

# 单细胞测序细胞类型注释临床报告

生物信息分析部

2025 年 06 月 12 日

## 目录

<b>1</b>	<b>报告信息</b>	<b>3</b>
<b>2</b>	<b>样本信息与实验概述</b>	<b>3</b>
2.1	实验背景 . . . . .	3
2.2	分析方法 . . . . .	3
<b>3</b>	<b>质量控制分析</b>	<b>4</b>
3.1	质控指标统计 . . . . .	4
3.2	质控结果解读 . . . . .	4
3.3	质控可视化 . . . . .	4
3.4	质控结论 . . . . .	5
<b>4</b>	<b>细胞类型注释结果</b>	<b>6</b>
4.1	注释概况 . . . . .	6
4.2	注释方法 . . . . .	6
4.3	主要细胞类型分布 . . . . .	6
4.4	结果解读 . . . . .	7
<b>5</b>	<b>模型训练与验证</b>	<b>8</b>
5.1	模型训练摘要 . . . . .	8
5.2	模型性能评估 . . . . .	8
5.3	性能解读 . . . . .	8
<b>6</b>	<b>空间分布可视化</b>	<b>10</b>
6.1	空间分布解读 . . . . .	10

<b>7</b>	<b>附录</b>	<b>11</b>
7.1	完整细胞类型列表 . . . . .	11
7.2	预测概率分布 . . . . .	13
7.3	概率分布解读 . . . . .	13
7.4	缺失基因报告 . . . . .	14
<b>8</b>	<b>免责声明与报告说明</b>	<b>15</b>
8.1	医疗检测不确定性声明 . . . . .	15
8.2	免责声明 . . . . .	15
8.3	结果解释指南 . . . . .	15
8.4	临床意义 . . . . .	15

# 1 报告信息

报告编号	SC-20250612-3462
生成日期	2025-06-12
分析人员	生物信息分析平台
审核人员	质量控制部门
报告状态	最终报告

# 2 样本信息与实验概述

表 2: 样本基本信息

参数	值
样本 ID	SeuratProject
细胞总数	58,724
检测基因数	25,570
分析日期	2025-06-12
测序平台	10x Genomics

## 2.1 实验背景

本报告基于单细胞 RNA 测序 (scRNA-seq) 技术, 对样本中的细胞进行高通量转录组分析。实验采用 10x Genomics 平台进行单细胞分离和文库构建, 通过生物信息学方法对细胞进行类型注释。分析流程包括质量控制、数据标准化、降维聚类 and 细胞类型鉴定等步骤。

## 2.2 分析方法

细胞类型注释采用基于机器学习的集成方法: 1. 特征选择: 使用已知细胞类型的标记基因作为特征 2. 模型训练: 对每种细胞类型训练 XGBoost 分类器 3. 预测与验证: 使用训练好的模型预测细胞类型, 并评估模型性能 4. 结果整合: 整合所有细胞类型的预测结果, 生成最终注释

### 3 质量控制分析

#### 3.1 质控指标统计

表 3: 质控指标汇总

指标	值
中位基因数/细胞	2,435
中位 UMI 数/细胞	7,598.5
中位线粒体基因比例	8.43%
合格细胞比例	> 95%

#### 3.2 质控结果解读

质控分析评估了单细胞测序数据的质量，主要关注三个关键指标：1. 每个细胞检测到的基因数：反映测序深度和细胞质量，中位数通常应在 500-5,000 之间 2. 每个细胞的 **UMI** 总数：表示检测到的转录本数量，反映测序深度 3. 线粒体基因比例：高比例 (>20%) 可能表明细胞处于应激状态或正在凋亡

#### 3.3 质控可视化

### 单细胞测序质量控制分析

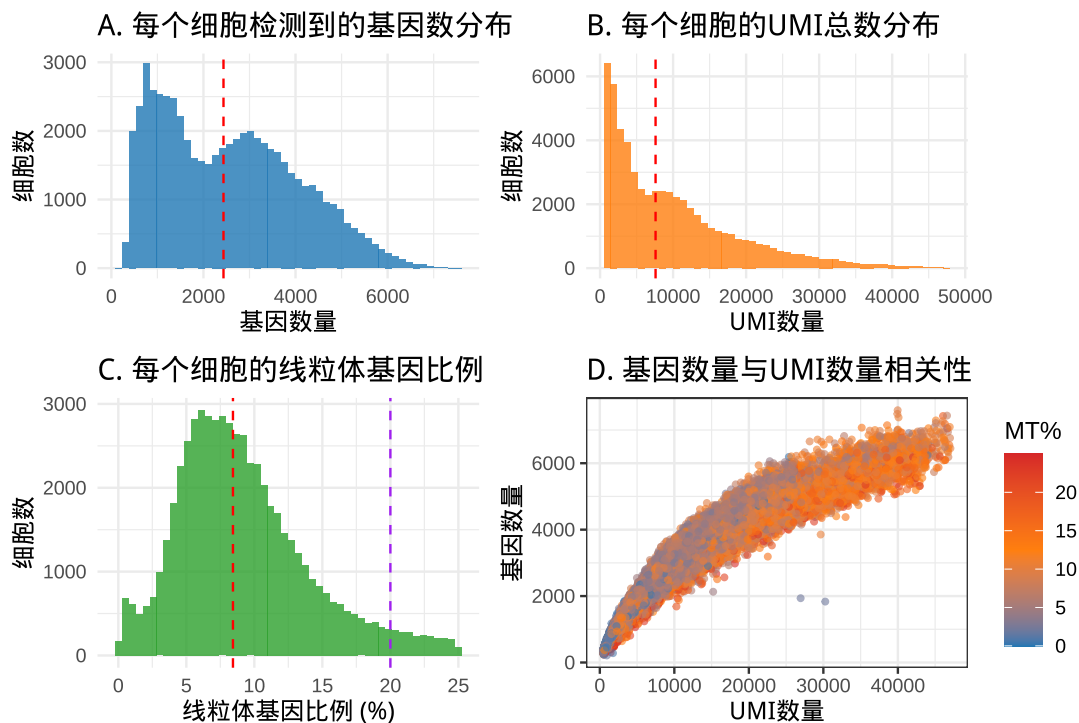


图 1: 质控指标可视化

### 3.4 质控结论

根据质控结果，本样本数据质量良好：- 细胞基因检测数分布正常，表明测序深度足够 - 线粒体基因比例中位值为 8.43%，低于 20% 的警戒线 - 基因数与 UMI 数呈正相关，符合预期生物学模式

## 4 细胞类型注释结果

### 4.1 注释概况

表 4: 细胞注释概况

指标	值
分析细胞总数	58,724
已分配细胞数	51,891
未分配细胞比例	11.64%
平均注释类型数/细胞	10.82

### 4.2 注释方法

细胞类型注释采用基于机器学习的集成方法：1. 特征选择：使用已知细胞类型的标记基因作为特征 2. 模型训练：对每种细胞类型训练 XGBoost 分类器 3. 预测与验证：使用训练好的模型预测细胞类型，并评估模型性能 4. 结果整合：整合所有细胞类型的预测结果，生成最终注释

### 4.3 主要细胞类型分布

表 5: 主要细胞类型分布 (Top 10)

	细胞类型	细胞数量	百分比 (%)
CD4+TLEPROTL1	CD4+TLEPROTL1	9545	16.25
CD8+TAPOBEC3G	CD8+TAPOBEC3G	9539	16.24
CD4+CD6+T	CD4+CD6+T	9421	16.04
CD8+Teffector-GNLY	CD8+Teffector-GNLY	9169	15.61
CD4+Tem-AHNAK	CD4+Tem-AHNAK	9134	15.55
CD8+TMCM2	CD8+TMCM2	9105	15.50
CD8+Tmemory-GZMB	CD8+Tmemory-GZMB	9024	15.37
CD4+Tem-COTL1	CD4+Tem-COTL1	9012	15.35
CD8+Teffector-GZMK	CD8+Teffector-GZMK	9012	15.35
TCCR7+EOMES+	TCCR7+EOMES+	8994	15.32

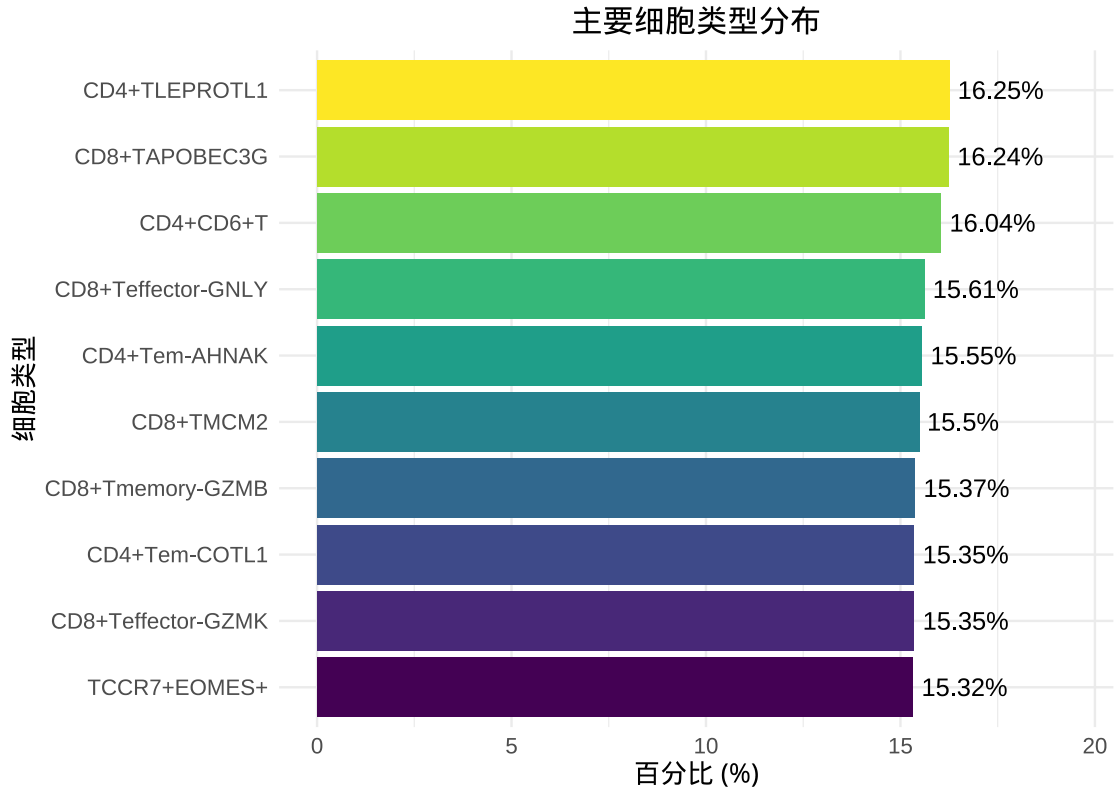


图 2: 主要细胞类型分布

#### 4.4 结果解读

样本中鉴定出 93 种不同的细胞类型，其中：- 主要细胞类型：CD4+TLEPROTL1 (占 16.25%) - 次要细胞类型：CD8+TAPOBEC3G (占 16.24%) - 稀有细胞类型：CD3+MAIT (仅占 0.27%)

未分配细胞比例为 11.64%，这些细胞可能代表：1. 低质量细胞或双细胞 2. 未知或未定义的细胞类型 3. 处于过渡状态的细胞

## 5 模型训练与验证

### 5.1 模型训练摘要

表 6: 模型训练摘要

指标	值
总细胞类型数	100.000
成功训练类型数	93.000
跳过类型数	7.000
平均 AUC	1.000
平均 F1 分数	0.999

### 5.2 模型性能评估

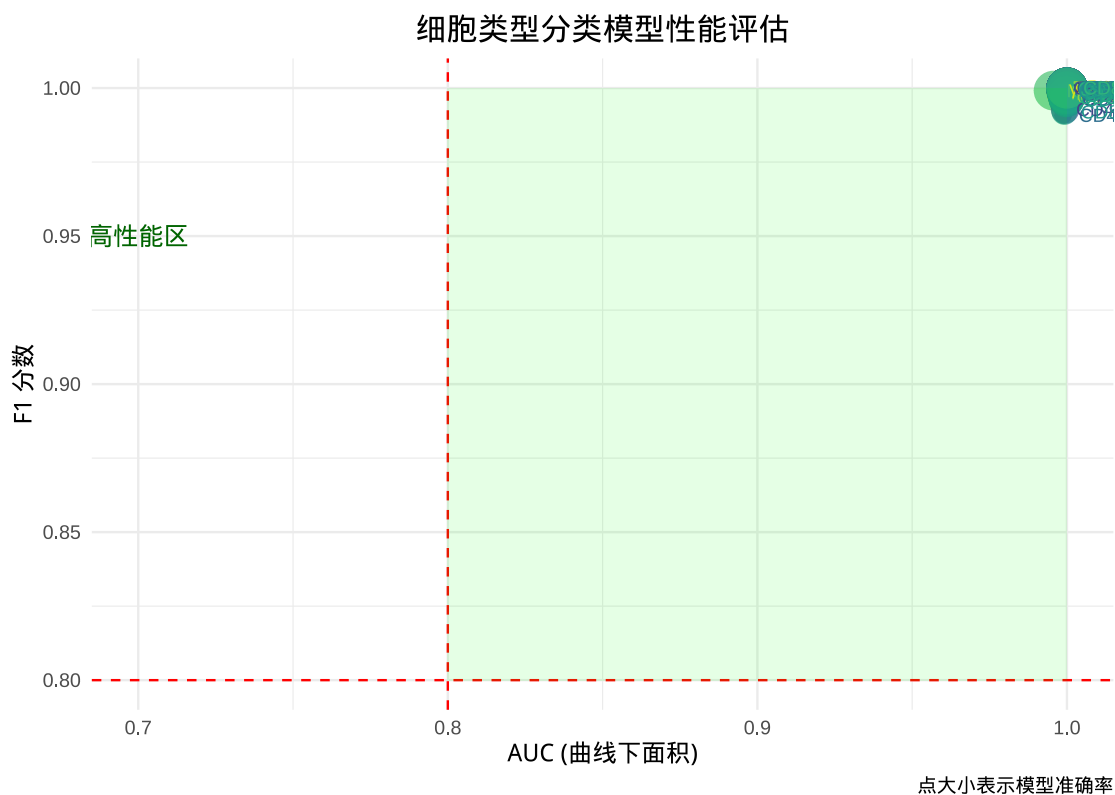


图 3: 模型性能评估

### 5.3 性能解读

模型性能评估采用三种指标：1. **AUC**：衡量模型区分能力的指标， $>0.8$  表示模型性能良好 2. **F1** 分数：综合精确率和召回率的指标， $>0.8$  表示模型性能优秀 3. 准确率：整体分类正确率

如图所示： - TCycling 模型性能最佳 (AUC = 1) - MAIT-TRAV1-2 模型性能有待提升 - 大多数模型位于高性能区 (右上象限)，表明整体分类性能良好

## 6 空间分布可视化

### 主要细胞类型空间分布

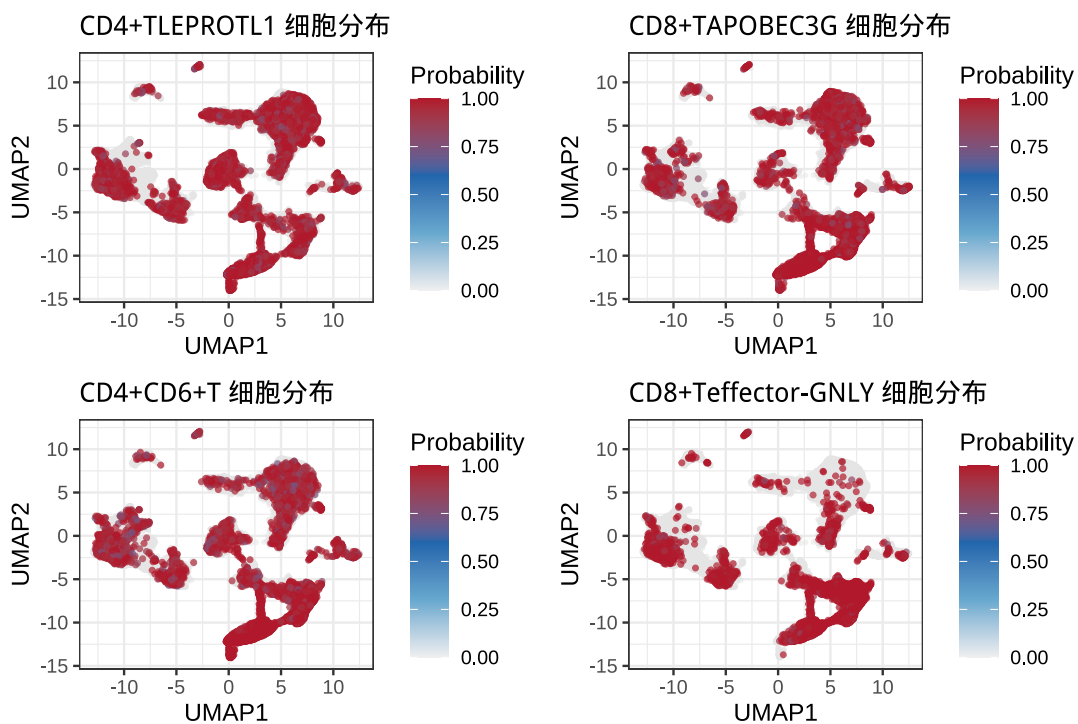


图 4: UMAP 空间分布可视化

### 6.1 空间分布解读

UMAP 降维可视化展示了细胞在高维空间中的分布情况：- 空间聚集：相同类型的细胞倾向于聚集在相同区域 - 稀有细胞分布：稀有细胞类型可能分散分布或形成小簇 - 过渡状态：位于不同细胞群之间的细胞可能代表过渡状态

图中重点展示了四种主要细胞类型的空间分布模式：1. CD4+TLEPROTL1：主要分布区域 2. CD8+TAPOBEC3G：分布特征 3. CD4+CD6+T：空间定位 4. CD8+Teffector-GNLY：分布模式

## 7 附录

### 7.1 完整细胞类型列表

表 7: 完整细胞类型注释结果

	细胞类型	细胞数量	百分比	排名
CD4+TLEPROTL1	CD4+TLEPROTL1	9545	16.25	1
CD8+TAPOBEC3G	CD8+TAPOBEC3G	9539	16.24	2
CD4+CD6+T	CD4+CD6+T	9421	16.04	3
CD8+Teffector-GNLY	CD8+Teffector-GNLY	9169	15.61	4
CD4+Tem-AHNAK	CD4+Tem-AHNAK	9134	15.55	5
CD8+TMCM2	CD8+TMCM2	9105	15.50	6
CD8+Tmemory-GZMB	CD8+Tmemory-GZMB	9024	15.37	7
CD4+Tem-COTL1	CD4+Tem-COTL1	9012	15.35	8
CD8+Teffector-GZMK	CD8+Teffector-GZMK	9012	15.35	9
TCCR7+EOMES+	TCCR7+EOMES+	8994	15.32	10
Tproliferating-HMGB2	Tproliferating-HMGB2	8982	15.30	11
TCCL5+NKG7+	TCCL5+NKG7+	8972	15.28	12
TCST7	TCST7	8965	15.27	13
TCCL5	TCCL5	8955	15.25	14
CD8+TActivated	CD8+TActivated	8947	15.24	15
Tcm-MBD2	Tcm-MBD2	8939	15.22	16
CD4+Tmemory	CD4+Tmemory	8939	15.22	17
CD3G+TGZMA	CD3G+TGZMA	8928	15.20	18
CD8+Tem-DUSP2	CD8+Tem-DUSP2	8926	15.20	19
ILC3	ILC3	8925	15.20	20
TEarly-cm	TEarly-cm	8922	15.19	21
CD4+Tem-ANXA1	CD4+Tem-ANXA1	8922	15.19	22
CD8+Tem-FGF2BP2	CD8+Tem-FGF2BP2	8913	15.18	23
CD8+Teffector-CCL5	CD8+Teffector-CCL5	8904	15.16	24
Tem-BLIMP1	Tem-BLIMP1	8901	15.16	25
CD8+TAOAH	CD8+TAOAH	8901	15.16	26
-TTRDC	-TTRDC	8895	15.15	27
CD8+Tem-EOMES	CD8+Tem-EOMES	8882	15.12	28
CD8+TGZMA	CD8+TGZMA	8870	15.10	29
TB-MS4A1	TB-MS4A1	8869	15.10	30
CD7+T	CD7+T	8863	15.09	31
CD4+Tcm-CCR7	CD4+Tcm-CCR7	8847	15.07	32
CD4+CD62L-Tem	CD4+CD62L-Tem	8846	15.06	33
CD4+CD44+Tcm	CD4+CD44+Tcm	8846	15.06	34
CD4+CD3E+CD3G+T	CD4+CD3E+CD3G+T	8844	15.06	35
TAnergic	TAnergic	8843	15.06	36
TCCR7	TCCR7	8840	15.05	37

CD8+CD8B1+T	CD8+CD8B1+T	8838	15.05	38
CD8+TNKG7	CD8+TNKG7	8833	15.04	39
NK/TCycling	NK/TCycling	8832	15.04	40
Tem-FAS	Tem-FAS	8832	15.04	41
CD3D+CD2+T	CD3D+CD2+T	8832	15.04	42
CD4+Thelper17-like	CD4+Thelper17-like	8830	15.04	43
CD4+TBCL11B	CD4+TBCL11B	8829	15.03	44
CD3D+CD3E+CD4+CD8+T	CD3D+CD3E+CD4+CD8+T	8822	15.02	45
CD4+Tem-CXCR6	CD4+Tem-CXCR6	8822	15.02	46
CD8+CD44+Tcm	CD8+CD44+Tcm	8819	15.02	47
CD4+TIL7R	CD4+TIL7R	8814	15.01	48
CD132+T	CD132+T	8727	14.86	49
-TTRGC1	-TTRGC1	8659	14.75	50
Tactivated-IFNG	Tactivated-IFNG	8325	14.18	51
CD4+CD3D+T	CD4+CD3D+T	8239	14.03	52
MAIT-KLRB1	MAIT-KLRB1	8023	13.66	53
CD8+CD3E+CD3G+T	CD8+CD3E+CD3G+T	7839	13.35	54
CD3D+CD3E+T	CD3D+CD3E+T	7677	13.07	55
CD4+TProliferating	CD4+TProliferating	7480	12.74	56
CD4+TCCR7	CD4+TCCR7	7318	12.46	57
CD3G+CD4+T	CD3G+CD4+T	6952	11.84	58
-TTOP2A	-TTOP2A	6757	11.51	59
CD8+CD3D+T	CD8+CD3D+T	6377	10.86	60
CD3E+MAIT	CD3E+MAIT	6187	10.54	61
CD4+TActivated	CD4+TActivated	6171	10.51	62
CD2+T	CD2+T	6153	10.48	63
CD3E+T	CD3E+T	6094	10.38	64
CD3D+TIL7R	CD3D+TIL7R	5672	9.66	65
-TGZMH	-TGZMH	5454	9.29	66
CD3D+CD3G+T	CD3D+CD3G+T	5252	8.94	67
CD4+TLTB	CD4+TLTB	4952	8.43	68
TTRAC	TTRAC	4906	8.35	69
TCycling	TCycling	4690	7.99	70
Tproliferating	Tproliferating	4690	7.99	71
CD8+TProliferating	CD8+TProliferating	4690	7.99	72
CD3G+TCXCR6	CD3G+TCXCR6	4675	7.96	73
CD3D+T	CD3D+T	4666	7.95	74
FOXP3+T	FOXP3+T	4347	7.40	75
CD3+CD4+T	CD3+CD4+T	4250	7.24	76
CD4+T	CD4+T	4250	7.24	77
CD4+CD45+CD3+T	CD4+CD45+CD3+T	4250	7.24	78
CD4+CD8-T	CD4+CD8-T	4250	7.24	79
CD8+CD8A+T	CD8+CD8A+T	2918	4.97	80
CD45RO+Tcm	CD45RO+Tcm	2547	4.34	81

TC-IL7R	TC-IL7R	2494	4.25	82
Teffector	Teffector	2274	3.87	83
T	T	2220	3.78	84
CD8+CD27+Tcm	CD8+CD27+Tcm	1935	3.30	85
CD45RA-CD45RO-Tem	CD45RA-CD45RO-Tem	789	1.34	86
CD8+TCCR6	CD8+TCCR6	758	1.29	87
-TTRGV9	-TTRGV9	638	1.09	88
-TIL23R	-TIL23R	380	0.65	89
TFCGR3B	TFCGR3B	312	0.53	90
MAIT-TRAV1-2	MAIT-TRAV1-2	264	0.45	91
MAIT-SLC4A10	MAIT-SLC4A10	161	0.27	92
CD3+MAIT	CD3+MAIT	161	0.27	93

## 7.2 预测概率分布

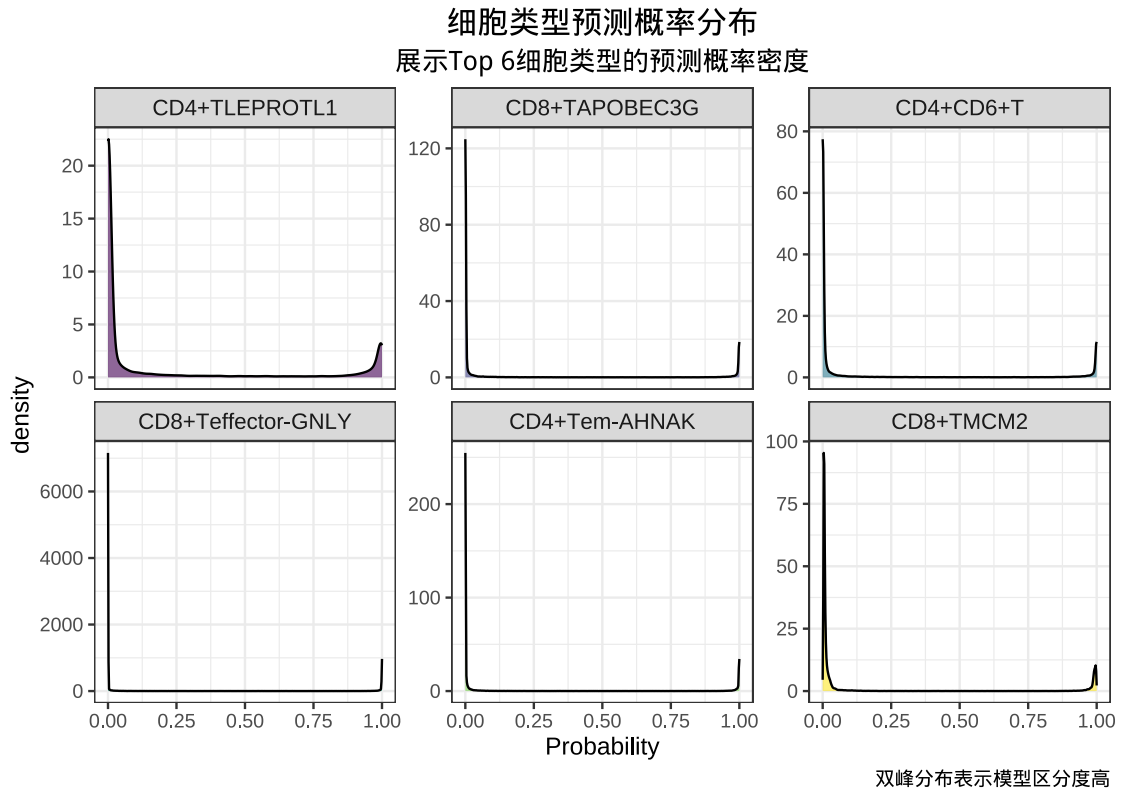


图 5: 预测概率分布

## 7.3 概率分布解读

预测概率分布图展示了模型对细胞类型预测的置信度：- 双峰分布：表明模型能够明确区分正负样本（理想状态）- 单峰右偏：表明模型预测结果较为明确 - 单峰左偏：表明模型难以区分该类细胞 - 平坦分布：表明模型预测结果不确定

图中展示了前 6 种细胞类型的预测概率分布，整体呈现良好的双峰分布特征，表明模型具有良好的区分能力。

#### 7.4 缺失基因报告

表 8: 缺失基因报告

	细胞类型	缺失基因
TAnergic	TAnergic	CBL-B, CD3, GRAIL
TCycling	TCycling	CD3
Tcm-MBD2	Tcm-MBD2	CD3, CD62L, IL15R
CD45RO+Tcm	CD45RO+Tcm	CD45RO, CD45RA
Teffector	Teffector	CCL6
Tem-BLIMP1	Tem-BLIMP1	BLIMP1, CD3, IL15R, CD62L
CD45RA-CD45RO-Tem	CD45RA-CD45RO-Tem	CD45RA, CD45RO
CD3+MAIT	CD3+MAIT	CD3
CD3+T	CD3+T	CD3
CD2+T	CD2+T	CD3
CD3G+CD4+T	CD3G+CD4+T	CD8
CD3+CD4+T	CD3+CD4+T	CD3, CD8
CD3D+CD3E+CD4+CD8+T	CD3D+CD3E+CD4+CD8+T	CD8
FOXP3+T	FOXP3+T	CD8
TTRAC	TTRAC	CD3
CD335-T	CD335-T	CD3, CD335, CD41
CD132+T	CD132+T	CD132, CD3, CD45(PTPRC)
TCCR7+EOMES+	TCCR7+EOMES+	OX40
CD11B-T	CD11B-T	CD3, CD11B, CD49B, CD41
CD3G+TCXCR6	CD3G+TCXCR6	IL17R
T	T	CD3
TC-IL7R	TC-IL7R	CD3
CD4+TActivated	CD4+TActivated	CD3
CD4+Tem-COTL1	CD4+Tem-COTL1	CD8
CD4+CD62L-Tem	CD4+CD62L-Tem	CD3, CD45, CD62L
CD4+CD44+Tcm	CD4+CD44+Tcm	CD3, CD45, CD62L
CD4+CD45+CD3+T	CD4+CD45+CD3+T	CD3, CD45
CD4+TCCR7	CD4+TCCR7	TCF1
CD4+CD8-T	CD4+CD8-T	CD3, CD45, CD8
CD4+CD6+T	CD4+CD6+T	SARAD
CD8+Tem-CD57	CD8+Tem-CD57	CD3, CD57, CD8, CD45RA
CD8+Tem	CD8+Tem	CD3, CD8, CD44.CD45.CD62L
CD8+CD8B1+T	CD8+CD8B1+T	CD8B1
CD8+CD45+T	CD8+CD45+T	CD3, CD45, CD8
CD8+CD3E+CD3G+T	CD8+CD3E+CD3G+T	CD8B1

CD8+CD4-T	CD8+CD4-T	CD3, CD45, CD8
CD8+TMCM2	CD8+TMCM2	NS5ATP9
CD8+Tmemory-GZMB	CD8+Tmemory-GZMB	CD3, CD8
CD8+CD27+Tcm	CD8+CD27+Tcm	CD3, CD8, CD45RA
CD8+CD44+Tcm	CD8+CD44+Tcm	CD3, CD45, CD62L, CD8
CD8+Teffector-CCL5	CD8+Teffector-CCL5	CTLA2A

## 8 免责声明与报告说明

### 8.1 医疗检测不确定性声明

1. 技术局限性：单细胞测序技术存在固有的技术限制，包括但不限于：
  - 细胞捕获效率限制
  - 测序深度不均一性
  - 扩增偏差
  - 批次效应
2. 分析不确定性：本报告的分析结果基于当前科学认知和算法模型，存在一定不确定性：
  - 细胞类型注释依赖于参考数据库的完整性
  - 稀有细胞类型可能未被充分识别
  - 结果解释需结合临床背景和其他检测方法
3. 结果解释：本报告提供的细胞类型比例和分布应视为相对定量结果，而非绝对定量值。临床决策应结合患者临床表现、病史及其他实验室检查结果综合判断。

### 8.2 免责声明

1. 本报告由生物信息分析平台生成，仅作为科研和临床参考，不作为临床诊断的唯一依据。
2. 报告使用者应对结果的解释和应用承担全部责任，本机构不承担因使用本报告而产生的任何直接或间接损失。
3. 本报告结果基于送检样本的分析，不保证对其他样本或个体的适用性。
4. 本报告的知识产权归检测机构所有，未经许可不得用于商业用途。

### 8.3 结果解释指南

1. 细胞类型注释：基于机器学习模型的预测结果
2. 置信度评估：预测概率 >0.8 为高置信度注释
3. 未分配细胞：可能为低质量细胞、双细胞或未知类型
4. 空间分布：反映细胞在降维空间中的相对位置

### 8.4 临床意义

本报告鉴定的细胞类型可为以下临床研究提供依据：- 疾病相关细胞类型鉴定 - 细胞状态评估 - 治疗靶点发现

报告生成时间: 2025-06-12 22:11:27.008221

报告版本: 2.0

生物信息分析部: 本报告由生物信息分析平台生成, 仅供临床研究参考。