattern of T cell development in tumor, TCR repertoire and origin of tumor cells. In addition, we uncovered the mechanisms that led to the discrepancy of T cell development in each tumor type. Based on our findings, we proposed a tripartite model of tumorigenesis and internal tumorous T cell developmental dysfunction, which was orchestrated by the epithelial origin of human TETs. Finally, according to the epithelial origin of tumors, we established a new molecular classification of human TETs which had shown advantage in prognostic prediction.

Results

Immune landscape reclassifies human TETs.

It is reported that lymphocyte infiltration level in tumor was correlated with the histological classification of TETs for decades\(^{13}\). However, details about the immune landscape of human TETs need to be decoded. To this end, we used mass cytometry, single-cell RNA sequencing, TCR profiling, FCM, HE and IF staining to investigate the immune landscape and its underlying mechanism of human TETs (Fig. 1A). Firstly, we established a panel of more than 40 markers to uncover the immune landscape of human thymus and TETs by CyTOF (Supplementary Table S1). Three normal thymus and twelve TET tissues were performed mass spectrometric flow tests (Supplementary Table S2). Single cell analysis showed that the cell composition of TETs was not only obviously different from normal thymus tissues, but also had a significant internal difference among tumors (Fig. 1B). Importantly, we found the tumor microenvironment of TETs, defined by cellular composition, can be further classified into three types (type 1, 2, 3) (Fig. 1B). Furthermore, each type of human TETs showed a unique histological morphology and lymphocytes infiltrating pattern (Fig. 1C).

Next, we annotated the composition of immune cells in TETs and normal thymus (Fig. 1D and 1E). The immune cell compartment of TET samples similar to normal thymus comprised major immune lineages (Fig. 1D and Supplementary Fig. S1A), among which CD3\(^{-}\)CD4\(^{+}\)CD8\(^{+}\) and CD3\(^{+}\) T cells were most abundant (Supplementary Fig. S1B). Through inter-group comparisons among normal thymus tissue and each type of TETs, we found obvious differences in the major cellular components between groups, especially in CD3\(^{-}\)CD4\(^{+}\)CD8\(^{+}\) and CD3\(^{+}\) T lymphocytes (Fig. 1F, 1G, and Supplementary Fig. S1C). The proportion of CD3\(^{-}\)CD4\(^{+}\)CD8\(^{+}\) lymphocytes was highest in type 1 and lowest in type 3 (Fig. 1G). Moreover, from type 1 to type 3, there was an obvious decreasing tendency of CD3\(^{-}\) CD4\(^{+}\) CD8\(^{+}\) lymphocytes in tumor (Fig. 1G). In contrast, the proportion of CD3\(^{+}\) T cells, B cells, and NK cells in tumor was lowest in type 1 and highest in type 3, which had a gradual increasing tendency among three types (Fig. 1G and Supplementary Fig. S1C). Monocytes and granulocytes were more abundant in tumor of type 3, and the
proportions of DCs in tissues of both type 2 and type 3 were more abundant than normal thymus and type 1 (Fig. 1G and Supplementary Fig. S1C). It is interesting that except CD3−CD4+CD8+ lymphocytes, other immune cells which were related to peripheral immune response in tumor were more abundant in type 3 TETs than that in type 1 and type 2 (Fig. 1G and Supplementary Fig. S1C). In addition, the discrepant abundance of CD3−CD4+CD8+ lymphocytes and CD3+ T cells in tumors among three types of TETs were further validated by flow cytometry (Fig. 1H). Our results showed that the immune landscape of tumor dramatically changed during the development of TETs. Moreover, these findings proposed that human TETs could be reclassified according to the change of immune landscape in tumor.

**T cell developmental pattern of TETs.**

The discrepant amount of CD3−CD4+CD8+ and CD3+ T lymphocytes among tumors suggested that there was a discriminatory T cell developmental pattern in each TETs type. In order to further investigate the change of T cell development in TETs, we used well accepted markers, which were highly correlated with T cell development in thymus 21, to annotate the cell subsets of CD3+ T cells in tumor and normal thymus (Fig. 2A, Supplementary Fig. S2A and S2B). Major subsets of T cells involved in thymic T cell development were represented in the dataset 12, including double negative (DN, CD4−CD8−), double positive (DP, CD4+CD8+), CD4+ single positive (CD4+ SP), CD8+ single positive (CD8+ SP), and FOXP3+ regulatory T (Treg) cells (Fig. 2A). Phenotypic analysis showed that T cell subsets in tissues of thymus and tumors which we defined were representative (Fig. 2B). We found the T cell composition in tumor of TETs was observably different from that in normal thymus (Fig. 2C and 2D). Importantly, the T cell compositions in tumor were significantly different among TETs types (Fig. 2C and 2D). Among tumor types, immature DP cells dominated in type 1, while mature SP cells dominated in the tumor of type 2 and type 3 (Fig. 2D and 2E). Further analysis showed that the proportion of DN and DP cells which were at the early stage of T cell development in tumor was the highest in the tumor of type 1 and lowest in type 3 (Fig. 2E and Supplementary Fig. S2C). It was interesting that, the proportion of DN and DP cells in tumor of TETs, especially DP cells, showed a tendency to decrease from type 1 to type 3 (Fig. 2E and Supplementary Fig. S2C). On the contrary, the proportion of CD4+ and CD8+ T cells which were at the late stage of T cell development was the lowest in tumor of type 1 and highest in type 3 (Fig. 2E). Moreover, the proportion of CD4+ and CD8+ T cells in tumor of TETs showed a gradual increasing trend from type 1 to type 3 which was contrary to DN and DP cells (Fig. 2E). Flow cytometry detection further confirmed the discriminatory amount of DN, DP and SP cells in each TETs type (Fig. 2F). Furthermore, immunofluorescence staining showed that DP and SP cells were located in the cortex or medulla like structure in tumor of type 2 TETs respectively. However, the typical cortex and medulla like structures were not observed and DP and SP cells were distributed in the stroma of tumor in type 1 and type 3 TETs disorderly (Fig. 2G).

The proportion of naïve SP cell (naïve CD4+ and naïve CD8+ cell) and memory SP cell (memory CD4+ and memory CD8+ cell) which have been fully-developed was lowest in type 1 and highest in type 2 TETs (Fig. 2E and Supplementary Fig. S2C). Interestingly, the tissue resident CD4+ T (CD4+ T<sub>Rm</sub>) cells, tissue
resident CD8+ T (CD8+ T_{RM}) cells and Treg cells, which generally accumulated in peripheral tissues\textsuperscript{22}, showed a higher tendency in the tumor of type 3 TETs than the other tumor types (Supplementary Fig. S2C). Besides, there is no significant bias in the ratio of CD4+ and CD8+ T cells among tumor types (Supplementary Fig. S2C). Taken together, these findings uncovered the unique pattern of T cell development in each type of tumor which further supported our reclassification and suggested that T cell developmental dysfunction occurred in the tumors of human TETs.

**Dysfunction of T cell development in TETs.**

To further uncover the T cell developmental dysfunction in tumor of human TETs, we performed scRNA-seq and TCRα β profiling on six tumor samples which had been detected the immune landscape by mass cytometry using the droplet-based 10x Genomics platform (Fig. 1A). The data of a normal thymus sample published by Jong-Eun Park et al. was also reanalyzed as normal control\textsuperscript{21}. After quality control including doublet removal, a total of 52,788 cells from tumors and 2,845 cells from normal thymus were included in our study (Supplementary Table S3). Uniform manifold approximation and projection (UMAP)-based clustering analysis of our samples resulted in 26 cell clusters (Supplementary Fig. S3A). Based on known lineage markers, we annotated cell clusters into nine major cell types (Supplementary Fig. S3B-S3E). These cell types included T cell (CD3E, CD3D), B cell (CD19, MS4A1, CD79A), epithelial cell (EPCAM, FOXD1), dendritic cell (DC) (ITGAX, CD86, CLEC9A, LILRA4), macrophage (CD14, C1QA), monocyte (CD86, CD14, S100A9), endothelial cell (CDH5, PECAM1), fibroblasts (PDGFRA) as well as vascular smooth muscle cell (VSMC) (ACTA2, RGS5) (Supplementary Fig. S3C-S3E, Supplementary Table S4). Consistent with the results of CyTOF, cellular atlas showed that T cells dominated in the discrepancy of cell types between tumor and normal tissue as well as among tumor types of TETs (Supplementary Fig. S3F and S3G).

In order to further investigate the T cell developmental dysfunction in TETs, T cells were re-clustered into 18 cell clusters (Supplementary Fig. S4A). These cell clusters were further annotated (Fig. 3A), according to the marker genes related to T cell development in thymus (Supplementary Fig. S4B and 3B). We defined the major lymphocyte subsets involved in thymic T cell development, including DN, DP, qβT (entry), SP, CD8+ resident memory T (CD8+ T_{RM}) (ZNF683, ITGAE, XCL1, XCL2), Treg (FOXP3, IL2RA), NKT (NKG7, KLRB1, ZBTB16) and Th-like (STAT3, STAT4, RORA, AHR) cells (Fig. 3B, Supplementary Table S5).

Previous CyTOF results showed there were two groups (CD3+ and CD3−) of DP cells in tumor of TETs (Fig. 1D and 2A). However, solely based on the transcriptional level of CD3 expression we were unable to define the CD3− subset of DP cells. To solve this divergence, we further annotated DP cells into proliferating (P) and quiescent (Q) subsets according to the expression level of MKI67 and CDK1 as previous study\textsuperscript{21} (Supplementary Fig. S4C and S4D). Consistently, the marker genes of CD3− DP cells, which were obtained from the protein marker-based sorting strategy in another study\textsuperscript{23}, were obviously highly expressed in DP (P) cells (Supplementary Fig. S4E). These results suggested a close match between the DP (P) cells defined by scRNA-seq and CD3− DP cells detected by CyTOF in our study.
Previous studies reported that thymic T cell development started from CD4\(^{-}\)CD8\(^{-}\) DN cells, which gradually expressed CD4 and CD8 to become CD4\(^{+}\)CD8\(^{+}\) DP cells, and then transitioned through a CCR9\(^{\text{high}}\) αβT (entry) stage to diverge into mature CD4\(^{+}\) or CD8\(^{+}\) SP cells\(^{21}\). Inter-group comparative analysis in our study showed that the composition of T cell subsets involved in T cell developmental stage was significantly different among tumor types of TETs (Fig. 3C-3E). Specifically, DP cells, which were in the early developmental stage, constituted the largest T cell subset in type 1, but least in type 3 (Fig. 3D and 3E). In contrast, SP cells, which were in the late stage of T cell development, constituted the largest T cell subset in type 3, but least in type 1 (Fig. 3D and 3E). Compared with normal thymus, the proportion of DP cells increased in type 1 TETs but decreased in type 2 and almost absent in type 3 (Fig. 3D and 3E). In contrast, SP cells were significantly increased in type 2 and type 3 TETs (Fig. 3D and 3E). It was interesting that naïve T cells were more abundant in the tumor of in type 2 than type 3 TETs, whereas memory T cells were more abundant in the tumor of in type 3 than type 2 TETs (Fig. 3D and 3E). These findings indicated that T cell developmental dysfunction in the tumor of each TETs type was different.

To further decode the details of developmental dysfunction of T cells in tumors, we performed trajectory analysis of the T cell subpopulations defined previously (Fig. 3F). Consistent with previous findings, trajectory analysis showed an obvious T cell developmental dysfunction in each type of TETs compared with normal thymus (Fig. 3G). Through inter-group comparison, we found T cells in type 1 TETs were enriched in the early stage of development, while T cells in type 2 TETs were abundant in both early stage and late stage of development (Fig. 3G). Interestingly, almost all T cells were concentrated in the late stage of development in type 3 TETs (Fig. 3G). The expression levels of Notch and Wnt signaling pathways related genes, which regulated the early stages of T cell development\(^{24,25}\), were higher in the T cells isolated from tumor of type 1 TETs than type 2 and type 3 TETs (Fig. 3H and 3I). In contrast, IL-7 signaling related molecules, which could drive intrathymic expansion of positively selected thymocytes prior to their export to the peripheral T cell pool\(^{26}\), were expressed higher in the T cells isolated from type 2 TETs than thymus and other tumor types (Fig. 3J). These results further demonstrated that T cell development was restrained in early stage in type1, increased in type 2 and almost absent in type 3 of human TETs.

Bias of TCR repertoire and diversity in TETs.

To investigate TCR repertoire and clonotype of T cells in tumor of TETs, TCR chains detected from the TCR-enriched 5’ sequencing libraries were filtered for full length recombinants and were associated with the cell type annotation. To analyze the usage and pairing of VJ genes in T cells, we ranked VJ genes according to genomic positions as previous study\(^{21}\). For TCRβ, we found an obvious bias in VJ gene usage among tumor types of TETs (Fig. 4A), which resembled VJ gene usage of DP and SP cells respectively (Supplementary Fig. S5A). Consistently, VJ pairs of T cells were also significantly different among tumor types of TETs (Fig. 4B), and consistent with DP and SP respectively (Supplementary Fig. S5B). As for TCRα locus, we found a clear association between developmental timing and V-J pairing, as
previously described\textsuperscript{21,27}. During the T cell developmental stage, recombination of proximal pairs on TCRα was mainly observed in DP stage, whereas distal pair recombination was mainly observed in SP stage (Supplementary Fig. S5C). This bias in recombination in turn restricted the V-J pairing of TCRα in DP and SP cells (Supplementary Fig. S5D). Consistently, we observed that the VJ gene usage and pairing of TCRα in each tumor type were in a similar pattern as DP and SP cells respectively (Fig. 4C and 4D).

Furthermore, we found the degree of expansion observed among the clonotypes, which were the number of cells sharing the same clonotype in the dataset, was strongly associated with the T cell developmental states defined by scRNA-Seq (Fig. 4E). It was worth noting that clonotype amplification was mainly observed in mature SP cells, but less in immature DN and DP cells (Fig. 4E). This also indicated that immature T cells (DN and DP cell) had more abundant clonal diversity than mature T cells (SP cell), which was consistent with the finding of the least clonal diversity and the most SP cells in type 3 TETs (Fig. 4F). Together, these findings uncovered the consequence in TCR repertoire and diversity of T developmental dysfunction in each tumor type.

**Epithelial origin decodes T cell developmental dysfunction in TETs.**

Thymic T cell development was orchestrated by heterogeneous TECs\textsuperscript{10}. To further investigate the epithelial origin of TETs and its impact on T cell development in TETs, the subsets of epithelial cells (tumor cells) in thymus and tumors were further analyzed (Fig. 5A and Supplementary Fig. S6A). UMAP-based clustering analysis showed a different composition of epithelial cells in tumor of human TETs, reflected the heterogeneity of tumor cells among different types (Fig. 5A). The epithelial cells were most abundant in type 3 TETs among all tumor types, while similar in type 1 and type 2 (Fig. 5A and 5B). Gene expression analysis showed the epithelial cell in type 3 expressed high level of cancer stem cell (CSC) related markers (\textit{NOTCH2}, \textit{CD24}, \textit{ALDH1B1} and \textit{PROM1}) (Fig. 5C). At the protein level, the epithelial cell in type 3 TETs expressed higher level of Ki67 further suggesting its malignancy (Supplementary Fig. S6B). In addition, trajectory analysis showed the epithelial cells in type 1 and type 2 TETs were at different developmental stages compared with type 3 TETs (Fig. 5D).

In order to further uncover the heterogeneity of tumor cells among TETs subtypes, epithelial cells in tumors of TETs were re-clustered into 17 cell clusters (Supplementary Fig. S6C). The cellular clusters were annotated into seven subsets based on a combination of known TEC markers, including cTEC-like cells (\textit{PRSS16}, \textit{CCL25}, \textit{ZBED2}), \textit{KRT14}\textsuperscript{+} mTEC-like cells (\textit{KRT14}), \textit{CHI3L1}\textsuperscript{+} TEC-like cells (\textit{CHI3L1}), \textit{GNB3}\textsuperscript{+} mTEC-like cells (\textit{POU2F3}, \textit{GNAT3} and \textit{GNB3}), \textit{MYOG}\textsuperscript{+} TEC-like cells (\textit{MYOG}), \textit{CHGA}\textsuperscript{+} TEC-like cells (\textit{CHGA}) and mcTEC-like cells (\textit{PRSS16}, \textit{CCL25}, \textit{KRT14} and \textit{DLK2}) (Fig. 5E and 5F, Supplementary Fig. S6D, Supplementary Table S6). Based on this definition, we found that \textit{CTSL}, encoding the proteases cathepsin L1, and \textit{PRSS16}, encoding thymus-specific serine protease which was reported to play an important role in thymic T cell positive selection\textsuperscript{28}, were mainly expressed in cTEC-like cells and mcTEC-like cells in TETs (Supplementary Fig. S6E). The genes, \textit{FEZF} and \textit{AIRE}, responsible for thymic T cell negative selection were mainly expressed in \textit{CHI3L1}\textsuperscript{+} mTEC-like cells, \textit{MYOG}\textsuperscript{+} TEC-like cells, \textit{GNB3}\textsuperscript{+} mTEC-like cells and \textit{CHGA}\textsuperscript{+} TEC-like cells (Supplementary Fig. S6E). Inter-group comparison showed the
composition of tumor cell subsets in TETs subtypes was significantly different (Fig. 5E and 5G). Besides, our results revealed the major tumor cell subsets in each type of TETs, such as *KRT14/GNB3*+ mTEC-like cells in type 1, *CCL25* cTEC-like cells in type 2 and *CHI3L1*+ mTEC-like cells in type 3 (Fig. 5E and 5G). The main tumor cell subsets in each TETs type were further confirmed by immunofluorescence (Fig. 5H). Interestingly, we found the mcTEC-like cells presented in both type 1 and type 2 tumors had more molecular characteristics of thymic epithelial progenitor cells (TEPCs) compared to *KRT14*+ mTEC-like and *CCL25* cTEC-like cells (Supplementary Fig. S6F), indicating *KRT14*+ mTEC-like and *CCL25* cTEC-like cells might be differentiated from mcTEC-like cells. Through pathway analysis, we found that NF-κB signaling pathway of mcTEC-like cells in type 1 was obviously activated, while that in type 2 was inhibited (Supplementary Fig. S6G). This meant the differentiation of mcTEC-like cells in type 1 was more inclined to be mTEC-like cells, while the differentiation of mcTEC-like cells in type 2 was more inclined to be cTEC-like cells, which was consistent with the result that mTEC-like cells increased in type 1 and cTEC-like cells increased in type 2 (Fig. 5G). Moreover, immunofluorescence staining result showed the AIRE expression in type 2 TETs was significantly less than the other two types (Supplementary Fig. S6H), which was correlated with the incidence of autoimmune disease in TETs.

CXCL12 and CCL25, the key cytokines promoting homing of blood-borne lymphoid progenitor cells into thymus, were almost not expressed in *CHI3L1*+ mTEC-like cells (Fig. 5I), the main component of the tumor cells of type 3 TETs. This observation suggested the tumor cells of type 3 TETs lost the function for recruitment of thymus progenitor cells, resulting in the T cell developmental dysfunction in initial period. Having identified the major TEC-like tumor cells in TETs subtypes, we used CellPhoneDB analysis to investigate the interactions between TECs-like tumor cells and T cells as previous study. T cell development in thymus was a complex process involving TEC-lymphocyte cell interactions, lymphocyte cell migration and lymphocyte cell localization. Here, we focused on the tumor cells and lymphocyte cell interactions mediated by chemokines, which enabled lymphoblast migration and anatomical localization. Our results demonstrated that mTEC-like and mcTEC-like tumor cells in TETs could induce DP cells migration for positive selection through *CCL25:CCR9* (Fig. 5J, Supplementary Table S7) and *CXCL12:CXCR3* interaction (Supplementary Fig. S6I, Supplementary Table S7). Interestingly, cTEC-like tumor cells also had a crucial role in inducing SP cell migration through *CCL19:CCR7* interaction (Fig. 5J), which differed from the cTEC in normal thymus.

To further illustrate the differences of T cell development among TETs subtypes in molecular level, we used Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis to discover the difference of signaling pathway enrichment in DP cells among tumor types. Compared with type 1 TETs, the genes high-expressed in DP cells of type 2 were significantly enriched in signaling pathways related to T cell expansion and positive selection, including Oxidative phosphorylation, NF-κB signaling pathway, Th17 cell differentiation, Th1 and Th2 cell differentiation and TNF signaling pathway (Fig. 5K). In the process of positive selection, most cells would be eliminated by apoptosis, so the enrichment of highly-expressed genes in apoptosis-related pathways also indicated that more DP cells in type 2 were undergoing positive selection (Fig. 5K). Compared with normal thymus tissue, genes high-expressed in
DP cells of type 2 were also enriched in the relevant signal pathways for positive selection (Supplementary Fig. S6J). Moreover, we analyzed the expression of marker genes which indicated the pre- or ongoing state for positive selection of DP cells. We found that DP cells in type 1 TETs mainly expressed pre-selection related genes, while DP cells in type 2 TETs highly expressed genes related to ongoing state of positive selection (Fig. 5L). These results indicated the positive selection of T cell development in type 2 TETs was promoted by tumor cells, while blocked in type 1. This was further supported by the observation that the expression level of HLA-DR molecule of tumor cells in type 1 TETs was lower than other types (Supplementary Fig. S6K). Taken together, these findings demonstrated that unique epithelial origin of each type of TETs led to a different dysfunction of epithelium and T cell interaction and consequence in the discrepancy T cell developmental dysfunction among tumor types.

**Tumor cells induce CD8$^+$ T$_{RM}$ cell-mediated immune response in type 3 TETs.**

Among three types of TETs, it was interesting that the thymic T cell development seemed absent in type 3 TETs, whereas mature T cells were still abundant in the mesenchyme of tumor tissue (Fig. 2C-2E). A possible reason for this phenomenon was the infiltration of immune cells induced by malignant cells of cancer, which could be commonly observed. We confirmed that the CD8$^+$ T$_{RM}$ (CD103$^+$CD69$^+$) cells, playing a central role in immune sensing network of peripheral tissues especially in the recruitment of various types of immune cells when they are activated, were significantly enriched in type 3 TETs (Fig. 6A and 6B). The unique abundance of CD8$^+$ T$_{RM}$ cells in type 3 TETs was further confirmed by flow cytometric analysis (Supplementary Fig. S7A). Compared with other epithelial subpopulations, CHI3L1$^+$ mTEC-like cells, which were most highly enriched in type 3 TETs, were predicted to have a largest number of interactions with CD8$^+$ T$_{RM}$ cells (Fig. 6C and Supplementary Table S7). Some specific ligand-receptor interactions between CHI3L1$^+$ mTEC-like tumor cells and CD8$^+$ T$_{RM}$ cells were identified, including TNFSF9:HLA-DPA1, CCL20:CXCR3, CEACAM5:CD8A (Fig. 6D and Supplementary Table S7). Consistently, CD8$^+$ T$_{RM}$ cells in type 3 TETs expressed higher level of activation and proliferation related molecules, including CD38, CD39, Ki67 and CD137, while the inhibitory molecule was decreased (Fig. 6E and Supplementary Fig. S7B). Furthermore, genes related to activation and TCR signaling at the transcriptional level were also significantly upregulated in CD8$^+$ T$_{RM}$ cells compared with other types of T cells in the tumor of type 3 TETs (Fig. 6F). Pathway analyses revealed the upregulated gene of CD8$^+$ T$_{RM}$ cells significantly enriched in TCR signaling pathway, antigen processing and presentation, cytokine-cytokine receptor interaction and chemokine signaling pathway (Fig. 6G). Further analysis demonstrated CD8$^+$ T$_{RM}$ in tumor of type 3 TETs could potentially recruit T cells, B cells, DCs and macrophage via CXCL12:CXCR4, CCL21:CCR7, CXCL13:CXCR5, XCL1/2:XCR1 and CX3CL1:CX3CR1 interactions, respectively (Fig. 6H and Supplementary Table S7). IF staining further validated the co-localization of epithelial cells and CD8$^+$ T$_{RM}$ cells, as well as CD8$^+$ T$_{RM}$ cells and DCs, B cells in tumor of type 3 TETs (Fig. 6I and 6J). Our results proved that CHI3L1$^+$ mTEC-like tumor cell induced CD8$^+$ T$_{RM}$ cell activation playing a central role in the immune response of type 3 TETs.
Epithelial origin reclassifies human TETs.

There were several kinds of histological classification and clinical stage of human TETs for decades. However, due to the limitation of current classification systems, novel classification which could perform more easily and have better prognostic predictive effect was urgent needed. To this end, we tried to establish a new classification system of human TETs based on our above findings, especially the unique epithelial origin of tumor cells in each type of TETs which played pivotal role in T cell developmental dysfunction and immune response in tumor microenvironment. Firstly, we selected representative genes based on scRNA-seq results which represented the unique epithelial cell type in each TETs types (Fig. SF) and found that GNB3 and CHI3L1 were expressed highly in tumor of type 1 and type 3 TETs respectively (Fig. 7A and 7B). Consistently, the expression of GNB3 and CHI3L1 in tumor was further validated by IF staining (Fig. 7C). Therefore, we reclassified 119 TETs patients of TCGA cohort into three groups according to the expression level of GNB3 and CHI3L1 (Type 1, GNB<sup>high</sup>; Type 2, GNB<sup>low</sup> CHI3L<sup>1low</sup>; Type 3, GNB<sup>low</sup> CHI3L<sup>1high</sup>) (Supplementary Fig. S8A). Kaplan-Meier survival analysis demonstrated that the classification we established closely correlated with survival of TETs patients (Fig. 7D; P = 0.014). Moreover, it was surprising that our classification based on GNB3 and CHI3L1 expression showed an advantage in the prognostic prediction of TETs patients compared with both Masaoka stage and WHO classification (Supplementary Fig. S8B and S8C).

Finally, based on above findings we proposed a tripartite frame to explain the origin of tumor cells and its impact on T cell development in tumor of human TETs as below (Fig. 7E): 1) For the type 1 TETs: risk factors induced the mutation of oncogenes in original TECs, which resulted in the activation of NF-κB signaling pathway in mcTECs-like cells and expansion of KRT14/ GNB3<sup>+</sup> mTECs-like tumor cells. The expanded KRT14/GNB3<sup>+</sup> mTECs-like tumor cells led to a bias of cellular composition and defect of cortex like structure in tumor. The lack of CCL25<sup>+</sup> cTECs-like cells and cortex like structure resulted in restraining T cell development, which was mainly in positive selection phase. Consequently, CD8<sup>+</sup>CD4<sup>+</sup> T cells (DP) accumulated, but CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cells (SP) decreased in tumor. 2) For type 2 TETs: risk factors induced the mutation of oncogenes in original TECs, which resulted in the inhibition of NF-κB signaling pathway in mcTECs-like cells and expansion of CCL25<sup>+</sup> cTECs-like cells. The expanded CCL25<sup>+</sup> cTECs-like tumor cells led to a bias of cellular composition and defect of medulla like structure in tumor. The accumulation of CCL25<sup>+</sup> cTECs-like cells accelerated the positive selection phase of T cell development in tumor. However, lack of KRT14/ GNB3<sup>+</sup> mTECs-like cells and medulla like structure resulted in suppression of T cell development in negative selection phase. Consequently, CD8<sup>+</sup>CD4<sup>+</sup> T cells (DP) decreased, but CD8<sup>+</sup> T cells and CD4<sup>+</sup> T cells (SP) increased in tumor. 3) For type 3 TETs: risk factors induced the mutation of oncogenes in original TECs, which resulted in the transforming and expanding of CHI3L<sup>1+</sup> mTECs-like cells in tumor. The malignant transformation of original TECs also led to the deficiency of functional mTECs, cTECs, cortex-like structure and medulla-like structure in tumor which resulted in deficiency of T cell development. However, CHI3L<sup>1+</sup> mTECs-like tumor cells of type 3 TETs
could activate the CD8$^+$ T$_{RM}$ cells and recruit other immune cells to form a unique immune microenvironment, such as effector T cells, B cells, DC cells and macrophages.

**Discussion**

TETs are common tumors originated from human thymus, the central lymphoid organ which is essential for T cell development. The particular origin of TETs suggests that the development of tumor may be involved in T cell developmental dysfunction. Through decoding the immune landscape and cell atlas of tumors, we successfully reclassified human TETs into three types. We linked the immune landscape and T cell development pattern of each tumor type to the dysfunction of T cell development which was controlled by its epithelial origin. Furthermore, we decoded the specific epithelial origin of tumors and revealed the underlying mechanisms of tumor cell-driven T cell developmental dysfunction in each tumor type of TETs. Next, we uncovered the consequence of tumor cell-mediated T cell developmental dysfunction in each tumor type through TCR repertoire, diversity and related signal pathway. In addition, we decoded a unique immune response pattern of type 3 TETs which was mainly orchestrated by tumor-induced CD8$^+$ T$_{RM}$ cell activation. Finally, we established a new classification of human TETs using the molecules which reflected the epithelial origin of each tumor type and has shown advantage in prognostic prediction.

TETs are heterogeneous tumors considered to originate from TECs. However, exactly cellular origin of TETs is largely unknown. Growing evidence has shown that TECs in human thymus are extraordinarily heterogeneous. Therefore, it is reasonable to speculate that the epithelial origin of human TETs may also be heterogeneous. The findings of discrepant tumor cell composition among human TETs types in our study supported this hypothesis (Fig. 5E). Besides, we identified the exact tumor cell subset of each human TETs type which represented its epithelial origin, for instance KRT14$^+$/GNB3$^+$ mTEC-like cells in type 1, CCL25$^+$ cTEC-like cells in type 2 and CHI3L1$^+$ mTEC-like cells in type 3 (Fig. 5E and 5G). These findings in our study suggested that there was heterogeneity in epithelial origin of human TETs, which may determine the biological nature and prognosis of human TETs.

Genetic aberrations were reported in human TETs a decade ago. Subsequently, a specific missense mutation in GTF2I was demonstrated to occur at high frequency in type A and type AB thymomas, but rarely in the aggressive subtypes. In contrast, thymic carcinomas carried a higher number of mutations in recurrent mutations of known cancer genes than thymomas, such as TP53, CYLD, CDKN2A, BAP1 and PBRM1. Consistently, another study identified the recurrent somatic mutations in TET2, CYLD, SETD2, TP53, FBXW7, HRAS and RB1 in thymic carcinoma, and no mutations in GTF2I. However, the details of how these driver oncogenes promote tumor developing and the key driver genes in each type of TETs are still unknown. Recently, GTF2I mutation was reported to induce the cell transformation and metabolic alterations of TECs, which provided direct evidence of GTF2I mutation in promoting tumorigenesis of TETs. Through decoding the epithelial origin of each human TETs type (Fig. 5E-5G), we identified GNB3 and CHI3L1 which specifically expressed in type 1 or type 3 tumors respectively in our study (Fig. 7A-7C).
These evidences suggest that \textit{GNB3} and \textit{CHI3L1} may be the driven oncogene in original TECs of type 1 and type 3 of human TETs respectively.

There are various histological classifications and clinical staging systems of human TETs in past decades, including WHO classification and Masaoka stage\textsuperscript{9}. Despite the upgrading of these classification systems in the past years, some controversies are long-term unresolved\textsuperscript{17,18,41}. One fundamental defect of current classification systems is that they are based on the morphology of the epithelial cells in tumors of TETs, but not biological characteristics. Recently, genomic analysis of tumors showed that according to the mutation of \textit{GTF2I}, T cell signaling mRNA signature and somatic copy number alterations (SCNA) levels, human TETs could be classified into four types which associated with disease-free and overall survival of patients\textsuperscript{19}. Using multi-platform omics analyses on 117 TETs, a pioneer study uncovered the integrated genomic landscape of TETs and established a new molecular classifications based on the gene mutation and expression level in tumor mass, which also could predict the outcome and incidence of autoimmune diseases of human TETs\textsuperscript{20}. In our study, we established a new classification of human TETs according to the molecules specifically expressed in the unique tumor cell subset which reflected the epithelial origin of each tumor type (Fig. 7A-7C and Supplementary Fig. S8A). We observed that our classification of TETs had advantage in prognostic prediction when compared with WHO classification and Masaoka stage in the same cohort of patients (Fig. 7D, Supplementary Fig. S8B and S8C). Our findings support the idea that the classification of TETs derived from biological characteristics of tumors, especially key molecular events occurred in epithelial origin of tumor, will predict the outcome and guide clinical practices more accurately.

High incidence rate of autoimmune disease is one of the most remarkable clinical features of human TETs\textsuperscript{10}. Recently, emerging evidences showed that acquired T cell deficiency also occurred in patients with TETs\textsuperscript{42}. In addition, the immune disorders of patients with TETs have obvious bias among tumor types, such as high incidence rate of autoimmune disease in type B tumors according to WHO classification\textsuperscript{14}. However, the cellular and molecular mechanisms of immune disorders in TETs were still largely unclear. We speculated the epithelial origin of human TETs, which play an essential role in thymic T cell development and central self-tolerance\textsuperscript{13}, accounted for the generation of immune disorders in patients. As expected, we found that cTEC-like cells increased in the tumor of type 2 TETs, in contrast mTEC-like cells decreased (Fig. 5G). Consistently, AIRE, the key molecule for negative selection of T cells, was not expressed in the major subsets of tumor cells in type 2 TETs (Supplementary Fig. S6D, S6E and S6H). These evidences suggested that the increased cTEC-like cells, decreased mTEC-like cells and decreased AIRE expression in tumor of type 2 TETs led to the dysfunction of central self-tolerance and generation of autoimmune disease in patients. In addition, the finding of relative lack of mature T cells in tumor (Fig. 2D and 2E), suggested the defects of T cell maturity in tumor may account for the generation of acquired T cell deficiency in type 1 TETs. These findings in our study provided a new clue to understand the generation of immune disorders in TETs patients.
Taken together, we reported a comprehensive study firstly using multiple omics to decode the biological characteristics of human TETs from immune phenotype to its underline mechanism orchestrated by the epithelial origin of tumors. We uncovered an epithelium-T cell loop which depicted the core discrepancy of biological characteristics among tumor types and reclassified human TETs. Finally, according to the unique gene expression in the identified epithelial origin of each tumor type, we successfully established a convenient molecular classification of human TETs which showed advantage in prognostic prediction. We believed that our findings provided insights into understanding the biology of human TETs and will shed more light on the translational research of human TETs.

Methods

Patients and specimens

Tumor tissues (T, homogeneous cellularity, without foci of necrosis) were obtained from patients with TETs who underwent surgical resection at the department of Thoracic Surgery, Second Affiliated Hospital, Zhejiang University School of Medicine between September 2020 and June 2021. None of the patients had received radiotherapy or chemotherapy before operation. The pathological results of all patients were identified as TETs. The normal human thymus samples were obtained from the department of Cardiac Surgery, Second Affiliated Hospital, Zhejiang University following cardiothoracic surgery from adult with heart disease, as the thymic tissue is routinely removed and discarded to gain adequate exposure of the retrosternal operative field. All samples were anonymously coded in accordance with local ethical guidelines (as stipulated by the Declaration of Helsinki), and written informed consent was obtained and the protocol was approved by the Review Board of the Second Affiliated Hospital of Zhejiang University School of Medicine.

Preparation of cell suspensions

Freshly excised tissues were stored in sterile RPMI (Corning) supplemented with 10% FBS (Life Technologies), 1% Streptomycin and penicillin (Life Technologies) and processed within 2 hours. The tissues were cut into small pieces and then digested in RPMI containing 10% FBS, type I collagenase (1 mg/ml), type IV collagenase (1 mg/ml) for 1 hour at 37°C using the gentleMACS™ Dissociator (Miltenyi Biotech) according to manufacturer’s instructions. The resulting single-cell suspension was filtered sequentially through sterile 70µm cell strainers. Then, cell suspensions were stored in complete media at 4°C for next experiments.

HE staining and image acquisition

For the HE analyses, TETs and normal thymus tissues were fixed in 10% formalin, embedded in paraffin and sectioned transversely. The 5µm-thick formalin-fixed paraffin-embedded (FFPE) sections on glass slides were backed at 37°C overnight, dewaxed in xylene, rehydrated through decreasing concentrations of ethanol, and washed in PBS. Then stained with hematoxylin and eosin. After staining, sections were
dehydrated through increasing concentrations of ethanol and xylene. The slides were scanned with the KF-PRO-120 scanner (KFBIO).

**Immunofluorescent staining and image acquisition**

The 5µm-thick FFPE sections were also backed at 37°C overnight, dewaxed in xylene, rehydrated through decreasing concentrations of ethanol, and washed in PBS. Antigen retrieval was performed in Target Retrieval Solution (TRS) citrate buffer (pH 6.0) using a pressure cooker. Then, they were blocked with a blocking buffer solution (5% FBS, 1% BSA and 0.2% Triton) for 2 h at room temperature and incubated with primary antibody in blocking buffer (BioLegend) at 4°C overnight. After washed by PBS (pH 7.4), secondary antibody staining was performed at room temperature for 1 h. Next, sections were washed three times with PBST and DAPI reagent was added for 10 min to detect cell nuclei. The antibodies of CD3 (ab16669, diluted at 1:100), CD8a (ab17147, diluted at 1:100), CD4 (ab133616, diluted at 1:500), CD103 (ab129202, diluted at 1:800), CD11c (ab52632, diluted at 1:500), CD20 (ab778237, diluted at 1:2000) and EPCAM (ab223582, diluted at 1:500) were procured from Abcam. The antibodies of CD45 (60287-1-IG, diluted at 1:500), CCL25 (25285-1-AP, diluted at 1:500), CHI3L1 (12036-1-AP, diluted at 1:400), AIRE (22517-1-AP, diluted at 1:500), KRT14 (60320-1-IG, diluted at 1:800) and GNB3 (12036-1-AP, diluted at 1:400) were procured from Proteintech Group. Epifluorescence multispectral whole slide images of all sections were acquired through the NIKON ECLIPSE C1 system (Nikon Corporation) and scanned at high resolution on a Pannoramic SCAN II system (3DHISTECH Ltd.).

**Flow Cytometry**

The antibodies CD45-BV510 (HI30), CD3-BV421 (UCHT1), CD4-APC (RPA-T4), CD8a-APCCY7 (RPA-T8), CD103-BV605 (Ber-Act8) and CD69-PECY7 (FN50) were procured from BioLegend. We preincubated fresh tissue cells (1×10^6/ml) in a mixture of PBS, 2% fetal calf serum, and 0.1% (w/v) sodium azide with FcγIII/IIIR-specific antibody to block nonspecific binding and stained with different combinations of fluochrome-coupled antibodies for 15 minutes at room temperature. Cells were washed by PBS, passed through a 70-µm filter, and data was collected on a FACSCanto II system and FACSFortessa system (BD Biosciences) and analyzed using FlowJo software (version 10.0.7).

**Mass Cytometry (CyTOF) sample preparation**

42 metal-conjugated antibodies used in this study were showed in supplementary Table 1. Briefly, the cells derived from the TETs samples were stained with 5µM cisplatin (Fluidigm) in PBS without BSA for viability staining. Then, samples were washed in PBS containing 2.5% BSA and blocked for 30 minutes at 4°C. After that, they were stained with cell-surface antibodies in PBS containing 5% goat serum and 30% bovine serum albumin (BSA) for 30 min at 4°C. Next, samples were washed, fixed and permeabilized using the Foxp3 fix and permeabilization kit (eBioscience) as well as 100nM Iridium nucleic acid intercalator (Fluidigm) according to the manufacturer's instructions at 4°C overnight. Cells were then washed twice with Foxp3 permeabilization buffer and incubated with intracellular antibodies in permeabilization buffer for 30min at 4°C. Finally, cells were washed twice with ddH₂O to prepare for analysis.
CyTOF Data Acquisition and analysis

Right before acquisition, samples were washed and resuspended at a concentration of 1 million cells/ml in water containing EQ Four Element Calibration Beads (Fluidigm). Samples were acquired on a Helios CyTOF System (Fluidigm) at an event rate of < 500 events/second. EQ beads (Fluidigm) were used for a loading control. All data were produced on a Helio3 CyTOF Mass Cytometer (Fluidigm). Mass cytometry data files were normalized using the bead-based normalization software, which uses the intensity values of a sliding window of these bead standards to correct for instrument fluctuations over time and between samples. CyTOF analyses were performed by PLTtech Inc (Hangzhou, China) according to the previously described protocol 43. The data were gated to exclude residual normalization beads, debris, dead cells and doublets for subsequent clustering and high dimensional analyses. The 42 immune cell markers were all applied for clustering and visualization. Phenograph algorithm 44 was used to cluster cells. 50,000 cells were selected randomly for visualizing by t-distributed stochastic neighbor embedding (t-SNE) dimension reduction algorithm 45 using R package cytofkit (version 0.13). Immune subset cells were defined by the median values of specific expression markers on Hierarchical clustering. Heatmaps of normalized marker expression, relative marker expression, and relative difference of population frequency were generated using pHeatmap R package and Python (https://www.python.org/) respectively. The comparisons between two groups were assessed by unpaired Student’s t-tests using GraphPad Prism (v8). The use of these tests was justified based on assessment of normality and variance of the distribution of the data.

Single-cell RNA Sequencing

Using Trypan Blue for the quality check of single cell suspension prepared as outlined earlier, the cell survival rate is generally above 80%. The cells that have passed the test are washed and resuspended to prepare a suitable cell concentration of 1000 cells/ul for 10x Genomics Chromium™. Approximately 10,000 cells were loaded onto the 10X Chromium Single Cell Platform (Single Cell 5’ library and Gel Bead Kit v.3) as described in the manufacturer’s protocol. Generation of gel beads in emulsion (GEMs), barcoding, GEM-RT clean-up, complementary DNA amplification and library construction were all performed as the manufacturer’s protocol. Qubit (Agilent 2100) was used for library quantification before pooling. The final library pool was sequenced on the Illumina NovaSeq 6000 instrument using 150-base-pair paired-end reads.

scRNA-seq clustering analysis

The LogNormalize method of the "Normalization" function of the Seurat software was used to calculate the expression value of genes. PCA (Principal component analysis) analysis was performed using the normalized expression value. Within all the PCs, the top 10 PCs were used to do clustering and t-SNE analysis. We used the Shared Nearest Neighbor (SNN) graph-based clustering algorithm from the Seurat package (FindClusters function with default parameters) to identify the clusters of cells. Of these, 8 clusters demonstrated high expression of genes specific to epithelial and T cell and thus were selected
for a final re-clustering analysis. These clusters were then identified as distinct cell subtypes based on literature review of highly expressed genes.

**Single-cell gene expression analysis**

Raw gene expression matrices were generated using CellRanger (10x Genomics) and analyzed using the Seurat v3 R package. All cells expressing < 500 genes were removed, as well as cells that contained < 500 unique molecular identifiers (UMIs) and > 25% mitochondrial counts. Samples were merged and normalized. Because every cell has a unique barcode, scRNA-seq data could be linked with the scTCR-seq data.

**Differential expression and pathway analysis**

DEGs were identified using the bimod test with FindMarkers and FindAllMarkers functions in Seurat with a threshold for log₂ (fold change) (log₂FC) ≥ 0.26 and for expression in more than 10% of cells. Besides, we selected 0.01 as the cutoff of p value. KEGG (10.1093/nar/gkw1092) analyses were performed using the OmicStudio tools (https://www.omicstudio.cn/tool) to identify gene set enrichment with hypergeometric test.

**TCR analysis**

Full-length TCR V(D)J segments were enriched from amplified cDNA via PCR amplification using a Chromium Single-Cell V(D)J Enrichment kit according to the manufacturer’s protocol (10X Genomics). 10X TCR-enriched libraries are mapped with the Cell Ranger Single-Cell Software Suite (version 4.0.0, 10x Genomics) to the custom reference provided by the manufacturer (version 2.0.0 GRCh38 VDJ reference). All assembled contigs were filtered to retain only those that were assigned a raw clonotype ID and categorized as being both full length and productive. Each clonotype was assigned a unique identifier, consisting of the predicted amino acid sequences of the CDR3 regions of these two chains, which was used to match clonotypes across samples. Clonality, which reflects the dominance of particular clones across the TCR repertoire, was calculated per sample. To visualize the degree of TCR clonotypes shared between T cell phenotypes, we used barcode information to project T cells with prevalent TCR clonotypes on UMAP plots.

**Pseudotime reconstruction and trajectory inference**

The R package Monocle (version 2) algorithm was used to reconstruct pseudotime trajectories to determine the potential lineage development among diverse T cell subsets and epithelial cell subsets. For each analysis, PCA-based dimension reduction was performed on DEGs of each phenotype, followed by two-dimensional visualization on UMAP. Graph-based clustering (Louvain) identified T cell into twelve subclusters. The cell differentiation trajectory was then captured using the orderCells function. As before for differential gene expression analyses, the Rpackage MAST was used to detect genes significantly covarying with pseudotime, based on a log-likelihood ratio test between the model formula including cell pseudotime and a reduced model formula. Additional model covariates were included in the residual model formula. Benjamini–Hochberg multiple testing correction was used to calculate FDR, and genes <
5% FDR were considered to vary significantly with pseudotime. For T cells from different groups, the same process with the same signature genes and Monocle parameters was used to construct the clone-based trajectories.

**Cell-to-cell communication of scRNA-seq data**

CellphoneDB ([www.CellPhoneDB.org](http://www.CellPhoneDB.org)) was used to assess putative receptor–ligand interactions between epithelial cell and T cell subsets, CD8+ T<sub>RM</sub> cell and other cell subsets (epithelial cell, B cell, DC, Fibroblast cell, monocyte, macrophage, VSMC). Briefly, the algorithm allows the detection of ligand–receptor interactions between cell types in scRNA-seq data using the statistical framework described in previous studies. The tool was run for 1000 iterations. The normalized interaction score has been calculated by multiplying the average expression level of ligand and receptors for all cell pairs, and maximum normalizing these values. The total number of pairwise paracrine interactions between CD8+ T<sub>RM</sub> cell and epithelial cell subsets obtained using the CellphoneDB scoring method were visualized as heatmaps in the R package pheatmap.

**Survival analysis**

The TCGA TETs data were used to evaluate the prognostic effect of individual genes or gene sets derived from specific cell clusters. The gene expression and survival data were downloaded from UCSC Xena ([http://xena.ucsc.edu/](http://xena.ucsc.edu/)). 119 patients with records of *GNB3* and *CHI3L1* gene expression information were included into the survival analysis (of these, 117 with the WHO classification and Mosaok stages). The expression profile was normalized by log2 (normalized_count + 1) to exclude potential bias. For each gene, patient cohorts were grouped into high and low expression groups by the top 20% value of the normalized average expression. Kaplan–Meier survival curves were plotted to show differences in survival time, and P values reported by the Log-rank test using GraphPad Prism (v8).

**Statistical Analysis**

Statistical analysis was performed using Prism version 8.3 (GraphPad Software). Unpaired Student’s t tests were used where comparisons were used to identify differences between two groups. All results are presented as mean ± SEM, and P values of < 0.05 were considered statistically significant (*P* < 0.05, **P** < 0.01, ***P*** < 0.001 and ****P*** < 0.0001).

**Declarations**

**ACKNOWLEDGMENTS**

We are grateful to the tissue donors and the clinical staff at The Second Affiliated Hospital of Zhejiang University School of Medicine who made tissue collection possible. We thank Jian Huang lab for technical assistance and helpful supports. We thank all the members of the Pin Wu lab for discussions and comments on the manuscript. This work was supported by funding from the National Natural Science Foundation of China (81572800 and 82073141 to PW, 82073142 to DW), the Fundamental
Research Funds for the Central Universities (2019QNA7025 to PW) and the Natural Science Foundation of Zhejiang Province (LY15H160041 to PW, LY19H160050 to DW).

AUTHOR CONTRIBUTIONS

PW designed the study. ZWX, MJL, HZX, DC, YYC and XKC performed experiments. ZWX, MJL and HZX collected the samples and clinical data. XX and JFL reviewed the histological type of samples. ZWX and PW analyzed the data, interpreted the data, and wrote the manuscript. YC and PW supervised the project, all authors reviewed and edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

References

1. Engels EA. Epidemiology of thymoma and associated malignancies. J Thorac Oncol 2010;5(10 Suppl 4):S260-5. doi: 10.1097/JTO.0b013e3181f162f2d [published Online First: 2010/10/05]
2. de Jong WK, Blaauwgeers JL, Schaapveld M, et al. Thymic epithelial tumours: a population-based study of the incidence, diagnostic procedures and therapy. Eur J Cancer 2008;44(1):123–30. doi: 10.1016/j.ejca.2007.11.004 [published Online First: 2007/12/11]
3. Marx A, Porubsky S, Belharazem D, et al. Thymoma related myasthenia gravis in humans and potential animal models. Exp Neurol 2015;270:55–65. doi: 10.1016/j.expneurol.2015.02.010
4. Huang B, Belharazem D, Li L, et al. Anti-Apoptotic Signature in Thymic Squamous Cell Carcinomas - Functional Relevance of Anti-Apoptotic BIRC3 Expression in the Thymic Carcinoma Cell Line 1889c. Front Oncol 2013;3:316. doi: 10.3389/fonc.2013.00316
5. Ehemann V, Kern MA, Breinig M, et al. Establishment, characterization and drug sensitivity testing in primary cultures of human thymoma and thymic carcinoma. International journal of cancer 2008;122(12):2719–25. doi: 10.1002/ijc.23335
6. Gokmen-Polar Y, Sanders KL, Goswami CP, et al. Establishment and characterization of a novel cell line derived from human thymoma AB tumor. Lab Invest 2012;92(11):1564–73. doi: 10.1038/labinvest.2012.115
7. Marx A, Chan JK, Coindre JM, et al. The 2015 World Health Organization Classification of Tumors of the Thymus: Continuity and Changes. J Thorac Oncol 2015;10(10):1383–95. doi: 10.1097/jto.0000000000000654 [published Online First: 2015/08/22]
8. Strobel P, Bauer A, Puppe B, et al. Tumor recurrence and survival in patients treated for thymomas and thymic squamous cell carcinomas: a retrospective analysis. J Clin Oncol 2004;22(8):1501–9. doi: 10.1200/JCO.2004.10.113 [published Online First: 2004/04/16]
9. Conforti F, Pala L, Giaccone G, et al. Thymic epithelial tumors: From biology to treatment. Cancer treatment reviews 2020;86:102014. doi: 10.1016/j ctrv.2020.102014
10. Abramson J, Anderson G. Thymic Epithelial Cells. *Annu Rev Immunol* 2017;35:85–118. doi: 10.1146/annurev-immunol-051116-052320 [published Online First: 2017/02/23]

11. Cumano A, Berthault C, Ramond C, et al. New Molecular Insights into Immune Cell Development. *Annu Rev Immunol* 2019;37:497–519. doi: 10.1146/annurev-immunol-042718-041319 [published Online First: 2019/04/27]

12. Starr TK, Jameson SC, Hogquist KA. Positive and negative selection of T cells. *Annual review of immunology* 2003;21:139–76. doi: 10.1146/annurev.immunol.21.120601.141107

13. Shelly S, Agmon-Levin N, Altman A, et al. Thymoma and autoimmunity. *Cellular & molecular immunology* 2011;8(3):199–202. doi: 10.1038/cmi.2010.74 [published Online First: 2011/02/15]

14. Weksler B, Lu B. Alterations of the immune system in thymic malignancies. *J Thorac Oncol* 2014;9(Suppl 2):S137-42. doi: 10.1097/jto.0000000000000299 [published Online First: 2014/11/15]

15. Zucali PA, Di Tommaso L, Petrini I, et al. Reproducibility of the WHO classification of thymomas: practical implications. *Lung Cancer* 2013;79(3):236–41. doi: 10.1016/j.lungcan.2012.11.015 [published Online First: 2013/01/03]

16. Kelly RJ, Petrini I, Rajan A, et al. Thymic malignancies: from clinical management to targeted therapies. *J Clin Oncol* 2011;29(36):4820–7. doi: 10.1200/JCO.2011.36.0487 [published Online First: 2011/11/23]

17. Roden AC, Yi ES, Jenkins SM, et al. Modified Masaoka stage and size are independent prognostic predictors in thymoma and modified Masaoka stage is superior to histopathologic classifications. *J Thorac Oncol* 2015;10(4):691–700. doi: 10.1097/jto.0000000000000482 [published Online First: 2015/01/30]

18. Weis CA, Yao X, Deng Y, et al. The impact of thymoma histotype on prognosis in a worldwide database. *J Thorac Oncol* 2015;10(2):367–72. doi: 10.1097/JTO.0000000000000393

19. Lee HS, Jang HJ, Shah R, et al. Genomic Analysis of Thymic Epithelial Tumors Identifies Novel Subtypes Associated with Distinct Clinical Features. *Clinical cancer research: an official journal of the American Association for Cancer Research* 2017;23(16):4855–64. doi: 10.1158/1078-0432.CCR-17-0066

20. Radovich M, Pickering CR, Felau I, et al. The Integrated Genomic Landscape of Thymic Epithelial Tumors. *Cancer Cell* 2018;33(2):244 – 58.e10. doi: 10.1016/j.ccell.2018.01.003 [published Online First: 2018/02/14]

21. Park JE, Botting RA, Dominguez Conde C, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science* 2020;367(6480) doi: 10.1126/science.aay3224

22. Masopust D, Soerens AG. Tissue-Resident T Cells and Other Resident Leukocytes. *Annual review of immunology* 2019 doi: 10.1146/annurev-immunol-042617-053214 [published Online First: 2019/02/07]

23. Dik WA, Pike-Overzet K, Weerkamp F, et al. New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J Exp Med* 2005;201(11):1715–23. doi: 10.1084/jem.20042524 [published Online First: 2005/06/02]
24. Taghon T, Waegemans E, Van de Walle I. Notch signaling during human T cell development. *Curr Top Microbiol Immunol* 2012;360:75–97. doi: 10.1007/82_2012_230 [published Online First: 2012/06/14]

25. Weerkamp F, van Dongen JJ, Staal FJ. Notch and Wnt signaling in T-lymphocyte development and acute lymphoblastic leukemia. *Leukemia* 2006;20(7):1197 – 205. doi: 10.1038/sj.leu.2404255 [published Online First: 2006/05/12]

26. Aspinall R. T cell development, ageing and Interleukin-7. *Mech Ageing Dev* 2006;127(6):572–8. doi: 10.1016/j.mad.2006.01.016 [published Online First: 2006/03/15]

27. Carico ZM, Roy Choudhury K, Zhang B, et al. Tcrd Rearrangement Redirects a Processive Tcra Recombination Program to Expand the Tcra Repertoire. *Cell Rep* 2017;19(10):2157–73. doi: 10.1016/j.celrep.2017.05.045 [published Online First: 2017/06/08]

28. Kadouri N, Nevo S, Goldfarb Y, et al. Thymic epithelial cell heterogeneity: TEC by TEC. *Nat Rev Immunol* 2020;20(4):239–53. doi: 10.1038/s41577-019-0238-0

29. Vento-Tormo R, Efremova M, Botting RA, et al. Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* 2018;563(7731):347–53. doi: 10.1038/s41586-018-0698-6

30. Takaba H, Takayanagi H. The Mechanisms of T Cell Selection in the Thymus. *Trends Immunol* 2017;38(11):805–16. doi: 10.1016/j.it.2017.07.010 [published Online First: 2017/08/24]

31. Bunting MD, Comerford I, McColl SR. Finding their niche: chemokines directing cell migration in the thymus. *Immunol Cell Biol* 2011;89(2):185–96. doi: 10.1038/icb.2010.142 [published Online First: 2010/12/08]

32. Chopp LB, Gopalan V, Ciucci T, et al. An Integrated Epigenomic and Transcriptomic Map of Mouse and Human alphabeta T Cell Development. *Immunity* 2020;53(6):1182 – 201 e8. doi: 10.1016/j.immuni.2020.10.024 [published Online First: 2020/11/27]

33. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature* 2017;541(7637):321–30. doi: 10.1038/nature21349

34. Masopust D, Soerens AG. Tissue-Resident T Cells and Other Resident Leukocytes. *Annu Rev Immunol* 2019;37:521–46. doi: 10.1146/annurev-immunol-042617-053214

35. Miller J. The function of the thymus and its impact on modern medicine. *Science* 2020;369(6503) doi: 10.1126/science.aba2429

36. Oramas DM, Moran CA. Thymoma: Histologically a heterogenous group of tumors. *Semin Diagn Pathol* 2021 doi: 10.1053/j.semdp.2021.06.002

37. Petrini I, Meltzer PS, Kim IK, et al. A specific missense mutation in GTF2I occurs at high frequency in thymic epithelial tumors. *Nature genetics* 2014;46(8):844–9. doi: 10.1038/ng.3016
39. Saito M, Fujiwara Y, Asao T, et al. The genomic and epigenomic landscape in thymic carcinoma. *Carcinogenesis* 2017;38(11):1084–91. doi: 10.1093/carcin/bgx094 [published Online First: 2017/10/03]

40. Kim IK, Rao G, Zhao X, et al. Mutant GTF2I induces cell transformation and metabolic alterations in thymic epithelial cells. *Cell death and differentiation* 2020;27(7):2263–79. doi: 10.1038/s41418-020-0502-7

41. Ahmad U. The eighth edition TNM stage classification for thymic tumors: What do I need to know? *The Journal of thoracic and cardiovascular surgery* 2021;161(4):1524–29. doi: 10.1016/j.jtcvs.2020.10.131

42. Christopoulos P, Fisch P. Acquired T-Cell Immunodeficiency in Thymoma Patients. *Critical reviews in immunology* 2016;36(4):315 – 27. doi: 10.1615/CritRevImmunol.2017018916 [published Online First: 2016/01/01]

43. Han G, Spitzer MH, Bendall SC, et al. Metal-isotope-tagged monoclonal antibodies for high-dimensional mass cytometry. *Nat Protoc* 2018;13(10):2121–48. doi: 10.1038/s41596-018-0016-7 [published Online First: 2018/09/28]

44. Levine JH, Simonds EF, Bendall SC, et al. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* 2015;162(1):184–97. doi: 10.1016/j.cell.2015.05.047 [published Online First: 2015/06/23]

45. Amir E-aD, Davis KL, Tadmor MD, et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature Biotechnology* 2013;31(6):545–52. doi: 10.1038/nbt.2594

46. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell* 2019;177(7):1888 – 902 e21. doi: 10.1016/j.cell.2019.05.031

47. Butler A, Hoffman P, Smibert P, et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36(5):411–20. doi: 10.1038/nbt.4096

48. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* 2017;14(10):979–82. doi: 10.1038/nmeth.4402

49. Efremova M, Vento-Tormo M, Teichmann SA, et al. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat Protoc* 2020;15(4):1484–506. doi: 10.1038/s41596-020-0292-x

Figures
Figure 1

Cellular composition of the human thymus and TETs. See also Supplementary Fig. S1, and Supplementary Tables S1 and S2. (A) Scheme of the overall study design and workflow. (B) Reduced-dimensional single-cell t-SNE maps of total cells from human TETs (n = 12) and normal thymus (n=3) by CyTOF, each color represents an independent sample. (C) Representative HE and immunofluorescence (red for EpCAM, green for CD45, pink for CD3 and blue for DAPI) stained sections of TETs and normal
human thymus (scale bar: 200μm). (N, normal human thymus). (D) t-SNE plot of immune cells (EPCAM-
CD45+) from all samples, colored and labeled by cell types. (E) t-SNE analysis of immune cells from all
samples colored by relative expression of CyTOF markers (CD45, CD3, CD4, CD8a, CD20, CD14, CD16,
CD56, CD11c, CD66b, HLA-DR and CD39). (F) t-SNE plots of immune cells from samples of each group
(N, T1, T2, T3), colored by cell types. (N, normal human thymus; T1, type 1; T2, type 2; T3, type 3; the
abbreviations below are consistent here). (G) Bar plots showing the frequencies of main immune cell
subsets among the four groups of samples (n=3 N, 4 T1, 4 T2 and 4 T3. Data are presented as mean ±
s.e.m. P values were determined by two-tailed Student’s t-test; ns, nonsignificant, *P < 0.05 and ****P <
0.0001). (H) Representative flow cytometric plot of CD3+ T cell and CD3-CD4+CD8+ lymphocytes of TETs
in each type of TETs.
Figure 2

T cell subsets in thymus and TET subtypes. See also Supplementary Fig. S2. (A) t-SNE plot of CD45+CD3+ T cells from all samples, colored by cell subsets. (DN, CD3+CD4-CD8-; DP, CD3+CD4+CD8+). (B) Heatmap showing relative normalized protein expression on the T cell subsets. (C) t-SNE plots of CD45+CD3+ T cells from samples of each group (N, T1, T2, T3), colored by cell subsets. (D) Composition of CD45+CD3+ T cell compartment showing average frequencies of major T cell subsets for each group.
The color same as (C). (E) Bar plots showing the frequencies of main T cell subsets among the four groups of samples (n=3 N, 4 T1, 4 T2 and 4 T3. Data are presented as mean ± s.e.m. P values were determined by two-tailed Student’s t-test; ns, nonsignificant, *P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001). (DP, CD3+CD4+CD8+). (F) Representative flow cytometric plot of CD3+CD4-CD8- T cell, CD3+CD4+ T cell, CD3+CD8+ T cell and CD3+CD4+CD8+ T cell of TETs in three types. (G) Representative immunofluorescence (IF) staining images showing CD3 (pale blue), CD8 (green), CD4 (red) and DAPI (nuclei, blue) in normal human thymus and TET samples.
Figure 3

T cell development dysfunction in thymus and TETs subtypes. See also Supplementary Fig. S3 and Supplementary Tables S3 and S5. (A) Uniform manifold approximation and projection (UMAP) visualization of total T cells from the normal human thymus (n=1) and TETs samples (n=6), colored by the identified cell subpopulation. (DN, double-negative T cells; DP, double-positive T cells). (B) Dot plot for marker genes expression in T cell types. Here and in later Fig., color represents maximum normalized mean expression of marker genes in each cell subgroup, and size indicates the proportion of cells expressing marker genes. (C) Same UMAP plot as (A), colored by groups. (D) Same UMAP plots of total T cells from samples of each group, colored by the identified cell subpopulation same as (A). (E) Composition of the T cell compartment showing average frequencies of major T cell subpopulations for each group (n=1 N, 2 T1, 2 T2 and 2 T3). (DN, double-negative T cells; DP, double-positive T cells). (F) Pseudotime trajectory for T cells from all samples in a two-dimensional state-space defined by Monocle2, colored by identified T cell subpopulations. (G) The same pseudotime plots for T cells from samples of each group, colored by group. (H-J) Dot plots depicting the relative level of expression of selected NOTCH (H), WNT (I), or IL-7 (J) signaling pathway genes of total T cell in each group.
Figure 4

VJ gene usage, pairing and TCR diversity in TETs subtypes. See also Supplementary Fig. S4. (A) Heat map showing the normalized proportion of each TCRβ V and J gene segment present at total T cells in each group. Gene segments positioned according to genomic location. (B) Frequency of V-J gene pairs in TCRβ locus of each group. (C) Same scheme as in (A) applied to TCRα V and J gene segments. (D) Same scheme as in (E) applied to TCRα V- J gene pairs. (E) UMAP plot of all T cells from TET samples (n=6),
colored by the number of TCR clonotypes detected. (F) Bar plot showing clonotype diversity among the three groups of TET samples (n=2 T1, 2 T2 and 2 T3).

**Figure 5**

The mechanism of tumor-driven T cell developmental dysfunction in each TETs type. See also Supplementary Fig. S5, Supplementary Table S6 and S7. (A) Same UMAP plot as (Fig. S3C) of cellular composition from samples of each group, colored by cell types of epithelial cell and other cell. (B) Bar
plot showing the frequencies of epithelial cell among the groups of TETs (n=2 T1, 2 T2 and 2 T3). (C) Violin plots showing expression of the marker genes for cancer stem cell (CSC) among different types of TETs. (D) Pseudotime trajectory for epithelial cell from all samples in a two-dimensional state-space defined by Monocle2, colored by groups. (E) UMAP plot of all epithelial cells from normal human thymus (n=1) and TETs samples (n=6), colored by the identified cell subpopulation (left). Same UMAP plot, colored by groups (right). (F) Dot plot for expression of marker genes in epithelial cell types. (G) Composition of the epithelial cell compartment showing average frequencies of major epithelial cell subpopulations for each group of TETs (n=2 T1, 2 T2 and 2 T3). (H) Representative immunofluorescence (IF) staining images showing CCL25 (pink), CHI3L1 (red), KRT14 (green) and DAPI (nuclei, blue) in TET samples. (I) Dot plot for expression of CCL25 and CXCL12 genes in epithelial cell subpopulations. (J) Heat map of chemokine interactions among subpopulations of T cell and epithelial cell, where the chemokine is expressed by the epithelial cell type and the cognate receptor by the T cell type. (DN, double-negative T cells; DP, double-positive T cells; P, proliferating; Q, quiescent). (K) KEGG pathway analysis of genes that are significantly (P < 0.01) upregulated in DP cell of T2 compared to T1. (DP, double-positive T cells; T1, type 1; T2, type 2). (L) Heatmap showing the normalized expression of pre-selection and selection associated genes within DP cell in each group by scRNA-seq.
Figure 6

The mechanism of tumor-driven CD8+ TRM cell activation in type 3 TETs. See also Supplementary Fig. S6 and Supplementary Table S7. (A) Same t-SNE plots as (Fig. 2C) of the CD45+CD3+ T cell from samples of each group by CyTOF, colored by cell types of CD8+ TRM cell and other cell. (B) Bar plot showing the frequencies of CD8+ TRM cell between the three groups of samples (n=3 N, 4 T1, 4 T2 and 4 T3. Data are presented as mean ± s.e.m. P values were determined by two-tailed Student’s t-test; ns, nonsignificant, *P < 0.05). (C) Heatmap of CellPhone DB analysis showing the number of interactions...
between CD8+ TRM cell and epithelial cell subpopulations. (D) Heat map of chemokine interactions among T cell and epithelial cell, where the chemokine is expressed by the outside cell type and the cognate receptor by the inside cell type. (E) Heatmap showing the normalized expression for activation associated proteins within CD8+ TRM cell in each group by CyTOF. (F) Dot plot for expression of cell activation and TCR activation associated genes within CD8+ TRM cell and other types of T cell. (G) KEGG pathway analysis of genes that are significantly (P < 0.01) upregulated in CD8+ TRM cells compared to other T cells. (H) Heat map of chemokine interactions among CD8+ TRM cell and other cell types (T cell, B cell, DC and macrophage cell), where the chemokine is expressed by the outside cell type and the cognate receptor by the inside cell type. (DC, dendritic cell) (I) Representative immunofluorescence (IF) staining images showing CD103 (red), EPCAM (green) and DAPI (nuclei, blue) in type 3 TET samples. The interactions between epithelial cell and TRM cell indicated by white arrow. Scale bar: 20 μm. (J) Representative immunofluorescence (IF) staining images showing CD103 (red), CD11c (green), CD20 (pale blue) and DAPI (nuclei, blue) in type 3 TET samples. The interactions between TRM cell and DC indicated by yellow arrow, between TRM cell and B cell indicated by white arrow. Scale bar: 20 μm.
Figure 7

Reclassifies TETs using representative gene markers of tumor cells. See also Supplementary Fig. S5 and S8. (A-B) Dot plots depicting the relative level of expression of GNB3 (A) and CHI3L1 (B) genes of total cell in each group. (C) Representative immunofluorescence (IF) staining images showing GNB3 (green), CHI3L1 (red), and DAPI (nuclei, blue) in TET samples. Scale bar: 50 μm. (D) Overall survival (OS) examined in each of the three types of TETs in the TCGA cohort. (E) Schematic model summarizing the developmental patterns of three types of TETs, including interactions between different types of cells. The ligand is secreted by the cell at the beginning of an arrow, and the receptor is expressed by the cell at the end of that arrow.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTables.rar
- InventoryofSupplementalInformation.docx
- SupplementalInformation.docx