

# 单细胞V(D)J测序技术深度解析：研究背景、前沿计算方法与公共数据集资源

## I. 引言：高分辨率免疫组库分析的必要性

### A. 适应性免疫系统的特异性引擎：V(D)J重组

**V(D)J基因的随机重组是适应性免疫系统产生海量多样性受体的分子基础，为识别几乎所有抗原提供了可能。** 适应性免疫系统是脊椎动物抵御病原体侵袭和清除体内异常细胞（如肿瘤细胞）的关键防线。其核心特征在于高度的特异性和记忆性，这主要由T细胞和B细胞介导。这些细胞通过其细胞表面的T细胞受体（TCR）和B细胞受体（BCR，即膜结合形式的抗体）识别抗原。V(D)J基因重组是适应性免疫系统产生惊人多样性受体的分子基础。在T细胞和B细胞发育过程中，编码TCR和BCR可变区的V（Variable）、D（Diversity，仅TCRβ链和BCR重链）和J（Joining）基因片段发生随机组合，同时在连接处引入非模板核苷酸（N/P核苷酸），从而产生理论上数量可达天文数字的独特受体序列。正是这种巨大的多样性，赋予了免疫系统识别几乎无限抗原的能力，使其能够应对不断出现的新型病原体和恶性肿瘤。每个T细胞或B细胞通常表达单一的、独特的TCR或BCR，这个独特的受体序列定义了该细胞的克隆型。

### B. 传统批量组库测序的局限性

**传统的批量测序方法无法解析受体的天然配对链，也无法将序列与单个细胞的功能关联起来，这极大地限制了对免疫应答的深入理解。** 在单细胞技术出现之前，免疫组库的研究主要依赖于批量测序方法。这些方法通过对大量T细胞或B细胞群体的TCR或BCR基因进行扩增和测序，能够提供关于免疫组库整体组成的信息，例如V(D)J基因片段的使用频率、CDR3（互补决定区3，抗原识别的关键区域）长度分布以及克隆型丰度等。然而，批量测序存在固有的局限性。首先，它无法解析单个细胞内TCRα链与β链（或BCR重链与轻链）的天然配对信息。由于TCR的功能依赖于α和β链的正确组合，BCR的功能依赖于重链和轻链的正确组合，丢失这种配对信息会严重影响对抗原识别特异性的准确理解。其次，批量测序将来自不同细胞的受体序列混合在一起，因此无法将特定的TCR或BCR序列与其来源细胞的表型或功能状态（如细胞活化状态、分化亚群、细胞因子表达谱等）直接关联起来。这种信息的丢失使得深入理解特定克隆在免疫应答中的具体作用变得非常困难。

### C. 单细胞V(D)J测序的兴起及其意义

**单细胞V(D)J测序通过在单细胞水平同时获取配对的受体序列和转录组信息，革命性地克服了传统方法的局限性，实现了对免疫应答前所未有的高分辨率解析。** 单细胞V(D)J测序技术的出现，为克服传统批量测序的局限性带来了革命性的突破。该技术能够在单细胞水平上同时获取成对的TCR α链和β链（或BCR重链和轻链）的完整编码序列，并且可以结合转录组测序（scRNA-seq）或细胞表面蛋白测序（如CITE-seq），从而将免疫受体序列信息与同一细胞的基因表达谱、表面蛋白表达谱等多种分子特征相关联。这种多模态、高分辨率的分析能力，使得研究者能够以前所未有的深度剖析免疫应答的复杂性。例如，可以精确识别出在特定疾病状态下（如癌症或感染）发生克隆性增殖的T细胞或B细胞，并同时了解这些细胞的功能状态（如效应细胞、记忆细胞、耗竭细胞等）和表型特征。正如一些研究强调的，获取“来自B细胞或T细胞的全长、成对的V(D)J序列”并结合其他数据模式是这项技术的关键优势。

从群体平均到单细胞精度的范式转变为研究稀有克隆和受体功能带来了可能，其中V(D)J序列与转录组的整合是其核心价值所在。 这种从群体平均到个体精度的转变，代表了免疫组库研究范式的根本性变革。它不再仅仅是对免疫组库进行统计学推断，而是实现了对单个免疫细胞及其抗原受体的直接、高分辨率测量。这种粒度的提升，使得研究稀有克隆、与特定受体相关的细微表型差异以及受体链的真实配对成为可能，而这些在过去是极具挑战性的。因此，单细胞V(D)J测序为我们更深入地理解免疫应答的机制开辟了新的道路。值得强调的是，将V(D)J序列与转录组数据在单细胞水平上整合的能力，不仅仅是一个附加功能，而是这项技术核心价值的驱动力。一个TCR序列本身可能揭示其潜在的抗原特异性，但结合该T细胞的转录组信息，我们才能判断它是一个效应T细胞、记忆T细胞、耗竭T细胞还是调节性T细胞。这种组合信息对于理解诸如肿瘤免疫治疗中某些T细胞克隆为何扩增而另一些则不然，或者在感染过程中T细胞状态如何演变等关键科学问题至关重要。

### D. 单细胞免疫组库研究的核心目标

**基于单细胞V(D)J测序技术的强大能力，研究的核心目标通常围绕克隆型鉴定、克隆性与多样性评估、V(D)J基因使用分析以及序列与功能关联这五个方面展开。** 这些目标共同构成了利用单细胞V(D)J测序技术深入探索免疫系统复杂性的基础框架。

- **克隆型鉴定（Clonotype Identification）**：精确识别每个单细胞的TCR或BCR序列，定义其克隆型。
- **克隆性评估（Clonality Assessment）**：分析免疫组库中不同克隆型的相对丰度和分布，评估克隆性增殖的程度。
- **多样性分析（Diversity Analysis）**：衡量免疫组库中独特克隆型的数量和分布均匀度，评估其广度。
- **V(D)J基因使用模式分析（V(D)J Gene Usage Patterns）**：研究不同V、D、J基因片段的使用频率及其组合偏好。
- **序列与细胞表型/功能关联（Linking Sequence to Phenotype/Function）**：将克隆型信息与细胞的转录组、表面蛋白表达或其他功能特性相关联，以揭示特定克隆的功能意义。

## II. “为何研究”：揭示免疫复杂性的动机与核心科学问题

单细胞V(D)J测序技术的根本动机在于解码适应性免疫应答的复杂机制，并将这些理解应用于人类健康与疾病的各个方面。

### A. 基础目标：理解抗原特异性与免疫记忆

**单细胞V(D)J测序的核心动机在于解码抗原特异性识别和免疫记忆的形成机制，通过追踪克隆动态变化来理解免疫系统的基本运作。** 适应性免疫的核心在于其对抗原的特异性识别和产生的持久免疫记忆。T细胞通过TCR识别由主要组织相容性复合体（MHC）呈递的抗原肽，而B细胞则通过BCR直接识别天然抗原。单细胞V(D)J测序能够帮助我们：

- **追踪抗原特异性克隆**：在抗原暴露（如感染或疫苗接种）后，识别并追踪那些特异性识别该抗原的T细胞和B细胞克隆的动态变化，包括它们的扩增、收缩以及分化为效应细胞或记忆细胞的过程。
- **研究免疫记忆的形成与维持**：通过分析记忆T细胞和B细胞的克隆型组成及其表型特征，揭示免疫记忆是如何建立、维持并在再次遇到相同抗原时被激活的。
- **探究TCR/BCR交叉反应性**：某些TCR或BCR可能识别多种结构相似但不同的抗原，这种交叉反应性在免疫保护（如针对病毒变异株）和自身免疫（如错误识别自身抗原）中都扮演着重要角色。单细胞分析有助于在克隆水平上研究这种现象。

**将受体序列（“词汇”）与细胞功能（“说话者特征”）相结合，使得我们能够全面解读特定免疫克隆在生物学情境下的确切作用。** 从更深层次看，适应性免疫的语言是由无数独特的TCR和BCR（如同“词汇”）及其所代表的抗原识别特异性（“语义”）构成的。单细胞V(D)J测序技术，尤其是当它与细胞表型和功能信息（“说话者”的特征）相结合时，使我们能够更全面地解读这些“词汇”在特定免疫情境下的确切含义和作用。例如，仅仅识别出一个肿瘤反应性TCR序列是第一步；而通过结合转录组学数据，了解表达该TCR的T细胞是具有杀伤功能的效应细胞还是功能耗竭的细胞，则是理解其在抗肿瘤免疫中真实角色的关键第二步。这种整合分析对于阐明疾病机制或评估治疗效果至关重要。

### B. 在疾病背景下的应用

**单细胞V(D)J测序在肿瘤、感染及自身免疫病等多种疾病的研究中均展现出巨大潜力。**

#### 1. 肿瘤学

**在肿瘤学中，该技术被用于识别和表征抗肿瘤的关键免疫细胞，追踪免疫治疗反应，并为开发新型细胞疗法提供候选受体。**

- **识别肿瘤浸润淋巴细胞（TILs）**：TILs是抗肿瘤免疫应答的关键效应细胞。单细胞V(D)J测序可以鉴定TILs中的TCR克隆型，揭示肿瘤微环境中T细胞的克隆性扩增情况，并结合转录组数据分析其功能状态（如效应、耗竭、记忆）。
- **追踪免疫治疗反应**：在免疫检查点抑制剂等免疫治疗过程中，追踪患者体内T细胞和B细胞克隆的动态变化，有助于理解治疗的应答机制，区分应答者与非应答者，并可能发现新的治疗靶点或生物标志物。
- **开发TCR工程化T细胞（TCR-T）疗法**：通过鉴定对肿瘤抗原具有高亲和力和高特异性的TCR，可以将其用于改造患者或健康供体的T细胞，开发过继性细胞治疗。

#### 2. 感染性疾病

**在感染性疾病研究中，单细胞分析帮助我们鉴定能够有效清除病原体的T细胞和产生中和抗体的B细胞，并评估疫苗效果。**

- **表征病原体特异性免疫组库**：在病毒（如HIV、流感病毒、SARS-CoV-2）或细菌感染过程中，分析TCR和BCR组库的变化，鉴定针对特定病原体抗原的克隆型。
- **理解保护性免疫应答**：研究康复患者或疫苗接种者体内产生的、与保护性免疫相关的中和抗体（BCR序列）和效应T细胞（TCR序列）的特征。
- **监测疫苗应答**：评估疫苗诱导的免疫应答的广度、强度和持久性，识别与保护效果相关的关键克隆型。

#### 3. 自身免疫性疾病

**对于自身免疫病，该技术是识别和理解那些错误攻击自身组织的“罪魁祸首”——自身反应性T/B细胞的强大工具。**

- **鉴定自身反应性T细胞和B细胞克隆**：在类风湿性关节炎、系统性红斑狼疮、1型糖尿病等自身免疫病中，识别那些错误攻击自身组织的T细胞和B细胞克隆。
- **理解免疫耐受破坏机制**：通过分析自身反应性克隆的特征及其调控环境，探究免疫耐受是如何被打破并导致自身免疫的。
- **追踪治疗效果**：评估旨在清除或调节自身反应性克隆的治疗方法的有效性。

### C. 疫苗开发与效力评估

**在疫苗开发领域，单细胞V(D)J测序能够精确评估疫苗诱导免疫应答的质量，从而指导和优化下一代疫苗的设计。**

- **评估免疫应答的广度和深度**：分析疫苗诱导的TCR和BCR组库的多样性、克隆性扩增程度以及V(D)J基因使用特征。
- **识别与保护性免疫相关的克隆型**：通过比较疫苗接种后受保护个体与未受保护个体的免疫组库差异，寻找与保护性免疫相关的特定TCR或BCR克隆型。
- **指导下一代疫苗设计**：基于对有效免疫应答特征的理解，设计能够更有效地诱导这些期望免疫特征的新型疫苗。

### D. 核心研究问题举例

**利用单细胞V(D)J测序技术，研究者们致力于回答一系列关键科学问题，例如特定抗原受体的特征、免疫组库在疾病中的动态变化、以及公共克隆的存在及其意义等。**

- **识别特定抗原（如肿瘤新抗原、病毒肽）的TCR/BCR具有哪些序列特征和结构特征？**
- **在疾病进展过程中，或在对治疗产生应答时，免疫组库的多样性和克隆性如何变化？**
- **是否存在与特定疾病或免疫应答相关的“公共”TCR/BCR（即在多个个体间共享的克隆型）？**

这些公共克隆的发现对于开发普适性的诊断工具或“货架式”细胞治疗产品具有重要意义。如果多个个体在应对同一病原体或癌症时，都利用了相似或相同的TCR/BCR，那么这些受体很可能靶向关键的、保守的表位。这使

得它们成为开发TCR-T疗法、设计旨在诱导这些特异性应答的疫苗，或开发用于评估免疫能力的诊断工具的理想候选。公共数据库在此类研究中发挥着至关重要的作用。

- 健康个体与患者之间，或不同疾病状态之间，V(D)J基因的使用模式有何差异？

对V(D)J基因使用模式的研究，虽然看似描述性的，但可以揭示由遗传易感性、慢性抗原暴露或疾病过程施加的选择压力或偏好，为理解免疫组库形成和选择的基本规则提供线索。人类拥有一套有限的V、D、J基因。在特定条件下（例如，特定的HLA类型、慢性感染）的优先使用或配对，表明存在选择性压力或重组机制/克隆选择中的固有偏见。这有助于我们理解对疾病的易感性或对疫苗的反应。

- 表达特定目标TCR/BCR的细胞具有哪些表型和功能特征？

### III. 单细胞V(D)J组库分析的前沿计算方法学

从原始测序数据中提取生物学洞见，依赖于一系列复杂的生物信息学分析流程和先进的计算方法。

#### A. 生物信息学分析流程概述

典型的单细胞V(D)J分析流程包括数据处理、序列比对、克隆型定义、指标计算、可视化和多模态整合六大步骤。

- 原始测序数据处理**：包括序列解复用、质量控制。
- V(D)J基因片段比对与序列拼接**：将测序读段比对到参考的V、D、J基因片段，并组装成完整的TCR/BCR可变区序列。
- 克隆型定义与分组**：基于序列特征（主要是CDR3序列）将具有相同或相似受体的细胞聚类为克隆型。
- 免疫组库指标计算**：量化免疫组库的各种特征，如克隆性、多样性、V(D)J基因使用频率等。
- 可视化**：通过图表形式直观展示免疫组库的结构和动态。
- 与其他单细胞模态数据整合**：将V(D)J信息与转录组、表面蛋白组等数据关联分析。

#### B. 原始数据预处理和质量控制（QC）

数据预处理和质控是保证后续分析准确性的基石，其核心任务是校正技术偏差并将测序读段准确分配到单个细胞。此步骤对于确保后续分析的准确性至关重要。

- 细胞条形码（Cell Barcode）解复用**：将混合测序的读段根据其细胞条形码分配到来源的单个细胞。
- 唯一分子标识符（UMI）处理**：利用UMI校正PCR扩增偏好，从而更准确地定量每个TCR/BCR转录本的原始数量。
- 低质量读段和细胞过滤**：去除测序质量较低的读段，以及那些细胞条形码模糊或UMI数量过低的“细胞”，以减少噪音。
- 平台特异性初始处理**：例如，广泛应用的10x Genomics平台提供了Cell Ranger软件包，其中的 `cellranger vdj` 命令可以执行初始的读段处理、细胞条形码分配、UMI计数以及V(D)J序列的拼接和初步注释。

#### C. V(D)J基因片段注释与序列拼接

准确鉴定V、D、J基因片段和高变区（尤其是CDR3）序列是核心步骤。

- 与生殖系V(D)J基因片段比对：

V(D)J基因片段的注释主要依赖于将测序读段与已知的生殖系基因参考数据库进行比对。经典的工具如IMGT/HighV-QUEST、MIXCR 和 IgBLAST 是鉴定V、D、J基因片段和提取CDR3序列的基础。它们依赖于与已知生殖系基因数据库的比对。一些较新的工具致力于提高比对的准确性和速度，例如SONG（Systematic Optimization of Next-generation sequencing for Gapped-alignment）据称在V(D)J基因比对方面具有高精度和高效率。

- 从头组装全长受体序列：

对于非靶向的scRNA-seq数据，可以利用TRUST4等工具从头组装TCR/BCR序列，从而在没有V(D)J富集的情况下也能获取免疫组库信息。对于非靶向捕获的单细胞转录组数据（scRNA-seq），一些工具可以从总RNA测序读段中从头组装出TCR/BCR转录本序列。例如，TRUST4可用于从批量RNA-seq和单细胞RNA-seq数据中进行从头组装。TraCeR 和 Platypus 专门设计用于从scRNA-seq数据中重建TCR序列，其中Platypus特别关注了TCR mRNA捕获效率较低的情况。10x Genomics的Cell Ranger也执行序列拼接任务。

#### D. 克隆型定义与分组

克隆型是指来源于同一个淋巴细胞祖先、表达相同TCR或BCR的细胞群体。

- 定义克隆型：

克隆型的定义通常基于共享的CDR3序列，但目前缺乏统一标准是该领域面临的一大挑战，这阻碍了不同研究间的直接比较。最常用的定义是基于共享的CDR3序列。对于成对的TCR/BCR数据，通常要求α链和β链（或重链和轻链）的CDR3序列均相同，有时还会加入V(J)基因使用一致性的要求。

AIRR（Adaptive Immune Receptor Repertoire）社区正在积极推动免疫组库数据标准的制定，其中包括克隆型定义的规范化。这一努力对于促进数据共享和荟萃分析至关重要。如果缺乏统一标准，例如研究A基于CDR3氨基酸序列定义克隆型，而研究B基于CDR3核苷酸序列和相同的V基因使用来定义，那么即使分析的是同一份原始数据，两者报告的克隆型数量和克隆共享水平也可能大相径庭。这种不一致性阻碍了构建大规模、可整合的克隆型及其相关表型/疾病状态数据库的努力。标准化是确保可重复性和促进合作发现的关键。

- 克隆型鉴定工具：

多种计算工具如scirpy和Dandelion等，提供了克隆型鉴定功能，并能与主流的单细胞分析框架集成。Cell Ranger 在其分析流程中提供克隆型鉴定功能。scirpy 提供了灵活的克隆型定义选项（如相同的CDR3核苷酸/氨基酸序列，或基于序列相似性的网络聚类），并能与scRNA-seq分析流程紧密集成。Dandelion 同样执行克隆型鉴定，并设计为与流行的scRNA-seq分析包 scanpy 整合。VDJtools 可以处理多种上游比对工具的输出，进行克隆型鉴定和下游的免疫组库统计分析。Immcantation框架（包括 Change-o, aLakazam 等工具）功能强大，尤其在BCR分析领域（如体细胞高频突变分析），其许多概念和方法也可应用于TCR分析。

#### E. 免疫组库指标量化

为了从宏观上描述免疫组库，需要计算克隆性、多样性、V(D)J基因使用频率等一系列量化指标。对克隆型进行鉴定和分组后，需要计算一系列指标来描述免疫组库的特征。

- 克隆性（Clonality）**：衡量克隆型频率分布的偏斜程度。例如，可以使用基尼系数（Gini coefficient）或香农熵（Shannon entropy）的标准化形式来评估。高克隆性通常表示免疫组库由少数几个高度扩增的克隆所主导。
- 多样性（Diversity）**：衡量免疫组库中独特克隆型的丰富程度（数量）和均匀程度（频率分布）。常用的多样性指数包括Chao1估计子（估计丰富度）、香农熵、辛普森指数等。
- V(D)J基因使用频率（V(D)J Gene Usage）**：统计不同V、D、J基因片段及其组合在免疫组库中出现的频率。
- CDR3长度分布（CDR3 Length Distribution）**：分析高变CDR3区的长度分布特征，这在不同免疫状态下可能发生改变。
- 趋同进化（Convergent Evolution）**：识别在不同细胞或不同个体中独立产生但具有相同或高度相似TCR/BCR序列的现象，这可能暗示了对共同抗原的趋同选择。

诸如scirpy、Dandelion、VDJtools 和 alakazam (Immcantation套件的一部分) 等工具都提供了计算这些免疫组库指标的功能。

#### F. 免疫组库可视化技术

有效的可视化技术，如克隆频率图和网络图，是直观理解复杂免疫组库数据结构和动态的关键。

- 克隆型频率图**：如条形图、饼图，展示丰度最高的克隆型及其占比。
- V(D)J基因使用图**：如热图、弦图（chord diagrams），展示不同V(D)J基因片段的使用频率和组合偏好。
- 多样性稀疏曲线（Rarefaction Curves）**：评估测序深度对观测到的克隆型多样性的影响。
- 网络图**：基于序列相似性（如CDR3序列的编辑距离）构建克隆型网络，可视化克隆型之间的亲缘关系和聚集模式，例如scirpy 和 Dandelion 均支持此类分析。
- 交互式浏览器**：例如10x Genomics提供的Loupe V(D)J Browser，允许用户交互式地探索V(D)J数据，并将其与基因表达数据（如果同时测定）关联起来进行可视化。

#### G. 与其他单细胞模态数据（如scRNA-seq, CITE-seq）的整合

单细胞V(D)J分析的核心优势在于能够将受体序列信息与转录组等其他单细胞模态数据整合，从而将克隆型与细胞的生物学状态直接联系起来。这是当前单细胞免疫组库分析的核心优势和发展方向，旨在将受体序列信息与细胞的生物学状态联系起来。

随着多组学技术的普及，分析工具已从独立的序列分析器演变为能够与Scanpy等框架集成的综合解决方案，这种整合能力是进行有意义生物学解释的基础。计算工具的发展也清晰地反映了测序技术的进步。随着像10x Genomics这样的平台使得单细胞V(D)J测序结合转录组测序变得更加普及，分析工具也从最初主要关注V(D)J序列本身的独立分析器，演变为能够与转录组等其他组学数据进行整合分析的解决方案（例如，scirpy 与 scanpy 的整合）。这种整合能力如今已成为进行有意义的生物学解释的事实标准。早期的V(D)J分析工具主要集中于比对、CDR3提取和基本的组库统计等任务。单细胞多组学技术的出现，创造了将V(D)J信息与其他细胞分子数据（如基因表达谱）联系起来的迫切需求。这推动了新一代工具的开发，这些工具能够处理例如 scanpy 的 AnnData 对象，并将V(D)J衍生的元数据添加到其中，或者反之。没有这种整合，研究者将不得不处理孤立的数据集，从而失去了单细胞多组学技术的主要优势——即在同一细胞内关联不同分子层面的信息。

- 将克隆型与基因表达关联：

scirpy 和 Dandelion 等工具被设计为可以与流行的scRNA-seq分析框架如 scanpy 紧密集成。另一个主流的scRNA-seq分析框架 Seurat 也可以通过自定义脚本实现类似整合。这种整合允许将克隆型信息（如克隆扩增程度、特定TCR序列）叠加到基于基因表达数据生成的UMAP或t-SNE等降维图上，从而直观地观察特定克隆型富集在哪些细胞亚群中。

- 细胞类型注释：

利用基因表达数据，可以使用 CellTypist 或 SingleR 等工具对单细胞进行类型注释（如CD4+初始T细胞、CD8+效应记忆T细胞等）。然后，可以在特定的细胞亚群内部分析其免疫组库特征。

3. 整合细胞表面蛋白数据（CITE-seq）：

CITE-seq技术通过寡核苷酸标记的抗体同时测量单细胞的RNA和表面蛋白表达。将V(D)J序列、转录组和表面蛋白组数据结合，可以实现对细胞状态更精细、更准确的定义。

**挑战：**在整合不同来源或不同批次的单细胞多组学数据时，需要妥善处理批次效应。进行有效的数据归一化，并确保不同组学层面信息能够准确对齐。

随着数据规模和复杂性的提升，对可扩展、稳健且用户友好的计算工作流程的需求日益突出。那些文档完善、易于安装、并且能够高效处理大规模数据的工具，将更容易获得广泛应用。这同时也关系到对更优可视化工具的需求，以便用户能够有效地探索和理解复杂的多组学分析结果。

**像immuneSIM这样的模拟工具对于基准测试新分析方法的准确性和性能具有不可估量的价值，为方法学验证提供了受控的环境。**

考虑到真实生物样本固有的复杂性以及往往未知的“真实情况”（ground truth），模拟数据为方法学验证提供了一个受控的环境。我们如何判断一个新的克隆型鉴定算法是否比旧的更准确？或者一个多样性指标是否对测序错误具有鲁棒性？在生物样本中，我们往往无法确知其真实的组库构成。模拟工具允许我们生成具有已知特性（如克隆数量、CDR3序列、错误率等）的组库，然后测试不同分析工具恢复这些已知特征的能力如何。这对于提高计算方法学的可靠性至关重要。

表1: 单细胞V(D)J分析计算工具比较概览

注：此表为示例性概览，工具的功能和特性可能随版本更新而变化。用户应查阅各工具的最新文档以获取详细信息。

工具名称 (Tool Name)	主要功能 (Primary Function(s))	关键算法特点/方法 (Key Algorithmic Features/Approach)	支持的受体类型 (Supported Receptor Types)	输入数据格式 (Input Data Format)	输出数据格式 (Output Data Format)	优势 (Strengths)	局限性/挑战 (Limitations/Challenges)	与其他工具/框架集成 (Integration)	发表文献/链接 (Publication/Link)
Cell Ranger (10x)	原始数据处理, 比对, 序列拼接, 初步克隆型鉴定, 与GEX整合	基于10x平台特有流程, 内置参考序列比对和组装算法	TCR, BCR	FASTQ (10x V(D)J + GEX)	filtered_contig_annotations.csv, clonotypes.csv, json, Loupe files	平台一体化解决方案, 用户友好, 与自家可视化工具紧密集成	主要针对10x数据, 克隆型定义相对固定, 高级定制能力有限	Loupe VDJ Browser, Seurat, Scanpy (通过读取输出文件)	<a href="#">10x Genomics</a>
TRUST4	从RNA-seq数据中从头组装 TCR/BCR序列	基于de Bruijn图的算法, 支持部分和全长 CDR3组装	TCR, BCR	FASTQ, BAM (scRNA-seq, bulk RNA-seq)	FASTA, TSV	无需特异性 V(D)J富集, 适用于多种 RNA-seq数据, 速度较快	准确性可能受RNA-seq覆盖深度影响, 对极低表达克隆可能仍有挑战	可作为上游工具为其他分析软件提供输入	<a href="#">Li et al., Nat Methods, 2021</a>
TraCeR	从scRNA-seq数据中重建配对 TCR序列	基于Trinity进行de novo转录本组装, 然后比对到IMGT参考序列	TCR	FASTQ (scRNA-seq)	TSV, FASTA	能够从无 V(D)J富集的scRNA-seq数据中恢复TCR, 开源	依赖Trinity, 计算资源消耗可能较大, 对低表达 TCR的敏感性有限	-	<a href="#">Stubington et al., Nat Methods, 2016</a>
Platypus	从scRNA-seq数据中重建TCR序列, 优化低表达TCR检测	结合k-mer计数和期望最大化算法	TCR	FASTQ (scRNA-seq)	TSV	针对低TCR表达优化, 提高检出率	主要关注TCR, 计算可能较复杂	-	<a href="#">Yermanos et al., NAR, 2021</a>
SONG	V(D)J基因片段快速精确比对	基于优化的gapped-alignment算法	TCR, BCR	FASTQ, FASTA	SAM-like, 自定义格式	高准确性和速度	主要是比对工具, 下游分析需其他软件	-	<a href="#">Sheng et al., Bioinformatics, 2022</a>
MiXCR	TCR/BCR序列比对, CDR3提取, 克隆定量	高度优化的比对算法, 支持UMI, 内置纠错机制	TCR, BCR	FASTQ, FASTA	TSV, VDJCA	准确性高, 功能全面, 支持多种物种和实验方案	商业软件 (有学术免费版), 命令行操作	VDJtools, Immunarch	<a href="#">Bolotin et al., Nat Methods, 2015</a>
scirpy	单细胞 V(D)J数据分析与整合	基于Scanpy AnnData对象, 提供灵活克隆型定义, 网络分析等	TCR, BCR	AnnData, Cell Ranger output, TraCeR output, etc.	AnnData, various plot formats	与Scanpy生态系统无缝集成, 功能丰富, 灵活, 强大的可视化	依赖Python和Scanpy环境, 对非Python用户有学习曲线	Scanpy	<a href="#">Sturm et al., Nat Commun, 2021</a>
Dandelion	单细胞 V(D)J数据分析与整合 (BCR侧重)	基于Scanpy AnnData, 强调BCR谱系树构建	TCR, BCR (尤其BCR)	AnnData, Cell Ranger output	AnnData, Newick	与Scanpy集成, 专注BCR谱系分析	相对较新, 用户社区可能不如scirpy大	Scanpy	<a href="#">Suo et al., Nat Commun, 2022</a>
VDJtools	免疫组库数据后处理和统计分析	Java程序, 支持多种输入格式, 提供丰富的组库统计和可视化	TCR, BCR	MiXCR, IMGT, Cell Ranger, etc.	TSV, various plot formats	支持格式广泛, 统计功能全面, 跨平台	命令行工具, 可视化非交互式, 不直接处理原始数据	MiXCR, Immunarch	<a href="#">Shugay et al., PLoS Comput Biol, 2015</a>
Immcantation	BCR/抗体序列分析框架	R包集合, 专注BCR体细胞高频突变, 克隆谱系等	BCR (主要), TCR	FASTA, AIRR Standard format	AIRR Standard format, various R objects	BCR分析功能非常强大和成熟, 遵循AIRR标准	主要基于R, 对非R用户有学习曲线	R ecosystem, AIRR community tools	<a href="#">immcantation.org</a>
Loupe VDJ Browser	10x Genomics V(D)J数据交互式可视化	桌面应用程序	TCR, BCR	10x .cloupe files	-	用户友好, 交互性强, 便于探索 V(D)J与GEX关联	仅支持10x Genomics数据格式, 分析功能有限	Cell Ranger	<a href="#">10x Genomics</a>

IV. 公共单细胞V(D)J数据集资源导航

公共数据集的积累和共享对于推动免疫组库研究的进展至关重要。它们不仅为计算方法的开发和基准测试提供了宝贵的资源，也为无法自行产生大规模数据的研究者提供了进行生物信息学分析和科学发现的机会。

A. 公共数据存储库的价值

公共数据集通过促进荟萃分析、为工具开发提供基准、赋能无实验室能力的研究者以及推动可重复性研究，成为加速整个领域发展的催化剂。

- 促进荟萃分析和普适性免疫学原理的发现：通过整合来自不同研究的数据，可以检验免疫学发现在不同人群、疾病背景或实验条件下的普适性，并可能发现新的、更广泛的规律。
- 为计算工具开发提供基准数据集：新的生物信息学算法和软件需要使用已知特征或经过充分注释的数据集进行测试和验证。公共数据集为此提供了便利。



- **赋能无湿实验室能力的研究者**：计算生物学家或生物信息学家可以利用公共数据进行纯粹的计算机模拟和分析研究，从而贡献于免疫学知识的生长。
- **推动可重复性和数据共享**：遵循FAIR原则（Findable, Accessible, Interoperable, Reusable），公开数据有助于提高研究的透明度和可重复性，促进科学合作。

B. 主要的数据门户和存储免疫组库数据的联盟

目前，数据主要存储在通用型存储库（如GEO）和专业化免疫组库数据库（如iReceptor、VDJdb）中，同时大型研究联盟（如HCA）也是重要的数据来源。

1. 通用型存储库：

- **\*\*NCBI Gene Expression Omnibus (GEO) 和 Sequence Read Archive (SRA)\*\***：这两个由美国国家生物技术信息中心（NCBI）维护的数据库是存储和共享各种高通量测序数据（包括单细胞V(D)J数据）的主要平台。
- **\*\*European Genome-phenome Archive (EGA)\*\***：欧洲生物信息学研究所（EBI）和基因组调控中心（CRG）合作维护的数据库，主要存储具有个体可识别性的基因型和表型数据，需要授权访问。

2. 专业化免疫组库数据库：

- **iReceptor**：一个重要的门户网站，旨在通过遵循AIRR社区制定的数据标准，整合和查询分布在不同存储库中的适应性免疫受体组库（AIRR-seq）数据。
- **VDJdb**：一个经过人工审阅和注释的TCR序列数据库，这些TCR序列具有已知的抗原特异性。
- **McPAS-TCR**：一个手动整理的TCR序列目录，这些序列与多种病理状况和特定抗原相关联。
- **TbAdb**：提供了T细胞和B细胞受体及其对应抗原信息的数据库。
- **\*\*Observed Antibody Space (OAS)\*\***：一个大规模收集已发表的成对和非成对BCR/抗体序列的数据库。
- **\*\*Immune Epitope Database (IEDB)\*\***：一个综合性的免疫学资源，包含大量识别免疫表位的TCR和BCR数据链接。

3. 来自大型研究联盟或特定项目的数据：

- **人类细胞图谱（Human Cell Atlas, HCA）**：这个国际性的大型合作项目旨在创建所有人类细胞类型的综合参考图谱，其中包含了大量的单细胞V(D)J数据。
- **特定疾病研究计划**：例如，针对COVID-19的大量研究产生了海量的公共V(D)J数据集。

C. 使用公共数据集时的注意事项

在使用公共数据时，必须注意不同研究在处理流程、克隆型定义上的差异，并警惕批次效应，以确保分析结果的准确性和可比性。

- **数据处理流程和克隆型定义**的差异：不同研究可能采用不同的生物信息学流程和克隆型定义标准，这可能影响跨研究比较的结果。
- **元数据质量和完整性**的重要性：高质量、详细的元数据（如物种、组织来源、疾病状态、细胞类型、实验条件、患者临床信息等）对于正确解读数据至关重要。
- **数据标准的遵循**：如果数据遵循了如AIRR社区制定的标准格式，将极大地提高其可用性和互操作性。
- **批次效应**：当合并来自不同研究的数据时，需要警惕并妥善处理可能存在的批次效应。

公共数据库的价值与其数据的标准化程度和元数据的丰富程度直接相关，这对于实现大规模荟萃分析和驱动预测算法的发展至关重要。

仅仅拥有原始序列数据，如果缺乏上下文信息，其科学价值是有限的。例如，一个研究人员如果想寻找对特定病毒肽有反应的TCR，他需要的不仅仅是TCR序列本身，还需要知道该序列来源的物种、组织、宿主的HLA类型、T细胞的功能状态、以及该特异性是否经过实验验证等等。如果这些元数据缺失，或者在不同数据库之间格式不统一，那么进行大规模的荟萃分析就会变得极其困难。iReceptor门户网站通过推广AIRR标准，正试图解决这一难题。

公共数据集的可用性，特别是那些关联了抗原特异性的数据集，极大地推动了机器学习预测算法的发展。

机器学习模型需要大量的标记数据进行训练。那些提供了TCR序列及其已知靶标表位配对信息的数据库，正是充当了这种训练数据的角色。随着更多高质量、经过实验验证的数据公开发布，这些预测模型的性能将会不断提升，从而可能加速治疗相关TCR的识别过程。

公共数据集不仅是免疫学家的资源，也为计算生物学家开发新算法提供了宝贵的原始素材。

V(D)J数据独特的结构（如成对链信息、与转录组的连锁）为新算法的开发带来了有趣的挑战和机遇。例如，基于转录组和受体序列共同定义细胞身份，或者利用BCR中的体细胞高频突变追踪B细胞谱系，这些都是计算研究的活跃领域。公共数据集为这些方法学的创新提供了原始素材。

表2: 部分可用的公共单细胞V(D)J数据集示例

注：此表为示例性概览，公共数据集的可用性和具体内容会不断更新。用户应查阅相关数据库和文献以获取最新和最详细的信息。

数据集标识/名称	主要发表文献/联盟	物种	组织/细胞类型	疾病/状况	研究重点	单细胞技术	细胞/组库数量	主要V(D)J相关发现/相关性	数据访问/链接
GSE150728	Wu et al., Cell, 2020	Human	PBMCs	COVID-19 (Severe vs. Mild)	SARS-CoV-2特异性T/B细胞应答特征	10x 5' scRNA-seq + VDJ	~15万细胞	揭示了重症患者中高度克隆扩增但功能耗竭的CD8+ T细胞	<a href="#">GEO: GSE150728</a>
GSE125970	Azizi et al., Cell, 2018	Human	Breast Tumor, Lymph Node, Blood	Breast Cancer	肿瘤微环境T细胞状态和TCR组库	10x 5' scRNA-seq + VDJ, SMART-seq2	~4.5万T细胞	描绘了乳腺癌T细胞亚群的转录组和TCR特征	<a href="#">GEO: GSE125970</a>
HCA (various)	HCA Consortium	Human	Various healthy tissues	Healthy	构建健康人体组织的细胞图谱	Various (10x VDJ)	Millions	提供健康个体不同组织TCR/BCR组库的基线数据	<a href="#">HCA Data Portal</a>
GSE110681	Guo et al., Nature, 2018	Human	Colorectal Cancer, Blood, etc.	Colorectal Cancer	结直肠癌TILs的TCR组库和新抗原反应	SMART-seq2 + TCRseq	~1.1万T细胞	鉴定了肿瘤特异性CD8+ T细胞克隆及其与新抗原反应的关联	<a href="#">GEO: GSE110681</a>
VDJdb	Bagaev et al., NAR, 2020	Human, Mouse	Various	Various	汇总已知抗原特异性的TCR序列	Various	>7万TCR条目	提供大量TCR-抗原配对信息，是开发预测工具的宝贵资源	<a href="#">vdjdb.cdr3.net</a>
OAS database	Olsen et al., NAR, 2022	Human, Mouse	Various	Various	大规模收集和标准化BCR/抗体序列	Various	>18亿序列	提供海量的BCR序列数据，支持抗体工程和机器学习模型开发	<a href="#">opentargets.github.io/oasis/</a>
COVID-19 Cell Atlas	Various	Human	PBMC, BALF, Lung, etc.	COVID-19	COVID-19免疫病理机制，T/B细胞应答	Various (with VDJ)	Large collection	整合大量COVID-19相关数据集，便于研究病毒感染后的免疫组库动态	<a href="#">covid19cellatlas.org</a>
iReceptor Gateway	Corrie et al., Front Immunol, 2018	Various	Various	Various	整合查询遵循AIRR标准的分布式免疫组库数据	N/A	N/A	通过AIRR标准促进跨研究、跨机构的数据检索和比较	<a href="#">gateway.ireceptor.org</a>

V. 连接序列与洞见：关联免疫组库特征与细胞状态及生物学背景

单细胞V(D)J测序的真正威力在于其能够将免疫受体的序列信息与同一细胞的其他分子特征以及更广泛的生物学背景联系起来。这种整合分析是从描述性组库分析迈向机制性免疫学理解的关键。

A. 整合分析的力量

单细胞V(D)J分析的威力在于能够将克隆型特性与同一细胞的表型直接关联，从而将研究从序列编目提升到对克隆功能角色的探究。现代单细胞V(D)J测序方法的核心优势在于能够同时捕获来自同一个细胞的V(D)J序列和其他细胞特征。这种内在的联系使得研究人员可以直接将克隆型特性（例如，特定的CDR3序列、克隆性扩增）与细胞表型（例如，活化状态、耗竭标志物、细胞因子表达谱、细胞类型身份）进行关联。这种能力将免疫组库的研究从单纯的序列编目，提升到了对特定克隆在特定生物学情境下功能角色的探究。例如，仅仅知道一个特定的TCR克隆在肿瘤样本中发生了扩增是有趣的，但如果能进一步通过整合转录组数据发现，这个扩增的TCR克隆主要存在于那些高表达PD-1、LAG-3和TOX（经典的T细胞耗竭标志物）的细胞中，那么这将提供一个更丰富、更具行动指导意义的洞见。这可能意味着这个克隆虽然扩增了，但其功能可能已经受损，从而提示了应用免疫检查点抑制剂（如抗PD-1抗体）来“唤醒”这些T细胞的潜在策略。

B. 关联克隆型与转录组特征

通过整合转录组数据，可以识别与特定克隆行为相关的基因表达程序和分子通路，从而揭示其功能状态。通过将V(D)J数据与同一细胞的scRNA-seq数据整合，可以：

- **识别与扩增克隆相关的基因表达程序**：例如，在肿瘤免疫治疗后，可以分析那些显著扩增的T细胞克隆是否表现出效应T细胞（如高表达GZMB、IFNG）、记忆T细胞（如高表达IL7R、CCR7）或耗竭T细胞（如高表达PDCD1、CTLA4、LAG3、TOX）的转录特征。
- **进行差异基因表达分析**：通过将细胞按照其克隆型（或克隆型大小）分组，然后比较不同组之间的基因表达谱，可以发现与特定克隆行为相关的分子通路和调控网络。

从整合数据中发现的V基因使用与细胞活化状态的关联模式，可以驱动关于免疫识别和应答规则的新假说。

从整合的单细胞数据中浮现出的模式，例如在疫苗接种后，某种特定的V基因使用总是与活化的T细胞表型（如高表达Ki67）相伴出现，可以驱动关于免疫识别和应答规则的新假说。如果跨多个接种了疫苗X的个体，使用Vβ5.1基因并表达Ki67的T细胞都一致性地扩增，这可能暗示Vβ5.1基因本身就特别适合识别疫苗X中的某个免疫优势表位。

### C. 关联克隆型与细胞表面蛋白表达 (CITE-seq)

CITE-seq技术提供了对细胞表面蛋白更直接的定量检测，能够实现对免疫细胞更精细的表型分析和功能状态确认。CITE-seq（Cellular Indexing of Transcriptomes and Epitopes by Sequencing）技术通过使用寡核苷酸标记的抗体，使得在单细胞水平同时测量RNA表达和细胞表面蛋白水平成为可能。

- **精细化免疫表型分析**：将特定的TCR/BCR序列与详细的免疫表型（例如，通过CD45RO、CCR7区分初始/记忆T细胞亚群；通过PD-1、TIM-3、LAG-3等评估T细胞耗竭程度）直接关联起来。
- **功能状态确认**：表面蛋白表达可以为转录组数据推断的细胞功能状态提供佐证或更精确的定义。

### D. 追踪克隆动态随时间或条件的变化

通过纵向和比较研究，可以追踪克隆动态随时间或条件的变化，从而发现与临床结局相关的免疫学标志物。

- **纵向研究**：在感染过程、疫苗接种后或肿瘤治疗期间，定期采集样本进行单细胞V(D)J和多组学分析，可以观察特定克隆型如何扩增、收缩，或者其细胞表型和功能状态如何随时间演变。
- **比较研究**：比较不同患者群体（例如，对某种治疗产生应答的患者 vs. 无应答的患者）之间的免疫组库特征和相关的细胞状态。这种比较有助于发现与临床结局相关的免疫学标志物。

将可追踪的特定克隆型与临床结果（如治疗效果、疾病进展、疫苗保护效果）联系起来，是单细胞V(D)J分析最具转化潜力的领域。如果一组TCR克隆型持续存在于对免疫治疗有反应的癌症患者中，而在无反应者中则缺失，那么这些克隆型本身就可能成为预测治疗反应的生物标志物。

### E. 建立因果关系面临的挑战

尽管整合分析能揭示强相关性，但确定因果关系通常需要进行进一步的功能实验来验证，这是从观察到机制理解的关键一步。

- **相关不等于因果**：例如，一个在肿瘤中扩增并表现出耗竭表型的T细胞克隆，它可能是因为长期暴露于肿瘤抗原而发生反应性扩增并最终耗竭，也可能这个克隆本身就具有易于耗竭的倾向。
- **功能验证的必要性**：对于那些被鉴定为具有重要生物学意义的TCR或BCR，通常需要通过功能实验来验证其特性，例如将其克隆并表达在报告细胞系中，测试其对抗原的反应性。

## VI. 免疫组库分析的新兴前沿与未来展望

单细胞V(D)J测序及其分析方法学正处于飞速发展之中，不断涌现的新技术、新算法和新生物学问题正在塑造该领域的未来。

### A. 技术进步

#### 1. 空间背景整合：

将单细胞分析与空间转录组学结合，是当前的技术前沿，它能在组织原位解析免疫克隆的空间定位，对于理解细胞互作至关重要。这将使得研究者能够了解具有特定TCR/BCR的克隆型在组织微环境中的精确空间定位。例如，一个细胞毒性T细胞克隆可能在肿瘤组织中数量丰富，但如果它被限制在基质区域而无法有效浸润到肿瘤细胞床，其抗肿瘤效果就会大打折扣。

#### 2. 更高通量的多组学联用：

未来的技术趋势是在单细胞水平整合更多维度的组学信息，以提供关于单个免疫细胞状态的更全面画像。未来有望在单细胞水平同时整合更多种类的组学信息，例如将V(D)J序列、转录组与ATAC-seq（分析染色质可及性）、代谢组学特征等结合起来。

#### 3. 长读长测序技术在V(D)J分析中的应用：

长读长测序技术有望更直接地获得完整的V(D)J序列，但其在单细胞应用中的通量和成本仍是待解决的挑战。长读长测序技术（如PacBio或Oxford Nanopore）在理论上能够更直接地获得全长的、无拼接错误的TCR/BCR转录本序列，但其在单细胞应用中的通量、成本和错误率仍是需要持续改进的方面。

### B. 计算挑战与创新

#### 1. TCR/BCR-抗原特异性预测：

利用AI预测TCR/BCR的抗原特异性是计算免疫学的“圣杯”，其成功将彻底改变个性化免疫疗法和诊断的开发模式。

利用机器学习和人工智能模型，基于TCR或BCR的序列和/或结构信息，预测其能够识别的抗原表位，是计算免疫学领域的“圣杯”之一。这需要大量高质量的、经过实验验证的受体-抗原相互作用数据作为训练集。如果TCR/BCR-抗原特异性预测取得成功，将使得我们能够仅从患者的免疫组库序列数据中就快速识别出与疾病相关的受体，从而极大地加速诊断方法的开发和个性化治疗策略的设计。

#### 2. 免疫组库演化的动态建模：

开发能够模拟和预测免疫组库动态演化的计算模型，是理解免疫系统复杂动力学的重要方向。这需要开发能够整合多因素影响并进行前瞻性预测的复杂模型。

#### 3. 先进机器学习与人工智能（AI）的应用：

AI/ML技术除了用于特异性预测，还可用于识别与临床相关的复杂免疫特征，并整合异构的多组学数据。例如，从高维免疫组库数据中识别与疾病状态或治疗反应相关的“免疫特征”，以及整合和解析来自不同组学层面的异构数据。

#### 4. 改进的克隆型定义和谱系追踪：

发展更精细的克隆型定义和谱系追踪方法，特别是针对复杂的BCR演化，是计算方法学持续创新的方向。特别是对于BCR，需要开发更准确、更高效的计算工具来追踪B细胞克隆在亲和力成熟过程中的演化谱系。

### C. 拓展生物学理解

未来的生物学研究将聚焦于更深入地理解交叉反应性、先天与适应性免疫的互作，并最终实现个性化的免疫治疗和疫苗设计。

- **深入理解TCR/BCR交叉反应性的规律**：结合多学科方法，揭示决定TCR/BCR交叉识别能力的分子机制。
- **阐明先天性与适应性免疫的互作**：通过整合的单细胞多组学研究，探究不同免疫细胞类型间的相互调节。
- **个性化免疫治疗和疫苗设计**：基于个体免疫组库的特征，设计更精准、更有效的个性化免疫治疗方案和预防性疫苗。

### D. 伦理考量

随着免疫组库数据变得日益强大，其涉及的数据隐私、安全共享和潜在歧视风险等伦理问题亟待建立负责任的规范。

- **数据隐私与可识别性**：个体的免疫组库具有高度独特性，理论上可能被用于重新识别个体身份。
- **数据安全与共享规范**：需要建立强大的数据匿名化技术、安全的数据存储和传输协议，以及明确的伦理和法律框架。
- **潜在的歧视风险**：需要警惕免疫组库特征信息被滥用用于例如保险、就业等方面的歧视。

## VII. 结论

单细胞V(D)J测序及其多组学整合策略，已成为洞察适应性免疫的革命性工具，其发展依赖于技术、计算和数据资源的协同进步。单细胞V(D)J测序技术及其相关的多组学整合策略，已经为我们以前所未有的分辨率和深度洞察适应性免疫应答的复杂性提供了革命性的工具。通过同时解析单个免疫细胞的抗原受体序列、转录组谱、表面蛋白表达乃至空间定位信息，研究人员能够更精确地鉴定功能性免疫克隆、追踪其动态变化，并将这些分子特征与细胞的生物学行为和临床表型紧密联系起来。

从原始数据中提取有意义的生物学洞见，高度依赖于复杂而精密的计算分析方法和稳健的生物信息学工具。本报告回顾了从数据预处理、V(D)J序列注释、克隆型定义到免疫组库特征量化、可视化以及多模态数据整合等关键计算环节的前沿方法。同时，也强调了标准化数据格式（如AIRR标准）和高质量公共数据集（如iReceptor、VDJdb、GEO等）对于促进领域发展、实现数据共享和荟萃分析的重要性。

该领域正处在技术创新、计算方法和数据资源增长的良好循环中，这种协同互动是推动领域发展的核心动力。贯穿本报告所描述的进展，并非源于单一因素，而是测序技术、计算方法和数据共享这些要素之间协同互动的结果。更先进的测序技术产生更多、更复杂的数据，这就需要更强大的计算工具来处理和分析；新的分析方法揭示出新颖的生物学模式，从而激发新的实验设计和数据生成策略；而公共数据的共享则通过允许多方研究者共同参与工具开发和科学发现，加速了这一良性循环。

该领域的最终目标是将海量数据转化为可操作的免疫学知识并应用于临床，这需要持续的跨学科合作才能实现。这一宏伟目标的实现，需要基础科学家、计算生物学家、临床医生以及伦理学家的持续跨学科合作。单细胞V(D)J测序技术无疑是驱动这一转化过程的强大引擎，其未来的发展和应用前景广阔，有望在感染性疾病、肿瘤、自身免疫病、疫苗开发和个性化医疗等多个领域带来深刻的变革。