

论著·基础研究

基于TCR组库分析流程的非小细胞肺癌特征分析

赵卓明^{1,2}, 刘振浩², 鲁曼曼², 张 钰², 许林锋², 谢 鹭^{1,2}

1. 上海海洋大学食品学院食品科学与工程学科(系), 上海 201306; 2. 上海市生物医药技术研究院基因组与生物信息研究所, 上海市疾病与健康基因组学重点实验室, 上海 200237

[摘要] **目的**·探究非小细胞肺癌(non-small cell lung cancer, NSCLC)免疫相关特征, 发掘V-J基因中的肿瘤潜在标志物, 为建立T细胞受体(T-cell receptor, TCR)-抗原识别预测模型提供基础。**方法**·收集704例NSCLC样本, 用于构建一套系统的TCR组库分析流程, 涵盖从原始数据到质控、过滤、TCR序列的识别和提取等上游分析部分, 以及组库克隆分布、克隆分型、V-J基因共享性、互补决定区3(CDR3)分布特征以及克隆追踪等下游分析流程。使用Shannon-Weiner指数、Chao1等指数分析样本克隆分布。根据克隆扩增数目进行克隆分型, 以探索不同分型间的差异。选取2例甲状腺乳头状癌(papillary thyroid carcinoma, PTC)样本进行去克隆扩增权重V-J基因共享分析, 以探究低频克隆型对V-J基因共享性的影响。最后, 通过分析V基因及高频克隆型CDR3分布特征和克隆追踪分析, 监测肿瘤免疫前后克隆型频率的变化, 寻找潜在的肿瘤标志物。**结果**·①NSCLC不同组织间克隆分布和克隆分型均存在显著差异, 且在不同年龄与性别间存在显著差异。②在V-J基因共享性分析中发现了特定高共享V-J基因, 同时在CDR3分布特征分析中存在非正态分布的高克隆V基因与氨基酸高频克隆型。③高频克隆型克隆追踪分析中, 观察到NSCLC中高表达或新表达的高频克隆型, 提示这些克隆型可作为潜在的肿瘤相关抗原或新抗原结合CDR3参考序列。④通过去克隆扩增权重V-J基因共享分析, 发现原本低表达的TRBJ2-5基因表达频率显著增加, 提示该基因可能为肿瘤免疫反应的潜在低频关键基因。**结论**·TRAV21和TRBV6.5基因在NSCLC中高克隆扩增, 可作为潜在肿瘤标志物。

[关键词] T细胞受体; TCR组库; 肿瘤免疫; 非小细胞肺癌

[DOI] 10.3969/j.issn.1674-8115.2023.12.006 **[中图分类号]** Q939.91; R730.3 **[文献标志码]** A

Analysis of tumor-related features of non-small cell lung cancer based on TCR repertoire workflow

ZHAO Zhuoming^{1,2}, LIU Zhenhao², LU Manman², ZHANG Yu², XU Linfeng², XIE Lu^{1,2}

1. Department of Food Science and Engineering, College of Food Science & Technology, Shanghai Ocean University, Shanghai 201306, China; 2. Shanghai-MOST Key Laboratory of Health and Disease Genomics & Institute of Geonome and Bioinformatics, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai 200237, China

[Abstract] **Objective**·To explore the immune-related characteristics of non-small cell lung cancer (NSCLC), discover potential tumor markers in V-J genes, and lay the foundation for establishing a TCR-antigen recognition prediction model. **Methods**·A total of 704 NSCLC samples were collected to establish a comprehensive T-cell receptor (TCR) repertoire analysis workflow. The upstream analysis included steps such as raw data processing, quality control, filtering, TCR sequence identification, and extraction. The downstream analysis included repertoire clone distribution, clone typing, V-J gene sharing, CDR3 distribution characteristics, and clone tracking. The sample clone distribution was analyzed by using indices such as Shannon-Weiner index and Chao1 index. Clone typing was performed based on the number of clone amplifications to explore differences among different types. The degree of V-J gene segment sharing was analyzed, and the sharing of low-frequency clone types was determined through clone amplification weight analysis of V-J genes by using two samples of papillary thyroid carcinoma. Finally, analysis of the distribution characteristics of V genes and high-frequency clone type CDR3, and clone tracking analysis were conducted to monitor changes in tumor immune clone frequencies before and after analysis, aiming to identify potential tumor markers. **Results**·① Significant differences were observed in clone distribution and clone typing among different NSCLC tissues, as well as among different ages

[基金项目] 国家自然科学基金(31870829)。

[作者简介] 赵卓明(1997—), 男, 硕士生; 电子信箱: PapyrusZhao@outlook.com。

[通信作者] 谢 鹭, 电子信箱: xielu@sibpt.com。

[Funding Information] National Natural Science Foundation of China (31870829)。

[Corresponding Author] XIE Lu, E-mail: xielu@sibpt.com。



and genders. ② Specific highly-shared V-J genes were identified in the analysis of V-J gene sharing, and non-normal distribution of high-clone V genes and amino acid high-frequency clone types were found in the CDR3 distribution analysis. ③ In the analysis of high-frequency clone type clone tracking, highly expressed or newly expressed high-frequency clone types were observed in NSCLC, suggesting that these clone types could serve as potential tumor-associated antigens or bind with CDR3 reference sequences of new antigens. ④ It was found that the expression frequency of *TRBJ2-5* gene, originally low-expressed, significantly increased, indicating its potential role as a key low-frequency gene in tumor immune response. **Conclusion** The *TRAV21* and *TRBV6.5* genes show high clone amplification in NSCLC and could serve as potential tumor biomarkers.

[Key words] T cell receptor (TCR); TCR repertoire; tumor immunity; non-small cell lung cancer (NSCLC)

非小细胞肺癌 (non-small cell lung cancer, NSCLC) 是一种常见的肺癌类型, 约占所有肺癌病例的 85%^[1]。在过去的几十年里, 针对 NSCLC 的研究取得了显著进展, 对其发病机制、诊断方法、治疗策略和预后因素都有了深入了解^[2]。通过 T 细胞受体 (T-cell receptor, TCR) 组库分析可以了解 NSCLC 免疫应答和免疫监视的特征, 从而揭示 T 细胞克隆多样性、免疫应答的活性以及肿瘤逃逸机制^[3], 对于指导免疫治疗策略和预测患者预后具有重要意义。TCR 组库分析可以评估肿瘤组织中 T 细胞克隆的多样性^[4], 以反映免疫系统对抗肿瘤的能力; 高克隆多样性通常与较好的预后结果相关^[5]。同时, 通过比较肿瘤组织和正常组织中的 T 细胞克隆组成, 可以了解肿瘤免疫应答的活性程度, 较高 T 细胞克隆丰度和活性可能意味着更强的抗肿瘤免疫应答^[6]。此外, TCR 组库分析可以识别免疫治疗相关的 T 细胞克隆群体或特定 TCR 序列, 以协助发展个体化的免疫治疗策略^[7], 筛选免疫反应较强的 TCR 用于患者治疗。

目前, 免疫组库测序技术可以较完整地获取患者 TCR 组库多样性等信息, 以促进 NSCLC 免疫治疗发展^[8]。免疫检查点抑制剂已成为 NSCLC 的主要治疗策略之一^[9]。这些抑制剂通过阻断肿瘤细胞与免疫细胞之间的抑制信号, 恢复 T 细胞对肿瘤的免疫攻击能力。ROBERT 等^[10] 研究发现 TCR 组库多样性变化可预测免疫检查点疗法的有效性。此外, T 细胞受体工程化 T 细胞 (T cell receptor-engineered T cells, TCR-T) 疗法^[11]、肿瘤疫苗^[12] 等方法也逐步应用于 NSCLC。ZHU 等^[13] 的研究将小鼠中特定 TCR 基因重定向至人类 T 细胞, 并通过体外实验观察到该基因导入对肝癌细胞的杀伤作用, 为 TCR-T 疗法提供了理论基础。本研究组已开发了 iTcep^[14], 可用于预测给定抗原肽的 TCR 结合特异性, 为肿瘤疫苗研发提供参考。

目前, 已存在一些免疫组库分析工具 (如 MiXCR^[15]、Imonitor^[16] 等), 但能实现从原始 TCR

组库测序数据到下游组库特征分析以及结果可视化完整流程的软件工具还很欠缺。为此, 本研究拟构建一套系统的 TCR 组库分析流程以探究肿瘤相关特征和发掘潜在肿瘤标志物, 涵盖从原始数据到质控、过滤、TCR 序列的识别和提取等上游分析部分, 以及组库克隆分布、克隆分型、V-J 基因共享性、互补决定区 3 (CDR3) 分布特征以及克隆追踪等下游分析流程, 并将其用于分析 NSCLC 患者的肿瘤相关特征。通过对免疫组库的深入研究和分析, 可以更好地理解免疫系统的工作原理, 为免疫治疗和药物研发提供有价值的信息, 也为后续封装成软件包并提供网页工具打好前期基础。

1 资料与方法

1.1 TCR 序列原始数据准备

从 NCBI BioProject 数据库获取 NSCLC (PRJNA544699) 和甲状腺乳头状癌 (papillary thyroid carcinoma, PTC; PRJNA664708) 患者样本 TCR 测序数据集, 用于流程方法测试及生物学特征案例分析和验证。其中, NSCLC 数据集包括 72 例患者的 220 个肿瘤区域样本、64 个匹配的非肿瘤肺组织样本和 70 个外周血单核细胞样本测序原始数据^[17]; 共计包含 704 个样本, 包括 T 细胞受体 α 链 (T-cell receptor α chain, TRA) 样本 353 例, T 细胞受体 β 链 (T-cell receptor β chain, TRB) 样本 351 例。PTC 数据集包括 2 例 PTC 患者肿瘤样本 TCR 测序数据^[18]。

1.2 数据预处理

使用 SRA Toolkit 工具将 SRA 格式原始数据转化为 fastq 格式文件, 作为流程起始输入文件。通过 FastQC 质量检查, 使用 Trimmomatic 程序去除接头序列、低质量序列、重复序列和低质量克隆, 对数据进行过滤以提高数据的准确性。

1.3 序列比对

基于MIXCR分析流程进行TCR序列比对。使用species参数设置物种为homo sapiens, 使用receptor type参数设置数据类型为TCR, 将质控后测序数据与MiXCR内置参考V、D、J和C基因片段数据库进行比对。

1.4 中间结果重注释

为方便下游分析, 对前述分析结果进行重注释, 筛选出cloneCount (clone数量)、cloneFraction (clone比例)、allVAlignments (V区基因比对)、allDAlignments (D区基因比对)、allJAlignments (J区基因比对)、nSeqCDR3 (CDR3区核苷酸序列)、aaSeqCDR3 (CDR3区氨基酸序列)、refPoints (序列比对参考位置) 等结果。

1.5 克隆分布及克隆分型分析

使用Shannon-Wiener指数、Chao1指数以及Pielou均匀度计算TCR组库中的多样性 (diversity)、丰富度 (richness) 与克隆性 (clonality)。

免疫组库的多样性使用Shannon-Wiener指数^[19]进行评估, 该指数是反映免疫组库多样性的经典指标。计算公式如下:

$$\text{Diversity} = -\sum_{i=1}^s (P_i \log_2 P_i)$$

其中, s 表示物种数目, P_i 表示该物种个体数占总个体数的比例。

免疫组库的丰富度使用Chao1指数^[20]计算。Chao1指数是丰富度的非参数渐近估计量, 计算中低频克隆性权重较高, 能较大程度反映免疫组库克隆型丰富度。计算公式如下:

$$\text{Richness} = S_{\text{obs}} + \frac{n_1(n_1 - 1)}{2(n_2 + 1)}$$

其中, S_{obs} 表示样本中观察到的克隆型数目, n_1 表示只有一条序列的克隆型数目 (single tons), n_2 表示只有2条序列的克隆型数目 (double tons)。

克隆性指的是一组具有相同或相似免疫受体, 即源自同一克隆的免疫细胞。克隆性是反映免疫组库中高克隆型扩增程度的指标, 可用于评估免疫系统中的克隆扩增情况和免疫应答的多样性。克隆性可用Pielou均匀度指数^[21]计算:

$$\text{Clonality} = 1 - \frac{-\sum_{i=1}^s (P_i \log_2 P_i)}{\log_2 s}$$

其中, P_i 表示该物种个体数占总个体数的比例。

为分析免疫组库克隆程度在不同组织间的表达差异, 将肿瘤、癌旁和外周血组织根据克隆扩增数目分为低、中、高克隆型, 克隆扩增数目区间分别为1~100、101~1 000和1 001及以上。此外, 为分析不同年龄对免疫组库克隆分布的影响, 将样本按年龄进行分组, 其中64岁及以下为中年组 (mid), 65~79岁为老年组 (old), 80岁及以上为高龄老人组 (eldest)。

克隆分布和克隆分型分析通过R语言 (4.2.0版) 计算, 不同分组样本分布频率比较分析采用 t 检验, $P < 0.05$ 表示差异有统计学意义。使用ggsignif包 (0.6.4版) 对 P 值进行布尔值转换, $P \geq 0.05$ 表示差异不显著, $P < 0.05$ 表示差异显著, $P < 0.01$ 表示差异非常显著, $P < 0.001$ 表示差异极显著。使用ggplot2包 (3.4.2版) 进行数据可视化。

1.6 V-J基因共享性分析

V-J基因共享性即V、J基因各连接点的频率, 可用于探索V、J基因间的关联性对其在不同组织间表达差异的影响。本研究选取 $P_i = 0.001$ 为频率阈值对V-J基因重要共享配对进行筛选, 使用R语言 (4.2.0版) 进行数据处理, 并使用circlize包 (0.4.15版) 进行可视化。在计算V-J基因共享性时, 一般需考虑克隆扩增对其频率的影响。本研究为探究肿瘤低频克隆型对V-J基因共享性的影响, 采用去克隆扩增权重的V-J基因共享性分析方法。此方法在进行频率计算前使用克隆分型替换克隆扩增数目, 以去除克隆扩增权重的影响, 能更好地反映低频克隆型的变化特征。

1.7 CDR3分布特征分析

正常情况下, 机体免疫组库中T细胞表现为多克隆状态, 其CDR3呈均匀的高斯分布。疾病状态下, 由于抗原的持续刺激, 会导致某一类型的T细胞偏好扩增, 从而引起CDR3分布特征的变化。因此, 本研究使用VDJtools对免疫组库中TCR可变区基因V以及氨基酸克隆型丰度分布进行分析, 以探究克隆型偏态分布与NSCLC的关系。

1.8 克隆追踪

克隆追踪分析从输入的免疫组库中选择最丰富的克隆型, 并对所有免疫组库进行批量追踪, 可用于监测癌症免疫学中克隆型频率的变化。此外, 以提供的

核苷酸或氨基酸序列作为参数,可用于跟踪特定克隆型序列。使用immunarch包(0.67版)对数据进行克隆追踪分析,并使用ggplot2包(3.4.2版)进行数据可视化。

2 结果

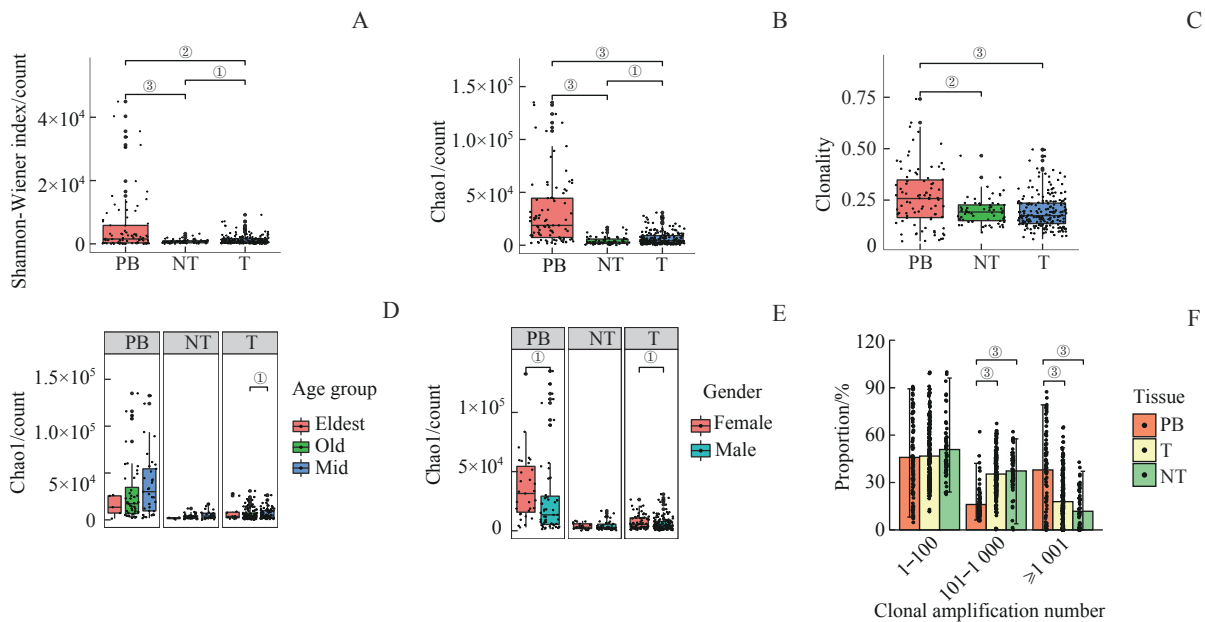
2.1 克隆分布与克隆分型

不同组间 Shannon-Wiener 指数均存在显著差异,其中外周血显著高于肿瘤和癌旁组织,肿瘤组织显著高于癌旁组织(图1A),提示 NSCLC 可能导致患者 TCR 组库多样性降低。丰富度上,外周血显著高于肿瘤和癌旁组织,肿瘤组织显著高于癌旁组织(图1B),结果与多样性分析契合,提示 NSCLC 同样会导致患者肿瘤及癌旁组织丰富度的降低。外周血的克隆性显著高于癌旁组织和肿瘤组织,而癌旁组织和肿

瘤组织无显著差异(图1C),提示肿瘤及癌旁组织高克隆扩增程度较低,NSCLC 可能通过外周血液形成免疫反应。

不同年龄分组及性别 TCR 组库的丰富度在 TRA 和 TRB 中观察到相似的结果。肿瘤组织中,中年组丰富度显著高于老年组(图1D),提示中年人群与老年人群相比,其 TCR 组库具备更强的肿瘤免疫潜力。此外,女性丰富度在外周血和肿瘤组织中显著高于男性(图1E),提示女性具有更强的肿瘤免疫潜力。

克隆分型分析结果显示:低克隆分型中,3组之间无显著差异;中克隆分型中,外周血显著低于癌旁和肿瘤组织;高克隆分型中,外周血显著高于癌旁和肿瘤组织(图1F)。该结果提示 NSCLC 可能引起患者外周血细胞高克隆,激活血液中相关分子的生物学功能,从而形成全局性免疫反应。



Note: A. Shannon-Weiner index of sample TCR β chain tumor, peripheral blood and paraneoplastic tissue. B. Abundance of TCR β chain among different tissues. C. Clonality of TCR β chain among different tissues. D. Difference in abundance of TCR β chain among different age groups in different tissues, respectively. E. Differences in the abundance of TCR β chains in different tissues by gender. F. Proportion of different tissues of TCR β chains in different clonal typing. T—tumor; PB—peripheral blood; NT—normal tissue (paraneoplastic tissue). ^① $P < 0.05$, ^② $P < 0.01$, ^③ $P < 0.001$.

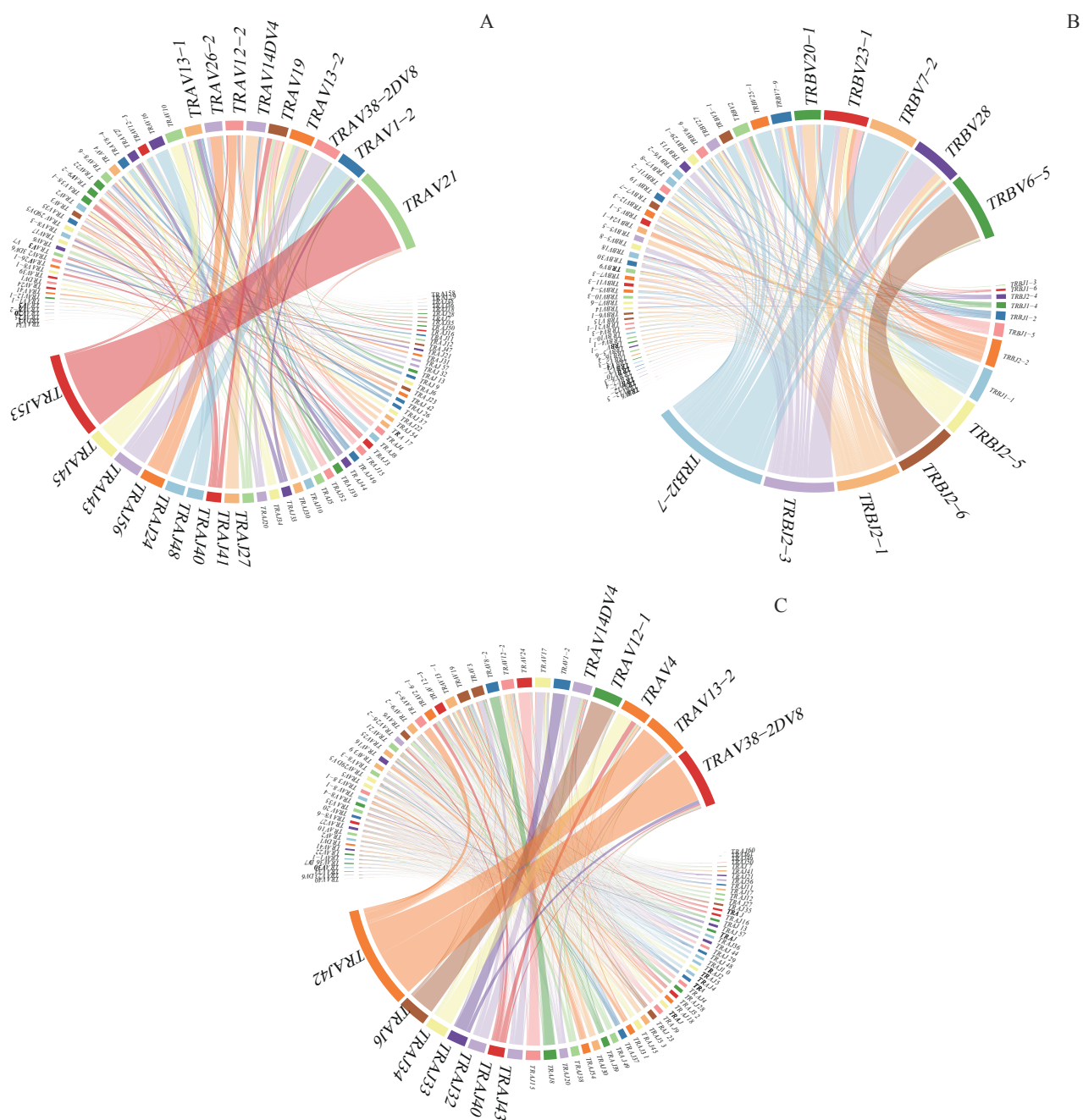
图1 NSCLC 克隆分布及克隆分型分析

Fig 1 NSCLC clonal distribution and clonal proportion analysis

2.2 V-J 基因共享性

为探究 V-J 基因在不同组织间表达的差异,筛选 4 例 NSCLC 患者测序数据分析其 TCR α 链和 β 链上 V、J 基因各连接点的频率,以探索 V、J 基因共享性的影响。结果显示患者 1 癌旁组织中存在 *TRAV21* 和 *TRAJ53* 基因高度共享(图 2A),外周血中存在

TRBV6-5 和 *TRBJ2-6* 高度共享(图 2B)。同时,在患者 2 外周血中发现 *TRAV38-2DV8*、*TRAV13-2* 基因与 *TRAJ42* 基因的高度共享(图 2C)。结果提示上述高共享性的 V-J 基因可能对 NSCLC 具有协同调控作用,且该调控可能存在多基因对单基因的协同作用。



Note: A. V-J gene sharing in paraneoplastic tissue with the degree of sharing of linkage thickness response of patient 1. B. V-J gene sharing in peripheral blood of patient 1. C. V-J gene sharing in peripheral blood of patient 2.

图2 *TRA* 和 *TRB* V-J 基因共享性

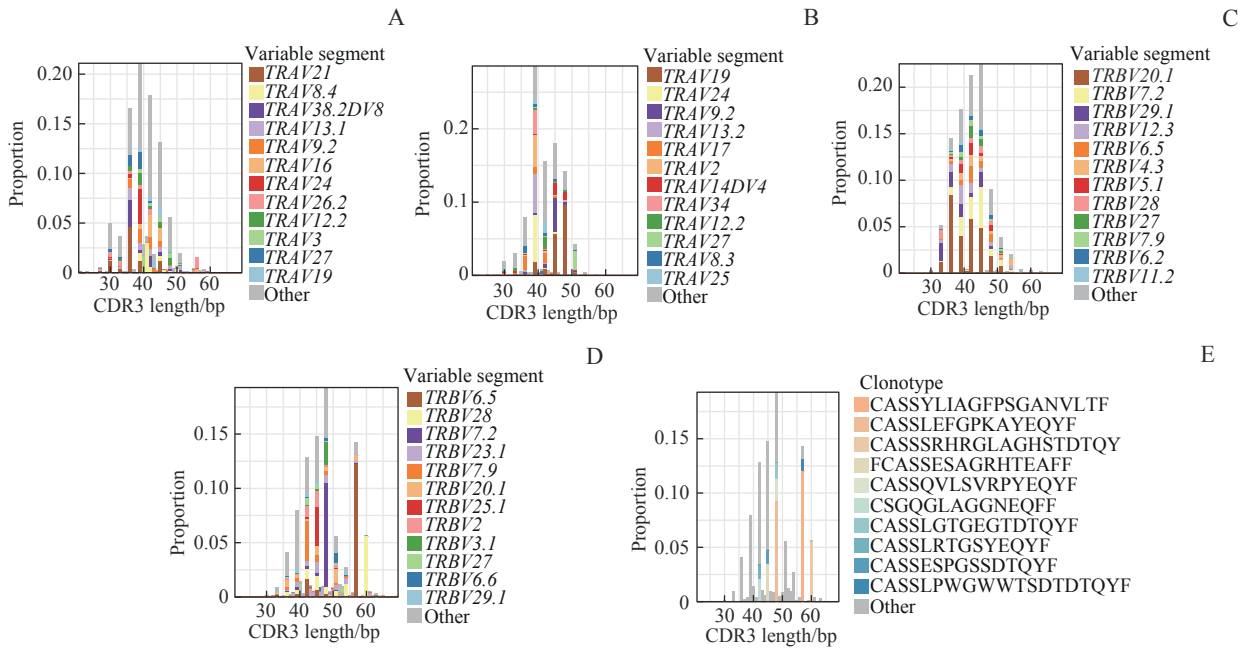
Fig 2 V-J gene sharing of *TRA* and *TRB*

2.3 CDR3 分布特征

在 TCR 双链中, 患者 1 的 *TRAV21* 和患者 2 的 *TRBV20.1* 在肿瘤中高克隆扩增出现非正态分布 (图 3A、C), 提示上述基因可能为 NSCLC 的潜在肿瘤标志物。此外, 患者 2 肿瘤组织中还出现了 *TRAV2*、*TRAV34*、*TRAV13.2* 及 *TRAV24* 等多个基因片段的非正态高频分布 (图 3B), 进一步验证了 NSCLC 免疫

反应可能存在的多基因协同调控作用。

患者 1 外周血出现 *TRBV6.5* 的非正态高克隆扩增 (图 3D), 提示 NSCLC 可能引起患者外周血细胞高克隆。进一步分析上述患者外周血高频克隆型 CDR3 分布, 结果发现 CASSYLIAGFPSGANVLTFF 的非正态高克隆扩增 (图 3E), 提示该高频克隆型可作为 NSCLC 中免疫反应相关的抗原结合参考 CDR3 序列。



Note: A. CDR3 distribution characteristics of tumor *TRA* V gene of patient 1. B. CDR3 distribution characteristics of tumor *TRA* V gene of patient 2. C. CDR3 distribution characteristics of tumor *TRB* V gene of patient 2. D. CDR3 distribution characteristics of peripheral blood *TRB* V gene of patient 1. E. Distribution characteristics of high-frequency clonotype CDR3 in peripheral blood of patient 1.

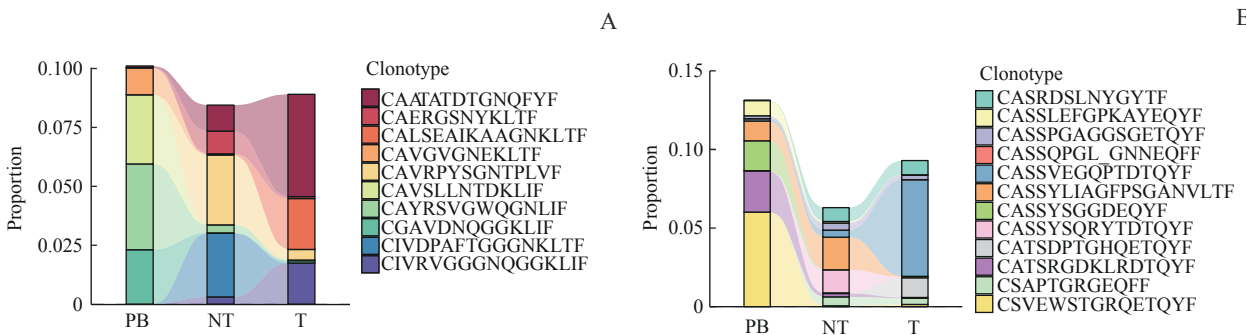
图3 NSCLC患者 *TRA* 和 *TRB* V 基因及高频克隆型 CDR3 分布特征

Fig 3 Distribution characteristics of *TRA* and *TRB* V gene and high frequency clonotype CDR3 in NSCLC patients

2.4 NSCLC 高频克隆型克隆追踪

通过 CDR3 分布特征分析, 发现了患者不同组织的高频扩增克隆型。为进一步探究这些高频克隆型在肿瘤发生过程中的动态变化, 分别选取了 TCR α 链和 TCR β 链中扩增程度最高的前 10 种与前 12 种克隆型进行克隆追踪分析。结果发现, CAATATDTGNQFYF、CIVRVGGNQQGKLIFF 和 CASSVEGQPTDTQYF 在癌旁组织表达量较低, 在肿瘤组织中表达量显著增加 (图 4), 提示上述克隆型可作为潜在的 NSCLC 肿瘤相关抗原识别参考序列。CAVSLNNTDKLIFF、

CGAVDNQQGKLIFF 和 CSVEWSTGRQETQYF 在外周血中高表达, 在癌旁和肿瘤组织中表达量显著降低; CAVRPYSGNTPLVFF、CASSYLIAGFPPSGANVLTF 和 CASSYSQRYTDTQYF 在癌旁组织高表达, 在肿瘤组织中的表达量显著降低 (图 4), 提示 NSCLC 可能抑制上述克隆型的正常表达, 进而影响相关生物学功能。CALSEAIKAAGNKLTF 和 CATSDPTGHQETQYF 在癌旁组织几乎不表达, 在肿瘤组织中表达量显著增加 (图 4), 提示上述克隆型可作为潜在的 NSCLC 肿瘤新抗原识别参考序列。



Note: A. Dynamics of the top ten clonotypes with the highest degree of amplification in the TCR α chain among different tissues. B. Dynamics of the top twelve clonotypes with the highest degree of amplification in the TCR β chain among different tissues.

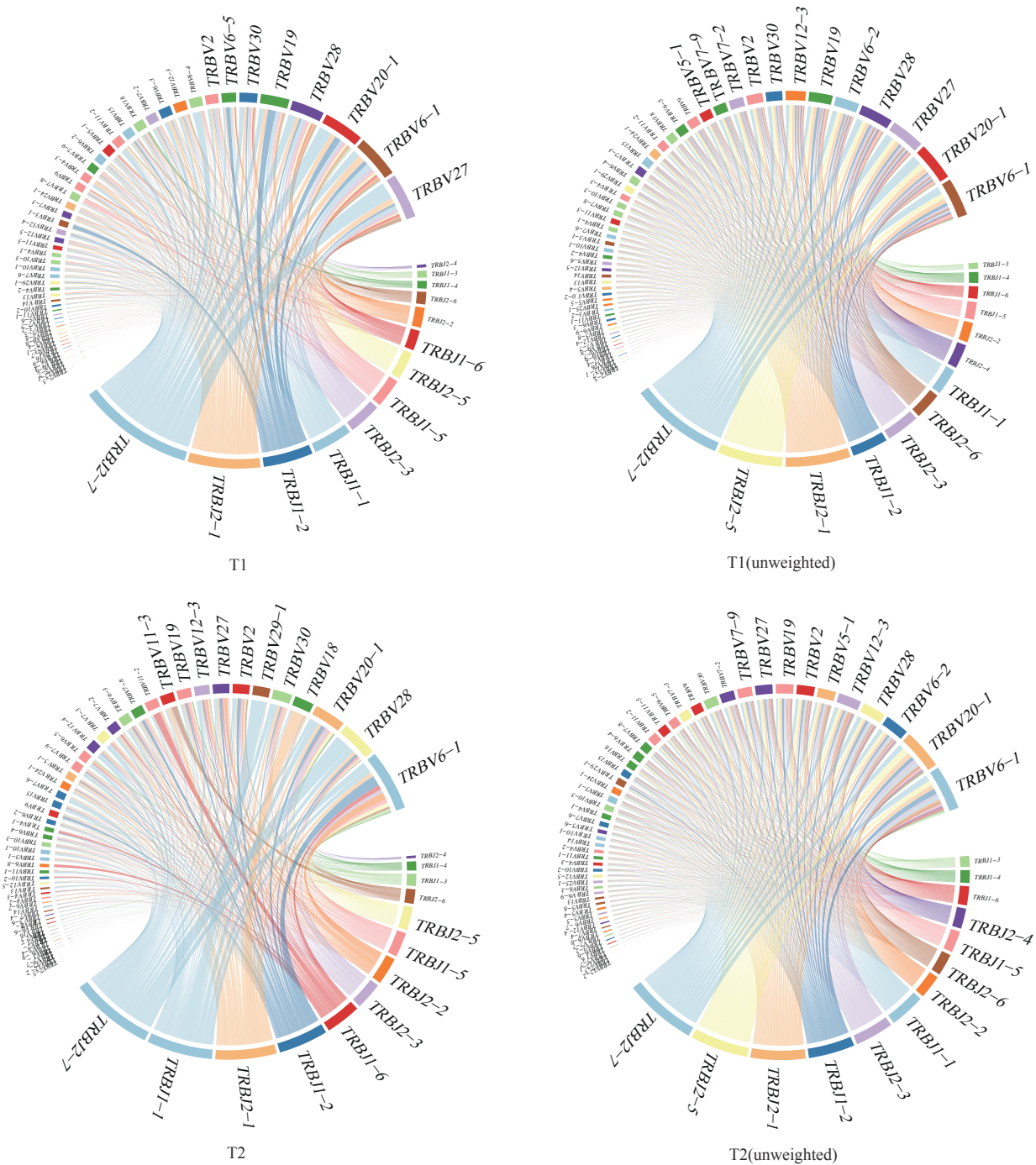
图4 NSCLC 患者高频克隆型克隆追踪

Fig 4 High-frequency clonotype clone tracking in NSCLC patients

2.5 去克隆扩增权重 V-J 基因共享性分析

克隆分布结果显示, 与外周血相比, 患者肿瘤与癌旁组织在多样性、丰富度以及克隆性上均呈显著低表达, 提示肿瘤中可能存在较多低频克隆突变。为进一步探索肿瘤相关低频克隆突变, 选取 2 例 PTC 样本

进行去除克隆扩增权重的 V-J 基因共享性分析。结果发现, 患者 1 与患者 2 原本表达较低的 *TRBJ2-5* 基因在去除克隆扩增权重后表达频率显著增加 (图 5), 提示该基因可能作为肿瘤免疫反应中潜在的低频关键基因。



Note: T1 and T2 are V-J gene sharing when clonal amplification weights were not removed for patient 1 and patient 2, and T1 (unweighted) and T2 (unweighted) are V-J gene sharing when clonal amplification weights were removed for patient 1 and patient 2.

图 5 PTC 肿瘤组织克隆扩增与克隆型基因共享性对比

Fig 5 Comparison of clonal amplification and clonotype gene sharing in PTC

3 讨论

免疫组库的研究可以为免疫学、生物学、医学和药物研发等领域提供很多有价值的信息。免疫组库的分析可以揭示生物体在不同免疫状态下抗体的多样性和特征,包括抗体的结构、亲和力、特异性、功能等。这些信息可以帮助我们更好地理解免疫系统的工作原理,以及在疾病和免疫治疗中的应用。为此,本研究建立了一套系统免疫组库分析流程,包括从原始数据到质控、过滤、TCR及BCR序列识别与提取等上游分析部分到免疫组库多样性分析、基因扩增偏好、V-J基因共享性、CDR3长度分布等下游分析及数据可视化流程。

基于建立的免疫组库分析流程,本研究分析了一套NSCLC数据集,其中包含72例患者的220个肿瘤区域、64个匹配的非肿瘤肺组织样本和70个外周血单核细胞样本。结果发现,NSCLC肿瘤和癌旁组织中,TCR组库多样性、丰富度和克隆性降低。克隆分型分析中,外周血高克隆分型显著高于肿瘤和癌旁组织,提示NSCLC可能引起患者外周血细胞高克隆,激活血液中相关分子的生物学功能从而形成全局性的免疫反应。

已有研究表明,性别与健康有重要关系^[22-24]。性别能影响患者的常用药反应,包括一些抗生素^[22];在较低血压下,女性心血管疾病风险较男性增大^[23];死于新型冠状病毒肺炎的男性更多,而女性似乎更易受到一系列持续症状的影响^[24]。因此,本研究分析了癌旁组织、外周血和肿瘤组织样本在不同性别分组间TCR组库丰富度的差异,结果显示女性外周血和肿瘤组织中的丰富度显著高于男性,提示女性具有更强的肿瘤免疫潜力。

为探究V-J基因在不同组织间表达的差异,本研究从NSCLC的TCR α 链和TCR β 链分别挑选了4例患者分析其V-J基因共享性。结果发现 *TRAV21* 和 *TRAJ53* 基因, *TRAV38-2DV8*、*TRAV13-2* 和 *TRAJ42* 基因高度共享,提示这些高共享性的V-J基因可能对NSCLC具有协同调控作用。在CDR3长度分布特征分析中,我们发现了 *TRAV21* 和 *TRBV6.5* 的非正态高克隆扩增,提示 *TRAV21* 和 *TRBV6.5* 的高克隆扩增可能是由NSCLC引起的,即上述基因可作为NSCLC的潜在肿瘤标志物。在高频克隆型CDR3特征分布分析中,我们提取了10个 *TRA* 高频克隆型和12个 *TRB* 高

频克隆型用于克隆追踪分析,并在克隆追踪分析中发现了潜在的NSCLC肿瘤相关抗原和新抗原结合CDR3参考序列。最后,为探索肿瘤相关低频基因,本研究选取2例PTC样本分析去除克隆扩增权重后克隆型间的V-J基因共享性,结果发现,患者1与患者2原本表达较低的 *TRBJ2-5* 基因在去除克隆扩增权重后表达频率显著增加,提示该基因可作为肿瘤免疫反应中潜在的低频关键调控基因。此外,已有研究^[25]发现乙型肝炎患者中同样存在 *TRBJ2-5* 基因的上调,进一步验证了本研究建立的TCR组库的可行性。

综上所述,本研究建立了一套系统的免疫组库分析流程,用于分析NSCLC的肿瘤相关特征,观察到NSCLC患者不同组织间克隆分布程度的差异和特定V-J基因的高共享性表达;随后,通过分析非正态克隆扩增的V基因和高频克隆型,发现了NSCLC潜在肿瘤标志物和抗原结合CDR3参考序列;最后,通过PTC患者去权VJ基因共享性分析发现了 *TRBJ2-5* 为潜在的低频关键调控基因。该免疫组库分析流程和肿瘤TCR组库特征分析方法还可以扩展到其他肿瘤和疾病。本研究组已建立流程预测病原体抗原蛋白序列中多肽疫苗候选表位^[26]。本研究流程可从适应性免疫系统特异性识别抗原肽的角度,通过预测高频CDR3序列为抗原表位识别工作提供免疫学参考,从而进一步评估T细胞表位被免疫细胞捕获并引发免疫应答的潜力。总体上,本流程可应用于如下场景:比较不同肿瘤或疾病间的TCR克隆型频率,发现潜在的广谱或特异的肿瘤标志物;分析结合特定抗原如肿瘤新抗原的CDR3结合序列特征,建立TCR-抗原识别预测模型;与其他组学数据如RNA-seq、单细胞测序数据联合,挖掘组学中的免疫组库特征,进行免疫学机制研究。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors disclose no relevant conflict of interests.

作者贡献/Authors' Contributions

谢鹭、赵卓明和鲁曼曼参与了研究设计与流程构建;谢鹭、赵卓明和刘振浩参与了论文写作和修改;张钰协助数据分析,赵卓明和许林锋参与了编程。所有作者均阅读并同意了最终稿件的提交。XIE Lu, ZHAO Zhuoming, and LU Manman participated in research design and process construction. XIE Lu, ZHAO Zhuoming, and LIU Zhenhao contributed to the writing and revision of the paper. ZHANG



Yu assisted in data analysis. ZHAO Zhuoming and XU Linfeng were involved in programming. All the authors have read the last version of paper and consented for submission.

• Received: 2023-07-20
• Accepted: 2023-11-14
• Published online: 2023-12-28

参·考·文·献

- [1] CHO J H. Immunotherapy for non-small-cell lung cancer: current status and future obstacles[J]. *Immune Netw*, 2017, 17(6): 378-391.
- [2] YANG T, XIONG Y, ZENG Y, et al. Current status of immunotherapy for non-small cell lung cancer[J]. *Front Pharmacol*, 2022, 13: 989461.
- [3] JACKSON L P, TJOA B A, MELLERT H, et al. Development of a TCR beta repertoire assay for profiling liquid biopsies from NSCLC donors[J]. *Cancer Drug Resist*, 2020, 3(3): 563-571.
- [4] JOSHI K, MILIGHETTI M, CHAIN B M. Application of T cell receptor (TCR) repertoire analysis for the advancement of cancer immunotherapy[J]. *Curr Opin Immunol*, 2022, 74: 1-8.
- [5] MCNEEL D G. TCR diversity: a universal cancer immunotherapy biomarker?[J]. *J Immunother Cancer*, 2016, 4: 69.
- [6] MEHTA N M, LI Y, PATEL V, et al. Prediction of peptide and TCR CDR3 loops in formation of class I MHC-peptide-TCR complexes using molecular models with solvation[J]. *Methods Mol Biol*, 2023, 2673: 273-287.
- [7] KIDMAN J, PRINCIPE N, WATSON M, et al. Characteristics of TCR repertoire associated with successful immune checkpoint therapy responses[J]. *Front Immunol*, 2020, 11: 587014.
- [8] 赵静静, 简星星, 王广志, 等. TCR组库及其在肿瘤免疫中的应用前景与挑战[J]. *生命科学*, 2021, 33(4): 466-471.
- ZHAO J J, JIAN X X, WANG G Z, et al. TCR library and its application prospect and challenges in tumor immunity[J]. *Chinese Bulletin of Life Sciences*, 2021, 33(4): 466-471.
- [9] BODOR J N, BOUMBER Y, BORGHAEI H. Biomarkers for immune checkpoint inhibition in non-small cell lung cancer (NSCLC)[J]. *Cancer*, 2020, 126(2): 260-270.
- [10] ROBERT L, TSOI J, WANG X, et al. CTLA4 blockade broadens the peripheral T-cell receptor repertoire[J]. *Clin Cancer Res*, 2014, 20(9): 2424-2432.
- [11] ZHANG Y, LIU Z, WEI W, et al. TCR engineered T cells for solid tumor immunotherapy[J]. *Exp Hematol Oncol*, 2022, 11(1): 38.
- [12] GARCÍA-PARDO M, GORRIA T, MALENICA I, et al. Vaccine therapy in non-small cell lung cancer[J]. *Vaccines*, 2022, 10(5): 740.
- [13] ZHU W, PENG Y B, WANG L, et al. Identification of α -fetoprotein-specific T-cell receptors for hepatocellular carcinoma immunotherapy [J]. *Hepatology*, 2018, 68(2): 574-589.
- [14] ZHANG Y, JIAN X, XU L, et al. iTcep: a deep learning framework for identification of T cell epitopes by harnessing fusion features[J]. *Front Genet*, 2023, 14: 1141535.
- [15] BOLOTIN D A, POSLAVSKY S, MITROPHANOV I, et al. MiXCR: software for comprehensive adaptive immunity profiling[J]. *Nat Methods*, 2015, 12(5): 380-381.
- [16] ZHANG W, DU Y P, SU Z, et al. IMonitor: a robust pipeline for TCR and BCR repertoire analysis[J]. *Genetics*, 2015, 201(2): 459-472.
- [17] WU Y, BISWAS D, USAITE I, et al. A local human V δ 1 T cell population is associated with survival in nonsmall-cell lung cancer[J]. *Nat Cancer*, 2022, 3(6): 696-709.
- [18] WANG Y, LIU Y, CHEN L, et al. T cell receptor beta-chain profiling of tumor tissue, peripheral blood and regional lymph nodes from patients with papillary thyroid carcinoma[J]. *Front Immunol*, 2021, 12: 595355.
- [19] SHANNON C E. The mathematical theory of communication. 1950 [J]. *MD Comput*, 1997, 14(4): 306-317.
- [20] CHAO A. Nonparametric estimation of the number of classes in a population[J]. *Scand J Statist*, 1984, 11: 265-270.
- [21] PIELOU E C. The measurement of diversity in different types of biological collections[J]. *J Theor Biol*, 1966, 13: 131-144.
- [22] PALA L, CONFORTI F. The effect of patient sex on the efficacy and safety of anticancer immunotherapy[J]. *Expert Opin Drug Saf*, 2021, 20(12): 1535-1544.
- [23] ZHANG Y, LIU B, ZHAO R, et al. The influence of sex on cardiac physiology and cardiovascular diseases[J]. *J Cardiovasc Transl Res*, 2020, 13(1): 3-13.
- [24] TADIRI C P, GISINGER T, KAUTZY-WILLER A, et al. The influence of sex and gender domains on COVID-19 cases and mortality[J]. *CMAJ*, 2020, 192(36): E1041-E1045.
- [25] YAN D, YANG J Z, JI Z K, et al. Profiling T cell receptor β -chain in responders after immunization with recombinant hepatitis B vaccine[J]. *J Gene Med*, 2021, 23(9): e3367.
- [26] 赵静静, 澹小秀, 王广志, 等. 生物信息学方法预测病原体抗原蛋白序列中多肽疫苗候选表位[J]. *生命科学研究*, 2021, 25(4): 363-371.
- ZHAO J J, ZHANG X X, WANG G Z, et al. Bioinformatics method for prediction of peptide vaccine candidate epitopes in pathogen antigenic protein sequences[J]. *Life Science Research*, 2021, 25(4): 363-371.

[本文编辑] 吴洋