

# T细胞运动轨迹的动态调控网络 分析Tce11状态转变的转录控制

敏燕,<sup>1,3</sup>胡静<sup>1,3</sup>袁华亭<sup>1,3</sup>徐丽雯<sup>1</sup>廖高明<sup>1</sup>江泽东,<sup>1</sup>朱嘉丽,<sup>1</sup>庞勃<sup>1</sup>平燕燕,<sup>1</sup>张云鹏<sup>1</sup>云晓<sup>1,2</sup>夏  
丽<sup>1,2</sup>

<sup>1</sup>哈尔滨医科大学生物信息学科学技术学院, 黑龙江哈尔滨, 150081; <sup>2</sup>黑龙江省寒区主要疾病高通量组学大数据重点实验室, 黑龙江省哈尔滨市150081

T细胞表现出异质性的功能状态, 这与对免疫检查点封锁的反应性和肿瘤患者的预后相关。然而, T细胞状态转变的动态过程背后的分子调控机制仍在很大程度上尚不清楚。基于非小细胞肺癌中T细胞的单细胞转录组数据, 我们结合细胞状态和伪时间, 提出了一个构建动态调控网络来解剖T细胞功能障碍过程的管道。在肿瘤浸润性T细胞功能障碍的过程中, 发现了不同阶段的候选调控因子。通过比较T细胞状态转换的动态网络, 我们发现了频繁的调节相互作用重布线, 并进一步细化了介导每个状态转换的关键调控因子。我们确定了几个已知的调控因子, 包括TCF7、EOMES、ID2和TOX。值得注意的是, 关键的调节因子之一, TSC22D3, 经常在从中间状态到功能障碍前和功能障碍前状态的状态转换中被发现, 通过调节相互作用重新布线在每个状态转换中发挥不同的作用。此外, TSC22D3的高表达与肿瘤患者的临床结局相关。我们的研究将转录因子(TFs)嵌入到时间动态网络中, 为控制T细胞状态转变过程的动态调控机制提供了一个全面的视角。

## 介绍

T细胞免疫检查点封锁(ICB)治疗已经彻底改变了癌症治疗, 通过重新激活肿瘤特异性T细胞反应来诱导患者持久的反应。<sup>1</sup>然而, 只有一部分患者能够获得长期有效的临床效益, 而我们对这些治疗方法的潜在反应或耐药性机制的理解仍然不完整。最近的单细胞分析研究为肿瘤浸润性CD8的异质性成分提供了证据<sup>2-4</sup>。肿瘤内CD8的主要成分<sup>+</sup>T细胞由以CCR7、LEF1和TCF7为特征的“幼稚样”细胞组成, 以CX3CR1、PRF1和KLRG1为特征的“细胞毒性效应”细胞组成,<sup>5</sup>以及“功能障碍”或“衰竭”的T细胞, 它们在肿瘤中所占的比例更高。<sup>6</sup>功能失调的T细胞, 其特征是共抑制受体的增加, 如PD1、LAG3、TIM3、CTLA4和TIGIT, 其效应有限

功能与效应CD8的比较<sup>+</sup>T细胞。功能失调的T细胞可进一步分为亚组, 如功能失调前、祖细胞耗尽和最终耗尽的T细胞。<sup>7</sup>此外, 这些异质状态组成表现出不同的内在效应功能和恢复活力的潜力,<sup>8</sup>表明形成治疗结果的决定因素的可能性。

的确, CD8的不同状态的组成和相对比例<sup>+</sup>肿瘤中的T细胞与临床结果相关, 如肿瘤分期,<sup>6</sup>对免疫疗法的反应,<sup>3</sup>以及患者的生存概率。<sup>9</sup>郑等人。<sup>6</sup>显示功能障碍CD8的比例增加<sup>+</sup>肝细胞癌(HCC)患者的晚期T细胞。乳腺癌中T细胞的单细胞图谱显示了一个组织驻留记忆子集, 它表达了功能障碍的特征基因, 并与改善预后相关。<sup>10</sup>重要的是, 表达tcf1的PD1的一个子集<sup>+</sup>功能障碍前的T细胞被发现对单药抗pd1治疗或抗pd1和抗ctla4联合治疗黑色素瘤的肿瘤控制至关重要。<sup>11</sup>此外, 在人类非小细胞肺癌(NSCLC)和结直肠癌(CRC)中也观察到T细胞功能障碍前亚群, 在ICB治疗的肿瘤中它们的细胞比例增加, 可能表明肿瘤患者有更好的生存率。<sup>3, 11-13</sup>鉴于不同状态的T细胞的不同功能和与临床结果和ICB治疗反应的关系, 迫切需要确定调节癌症中T细胞状态之间过渡的调节因子。

在本研究中, 我们提出了一个对T细胞功能障碍过程中的动态网络的全面分析, 并解剖了

2021年7月13日收到; 2021年10月6日接受;  
<https://doi.org/10.1016/j.omtn.2021.10.011>.

## <sup>3</sup>这些作者的贡献相同

通信地址: 李夏, 哈尔滨医科大学生物信息学科学技术学院, 黑龙江哈尔滨, 150081。

电子邮件: [lixia@hrbmu.edu.cn](mailto:lixia@hrbmu.edu.cn)

通讯地址: 中国哈尔滨医科大学生物信息学科学技术学院, 黑龙江省哈尔滨市, 邮编: 150081。

电子邮件: [小云@ems.hrbmu.edu.cn](mailto:小云@ems.hrbmu.edu.cn)

通信地址: 张云鹏, 哈尔滨医科大学生物信息学科学技术学院, 黑龙江省哈尔滨市, 邮编: 150081。

电子邮件: [zhangyp@hrbmu.edu.cn](mailto:zhangyp@hrbmu.edu.cn)



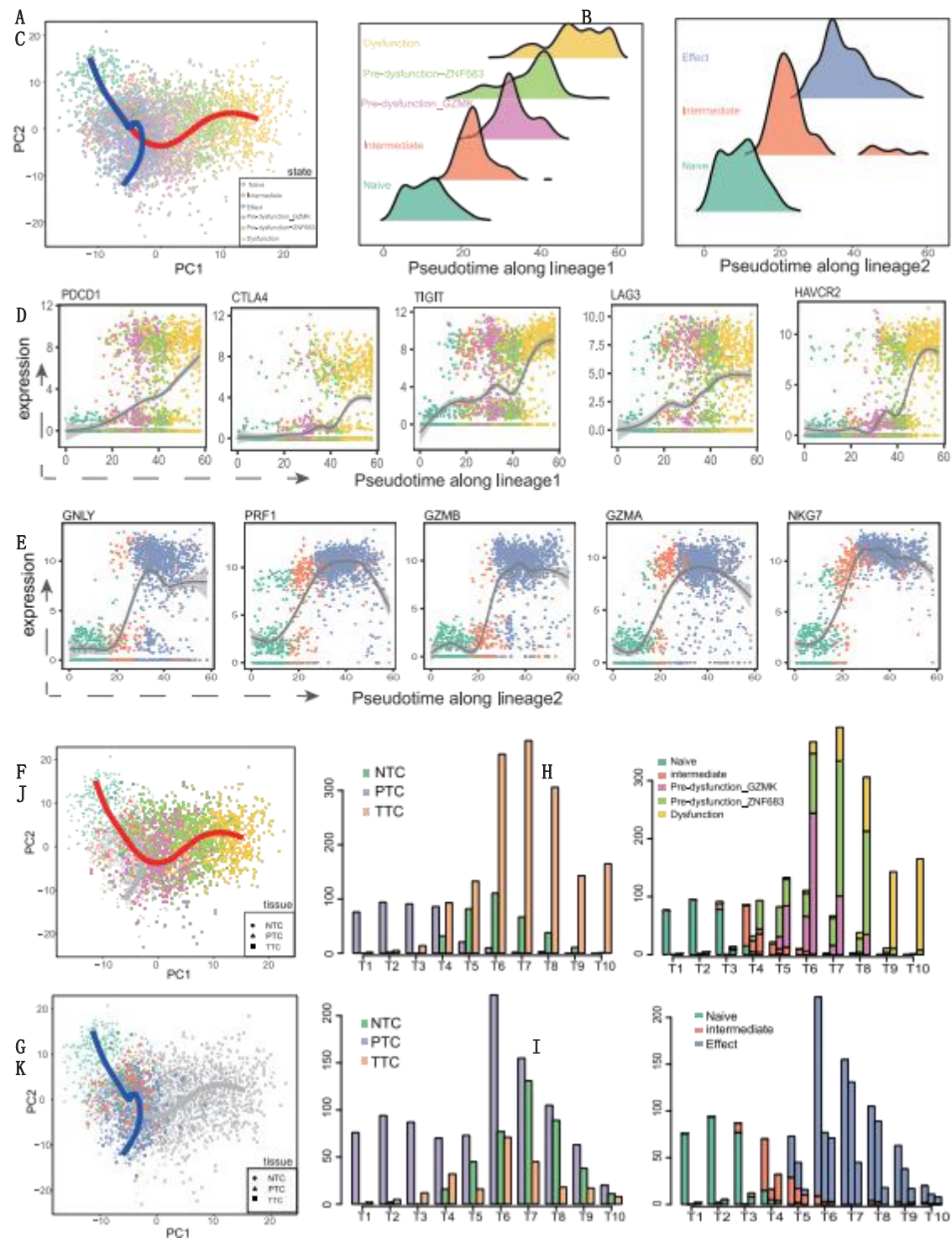


图1. scRNA-seq显示了CD8<sup>+</sup> T细胞的一个分叉轨迹

CD8<sup>+</sup>的(A)伪时间分析T细胞使用单细胞算法,在轨迹中产生两个谱系(彩色线)。轨迹上的细胞被标记为细胞状态,显示了功能障碍谱系(红线)和效应谱系(蓝线)。(B和C)功能障碍谱系(B)和效应谱系(C)上各状态的伪时间密度分布(D)沿功能障碍谱系伪时间的共抑制受体的表达水平。(E)效应分子沿伪时间的表达水平的影响谱系。(F和G)功能障碍谱系(F)和效应谱系(G)上的细胞起源组织。由椭圆、三角形和矩形等形状构成的点表示

(下页继续图例)



每个状态过渡阶段的关键转录因子 (TFs)。首先,我们构建了CD8的发展轨迹<sup>+</sup>T细胞和识别出在肿瘤环境诱导的功能障碍谱系中差异表达的动态基因。接下来,我们沿着轨迹将细胞分配到状态过渡阶段,然后为每个细胞的状态过渡阶段构建一个动态调控网络。然后,通过比较不同阶段之间的调控网络,我们发现了高度的重布线,并解剖了不同的重布线模式。此外,通过使用中心性度量,确定了每个阶段的关键调节器。

## 结果

构建CD8的状态转变轨迹<sup>+</sup> 肿瘤微环境中的T细胞

我们获得了3700个CD8<sup>+</sup>来自14例NSCLC患者的T细胞。<sup>13</sup>这些细胞是从邻近的正常肺组织 (NTC)、肿瘤组织 (TTC) 或外周血 (PTC) 中分离出来的 (图S1A)。在最初的研究中,所有的T细胞都进行了无监督的聚类,并被很好地标注为不同的状态,包括幼稚T细胞、中间T细胞、效应T细胞、分别带有GZMK和ZNF683标记的两个功能障碍前T细胞簇和功能障碍细胞 (图S1B)。

为了描述CD8之间的关系<sup>+</sup>在不同状态的T细胞下,我们使用弹弓算法进行伪时间分析来建立跨细胞状态的轨迹。<sup>14</sup>推断出的状态转变轨迹包含两个谱系,分别呈现出一个从幼稚状态到功能障碍再到效应状态的分叉结构 (图1A)。两种谱系都从原始状态开始,在中间状态之后分化。然后,将pre-dysfunction\_GZMK和pre-dysfunction\_ZNF683状态依次定位于谱系1 (i. e., , 最后以功能障碍状态结束 (图1B)。相比之下,谱系2 (即效应谱系) 直接以最终效应状态结束 (图1C)。在个体患者中也观察到了这两种谱系,这表明T细胞状态转变的轨迹在不同的患者中普遍存在 (图S2A和S2B)。

对于每个谱系,不同状态的细胞表现出预期的不同的伪时间分布<sup>5</sup> (图1B和图1C), 表示伪时间计算的准确性。此外,基因在T细胞功能障碍的早期和晚期显示功能障碍的伪时间动力学高度一致,与幼稚状态标记 (CCR7、SELL和IL7R) 活跃在早期伪时间和共抑制受体 (PDCD1、CTLA4、TIGIT、LAG3, HAVCR2) 后活跃的 (图1D; 图S2F), 这证实了功能障碍谱系的准确性。同样,效应分子GNLY、PRF1、GZMB、GZMA、NKG7的表达也不断增加

效应谱系的伪时间,进一步证实了沿着效应谱系的细胞排序的准确性 (图1E; 图S2G)。

从组织起源到状态转变轨迹的投影显示了两种谱系中组织分布的明显差异 (图1F和1G)。为了进一步清楚地解剖组织来源和组织运动轨迹之间的关系,我们根据伪时间将细胞平均分为10组,并比较了它们的组织来源/细胞状态组成。对于功能障碍谱系,外周血来源的原始T细胞和中间T细胞在前四组细胞中占主导地位,然后急剧下降 (图1H和1J)。同时,来自肿瘤和正常组织的功能障碍前T细胞逐渐增加。正常组织源性细胞减少,直到细胞组6。相反,肿瘤浸润性T细胞在随后的细胞组中进一步增加并占优势。在效应谱系中,血液来源的原始T细胞和中间T细胞也在起始组中占主导地位,正常组织来源的和肿瘤浸润效应T细胞随着伪时间逐渐增加,对任何组织来源都没有明显的偏差 (图1I和1K)。结合组织分布和细胞状态组成,这些结果表明,血液循环中的幼稚T细胞被招募到肿瘤中,持续的抗原刺激驱动中间状态细胞进入功能障碍前细胞,并进一步分化为功能障碍状态,这类似于CD8<sup>+</sup>由Andreatta等人提出的T细胞分化模型。<sup>15</sup>另一种谱系从中间状态细胞直接分化为正常效应细胞或瘤内旁观者细胞。

描述T细胞功能障碍轨迹的动态基因

为了确定哪些基因调控了T细胞功能障碍的进展,我们首先尝试通过应用tradeSeq算法来识别在功能障碍谱系中差异表达的基因。

<sup>16</sup>结果表明,共鉴定出1787个基因被动态调控 (错误发现率[FDR] < 0.05; 图2A)。除了众所周知的共抑制受体如PDCD1、CTLA4和HAVCR2外,还发现了许多参与T细胞分化的基因 (图2B–2F)。例如,我们观察到效应分子,如IFNG、PRF1、GZMB和GZMH (图2D) 和肿瘤坏死因子家族 (图2E) 的持续增加,这表明功能障碍的T细胞可能并没有完全失去它们的抗肿瘤效应潜能。<sup>17</sup>此外,与淋巴细胞迁移相关的基因沿着谱系动态表达,包括趋化因子CXCL13、CCL3、XCL1、XCL2、CCL4和CCL5 (图2F)。

接下来,我们将具有相似表达趋势的基因进行了分组。根据聚类结果,我们观察到三种不同的趋势,并影响了不同的生物学功能 (图2A)。集群1中的基因,在早期下调并逐渐上调

PTC、NTC和TTC,它们是CD8的缩写<sup>+</sup>T细胞分别来自外周血、邻近正常组织和肿瘤组织。(H和J) 伪时间分布(H)

以及在功能障碍谱系中,来自不同组织来源的细胞的状态组成(J)。根据伪时间,该谱系被平均分成10个箱子。条形图的高度代表了每个箱子中来自不同组织的细胞的数量。(I和K) 来自不同组织来源的细胞的假时间分布(I) 和状态组成(K) 对效应谱系的影响。



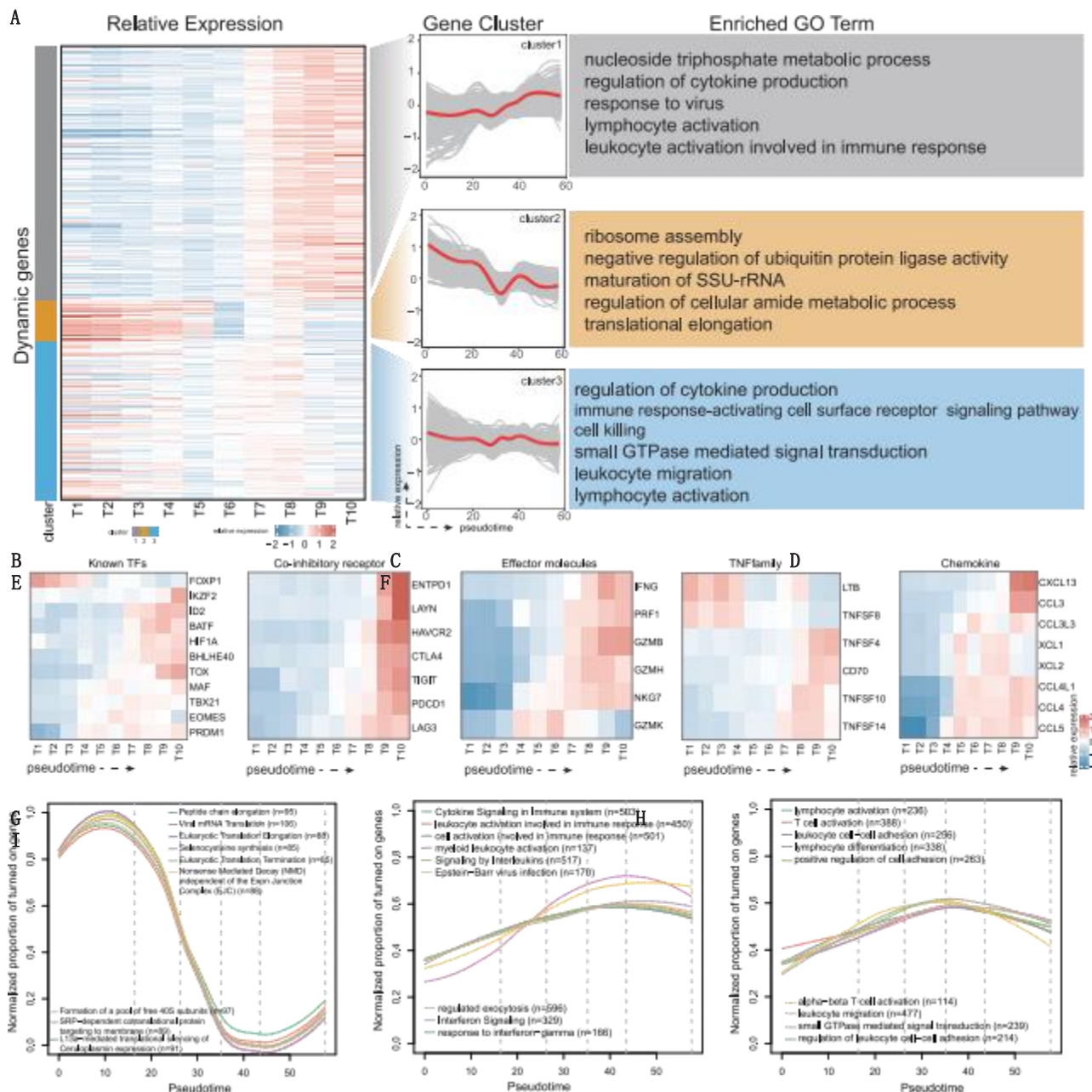
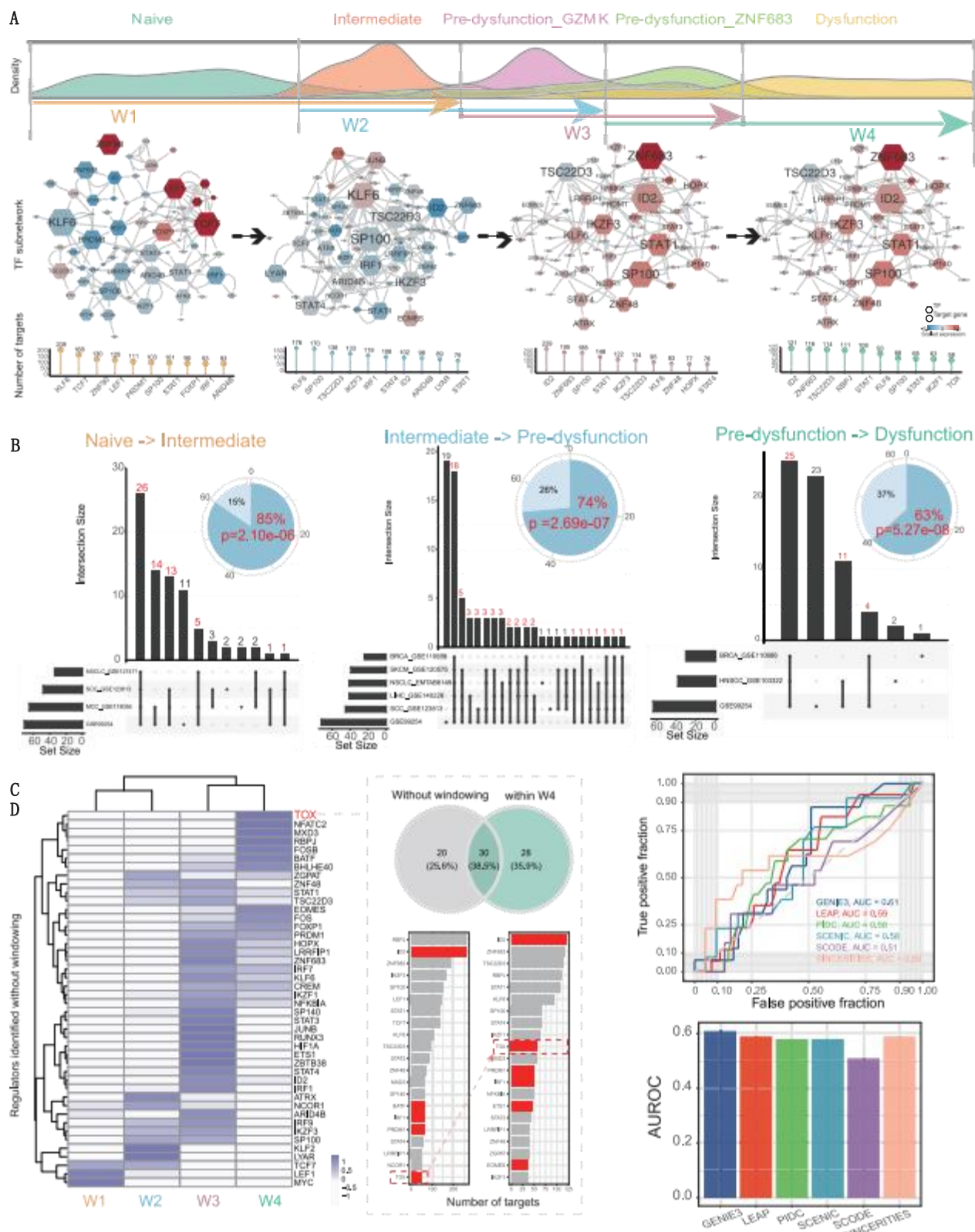


图2. 功能障碍谱系的动态基因分析

(A) (左) 热图显示10个细胞箱中动态基因的平均表达。所有的动态基因被分为三个簇。(中) 每个簇中所有基因(灰色线)和簇平均(红色红线)的功能障碍谱系的表达趋势。(右) 每个簇中具有代表性的丰富基因本体论(GO)术语。(B-F) 动态基因中包含的代表性基因表达的动态变化, 包括已知的TFs (B)、共抑制受体(C)、效应分子(D) 肿瘤坏死因子家族(TNF家族)(E)和趋化因子(F)。(G-I) 集群2 (G)、集群1 (H)和集群3 (I)的功能实体活动的动态变化。显示了每个簇的前10个富集术语。

在伪时间的后期, 在与T细胞细胞毒性相关的生物过程中高度富集, 包括免疫系统中的细胞因子信号和干扰素信号(图2A; 图S3A)。集群2中的基因具有早期激活和晚期下调

伪时间涉及核糖体、针对膜的翻译起始蛋白和肽链延伸(图2A; 图S3B)。然而, 第三个簇包括了许多基因, 它们没有明显的一致动态趋势, 并且是随机幸运的



(下一页上的图例)

随着时间的推移，它在淋巴细胞活化、白细胞间的细胞粘附和淋巴细胞分化中发挥了广泛的作用（图2a；图S3C）。

为了进一步分析整个功能障碍过程中生物功能的时间关系，我们进行了高分辨率的功能活性分析（见材料和方法）。与细胞翻译相关的信号通路在幼稚期高度活跃，从中间活跃期开始迅速减少（图2G），这与最近提出的即时翻译加速T细胞激活过程的观点一致。<sup>18</sup>翻译相关功能的下降伴随着细胞毒性效应和白细胞迁移相关功能的持续增加。细胞毒性功能持续上调，并在功能障碍阶段达到最高活性（图2H）。然而，T细胞的活化和细胞迁移的活性在功能障碍前阶段达到峰值，然后在随后的阶段略有下降（图2I）。综上所述，沿着肿瘤微环境诱导的功能障碍谱系的基因表达分析揭示了T细胞功能障碍过程的分子动力学，并描述了生物事件的顺序动态。

沿着T细胞功能障碍的轨迹构建动态网络

T细胞在功能障碍过程中的状态转变受到TFs及其相关辅助因子的调控，它们共同作用，动态调控靶细胞的表达。为了更好地理解驱动T细胞功能障碍的分子机制，我们提出了一个动态调控网络构建的管道，采用可变长度的滑动窗口方法来分裂细胞（见材料和方法；图S4）。简而言之，我们根据细胞状态的伪时间分布，将功能障碍谱系上的细胞分成四个重叠的窗口，每个窗口主要由两个连续状态的细胞组成。例如，窗口1（W1）主要由初始状态和中间状态的细胞组成，而W2主要由中间状态和pre-dysfunction\_GZMK状态的细胞组成（图3A；图S5A）。然后，分别为每个窗口构建基因调控网络（grn），以反映控制相应两种状态之间过渡的调控链接。总的来说，在至少一个网络中可以推断出82个调控因子和1641个靶点之间的7117个相互作用（图3A），每个调控因子的大小从1到229个基因不等（图S5B）。W1、W2和W3的节点和相互作用的数量相似，

而最后一个窗口W4的网络包含相对较少的节点（1143个）和更少的交互作用（1886个）（图S5C）。

我们还将动态GRN构建管道应用于另外8个单细胞RNA测序（scRNA-seq）数据集（表S1），包括NSCLC、黑色素瘤、乳腺癌、头颈部癌、HCC、默克尔细胞癌和鳞状细胞癌。我们使用了单线<sup>19</sup>方法绘制这些数据集的单元状态，然后应用变长滑动窗口方法和GENIE3为每个数据集构建动态网络（图S6；参见材料和方法）。我们比较了这些数据集之间每个状态转换的调节器。我们发现显著的重叠监管机构的核心数据集和独立的数据集，有85%（从幼稚到中间状态），74%（从中间到功能障碍），和63%（从功能障碍）的监管机构被确定在至少一个额外的数据集（图3B， $p = 2$ 。分别为 $10e-06$ 、 $2.69e-07$ 和 $5.27e-08$ ）。

为了说明在动态GRN构建管道中使用可变长度滑动窗口方法的优点，我们在没有窗口细胞的情况下实现了GENIE3算法。我们证明，在没有窗口的情况下识别的TFs与我们的方法显示出靶基因的重叠（图3C，左图），这表明使用所有细胞构建的GRN（GRN\_all）可以捕获整个谱系的某些信息。然而，GRN\_all可能偏向于大的细胞亚群，并有一些来自其他细胞的信息丢失。事实上，这些TFs主要捕获W3中的靶基因（图3C，左图），表明来自W3细胞的信息倚重，W3的细胞数量最多。此外，我们也不能仅根据GRN\_all来区分这些TFs的分数阶段。例如，我们在GRN\_all中鉴定了50个TOX靶基因，其中30个靶基因与W4中鉴定的靶基因共享（图3C，右上面板）。TOX是从前功能失调状态到晚期功能失调状态过程中的关键转录调控因子，<sup>5</sup>目前还不能确定它是否在GRN\_all的功能障碍阶段发挥作用。此外，TOX在GRN\_all中的排名低于在W4中（图3C，右下面板）。这些结果表明了变长滑动窗口方法的优点和必要性。

为了进一步估计不同的GRN构建算法对所得到的GRN的影响，我们选择了晚期功能障碍窗口，

图3. 沿着功能障碍谱系构建和验证动态网络

(A) (上) 不同状态下细胞的假时间密度（幼稚、中间、pre-dysfunction\_GZMK、pre-dysfunction\_ZNF683、功能障碍）。对连续两个相邻状态的细胞进行加窗，构建基因调控网络（GRN），反映了两种状态之间的过渡。(中) 每个窗口中TFs的子网。对于每个网络，节点在相应的窗口中用平均表达式着色。节点的大小根据其目标基因的数量进行调整。(下) 棒棒糖图显示了前10个调控因子的靶基因数量。(B) 在每个状态转换的不同数据集中识别的调节器的重叠。饼状图显示了仅在核心数据集中识别的监管者（浅蓝色）的比例，以及在至少一个独立数据集中识别的监管者的比例（蓝色）。通过超几何检验来评估其显著性。(C) 热图显示按比例Jaccard指数，估计每个窗口中每个TF的目标重叠。维恩图显示了网络W4中TOX和网络无窗口的靶基因的重叠。列出了两个网络中顶级调控因子的靶基因数量，与T细胞功能障碍相关的策划TFs被染成红色。(D) 受试者工作特征（ROC）曲线和ROC下面积（AUC）描述了六种GRN构建算法对功能障碍相关调节因子的预测性能。



W4, 以执行各种算法, 因为我们可以从文献中管理调节T细胞功能障碍的TFs (表S2)。这些算法包括LEAP,<sup>20</sup> PIDC,<sup>21</sup> 风景优美,<sup>22</sup> SCORE,<sup>23</sup> 和罪恶感。<sup>24</sup> 在选择本研究中使用的GENIE3时, 我们没有观察到其他算法的性能提升 (图3D; 图S7A)。虽然风景算法也实现了GENIE3+考虑motif富集, 但它可以比其他算法揭示更少的调节因子 (图S7B), 可能导致更高的假阴性。例如, 在使用风景景观构建的GRN中, 完全遗漏了众所周知的与功能障碍相关的调节器ID2和TOX。因此, 我们最终选择了GENIE3算法来构建grn。

接下来, 我们提取了窗口中连续两个相邻状态之间的调节活动存在显著差异的候选调节因子 (图S7C; 表S3; 见材料和方法)。我们检测到了许多众所周知的T细胞状态转变的调节因子 (图S7D)。例如, 在W1的早期调控网络中, LEF1、MYC、PRDM1和TCF7在状态转变过程中是最差的TFs。先前的研究表明, LEF1、TCF7和MYC在细胞周期调控、转录激活和代谢重编程中的重要作用,<sup>25</sup> 证实了它们在T细胞早期激活中的关键作用。综上所述, 我们的计算分析揭示了T细胞功能障碍过程中复杂调节相互作用的时间动态, 并强调了我们的管道在描述状态转变过程中的调节机制方面的有效性。

#### 实质性的监管重新布线控制着状态的过渡

为了复杂地表征T细胞功能障碍过程中调节相互作用的动态, 我们系统地比较了不同过渡过程中的调控网络。我们观察到不同网络之间存在大量的节点重叠, 在所有4个网络中都存在44.4% (图4A), 而TFs甚至更高 (63.4%) (图4B)。然而, 通过检查四个网络之间的边缘重叠, 我们发现在每个窗口中都有大量特定的边 (图4C)。此外, 我们计算了在任意两个网络中存在的每个TF的Jaccard指数 (JI), 以测量其规则的重叠。两个连续窗口 (W1\_W2、W2\_W3、W3\_W4) 之间的平均JI值为0.11, 0.14和0.12 (图4D), 分别表明TFs沿着T细胞的轨迹基本上重新连接到不同的目标 (即“重新布线”)。PRDM1在第一个窗口中表达上调, 并在随后的状态下持续表达 (图3B-3E)。而PRDM1在不同的窗口中调控特异性基因, 并参与了窗口特异性功能。例如, PRDM1在W1中参与了干扰素- $\gamma$ 相关的功能, 同时它分别参与了W3和W4的T细胞活化和去磷酸化 (图4E)。另一个例子是ID2, 它是终末分化T细胞形成的重要调节因子。<sup>26</sup> 我们发现ID2也在四个网络中重新连接, 并发挥了不同的作用 (图4F)。

T细胞衰竭表现为共抑制分子的高表达。我们提取了每个窗口中的共抑制受体的子网络, 以研究它们的调控重接线 (图4G)。最初的事件是LYAR、KLF6、JUNB和PRDM1在早期激活TIGIT (图4G左图), 其表达量明显高于其他共抑制受体 (图S8A和S8B)。同时, LAG3被KLF6、SP100和HOPX激活 (图4G, 第二面板), 导致中间状态的表达上调 (图S8A和S8C)。随后, 在W2和W3的网络中, 由PRDM1、IKZF3、STAT1和ID2依次激活PDCD1和HAVCR2 (图4G, 第二和第三面板; 图S8A)。最后, TOX和RBPJ在功能障碍晚期上调了CTLA4 (图4G, 第四面板; 图S8A)。共抑制受体的动态重组显示, 它们在T细胞功能障碍过程中受到不同TFs的调控, 导致激活级联。

#### 解剖TFs的动态重新布线模式

在状态转换网络中存在大量的特定边缘, 通过四个窗口内tf-目标对的Spearman相关性进一步证实了这一点。我们如预期的那样, 在相应的窗口中观察到调控对的相关性最高 (图5A), 并且我们还注意到了各个窗口中相关性的逐渐变化。在状态转换过程中, 窗口间调控相互作用的波动和重新连接表明, TFs可能在某些阶段发挥主导调控作用。为了阐明每个TF的主导调控阶段, 我们将其分配到TF的目标数量高于所有窗口的平均数量的窗口。共获得了11种调控模式, 其中包括4个窗口特异性调控模式和7个持续调控模式 (图5B和5C)。例如, 与幼稚状态相关的TFs BACH2、LEF1和MYC<sup>27-29</sup> 在W1中发挥特定作用, 而功能障碍相关的TOX和BATF<sup>30</sup> 在W4中起主导作用。此外, 一些TFs在两个或三个窗口中持续发挥主导作用。例如, TCF7在W1和W2中持续发挥作用, 这与它在从幼稚状态到功能障碍前状态的过渡中发挥的关键作用相一致。

<sup>31</sup>SP140通过调节病毒感染中的共抑制受体LAG3和TIGIT与T细胞功能障碍相关,<sup>32</sup> 不断发挥监管作用W1至W3。

为了探究TFs的协同效应, 我们基于辛普森指数估计了每个窗口中TFs之间的调控相似性<sup>33</sup> 并分别确定了W1和W2两个模块, W3和W4的一个模块 (图5D-5G; 图S9A和S9B)。引人注目的是, 同一模块中的tf倾向于具有相同的调控模式, 特别是窗口特定的模式。W1中的模块M1包含TCF7、LEF1、MYC、SATB1和BACH2, 除TCF7外, 这些模块都在W1中起主导作用 (图5D), 该模块参与了真核翻译延伸 (图S9C)。W2中的M2含有w2特异性调控因子ZBTB38、EOMES和ZEB2 (图5E), 在参与免疫应答的细胞杀伤和白细胞激活中发挥作用



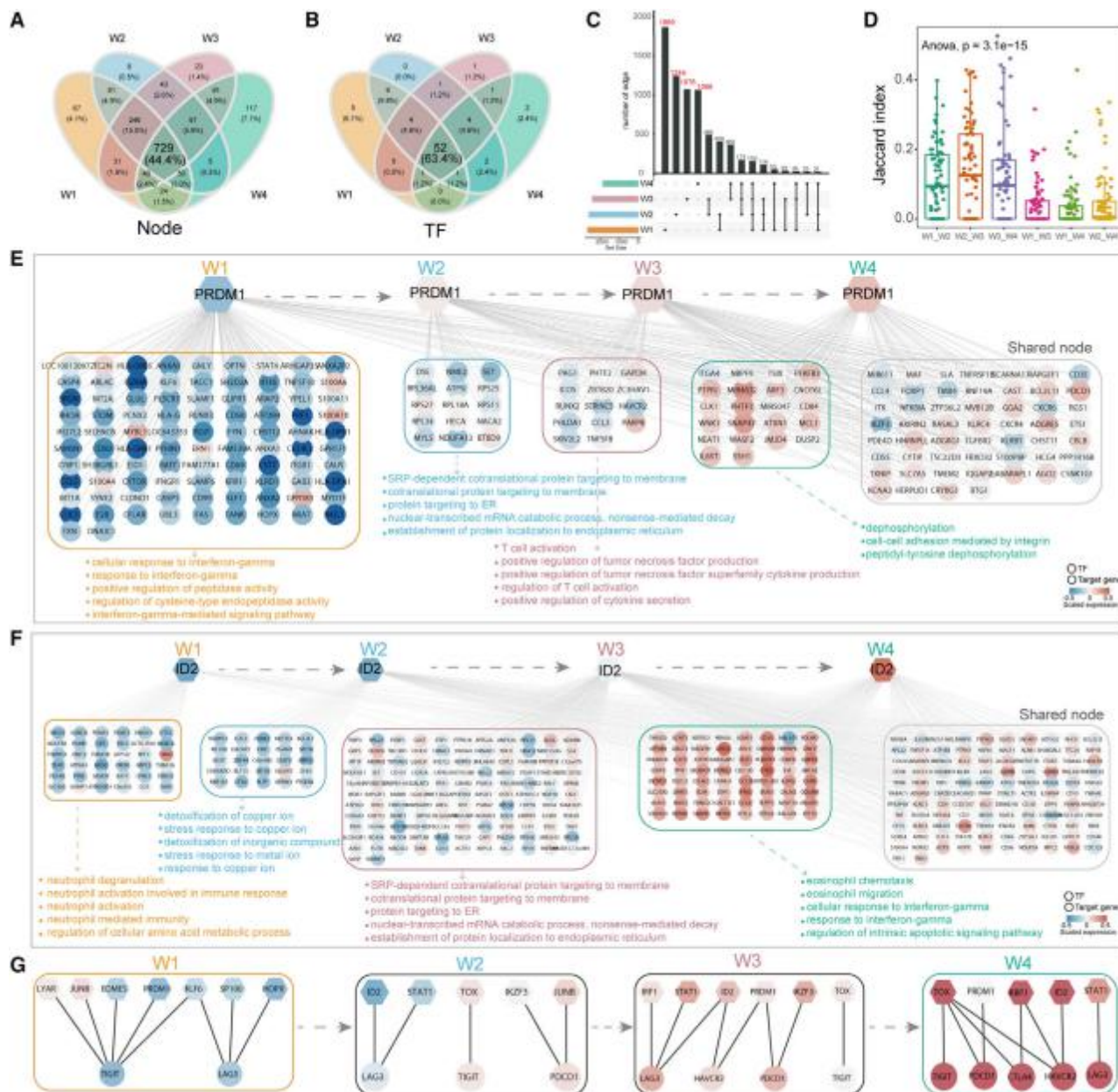


图4. T细胞功能障碍期间的大规模网络重新布线

(A和B) 维恩图，显示了节点(A)或TFs (B)在四个窗口上的重叠。(C) UpSet图显示了跨越四个窗口的网络边缘的交集。特定的边的数量用红色标记。(D) 箱线图显示了任意两个窗口中TFs的JI。采用方差分析检验评估不同组间差异的显著性。(E和F) 转录因子PRDM1 (E)和ID2 (F)的调控重新布线。前四组靶点分别代表W1、W2、W3和W4中的窗口特异性靶点，最后一组基因包含至少由两个窗口共享的靶点。窗口特异性组的代表性功能如下基因所示。(G) 每个窗口中共抑制受体 (TIGIT、LAG3、PDCD1、HAVCR2和CTLA4) 的子网络。

(图S9D)。W3中的M1包含w3特异性调控因子HOPX、ZNF683和ID2，以及两个持续的调控因子，它们只调控少量的靶点(图5F)，与免疫系统中的细胞因子信号传导和参与免疫应答的白细胞激活有关(图S9E)。W4中的M1包含TOX、ETV1和RBPJ，它们在W4中起主导作用(图5G)，

在淋巴细胞活化和T细胞共刺激中富集(图S9F)。

关键的调节因子导致了T细胞的功能障碍

为了进一步衡量不同tf在状态过渡网络中的重要性，我们建立了一个综合中心性度量

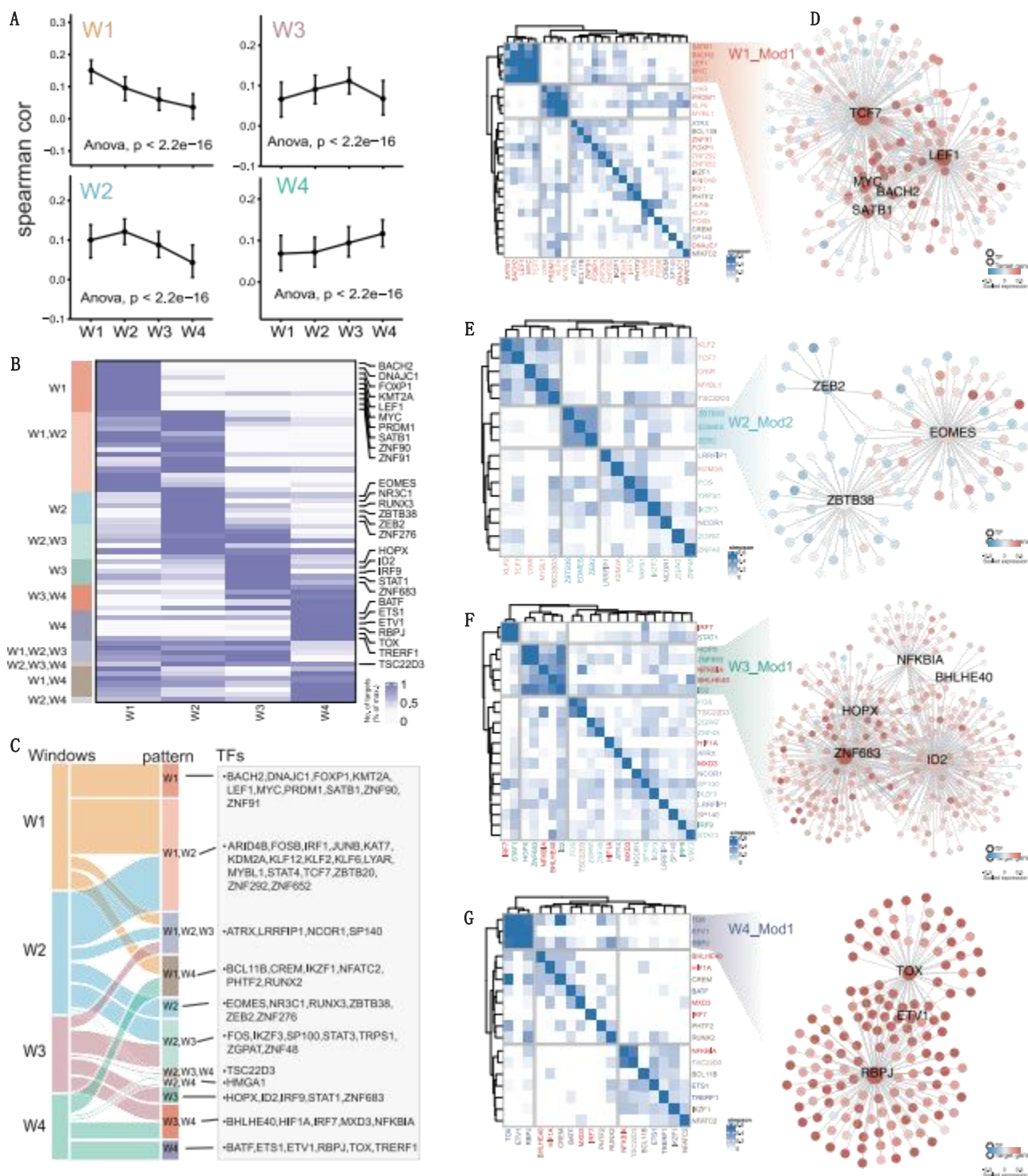


图5. 转录因子的动态调控模式

(A) 分别计算四个窗口中每个窗口内tf-目标对的中位数斯皮尔曼相关性。胡须表示第25百分位值和第75百分位值。采用方差分析检验评价其显著性。(B) 热图显示每个动态TF的目标数量归一化其最大目标数量。(C) Sankey图显示网络W1-W4和网络11调控模式之间的关系。右表中列出了每个监管模式对应的TFs。(D-G) 基于每个窗口的Simpson索引确定的TF模块。(左) 显示TF对的辛普森指数的热图。(右) 相应的TF模块的子网, 其节点被表达式级别着色。

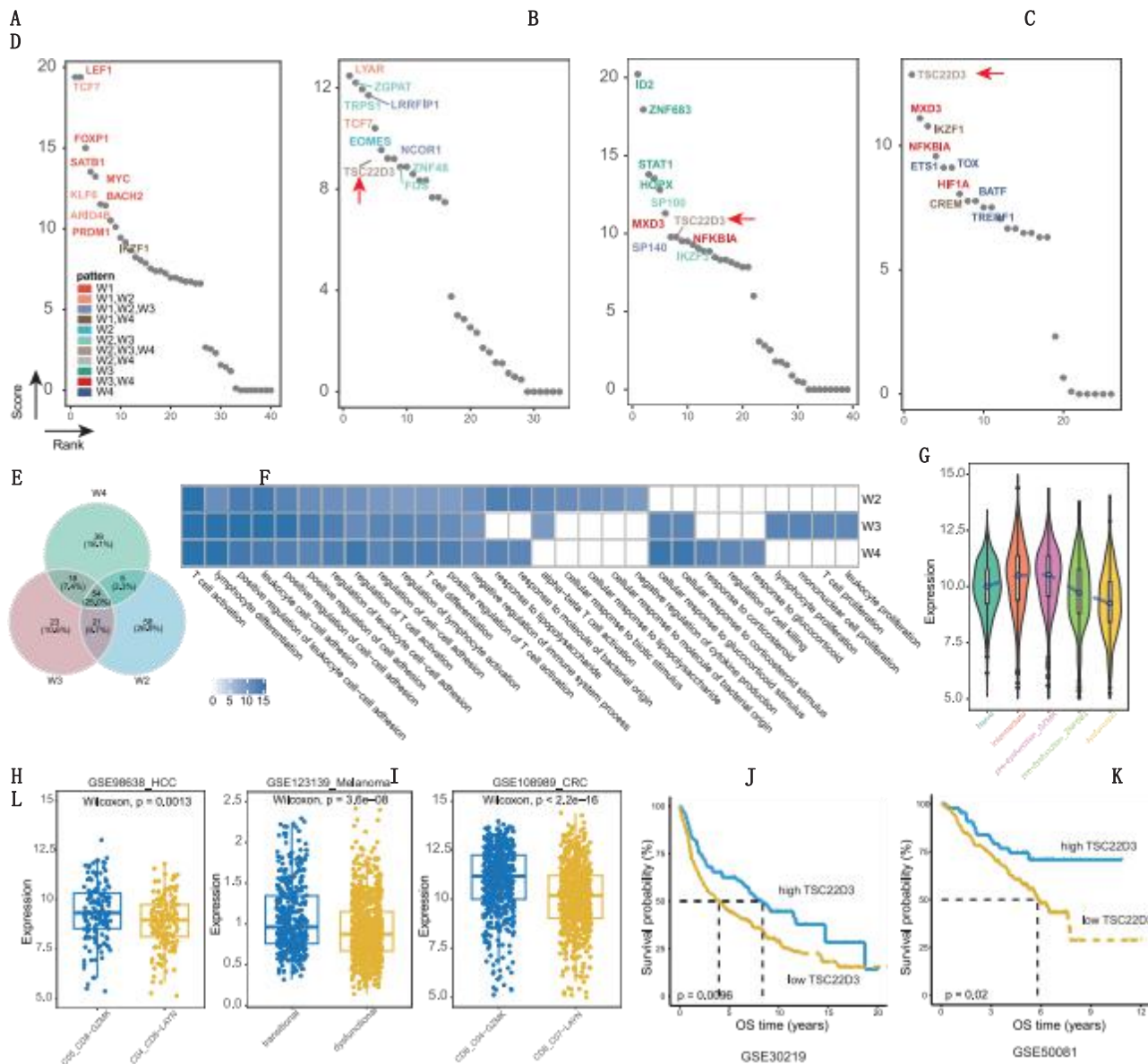


图6. T细胞功能障碍过程中四个网络的关键调节因子

(A-D) 基于综合中心性度量的四个网络中的TFs排名。对于每个网络，前10个tf都被它们的调控模式标记和着色。箭头表示在W2、W3和W4中存在的临界TF TSC22D3。(E) TSC22D3在W2、W3、W4中目标重叠的维恩图。(F) 每个窗口中tsc22d3调控基因的富集GO术语。显示-日志的颜色10富集显著性的转换p值。(G) 小提琴图显示各细胞状态下TSC22D3的表达水平。只有表达量大于5的细胞被用于可视化。(H-J) Boxplot显示了肝细胞癌(H)、黑色素瘤(I)和结肠直肠癌(J)在功能障碍前和功能障碍状态下TSC22D3的表达水平。采用Wilcoxon检验评价其显著性。同样，低表达细胞不被纳入可视化，Smart-Seq2数据集(GEO: GSE98638和GSE108989)的阈值为5，MARS-seq数据集(GEO: GSE12313139)的阈值为0.1。(K和L) 基于GEO: GSE30219 (K)和GEO: GSE50081(GEO: GSE50081(L)表达的Kaplan-Meier生存曲线显示患者分层生存概率存在显著差异。

结合度、接近度、中间性、特征值和页面排名(见材料和方法)。在四个状态过渡网络中显示的前10个tf如图6A-6D所示。这一组关键的调控因子包括一些著名的介导T细胞状态转变的调控因子，包括TCF7，<sup>31</sup>埃莫斯，<sup>34</sup>ID2，<sup>26</sup>和TOX。<sup>35</sup>值得注意的是，TSC22D3不断地播放

在W2、W3和W4中的调节作用(图5C)，它是所有三个窗口中最重要调节作用之一(图6B-6D)。虽然TSC22D3在不同的窗口中调控不同的靶点(图6E)，但随着时间的推移，它可以不断调节一些生物学功能，如T细胞活化和细胞-细胞粘附(图6F)。Ayroldi等人。<sup>36</sup>是否报道了诱导



TSC22D3的表达可参与调节T细胞的活化和凋亡。尽管如此，我们也观察到了TSC22D3的窗口特异性功能。例如，它可以在W2中特异性地调节细胞对生物刺激的反应，而在W3中特异性地调节T细胞的增殖（图6F）。

在状态过渡时间轴上，我们发现TSC22D3的表达量逐渐增加，然后逐渐下降，在功能障碍前的状态中表达量最高（图6G）。我们进一步研究了TSC22D3在另外三个CRC的scRNA-seq数据集中的表达，<sup>17</sup>，HCC，<sup>6</sup>和黑色素瘤<sup>37</sup>并观察到在功能障碍前状态的表达始终显著高于功能障碍状态（图6H和6I），表明其在T细胞功能障碍前状态中的关键作用。先前的研究报道，更高比例的功能障碍前状态T细胞有利于患者的生存。<sup>8, 11</sup>在这里，我们试图探索TSC22D3的表达是否，一个潜在的功能障碍前T细胞状态的关键调控因子，was sufficient to predict prognosis of NSCLC patients. 基于GEO中NSCLC患者的两个队列：GSE30219和GSE50081，我们发现TSC22D3的高表达与总生存率显著相关（图6K和6L）。这些结果表明，TSC22D3可能作为T细胞功能障碍前的关键调控因子，并与NSCLC患者的临床结局相关。

## 讨论

在本研究中，我们将轨迹推断与基于单细胞转录组学的GRN构建相结合，以绘制T细胞功能障碍过程中细胞状态转变的基因调控景观。通过构建T细胞状态转变轨迹，我们检测到肿瘤微环境诱导的谱系中存在高度的转录异质性和表达动力学。动态网络分析揭示了控制状态转换的候选调节器和大量的网络重新布线。我们进一步优化了每个状态转变中的关键基因，证明TSC22D3在多个网络中具有高中心性，在功能障碍前状态中高表达，并与肿瘤患者的临床结局相关。

在不同的细胞环境下，TFs可以通过改变其目标来执行显著的不同功能。<sup>38</sup>调节相互作用的重新布线已被发现是状态转变的一个标志，而由网络重新布线产生的调节程序的改变已被报道具有强烈的表型影响。<sup>39</sup>在T细胞功能障碍的过程中，我们也观察到大量的调控重组，并揭示了调控程序的改变。例如，EOMES（图S5D）和PRDM1（图4E），据报道，它们在急性和慢性感染中调节具有重要的环境特异性功能的不同靶点，<sup>40</sup>在不同的T细胞状态下显示了大量的重组。我们还观察到在特定状态下，重布线TFs的不同目标的不同功能，这表明了网络重布线对于控制状态转变的重要性。重要的是，我们的研究解剖了共抑制受体的重新连接，并描述了它们的激活级联，这可能有助于未来的识别

在免疫治疗中的阶段特异性靶点。<sup>8</sup>

我们的分析允许对肿瘤中T细胞功能障碍过程中的动态调节机制进行广泛的、无偏见的研究。例如，与已发表的作品相一致，<sup>41</sup>我们确定TOX和BATF是从功能障碍前状态向功能障碍状态过渡的关键调控因子。我们的数据确定RBPJ是CD8的潜在调控因子<sup>+</sup>功能障碍细胞，因为T细胞激活后RBPJ表达上调，特别是在功能障碍高峰时，而RBPJ在T细胞功能障碍晚期与TOX协同靶向共抑制受体HAVCR2和CTLA4（图4G）。此外，另一个潜在的调节因子TSC22D3在功能障碍前状态后持续发挥多种功能，直到功能障碍状态。TSC22D3参与了细胞对皮质类固醇和糖皮质激素刺激的反应，这与之前的研究一致，<sup>42</sup>这表明内源性糖皮质激素信号可以塑造T细胞从幼稚状态到功能障碍状态的分化，提示从糖皮质激素刺激到功能障碍状态的调节级联反应是由TSC22D3介导的。TSC22D3的生存分析显示其在NSCLC患者中的预后意义。尽管我们注意到TSC22D3在各种免疫细胞类型中广泛表达<sup>43</sup>（图S10A），我们通过CD3基因表达水平（CD3D、CD3E和CD3G）的几何平均值对TSC22D3的表达水平进行了归一化，以只考虑TSC22D3在肿瘤样本中T细胞中的相对表达量。

我们还注意到CD4中TSC22D3的表达<sup>+</sup>这可能会混淆结果。而TSC22D3在T细胞中的表达水平是否与CD8中呈正相关<sup>+</sup>T细胞/功能障碍前T细胞中TSC22D3的表达水平与CD8成正比<sup>+</sup>T细胞/功能障碍前T细胞。事实上，我们观察到在总T细胞中TSC22D3的表达水平与CD8之间的强正相关<sup>+</sup>T细胞/功能障碍前的T细胞（图S10B）。因此，TSC22D3在T细胞中相对表达的预后关系可以反映其在CD8中表达的临床影响<sup>+</sup>T细胞/功能障碍前T细胞。此外，ID2、ZNF683、HOPX和IRF9是W3网络中TFs上调最显著的一个（图3a；图S7C和S7D）。这些TFs的动态表达表明，细胞在过渡的功能障碍过程中逐渐获得了长期的细胞毒性和记忆表型，<sup>26, 44 - 46</sup>允许功能失调的T细胞在肿瘤中发挥持续的效应作用。综上所述，我们认为这种识别细胞轨迹潜在调控因子的策略有希望，并将有助于阐明在不同阶段控制功能障碍T细胞分化的复杂转录网络。

## 材料和方法

### 数据收集和处理

我们从GEO：GSE99254下的GEO数据库中下载了14例患者的scRNA-seq数据。<sup>13</sup>我们提取了除黏膜外的CD8<sup>+</sup>T细胞的计数表达谱<sup>+</sup>相关不变T（MAIT）细胞，与其他CD8相比，它具有不同的T细胞受体（TCRs）和发育过程<sup>+</sup>细胞<sup>13</sup>去除平均计数小于1的基因后，进行计数归一化，首先将计数除以每个细胞的总计数，然后进行



与细胞间总数的中位数的乘法。然后，我们对表达式配置文件进行了对数转换。最后，我们总共保留了12306个蛋白质编码基因和3700个CD8<sup>+</sup>具有良好注释细胞状态的T细胞，包括303个幼稚细胞（CD8\_C1-LEF1）、206个中间细胞（CD8\_C2-CD28）、1192个效应T细胞（CD8\_C3-CX3CR1）、674个GZMK标记功能障碍前细胞（CD8\_C4-GZMK）、832个ZNF683标记功能障碍前细胞（CD8\_C5-ZNF683）和439个功能障碍细胞（CD8\_C6LAYN）。

#### CD8<sup>+</sup> T细胞状态过渡轨迹推断

推断CD8<sup>+</sup>T细胞的发育轨迹，我们使用了弹弓算法。<sup>14</sup>弹弓是一种快速、鲁棒的分支轨迹推理方法，在最近的基准测试研究中，它被证明是性能最好的方法之一。<sup>47</sup>简而言之，弹弓使用预先退出的集群来推断基于最小生成树的谱系层次结构，并在伪时间轨迹上对齐细胞。具体来说，我们首先下载了聚类特异性的签名基因<sup>13</sup>并对所有细胞进行主成分分析（PCA），以保持状态间的主要生物学变异。然后，我们在前三个主组件上运行弹弓，以原始状态作为初始集群。使用弹弓管道，我们获得了伪时间值和分配的单元格分支（图S2E）。此外，我们对每个患者进行了相同的低轨迹推理。通过轨迹推断，我们得到了三个分支，分支1包含几乎所有的功能障碍前和功能障碍T细胞，分支2富含效应T细胞，分支“1,2”由幼稚T细胞和中间T细胞组成（图S2F和S2G）。为了保证轨迹的准确性，我们从相应的分支中删除了不属于富集状态的模糊细胞。

#### 沿着状态转变的轨迹识别动态基因

我们使用的是贸易标准<sup>16</sup>版本1.2。1来识别沿着肿瘤环境诱导的T细胞功能障碍的轨迹动态表达的基因。差异分析仅限于高质量的基因，通过了三个过滤标准，即(1)平均表达量大于0.5，(2)表达细胞的比例大于0.05，(3)在超过100个细胞中检测到。对于每个基因，我们拟合一个参数K（结数）为6的一般加性模型（GAM）来建模基因表达和伪时间之间的关系，并使用关联检验函数检验它们的关联的显著性。我们选取了fdr校正后p值<0.05的显著基因作为动态基因。利用元景观对动态基因进行功能富集分析。<sup>48</sup>

#### 基因集合沿伪时间的功能活性分析

我们首先使用如前所述，使用隐马尔可夫模型（HMM）方法确定每个伪时间内基因的开/关二进制状态。<sup>49</sup>简而言之，我们将伪时间分为40个箱子，并平均每个箱子内每个基因的表达水平。我们将平均表达式值赋值给观察变量

并使用鲍姆-韦尔奇算法提取最可能的发射概率和转移概率。然后，应用维特比算法对每个基因的开/关双星状态进行预测。对于每个功能基因集，我们将每个箱中所有表达基因的二进制状态进行平均，并进一步归一化，以表示开启基因的比例。最后，利用开启基因沿伪时间的比例来估计功能基因集的活性转变。

#### 国家转型期动态监管网络的建设

许多计算方法已经发展出来，从单细胞基因表达数据中预测grn，<sup>50, 51</sup>但大多数细胞都忽略了细胞的时间序列。此外，细胞沿着运动轨迹的分布并不均匀，这使得调控网络偏向于在假时间的高密度区域内的细胞。为了避免这些偏差，我们将细胞状态和伪时间结合起来，用可变长度的滑动窗口方法重新分裂细胞，并构建了跨状态过渡阶段的调控网络。首先，基于不同状态下细胞的伪时间密度，计算两个相邻状态之间的密度曲线的交点，作为细胞分裂成多个伪时间区间的边界。然后，将每两个连续的时间间隔作为一个窗口。因此，功能障碍谱系的细胞分为4个窗口，代表不同的状态过渡阶段。每个窗口主要由两个连续状态下的单元格组成。然后，我们为每个窗口分别构建了一个GRN。

我们首先过滤每个窗口中的基因，以去除表达水平很低（平均表达量小于1）或在过少的细胞中（检测细胞数量小于10和检测细胞比例小于0.05）的基因。然后，使用GENIE3构建无监督grn，<sup>52</sup>这是在DREAM4和DREAM5 GRN重建挑战中性能最高的方法。GENIE3利用了随机森林（RF）机器学习算法，可以处理组合和非线性的交互作用，适用于单细胞数据。<sup>50</sup>遗憾的是，它训练随机森林模型来预测每个靶基因的表达方差，并使用TFs的表达作为输入。然后使用每个模型来获得TFs的权重。我们提取的权重高于预定义的阈值，该阈值被确定为所有权重的平均值加上标准差的两倍，作为高自信的监管链接。然后，对所有目标模型的排名进行汇总，得到所有监管链接的全球排名。0.02权重大于0的链接用于构建GRN。此外，我们将靶点分为正相关和负相关的靶点（根据斯皮尔曼相关性），以分离可能被激活和被抑制的靶点。最后，只保留了被激活的链接

遵循步骤。该管道的代码可以在<https://github.com/MinYan19940/DynamicGRNPipe>上找到。

#### 构建独立数据集的网络

我们从GEO和ArrayExpress数据库中下载了8个数据集，涉及7种癌症类型。GEO: GSE120575，

快速传递: E-MTAB-6149, GEO: GSE123813, 我们提取了CD8<sup>+</sup>基于原始细胞注释的T细胞。对于GEO: GSE110868、GSE140228、GSE118056、GSE103322和GSE127471, 我们提取了CD8<sup>+</sup>基于T细胞标记物的表达。经过质量控制, 我们得到了53,842个CD8<sup>+</sup>T细胞总数(表S1)。为了获得一致的细胞状态, 我们使用统一的管道对细胞进行注释, 并构建状态转换轨迹。首先, 使用Seurat R包来识别主要的细胞簇。生成了前2000个高度可变的基因, 并用于进行PCA。前5个主成分(PCs)用于基于图的聚类(参数res = 0.3)来识别不同的聚类。其次, 使用SingleR R包将集群映射到本研究使用的6种细胞状态。简而言之, 我们计算了每个聚类中基因的平均表达量。然后, 以细胞簇的平均表达谱作为输入, 我们使用核心数据集(GEO: GSE99254)作为参考, 使用SingleR将细胞簇映射到已知的细胞状态(图S6A)。接下来, 我们基于细胞的映射状态, 应用弹弓算法构建状态转换轨迹, 并计算细胞的伪时间(图S6B)。由于pre-dysfunction\_GZMK状态失踪的快速表达: E-MTAB-6149和地理: GSE110686数据和pre-dysfunction\_ZNF683状态缺失的地理: GSE103322数据, pre-dysfunction\_GZMK和pre-dysfunction\_ZNF683不区别时构建监管网络从功能障碍前功能障碍或中间状态。最后, 我们使用细胞状态、伪时间和动态基因表达谱作为管道的输入, 为每个数据集构建一个动态调控网络。采用超几何检验来评估独立数据集和核心数据集中调节器重叠的显著性。

#### 与其他GRN重建算法的比较

为了评估不同GRN重建算法对功能障碍相关TFs识别的影响, 我们还实现了其他算法在第四个窗口构建GRN, 包括LEAP,<sup>20</sup>PIDC,<sup>21</sup>风景优美,<sup>22</sup>SCODE,<sup>23</sup>和罪恶感。<sup>24</sup>对于风景, 我们使用了pySCENIC的Docker图像, 对于其他的, 我们使用了来自直管道的对应的Docker图像。<sup>53</sup>为了比较这些算法的性能, 我们通过整理文献中与T细胞功能障碍相关的TFs, 进行了受试者工作特征(ROC)分析(表S2)。ROC分析使用pROC软件包进行。<sup>54</sup>

#### 在州过渡期间精炼候选监管机构

为了细化每个窗口中的候选调控因子, 我们首先估计了细胞中每个TF调控因子的平均表达量, 对应于每个窗口中的两种状态。然后, 我们使用Wilcoxon秩和检验检验了两种状态之间调节活动的差异显著性。然后根据FDR方法对多重测试问题的P值进行校正。带FDR < 0的TF。05被认为是窗口内状态过渡期间的候选监管机构。

#### 测量节点中心性

节点中心性度量被用来衡量grn中TFs的重要性。在这里, 我们使用度、接近度、中间度、特征值和PageRank来评估节点的中心性。<sup>55</sup>然后, 提出了一种秩聚合方法(RRA)<sup>56</sup>被用于整合所有的中心性指标。最后, 将每个TF的秩积分得分计算为 $-\log_{10}(\text{p值})$ 表示最终的中心性度量。

#### NSCLC患者TSC22D3的生存分析

我们从GEO中收集了两个数据集(GEO: GSE30219和GSE50081), 以评估TSC22D3表达与NSCLC患者临床结局的相关性。为了校正肿瘤样本中T细胞浸润水平的影响, 我们通过CD3D、CD3E和CD3G表达水平的几何平均值对TSC22D3的表达进行归一化。<sup>35</sup>然后, 我们根据TSC22D3标准化表达水平的第三个四分位数, 将NSCLC患者分为两组。绘制Kaplan-Meier生存曲线。<sup>57</sup>采用log-rank检验检验两组间生存概率差异的显著性。

#### 统计分析

超几何检验用于识别每个分支上显著富集的T细胞状态。采用Wilcoxon秩和检验检验共抑制受体表达差异的显著性, 以及功能障碍前和功能障碍状态中关键调控因子表达的差异。采用单向方差分析来测试跨四个窗口的tf-目标对的斯皮尔曼相关性的动态。

#### 补充信息

补充信息可以在<https://doi.org/10.1016/j.omtn.2021.10.011>上找到。

#### 确认

国家重点研发计划项目(2018YFC2000100)、国家自然科学基金项目(61873075、32070673、31871336、31900478)、黑龙江省自然科学基金项目(YQ2019C012)、黑龙江省博士后基金项目(LBHQ18099)、黑龙江省创新人才青年学者计划项目(UNPYSCT-2017059)资助。

#### 作者贡献

X.L., Y.X.和Y.Z.构思和设计了这项研究。M.Y.和J.H.设计框架。M.Y., J.H., H.Y., 和L.X.对数据进行了分析, 并实施了该方法。X.L.和Y.X.修改了手稿。G.L.和Z.J.获取数据。J.Z.和B.P.组织的数字。B.P.和Y.P.提供了建设性的讨论。Y.Z.和Y.P.他帮助解释了研究结果。M.Y., J.H.和H.Y.起草了手稿。所有作者都阅读并批准了稿件。

#### 利益声明

作者声明没有任何相互竞争的利益。

参考文献

1. 考肯, K. E., 和惠利, E. J. (2015). 克服感染和癌症中的T细胞衰竭. 趋势免疫醇。36, 265 – 276.

2. Azizi, E., 卡尔, A. J., Plitas, G., 康沃尔, A. E., 科诺帕奇, C., 普拉巴卡兰, S., 奈尼斯, J., 吴, K., 塞塞利奥瓦斯, V., 塞蒂, M., 等。(2018). 乳腺肿瘤微环境中不同免疫表型的单细胞图谱. 174号牢房, 1293–1308. e36.

3. 萨德-费尔德曼, M., 伊扎克, K., 比约加尔德, S. L., 雷, J. P., 德波尔, C. G., 詹金斯, R. W., 利布, D. J., 陈J. H., 弗雷德里克, D. T., Barzily-Rokni, M., 等人。(2018). 定义与黑色素瘤对检查点免疫治疗的反应相关的T细胞状态。单元格175、998–1013. e20。

4. 我, 伊沙尔, B., 普拉卡丹, 西南州, 沃兹沃斯, M. H., 第二, 崔西, D., 特伦贝塔, J. J., 罗特姆, A., 罗德曼, C., 连, C., 墨菲, G., 等人。(2016). 用单细胞rna测序技术解剖转移性黑色素瘤的多细胞生态系统. 科学352、189–196。

5. 范德伦, A. M., Thommen, D. S. 和舒马赫. N. (2020). CD8<sup>+</sup>人类癌症中的T细胞状态: 来自单细胞分析的见解. Nat. 发动机的旋转癌症20、218–232。

6. 郑氏, C., 郑, 柳, C. 郭、H、张、Y、郭、X、康, B., 胡, R., 黄, J. Y., Zhang, Q., 等人。(2017). 通过单细胞测序显示的肝癌中浸润性T细胞的景观. 单元格169、1342–1356. e16。

7. 空白, C. U., 海宁, W. N., 赫尔德, W., 霍根, P. G., 卡利斯, A., Lugli, E., 林恩, 菲利普, 饶, a., 雷斯托夫, n. P. 等人。(2019). 定义“T细胞衰竭”。Nat. 发动机的旋转免疫醇。19, 665 – 674.

8. 卡, A., Zehn, D., 和乌兹施耐德德汽车公司, D. T. (2020). 前体耗尽的T细胞: 免疫治疗成功的关键? Nat. 发动机的旋转免疫醇。20, 128 – 136.

9. 克莱门特, C. G., Mihm, M. C., 小法官, 布法利诺, 祖里达, 科里尼, P. 和卡西内利, N. (1996). 肿瘤浸润淋巴细胞在原发性皮肤黑色素瘤垂直生长长期的预后价值. 癌症77、1303–1310。

10. 萨瓦斯, 维拉萨, B., 也门 C., 萨利姆 A., Mintoﬀ, 中央总统, 卡拉米亚, 萨加多, 伯恩, D. J., Teo, Z. L., 杜什扬滕, S., 等人: 凯瑟琳·库宁汉姆基金会家族性乳腺癌研究联盟 (kConFab) (2018年)。乳腺瘤T细胞的单细胞图谱显示了一个与改善预后相关的组织驻留记忆子集. Nat. 医学24, 986 – 993.

11. 厂主公元前, 参议员, 阿博西, R., 比, K., 维尔库德, Y. V., 拉弗勒, M. W., 耶茨, K. B., Lako, A., Felt, K, Naik, g. s. 等。(2019). 耗尽的CD8子集<sup>+</sup>T细胞不同程度地介导肿瘤控制和对检查点封锁的反应. Nat. 免疫醇。20, 326 – 336.

12. 西迪基, 朔伊布勒, 陈努帕蒂, 富尔特斯, 马拉科, 卡尔德隆, 科皮特, 佩斯, 费雷拉, D., 卡莫纳岛, S. J., 斯卡佩利诺, L., 格菲勒, D., 普拉德万德, S., 等。(2019). 瘤内Tcf1<sup>+</sup>第1页<sup>+</sup>CD8<sup>+</sup>具有干细胞样特性的T细胞在疫苗接种和检查点封锁免疫治疗中促进肿瘤控制. 豁免50, 195 – 211. e10。

13. 郭, 十, 张, 年, 郑, 左., 郑、C., 宋, J., 张, Q., 康, B., 刘, z., 金, l., 邢, R., 等。(2018). 通过单细胞测序, 非小细胞肺癌中T细胞的整体特征. Nat. 医学24, 978 – 985.

14. 街道, K., 里索, D., 弗莱彻, R. B., 达斯, D., 恩盖, J., 优素福, Purdom, 和Dudoit, S. (2018). 弹弓: 单细胞转录组学的细胞谱系和假时间推断. BMC基因组学19477。

15. 安德烈塔, M., 科里亚-奥索里奥, 穆勒, 古巴, 库科斯, G. 和卡莫纳, S. J. (2021). 使用参考图谱从单细胞转录组学数据中解释T细胞状态. Nat. 通勤。12, 2965.

16. 范登贝尔热, 贝齐耶斯路, K街, 华盛顿, 纽约, 杜多伊特和克莱门特。(2020). 基于轨迹的单细胞测序数据的差异表达分析. Nat. 通勤。11, 1201.

17. 张, L., 于, X, 郑, L, 张, Y, 李, Y, 方, Q, 高, R., 康, B., 张, Q, 黄, J. Y. 等人。(2018). 谱系追踪揭示了结直肠癌中T细胞的动态关系. 自然564、268–272。

18. 沃尔夫, 金, W., 佐皮, G., 沃格尔, 我, 阿赫梅多夫, M., 布莱克, C. K. E., Beltraminelli, T., Rieckmann, J. C., 拉米雷斯, N. J., 贝内韦托, M., 等人。(2020). 维持T细胞准备的蛋白质翻译的动态. Nat. 免疫醇。21, 927 – 937.

19. 阿兰 D., 发狂的刘、刘、吴、e、冯、五、徐、A., 查克, 奈卡瓦迪, R. P., 沃尔特斯, P. J., Abate, A. R., 等人。(2019). 肺单细胞的参考分析

测序显示了一个过渡性促纤维化巨噬细胞. Nat. 免疫醇。20, 163 – 172.

20. 斯佩克特, A. T., 和李, J. (2017). LEAP: 利用伪时间排序构建单细胞rna测序数据的基因共表达网络. 生物信息学33、764–766。

21. 陈, T. E., 斯腾普夫, M. P. H., 和巴布蒂, A. C. (2017). 基因调控网络推断从单细胞数据使用多变量信息测量. 细胞囊肿。5, 251 – 267. e3.

22. 范德桑德, B., Flerin, C., 戴维, K., 德韦格纳, M., 赫尔塞尔曼, G., 艾巴尔, S., 瑟林克, R., 塞伦斯, W., 康顿, 罗松, Q., 等。(2020). 一个可扩展的单细胞基因调控网络分析的风景工作方案. Nat. 协议。15, 2247 – 2276.

23. 松本, H., 柯字, H., 古川, C., 高, M. S. H., 高, S. B. H., 豪达, 海东, 和尼北海道, I. (2017). SCODE: 一种基于分化过程中单细胞rna测序的高效调控网络推理算法. 生物信息学33、2314–2321。

24. 高, 北, 南. M. M., 甘德里龙, O., 和古纳万, R. (2018). 准确性: 从带有时间戳的单细胞转录表达谱推断基因调控网络. 生物信息学34、258–266。

25. 威林格, 弗里曼, 赫伯特, 长谷川, H., 麦克迈克尔, A. J. 和Callan, M. F. (2006). 人幼稚CD8<sup>+</sup>T细胞在体内外遇到抗原后下调WNT通路转录因子淋巴增强因子结合因子1和转录因子7 (T细胞因子1) 的表达. J. 免疫醇。176, 1439 – 1446.

26. Kaech, S. M. 和Cui, W. (2012). 效应器和记忆CD8的转录控制<sup>+</sup>T细胞分化. Nat. 发动机的旋转免疫醇。12, 749 – 761.

27. 津久, 南, 元, 元, 木藤, A., 竹内公司, A., 小谷, 黑明, 岩岚, 斋藤。(2013). Bach2通过抑制效应记忆相关基因来维持T细胞处于幼稚状态. 过程Nat1。阿卡德. 科学. 美国110, 10735–10740。

28. 普雷斯顿, 辛克莱, 卡斯卡尔, 赫克尔曼, J. L., 纳瓦罗, M. N., 费雷罗, 我, 麦克唐纳, 考林, 考林, 坎特雷尔。(2015). 通过T淋巴细胞中抗原受体信号强度和白细胞介素-2对Myc表达的单细胞调节. EMBO J. 34, 2008 – 2024.

29. 邢、S、李、F、曾、Z、赵、Y、于、S、山、Q、李、Y、菲利普斯、F. C、美纳、P. K、齐、H. H. 等。(2016). Tcf1和Lef1转录因子建立CD8<sup>+</sup>通过内在的HDAC活性来识别T细胞. Nat. 免疫醇。17, 695 – 703.

30. 利维, 哦。(2010). 通过BATF筋疲力尽. Nat. 发动机的旋转免疫醇。10, 747.

31. 陈, Z., 季、Z、乔、S. F., 曼内、S、蔡、Z., 黄, 约翰逊, 约翰逊, 约翰逊, 斯塔普, 约翰逊, B., 徐勇明, C., 以及其他(2019). 以tcf1为中心的转录网络驱动效应与耗尽CD8<sup>+</sup>T细胞命运决定. 豁免51、840–855. e5.

32. 苏密达, T. S., 杜尔伯格, S., 舒普, J., 史迪威, H. A., Axisa, P. P., 科米, 林肯, 昂特曼, A., 卡明斯基, 北州, 马迪, A., 以及其他(2020). I型干扰素转录网络调控人类T细胞中共抑制受体的表达. bioRxiv. <https://doi.org/10.1101/2020.10.30.362947>.

33. 福克斯曼贝斯, J. I., Diallo, A., 尼尔逊, 索托, 迈尔斯, C. L. 和沃尔霍特, A. J. (2013). 使用网络来衡量基因之间的相似性: 关联指数选择. Nat. 方法10、1169–1176。

34. 李, J., 他, Y. 郝强、倪李、董C. (2018). 高水平的Eomes促进抗肿瘤CD8的衰竭<sup>+</sup>T细胞. 前面免疫醇。9, 2981.

35. 金, K., 公园, 公园, 公园, S. Y., 金, 公园, S. M., 赵, J. W., 金, D. H., 公园, Y. M., Koh, Y. W., 金, H. R. 等。(2020). 单细胞转录组分析显示, TOX是T细胞衰竭的促进因子, 也是人类癌症中抗pd1反应的预测因子. 基因组医学。12, 22.

36. 阿伊罗尔迪, 米格里奥拉蒂, 布鲁斯科利, 马尔凯蒂, C., 佐罗, 坎纳里尔, 达达米奥, F. 和里卡尔迪, C. (2001). 糖皮质激素诱导的亮氨酸拉链因子通过抑制核因子kB来调节t细胞的激活. 血液98743–753。

37. 李, H., 范德伦, 上午, 尤夫, 我, 卢布林, 格尔巴德-索洛德金, D., 范阿库伊, A. C. J., 范登·布拉伯尔, M., 罗兹曼, E. A., 哈南, J. B. A. G., Blank, C. U., 等人。(2019). 功能失调的CD8<sup>+</sup>T细胞在人类黑色素瘤中形成了一个增殖的、动态调节的隔壁. 细胞176、775 – 789. e18。





38. 楼, 李, 孔, X, 张, 刘, 李, D, 格斯坦, M. (2020). 主题网: 一个测量转录调控网络变化的框架。生物信息学36 (Suppl\_1), i474-i481。
39. 金, H. J., Osteil, P., 汉弗莱, S. J., Cinghu, S., 奥德菲尔德, A. J., 帕特里克, E., 威尔基, E. E., 彭, G., Suo, S., Jothi, R., 等。 (2020). 整合统计学习揭示了多能性进展过程中的转录网络动态。核酸。48, 1828 - 1842。
40. Doiling, T. A., 克劳福德的变体 A., Angelosanto, J. M., 佩利, M. A., 齐格勒 C. G. 和惠利, E. J. (2012). 网络分析揭示了中心连接的基因和路径-CD8中涉及的方法<sup>+</sup>T细胞衰竭与记忆的关系。免疫力37、1130-1144。
41. 曼恩, T. H., 和凯奇, S. M. (2019). 抽搐一下, 是时候让T细胞衰竭了。Nat. 免疫。20, 1092 - 1094。
42. Cannarile, L. 德尔菲努法学博士, 奥多里西奥, 里卡尔迪, C., 和爱尔兰, E. (2019). 暗示了GILZ在糖皮质激素调节t细胞激活中的作用。前面免疫。10, 1823。
43. 太阳 D., 王, J., 韩, Y., 东, X., 葛, J., 郑, r., 石, X., 王, B., 李, z, , 任, p., 等。 (2021). TISCH: 一个全面的网络资源, 使交互式的单细胞转录组可视化的肿瘤微环境。核酸。49 (D1), D1420 - D1430。
44. 布劳恩, 弗伦奇, M., 和蒂尔, A. (2015). 业余爱好和人类效应t细胞的分化: 一个漫长的旅程的开始。欧洲人J. 免疫。45, 2762 - 2765。
45. 刘, Y., 和张, W. (2020). HOPX在正常组织和肿瘤进展中的作用。生物科学。棱纹平布40, BSR20191953。
46. 胡贝尔, 苏普鲁曼科, 阿什赫斯特, 马尔巴赫, 雷费, 沃尔夫, S, 斯特雷克, T., 北京, 荣格, 奥伯曼, 约翰逊, 等。 (2017). IRF9防止CD8<sup>+</sup>急性淋巴细胞性脉络丛脑膜炎病毒感染期间T细胞的外部方式衰竭。J. 维罗尔。91, e0121917。
47. 萨伦斯, W., 坎努特, 托多罗夫, H., 和萨伊斯, Y. (2019). 单细胞轨迹推理方法的比较。Nat. 生物技术。37, 547 - 554。
48. 周, Y., 周, b., 帕切, L., 张, M., 科达巴赫什, A. H., 塔纳西胡克, 本纳, C. 和昌达, 美国。 (2019). 元景观为系统级数据集的分析提供了一个面向生物学家的资源。Nat. 通勤。10, 1523。
49. 辛, J., 伯格, D. A., 朱, Y., 申, J. Y., 宋, J., 博纳吉迪, M. A., 埃尼科洛波夫, G., 诺恩, D. W., 基督教, K. M., 明, G. L. 和Song, H. (2015). 单细胞RNAseq与瀑布揭示了成人神经发生背后的分子级联反应。细胞干细胞17, 360-372。
50. Aibar, S., 冈萨雷斯-布拉斯, C. B., 采购产品莫尔曼, 图, 图, 图, 硕士, 硕士, 等, 等, 等。 (2017). 风景优美: 单细胞调控网络的推理和聚类。Nat. 方法14、1083-1086。
51. 金, S. (2015). 一种快速计算半偏相关系数的R包。通勤。斯达应用程序。方法22、665-674。
52. 嗯, 我, 嗯, A., 温克尔, L., 和盖尔茨, P. (2010). 使用基于树的方法从表达式数据中推断调控网络。PLoS ONE 5, e12776。
53. Pratapa, A., Jalihal, 美联社, 劳, J. N., 巴拉德瓦吉, A., 和穆拉利, T. M. (2020). 从单细胞转录组数据推断基因调控网络的基准算法。Nat. 方法17、147-154。
54. 萨克斯, M. C. (2017). plotROC: 一个绘制ROC曲线的工具。J. 斯达软。79, 2。
55. 周, s., 黄, 叶叶, 刘, H., 周, X., 袁, M., 侯, F., 王, L., 江, W. (2021). 单细胞RNA-seq基于基因调控网络分析了三阴性乳腺癌的瘤内异质性。摩尔每一个。核酸23, 682-690。
56. 科尔德, 劳尔, 阿德勒, P. 和维洛, J. (2012). 基因列表整合和元分析的鲁棒秩聚合。生物信息学28、573-580。
57. 卡桑巴拉, 科辛斯基, M., 和比切克, P. (2021). 使用“ggplot2”绘制生存曲线。R软件包版本0.4.9, <https://cran.r-project.org/web/packages/survminer/index.html>。

OMTN, 第26卷

## 补充信息

T细胞运动轨迹的动态调控网络

分析T细胞状态转变的转录控制

闵燕、胡静、袁华亭、徐立文、廖高明、毛泽东、朱嘉丽、薄庞博、平燕燕、张云鹏、萧云云、夏丽

**表 S1: 数据集 习惯于 为了 修建 精力充沛的 网络**

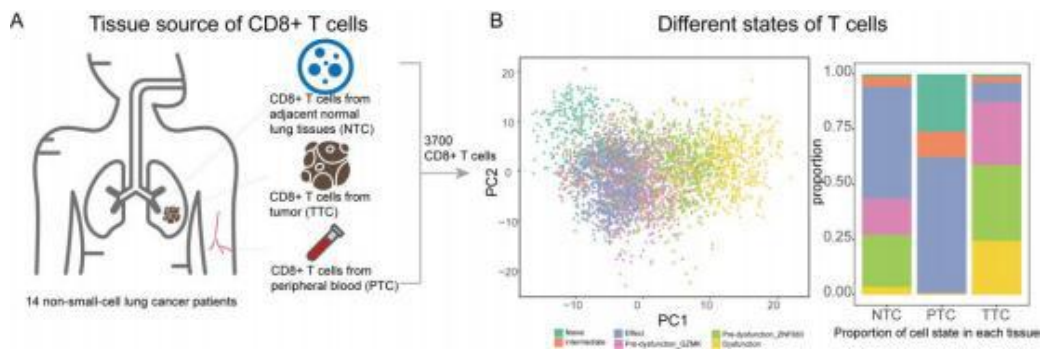
资料组	癌 类型	#CD8+ T 细胞	平台发布	
GSE127471	非小细胞 肺癌	300	10X	纽曼AM等人。Nat生物技术2019
GSE120575	斯克姆	6695	智能塞克2	Sade-Feldman M等。单元格2018
GSE140228	利希克	16532	10X	Zhang Q等。单元2019
GSE118056	保持密切 接触	756	10X	Paulson KG等。2018年 纳特委员会
EMTAB6149	非小细胞 肺癌	12040	10X	Lambrechts D等。Nat Med 2018
GSE103322	汉斯科	434	智能塞克2	Puram SV等。单元格 2018
GSE123813	短路电流	15560	10X	Yost KE等人。Nat Med 2019
GSE110686	乳腺癌易 感基因	1525	10X	Savas P等人。Nat Med 2018

**表S2。与T细胞功能障碍相关的治疗TFs。**

表S3。在每个窗口中具有显著不同的调控活动的TFs。

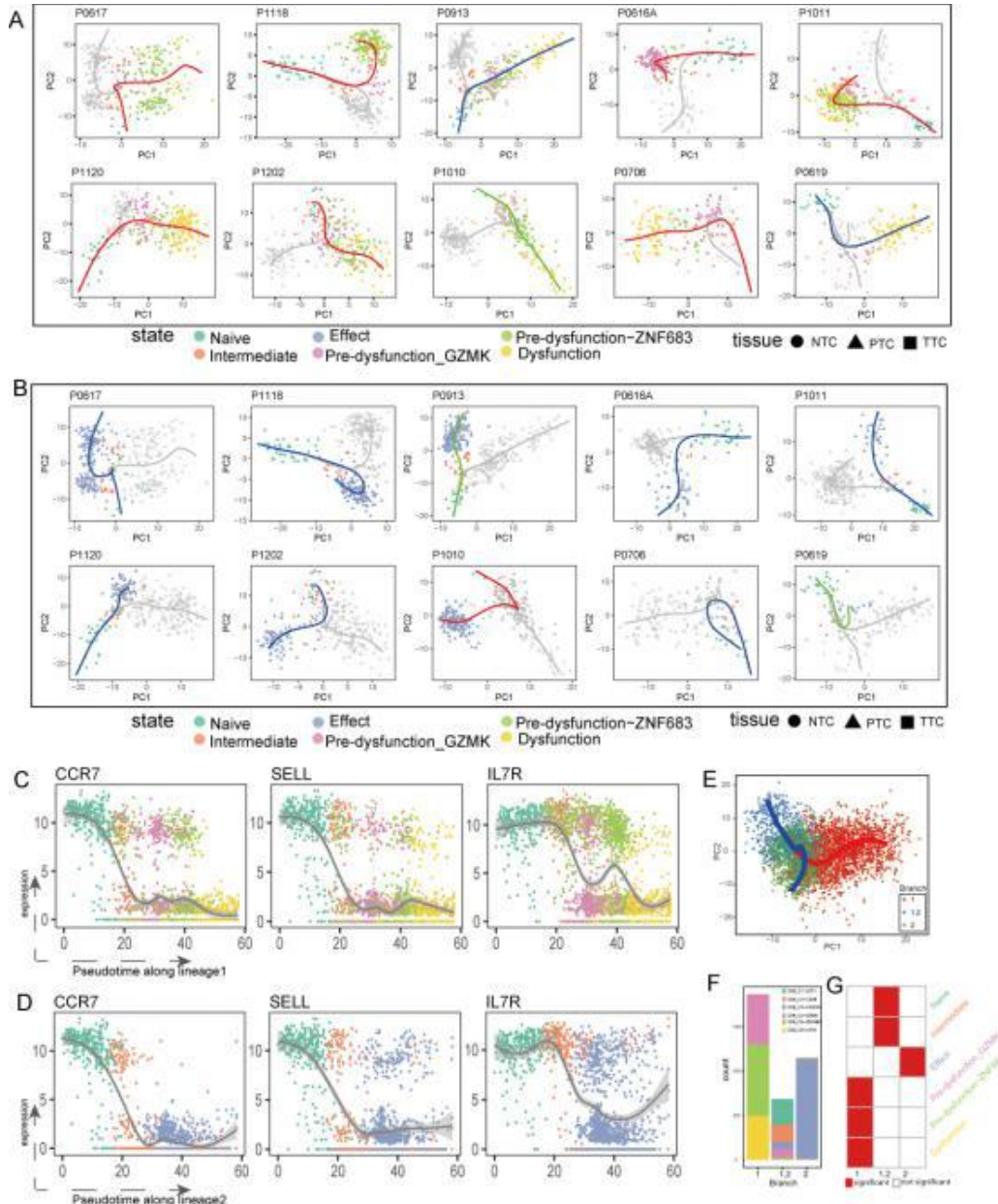
窗户	方向	数量	TF
W1	往...的下 端	23	ARID4B, ATF6B, ATRX, BACH2, BCL11B, ETS1, FOXP1, HIVEP2, IKZF1, IRF9, KLF2, KLF3, LEF1, MYC, SATB1, TCF7, TRERF1, ZGPAT, ZNF292, ZNF48, ZNF652, ZNF83, ZNF91
W1	在上面	24	b、b、1、1、fd2、3、3、prtf1、rbpj、runx3、spl40、zbtb38、znf6 83
W2	往...的下 端	19	HOPX, ID2, KLF2, LEF1, LRRFIP1, LYAR, MXD3, MYBL1, MYC, NFKBIA, PRDM1, RUNX2, SATB1, STAT1, TCF7, TRERF1, TSC22D3, ZGPAT, ZNF683
W2	在上面	13	奶油、奶油、FOS、IKZF3, KDM2A, NCOR1, NFATC2, RBPJ、TOX、trps1、zbtb38、zeb2、znf48
W3	往...的下 端	4	EOMES, FOS, TCF7, TSC22D3
W3	在上面	52	ARID4B, ATRX, BATF, BCL11B, BHLHE40, CREB3L2, CREM, DNAJC1, ETS1, FOXP1, HIF1A, HIVEP3, HMGA1, HOPX, ID2, IKZF1, IKZF3, IRF1, IRF7, IRF9, KAT7, KDM5B, KLF13, KLF6, KMT2A, LRRFIP1, LYAR, MXD3, NCOR1, NFATC2, NFKBIA, NR3C1, PHTF2, PRDM1, RBPJ, RUNX2, RUNX3, SP100, SP140, STAT1, STAT3, STAT4, TRERF1, ZBTB20, ZBTB38, ZEB2, ZGPAT, ZNF276, ZNF292, ZNF48, ZNF652, ZNF683
W4	往...的下 端	27	ATRX、bcl11b、b、1、1、1、1、1、i、f1、KLF12, KLF6, KMT2A, LRRFIP1, LYAR, 异氰酸盐r1、2、2、RUNX3, SP100, STAT4, TRERF1, TSC22D3、zbtb38、zgpata、znf276、znf91
W4	在上面	25	干旱4b、蝙蝠、蝙蝠40、信条2、奶油、奶油1、等V1, HIF1A, ID2, IKZF3, IRF7, IRF9, MAF, MSC、mxd3、nr3c1、phtf2、rbpj、RUNX2, SP140, STAT1, STAT3, TOX, znf29 2、znf48





图S1。NSCLC scRNA-seq的数据收集和统计信息。

(A) NSCLC数据集的详细信息，包括患者数量，组织取样。(B) (左图) T细胞的PCA图，根据其状态着色。(右图) 具有不同来源的细胞的状态组成。

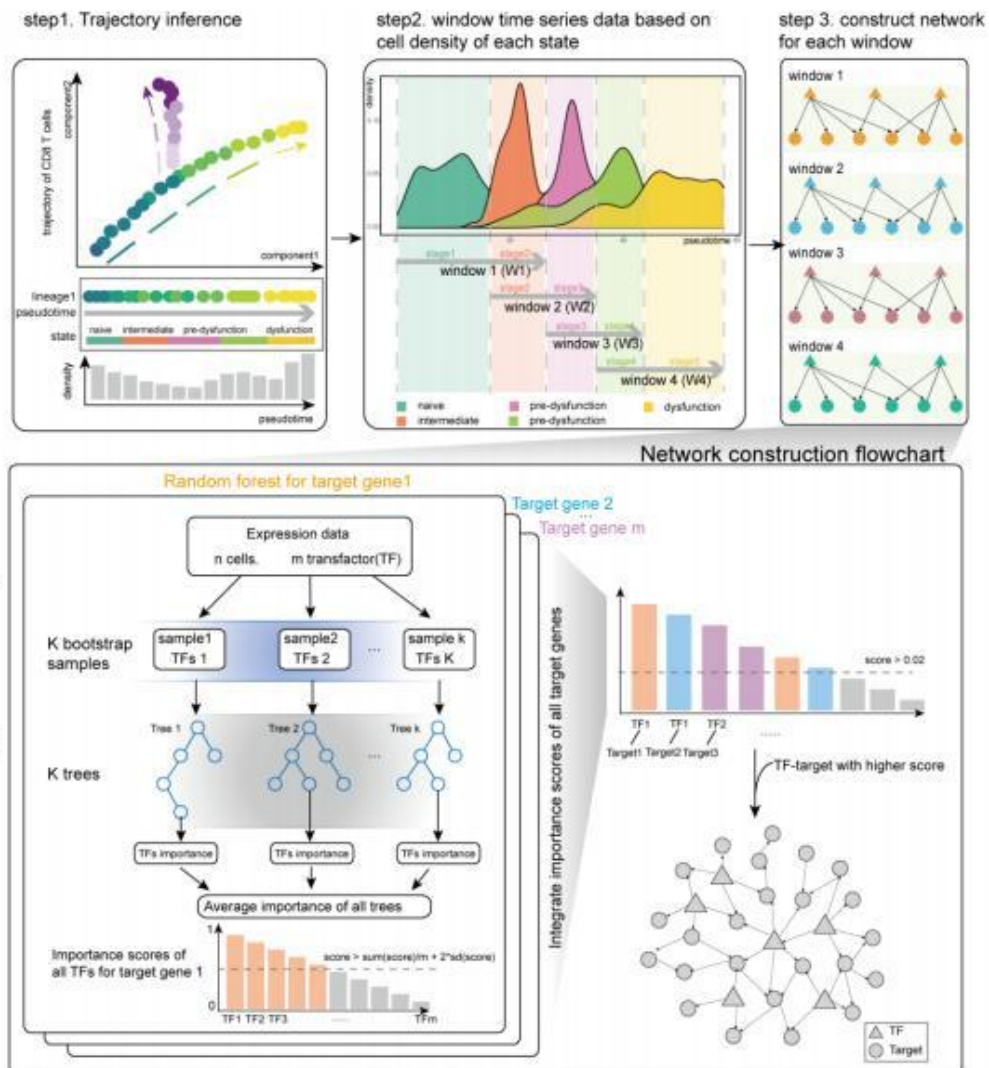


图S2。朴素标记物沿伪时间的表达分布。

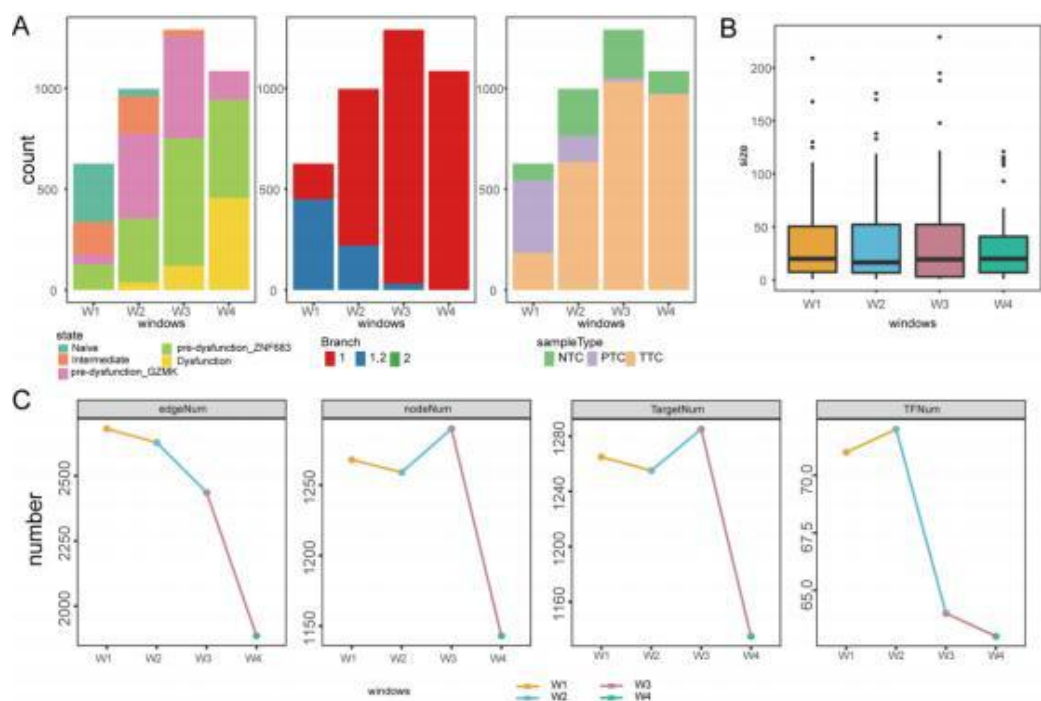
(A-B) 对单个患者的CD8<sup>+</sup> T细胞的假时间分析。彩色系为功能障碍系(A)和效应系(B). 由组织来源表示的点形状。(C-D) 在功能障碍谱系(C)或效应谱系(D). 上的幼稚标记物沿伪时间的表达分布(E)所有细胞的T细胞轨迹, 按分支着色。(F)这三个分支上的细胞状态组成。(G)每个分支中细胞状态的富集。填充红色的盒子表示超几何检验 $p$ 值 $<0.05$ 的显著结果。



The pipeline of dynamic network construction



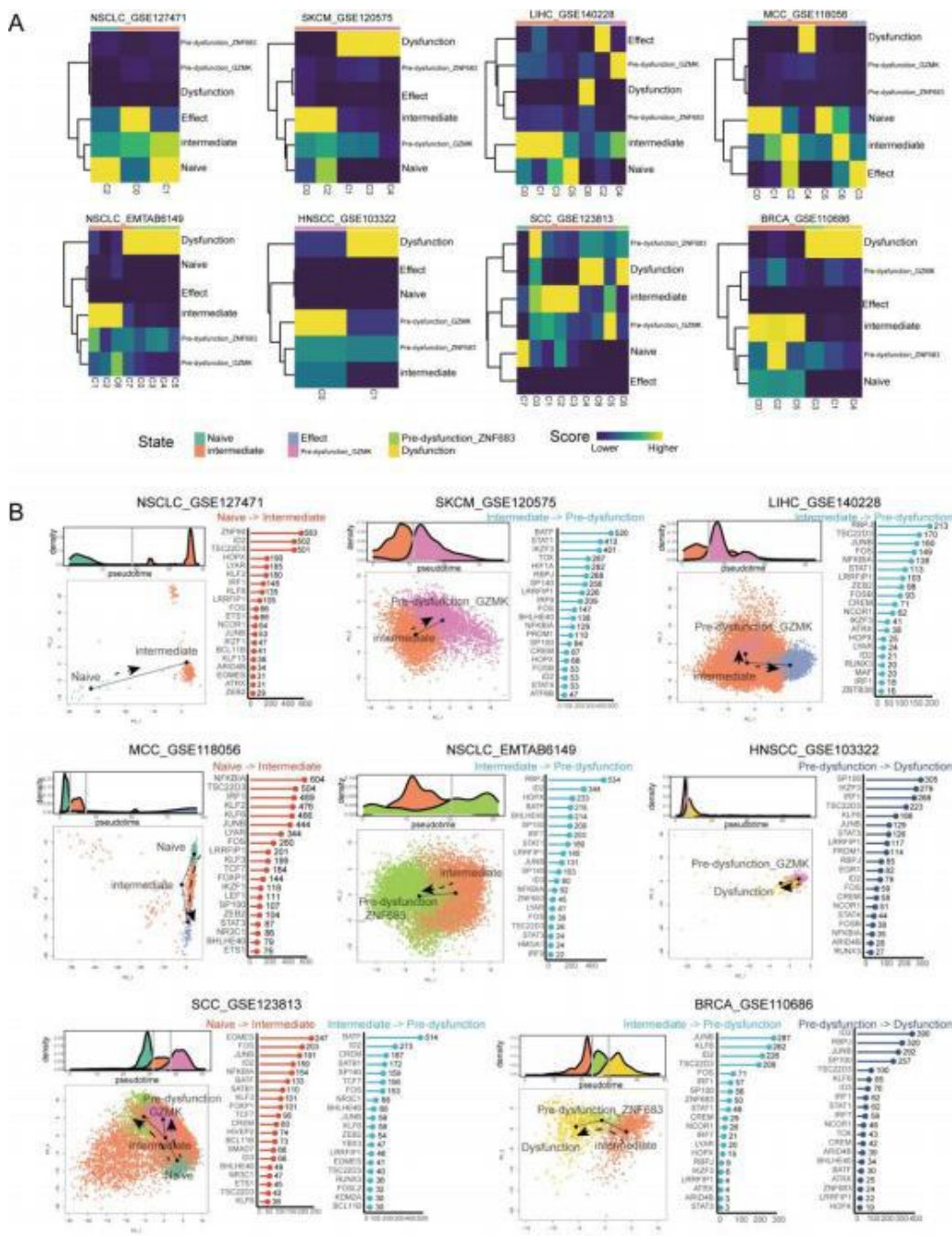
图S4。动态网络建设的管道。



图S5。动态网络的基本统计数据。

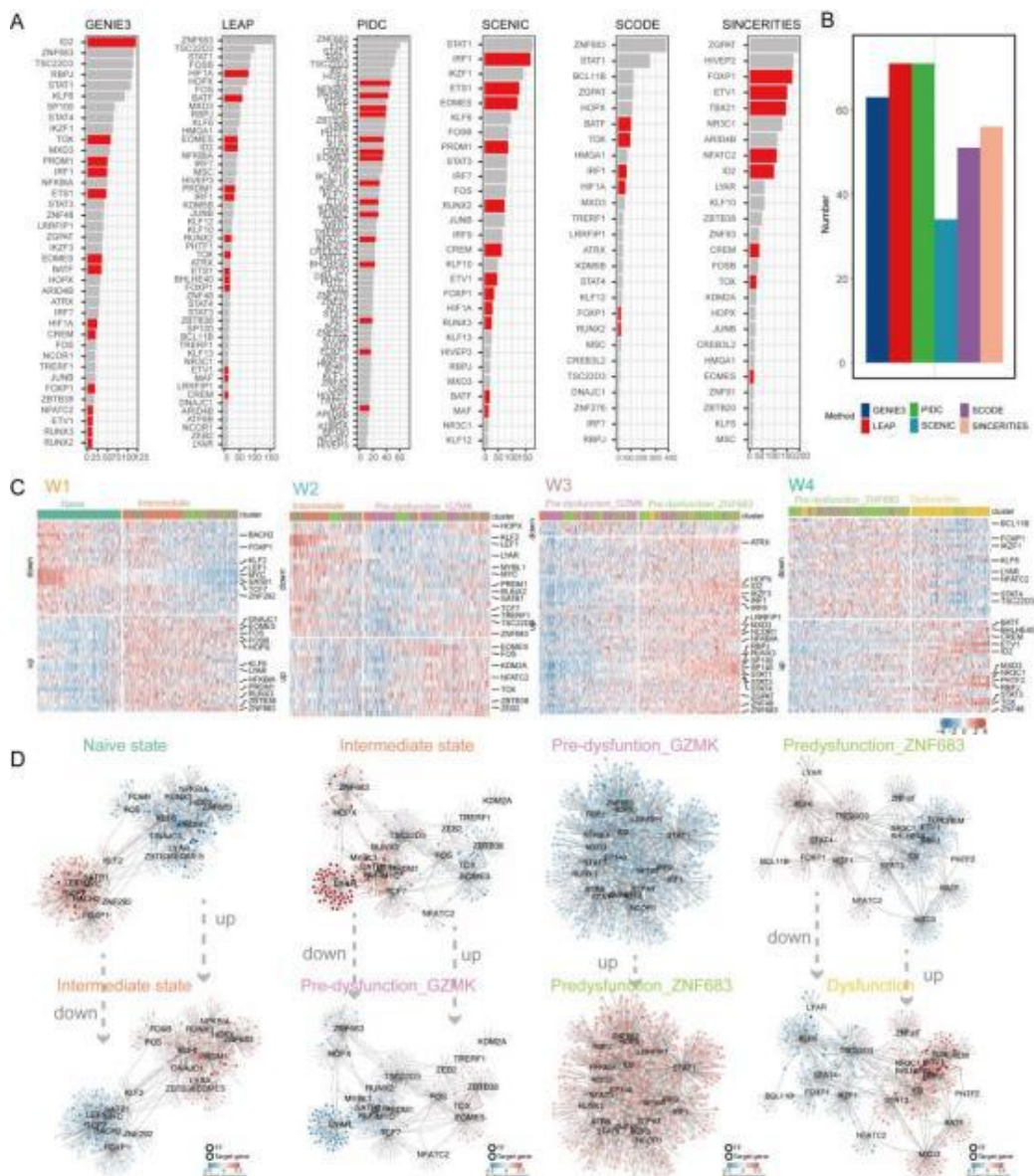
(A) 每个窗口中细胞在状态、分支和组织来源方面的分布。(B) 每个网络中规则大小的分布。(C) 有四个窗口的网络中的边、节点、目标和TFs的数量。





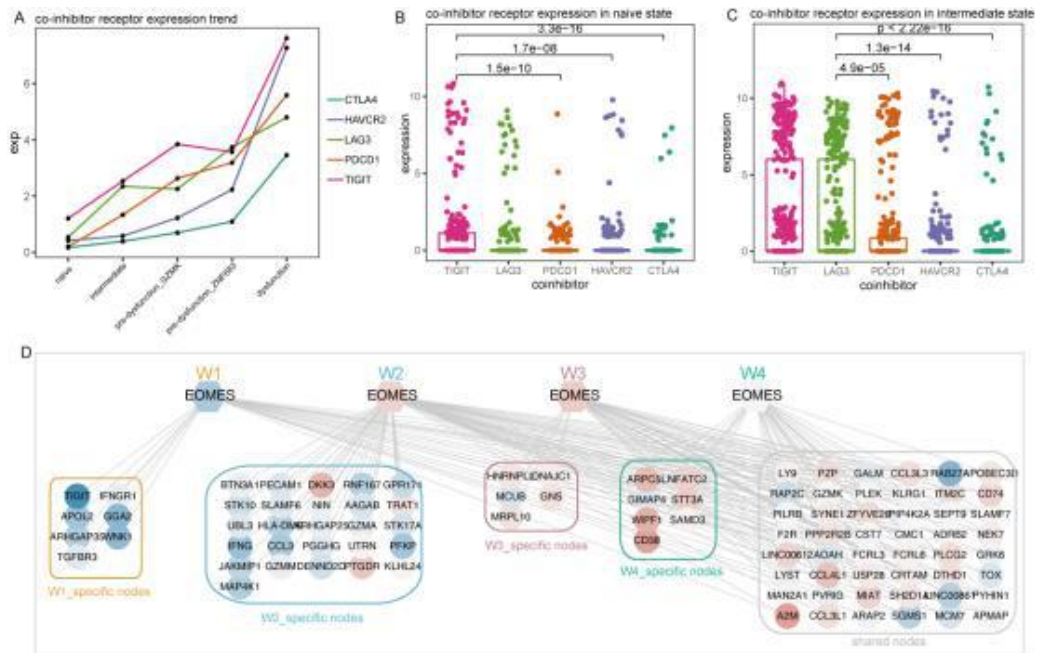
图S6。为独立的数据集构建动态网络。

(A) 跨所有参考状态 (CD8\_C1-LEF1、CD8\_C2-CD28、CD8\_C3-CX3CR1、CD8\_C4-GZMK、CD8\_C5-ZNF683和CD8\_C6-LAYN) 的所有集群的singleR评分。(B) 状态为每个数据集的过渡轨迹和前20个调节器。密度图显示了不同状态的伪时间分布。散点图显示了状态转变的轨迹。棒棒糖图显示了前20个调控因子的靶基因的数量。



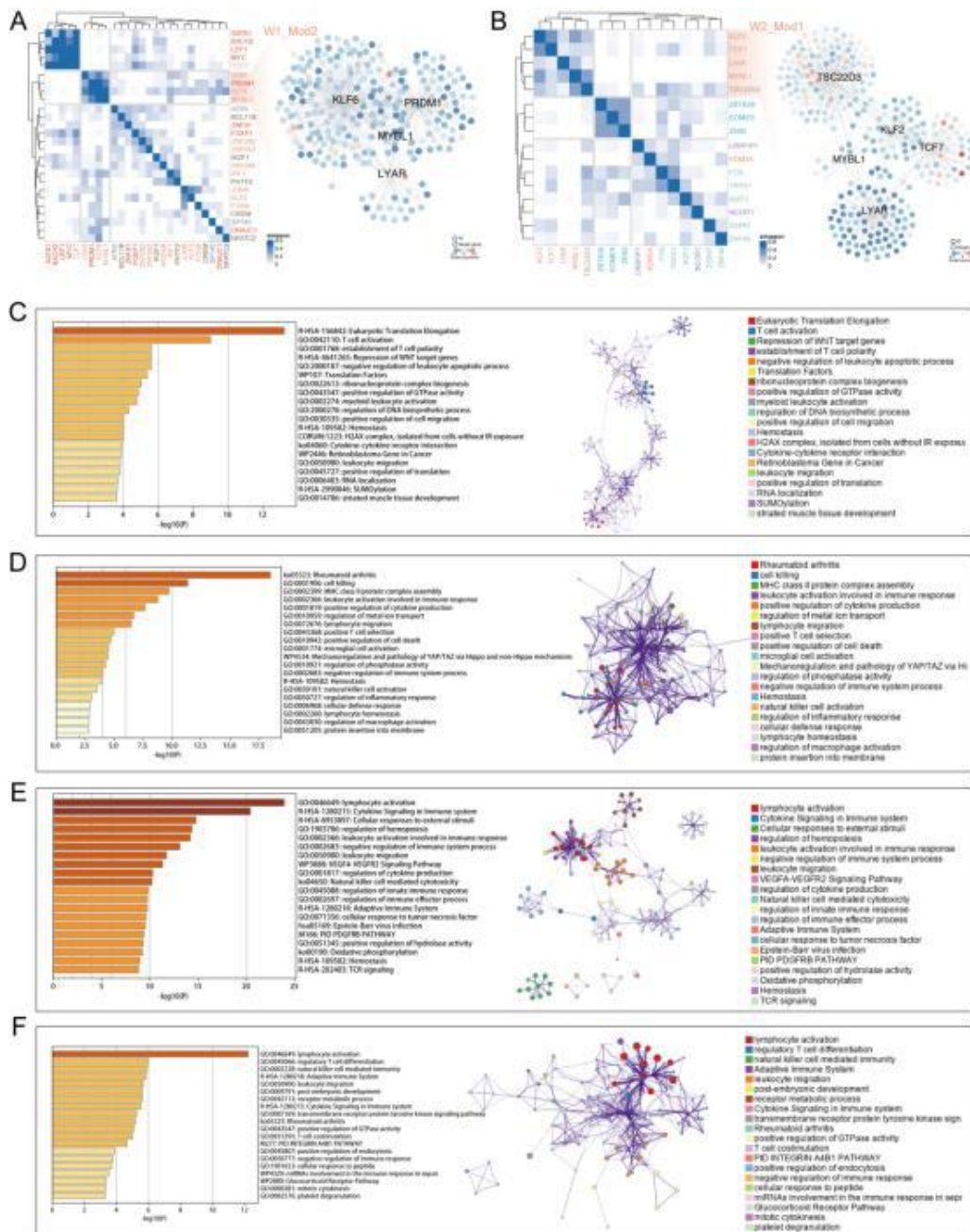
图S7。TFs的活动在不同的状态下发生动态变化。

(A) W4窗口中不同方法识别的(A)调节器。这些调控因子按靶基因的数量排序。标记与T细胞功能障碍相关的副酸TFs。(B)在窗口W4中通过不同方法识别的调节器的数量。(C)各窗口细胞中显著不同的TF调控因子的活性(Wilcoxon秩和检验,  $FDR < 0.05$ )。前20个显著的TFs被标记。(D)每个窗口前20个TFs对应两种状态之间的基因表达变化。节点用相应状态下的平均表达式来着色。

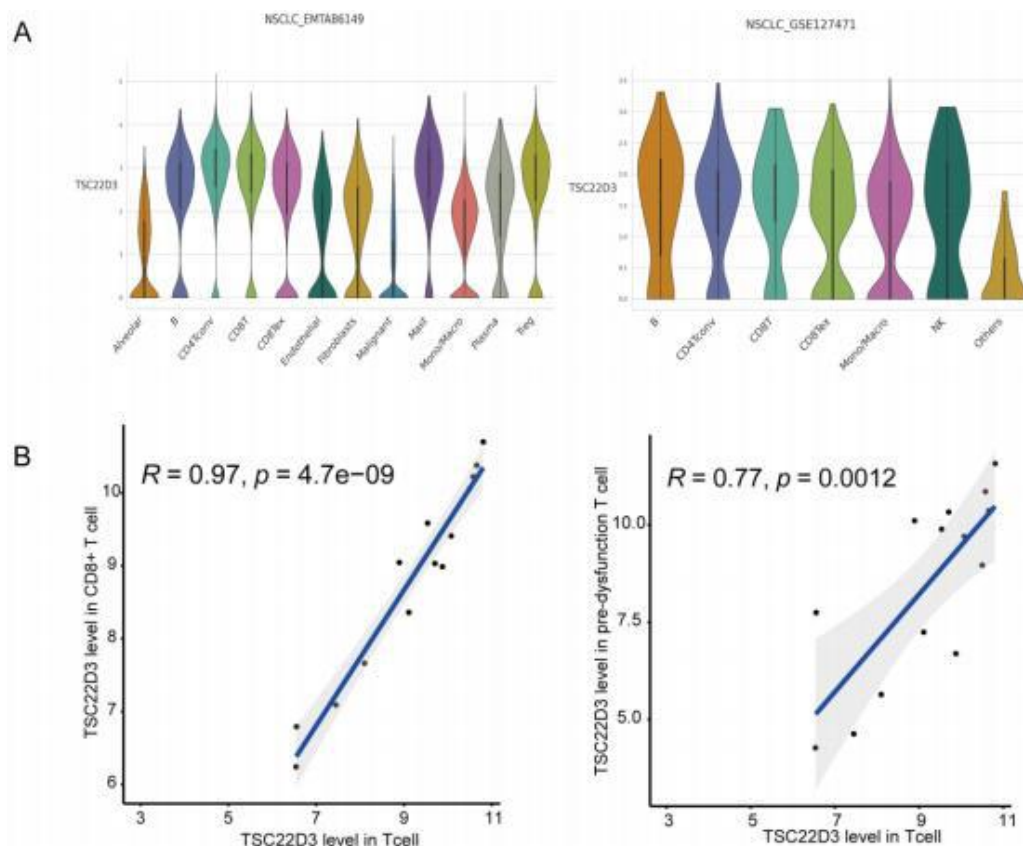


图S8。共抑制受体的表达动态，以及转录因子EOMES的调控重组。(A)不同状态下细胞中各共抑制受体的平均表达量。(B-C)在W1窗口内的幼稚细胞(B)或中间细胞(C)中共抑制受体的表达。(D)前四组靶基因分别代表W1、W2、W3和W4的窗口特异性靶点，最后一组基因包含至少有两个窗口共享的靶点。





图S9。窗口特定的监管模块的丰富条款。(A) 模块M2在窗口W1中。左。W1网络中TFs之间的辛普森指数。右边的M2中基因的表达。(B) 模块M1在窗口W2中。(C-F) W1(C) 模块M1、W2 (D) 模块M2、W3 (E) 模块M1和W4 (F) 模块M1的富集结果可视化。



图S10。TSC22D3在不同细胞类型中的表达。

(A) 在EMTAB6149和GSE127471数据集中，TSC22D3在不同细胞类型中的表达。(B) 在核心数据集GSE99254中，T细胞中TSC22D3的表达水平与功能障碍前T细胞或CD8+ T细胞中TSC22D3的表达水平的相关性。