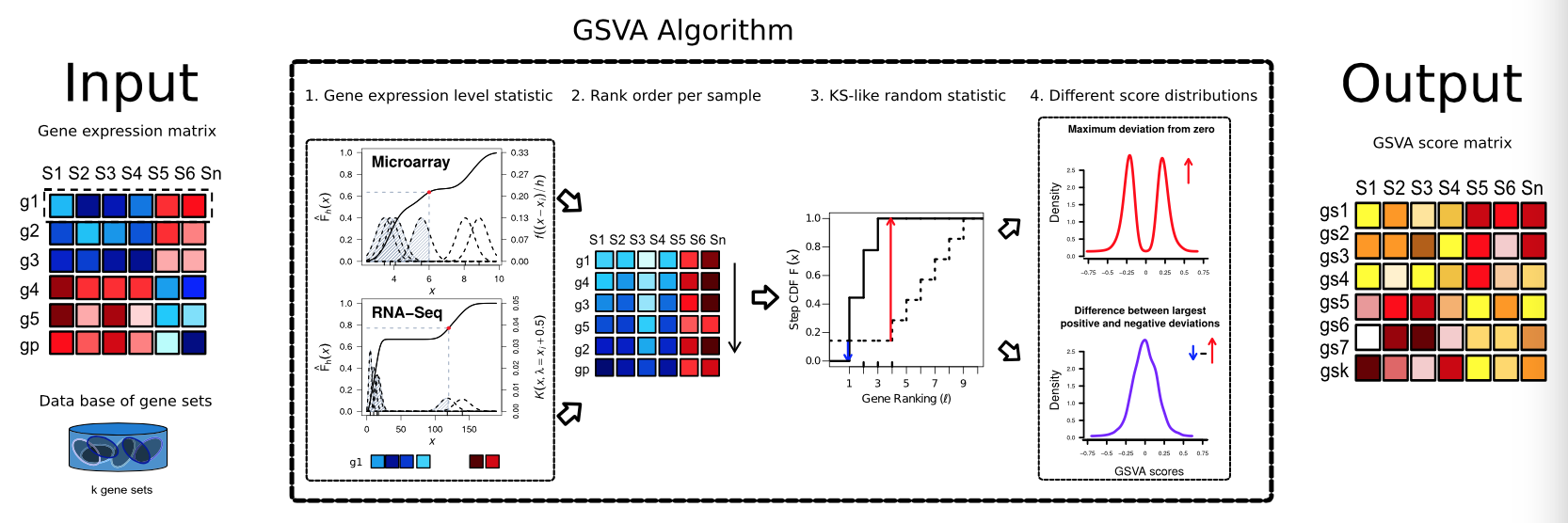
1. **GSVA应用**背景

GSE(Gene set enrichment)基因集富集分析通常使用预定义的基因集，如来自功能注释或先前实验的结果，将基因按照在两类样本中的差异表达程度排序，然后检验预先设定的基因集合是否在这个排序表的顶端或者底端富集。GSE从不同功能基因集(基因通路或者基因标签)的角度对基因表达谱整体层次上进行富集情况的研究，检测的是基因集合而不是单个基因的表达变化，因此可以包含细微的表达变化，从而得到更理想的生物解释结果。然而随着海量分子数据的积累，传统用于案例-对照研究的GSE分析方法已不能满足对大量异质性样本多样化分析(如TCGA的多表型、多层次大型队列研究)的需要，GSVA分析在这种情形下应运而生。GSVA(Gene Set Variation Analysis)基因集变异分析以一种非监督方式对样本群体的通路活性差异进行评估，把所有样本的基因表达量(芯片表达量或者RNA测序表达量)矩阵转化为所有样本的基因集富集分数矩阵，从而可以对任意标签的样本分组进行基因集富集程度差异评估。GSVA 分析由GSVA包(R包)的gsva()功能实现，gsva()功能除可以进行GSVA分析外，还可以进行ssGSEA(单样本基因集富集分析)、zscore分析、PLAGE分析。

1. **GSVA分析原理**

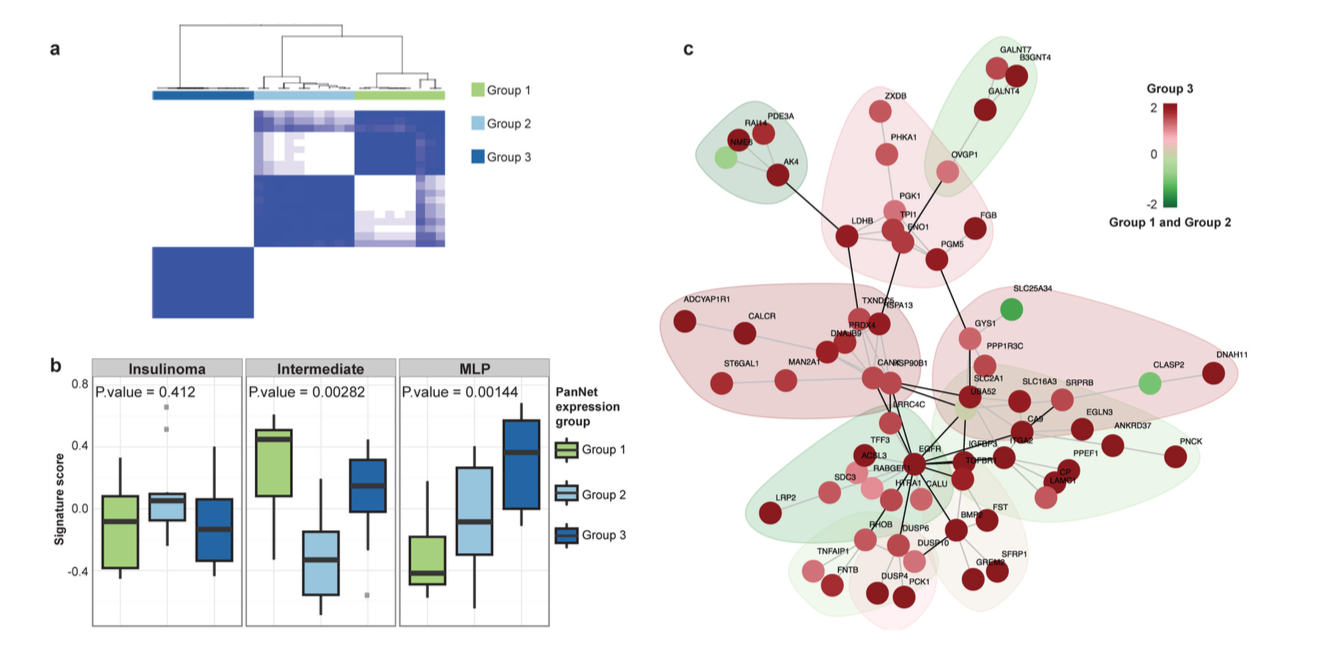
首先将表达量矩阵(行：样本；列：基因) 按照单个样本的基因表达量(经过ECDF核估算)由高到低进行排序，然后对单个样本的所有基因从头开始遍历，根据每个基因是否在预定义的目标基因集中出现(出现则富集统计量递增，不出现则富集统计量递减)累计计算得到每个基因位置的富集统计量，最终单个样本的目标基因集富集分数(ES, the enrichment score)有两种计算方法：(1)ES值直接是绝对值最大的富集统计量(类似于GSEA); (2)ES=｜最大的正富集统计量－最大的负富集统计量｜。其中第二种计算方法为默认，衡量的是目标基因集里的基因朝一个方向富集的程度(如高表达或低表达)。



**Figure 1 GSVA methods outline**

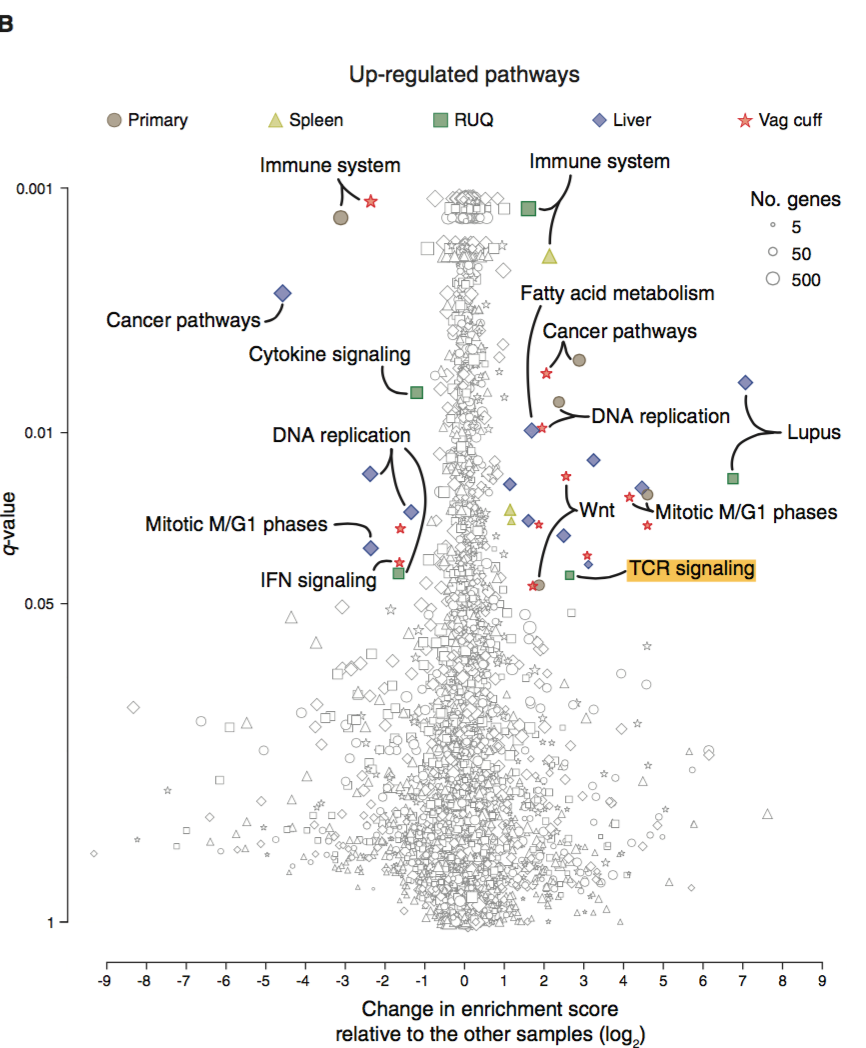
1. **GSVA应用简介**
2. **多病人多样本分析**

通过计算与已知分型有关的基因集(gene signature)在各个样本的gsva值，对新的样本分型与已有分型做出关联。如下图b所示，基于三个已知分型(Insulinoma/Intermediate/MLP)相关的基因集在所有新PanNET样本中的gsva分布结果，发现 Intermediate分型有关的基因集在Group1中显著富集，MLP分型有关的基因集在Group3中显著富集，即Group1分型与Intermediate分型有关，Group3分型与MLP分型有关。后又针对Group3与Group1+Group2的差异表达基因集进行了dNet功能互作基因网络构建(图c)，发现Group3(MLP亚型相关)与缺氧/缺氧诱导因子信号通路有关。



1. **单病人多样本分析**

通过GSVA包的单样本基因集富集分析功能(gsva(), method= ‘ssgsea’),  对一个stage IV级女性高级别浆液性卵巢癌病人的原发右上象限（RUQ）样本以及阴道口、肝、脾四处转移瘤样本进行分析发现，稳定/退化期的转移区域(肝样本和RUQ样本)中显著富集免疫相关的通路（该通路在vaginal-cuff样本中富集程度显著偏低）;系统性红斑狼疮通路在spleen和RUQ样本中显著高富集，TCR信号通路只倾向在RUQ样本中富集。而进展期的肿瘤（原发样本和阴道口样本）更富集的是癌症发生与增殖通路以及Wnt信号通路。



1. **参考文献**
2. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data[J]. BMC bioinformatics, 2013, 14(1): 7.
3. Scarpa A, Chang D K, Nones K, et al. Whole-genome landscape of pancreatic neuroendocrine tumours[J]. Nature, 2017, 543(7643): 65-71.
4. Jiménez-Sánchez A, Memon D, Pourpe S, et al. Heterogeneous tumor-immune microenvironments among differentially growing metastases in an ovarian cancer patient[J]. Cell, 2017, 170(5): 927-938. e20.