## 2.4 FUSION TWAS全转录组关联研究 (Dataset: AA)

* Content: [LDSC]
* lg Comment: 版本？

Reply: 已补充。

* Content: FUSION
* lg Comment: 版本？

Reply: FUSION 没有版本信息。

* Content: TWAS 预测，
* lg Comment: 这两处有没有什么限制阈值或者

Reply: 这里还只是按步就班的分析，后面表格的部分有注明 P 阈值

## 2.5 Seurat 单细胞数据分析 (Dataset: AA)

* Content: 一个细胞至少应有 1000 个基因，并且基因数量小于 7500。线粒体基因的比例小于 25%。
* lg Comment: 这个有没有参考文献依据

Reply: 分析的部分有两张图 (QC质控前和质控后)，是根据图调整参数的。(低质量的细胞或空液滴通常含有很少的基因; 细胞双联体或多联体可能表现出异常高的基因计数; 低质量/死亡细胞通常表现出广泛的线粒体污染) (请参考 <https://satijalab.org/seurat/articles/pbmc3k_tutorial.html#qc-and-selecting-cells-for-further-analysis>)

* Content: 去除 Monocyte 细胞。
* lg Comment: 为何去掉这个？有文献或者标准参考吗？

Reply: 这里是一处描述错误 (是之前调试的时候加入的) 数据集里没有注释到 Monocyte 细胞，所以没有去掉。

* Content: features
* lg Comment: 这两组features指的是AA和control之间吗？

Reply: 括号中说明了 features 是什么：基因集和代谢通量，这里是关联分析部分的方法说明。

* Content: ，**基因集** (NAT10, LINC02193, DDX10P1, …[n = 116]
* lg Comment: 这里就是AA和对照的差异基因数量是吧？ 这里是否放到结果中更好些。 还是说两组一共预测到了116 基因？

Reply: 不是差异表达基因，是 TWAS 的结果，你看括号中的说明：来自于FUSION TWAS全转录组关联研究。报告中出现的格式：**基因** (… 来自于 …) 都是对这些基因或者其他结果的说明。

* Content: [n = 10]
* lg Comment: 20多种细胞这里共注释到了10个细胞类群？

Reply: 报告中没有出现过20种细胞吧？…… 总共注释到 8 种，UMAP 图中的那些，然后又分析了 T 细胞的亚群，加起来刚好 10 种。

* Content: features
* lg Comment: 所以，features具体指的什么？

Reply: 加粗的：**基因集** 与 **代谢通量**。

## 2.7 scFEA 单细胞数据的代谢通量预测 (Dataset: AA)

* Content: scFEA 预测细胞的代谢通量
* lg Comment: 这里与2.5中的代谢通量区别吗？

Reply: 2.5 部分的代谢通量来自这里的分析 (我现在把 2.5 部分的关联分析的说明，单独分离出来了，以免误解) 。

## 2.8 Limma 代谢通量差异分析 (Dataset: AA\_FLUX)

* Content: 异分析：
* lg Comment: 差异分析的是基于的基因表达量还是细胞比例差异？

Reply: 代谢通量差异，和基因无关。

## 2.10 ClusterProfiler 富集分析 (Dataset: EPC)

* Content: 以 p.adjust 表示显著水平
* lg Comment: <0.05?

Reply: < 0.05。以图片为基准。

## 3.4 CompBioLab 数据获取 (<PMID:37908861>)

* Content: 单细胞数据集
* lg Comment: Aplastic\_anemia-KRICT-2019-08 数据集名称？

Reply: 这是来自于 PMID:37908861，已经附上链接了。

## 3.5 Seurat 单细胞数据分析 (AA)

* Content: **Fig.**
* lg Comment: 这个图如何解读？ 参考： 为更好地评估分析细胞类群，基于各细胞中基因的表达进行线性降维处理。ElbowPlot函数进行可用维度的分析筛选，结果如图3-4所示。由图可知，线性降维后的标准差随着维度的增加逐渐减小，当在第20个维度时标准差逐渐趋于平缓，故将细胞聚类的线性最佳维度值设置为20。

Reply: 这个图没有必要解读这么细，只是作为附加材料，不会有文献放到文中的。如果编辑不理解单细胞数据集的分析过程怎么写，我建议取一篇发表过的文章作为模板，因为前面分析流程都是差不多的。

* Content: SCSA
* lg Comment: 少了一个聚类图？

Reply: 已补充。

* Content: **Fig.**
* lg Comment: 这个结果能不能把后边的颜色变成一个和蓝色对调的颜色吗，这样显得显著性明显一些的 , 比如表达高的是红色或者紫色， 表达第的是现在的颜色，中间变化值设置为4？

Reply: 没有图会以 4 作为中心点的，只会设置 0。这个 Marker 已经可以了。

* Content: **(File path: F**
* lg Comment: 两个分组件的差异补充下。

Reply: 已补充 (AA 组细胞数量远远多余 Normal，这个图展示效果不好)

## 3.6 Seurat 细胞亚群分析 (T\_CELL)

* Content: 匹配 scsa\_cell 中
* lg Comment: 从图5到图10中间的变化中间做了哪些限制或者进一步筛选

Reply: 这里只是取了 T 细胞分析而已。这个报告读起来可能不是很流畅，因为我设置文字注释的时候，尽可能让它“泛化”，让下次做不同分析的时候，也能这么注释，这样就不用反复修改这些基本内容了。

* Content: **Fig.** **8** T CELL UMAP Clustering
* lg Comment: 基于什么标准分成了18；类？后续有归类成了后边的4类？

Reply: UMAP 聚类是无监督聚类，前面说明了设置参数，18 是这样得到的结果。后续分为 4 类是对细胞的注释。

* Content: pe annotation
* lg Comment: 后续为什么没有对这几类细胞进行展开分析？

Reply: 已经在 “Limma 代谢通量差异分析 (AA\_FLUX)” 和 “细胞群 features 关联分析 (AA)” 的部分做了分析了。

* Content: UMAP 图。
* lg Comment: UMAP 图。

Reply: T 细胞亚群的UMAP 图聚类图

* Content: **Fig.**
* lg Comment: 可以借鉴这种去做第一个图片

Reply: 这个是经过 Z-score 归一化的。第一个图归一化后，效果不明显。

## 3.8 Limma 代谢通量差异分析 (AA\_FLUX)

* Content: 对细胞的代谢通量
* lg Comment: 始终没有理解代谢通量是基于什么标准？

Reply: PMID: 34301623，“A graph neural network model to estimate cell-wise metabolic flux using single-cell rna-seq data”。像基因表达量一样，代谢也可以一个值，可以差异分析。

* Content: DMF
* lg Comment: 全称和中文？

Reply: Defferential Metabolic flux, 差异代谢通量。

## 3.9 细胞群 features 关联分析 (AA)

* Content: TWAS Genes
* lg Comment: 这里TWAS预测后直接就能得到关联统计的结果么？ 有没有什么工具和参数？

Reply: 我在 “2.5 Seurat 单细胞数据分析 (Dataset: AA)” 的部分做了方法说明。在总结的部分也做了说明 (这部分最详细，整篇文章的思路) 。

* Content: Fig. [**14**]
* lg Comment: 13 14 15的区别是什么？ 分别是不同细胞与各个代谢通量之间的关系么？ 那为什么有的通量少，有的多，这个相关性的入选阈值是什么？

Reply: 三张图代表 EPC (2085), Myeloid\_cell (391), NK\_cytotoxic\_T (69)，不同的细胞的关联分析结果。关联分析做的是，各个细胞群的 TWAS 与 代谢通量之间的相关性 (假如这个细胞群有表达 TWAS 筛选出来的基因，并且有差异代谢通量的话)。看看总结的部分吧。

#### 3.10 细胞差异表达分析 (AA\_EPC)

* Content: 细胞差异表达分析
* lg Comment: 这里之后只进行了EPC？ 像T、HSC、B细胞这些无法进行还是没必要呢？ , PMID: 38145351    PMID: 37908861  PMID: 37064094 , 从以往AA的单细胞数据分析结果来看，这几类细胞对于疾病的进展也很重要。 如果做EPC的话，前期应该有什么标准或者结果提醒了其变化的重要性？ 比如差异最大，或者与代谢通量的相关性最强？ 这些内容可以在合适的地方补充下。

Reply: 在关联分析的部分已经筛选剩下： EPC (2085), Myeloid\_cell (391), NK\_cytotoxic\_T (69) 三种细胞了，而 EPC 找到的关联性最强，所以主要分析 EPC 了。

* Content: ，筛选差异表达基因
* lg Comment: 统计结果大概描述下？ 多少个上下调？

Reply: 已补充。

## 3.11 汇总: EPC\_DEG + EPC\_sig (EPC)

* Content: **基因集** (S100A6, PRELID1, TIMM10, …[n = 37]
* lg Comment: 如何筛选了37个基因？

Reply: 关联分析，EPC 热图中的 37 个基因，筛选的算法写在方法的部分 (原来的 2.5，写有公式的部分，结合了 TWAS 和 代谢通量预测。

## 3.12 Monocle3 拟时分析 (EPC)

* Content: Moran’s I test (Graph Test) 差异表达基因。仅分析这些基因。
* lg Comment: 这里又能说明什么主要结果？这11个基因对于AA的发生至关重要？

Reply: 说明这些基因在 EPC 细胞分化过程中会发生显著变化。

* Content: **Fig.** **25** AA dimension plot of expression level genes
* lg Comment: 通过这些图片可以表明什么?

Reply: Fig. 24 EPC Set1 genes in pseudotime 中 4 个基因的 UMAP 图的表达量，与 Fig. 24 互为参考。

* Content: 来自于Monocle3 拟时分析[Section: EPC]
* lg Comment: 这4个基因的由来？和之前的11个基因并列吗？ 而且都没有差异吗？ 或者怎么由11到4筛选的步骤和标准？

Reply: 根据 Fig. 24 EPC Set1 genes in pseudotime，筛选的，表达量较高，并且趋势变化明显。