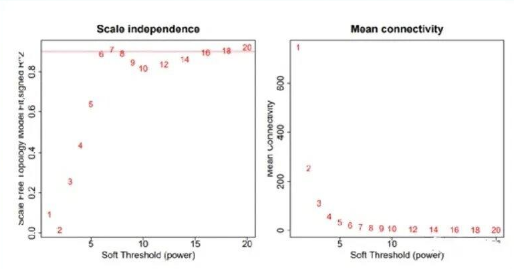
**WGCNA分析结果图怎么看**

**1. Power 图**

首先回顾一下power的意义:Power代表软阈值( Soft Threshold ) ,它是根据pick SoftThreshold函数计算出来的。在统计中它是一个检验研究工作有效性的指标，Power越大，研究结果越可靠。在WGCNA中是反应相关性数值是否或者接近无标度网络特征，具有重要意义。在WGCNA分析流程第2步中，需要通过阈值来识别module。首先需要筛选合适的阈值,使得样本变为一个无尺度分布，即少部分基因处于绝对优势的位置(表达量高)，大部分处于劣势(表达量低)。通过幂指数的方式，对基因之间的相关系数取一定的幂指数,最终确定一个合适的阈值。主要是power值的选择，通过pick SoftThreshold函数计算power值，再利用Power图确定合适的阈值。

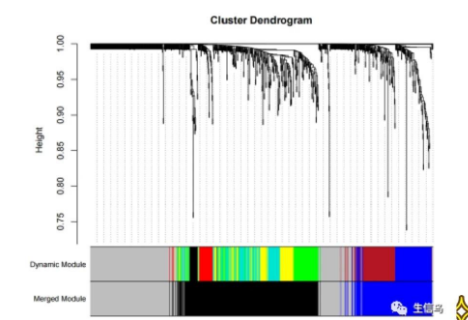


两张图的横坐标,均代表软阈值(power)的取值。左图的纵坐标是scale-free fit index，即signed R2，相关系数的平方越高,说明该网络越逼近无标度网络的分布。相关参考文献中有大量数据证明当signed R2大于0.85时，网络就已经符合无标度网络的分布。因此,WGCNA包中计算SoftThreshold的函数pick Soft Threshold中RsquaredCut 默认值为0.85，最佳的powers值保存在sft$powerEstimate。上图中,作者重新设定了一个阈值cex1=0.9，因此我们看到图中有一条红色的横线，表示第一个signed R2达到这条红线时的最佳powers值，此图中是6。

右图的纵坐标代表所有节点的平均连通数,越低越好(无尺度网络中，大部分节点的连通度较低,因此平均连通数越低越好）。通常我们对于power的选择主要参考左图，右图用于辅助参考。

**2. Cluster dendrogram图**

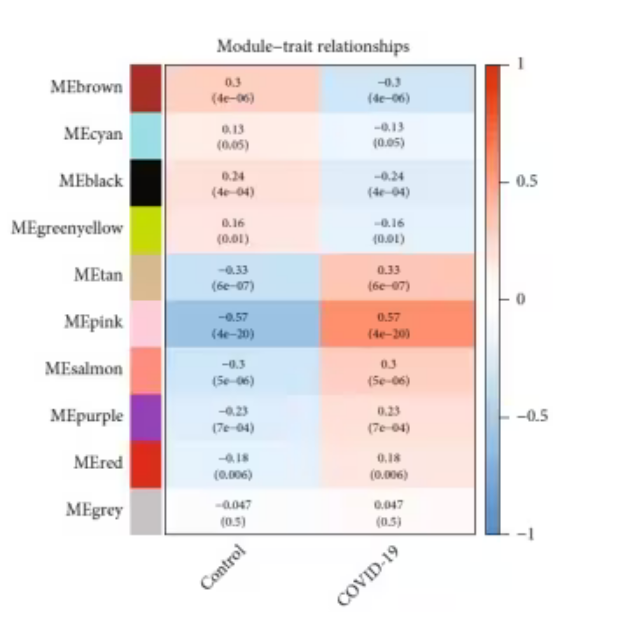
Cluster dendrogram(聚类树状图)，是用于层次聚类的树状图(也称为聚类树)。根据最佳软阈值和表达谱数据首先计算得到一个邻接矩阵adjacency，然后将邻接矩阵转换为TOM矩阵再计算一个距离矩阵dissTOM (ps:物体与物体之间进行聚类总是根据某一相似度指标来进行刻画的，那么这里的距离矩阵dissTOM就是这个指标，只不过是距离(非相似度))，这个距离矩阵就被用来画聚类树图。



我们可以把这个图分为两部分看:上半部分是基因的层次聚类树状图,下半部分是基因模块,也就是网络模块。上下对应,可以看到距离较近的基因(聚类到同一条分支)被划分到了同一模块。其中上半部分的height称为高度轴，一般表示的表示观测值和簇之间的距离。在当前的树状图中，y轴代表的是簇之间距离度量的值。如果小伙伴们看到上图中两个集群在高度x处合并，那么就表示这些簇之间的距离为x就是对y轴做垂线的值。再来看下图的下部分，Dynamic Module是动态的模块，与上面的height相对应，即不同的高对应不同的模块，就是用不同的颜色标记出来的部分，可以看到在动态模块中的颜色比较丰富,模块比较多。Merged Module是合并后的模块，如果仔细观察的话就会发现合并后的基因模块明显减少了，只有灰色、黑色和蓝色模块，也就说明聚类近的基因被划到了同一模块。

**3. Module-trait-relationship图**

先回顾下Module的意义，它本质是高度内连的基因集，模块内是高度相关的基因。在WGCNA分析中把基因聚类成模块后，对每个模块进行了功能富集分析、模块与性状进行关联分析和模块与样本进行关联分析，分别可以去查找基因的功能特征，筛选出表型与性状（比如肿瘤进展、预后良好、转移、xx疾病等）关联较高的模块和查找样本中特异表达基因。



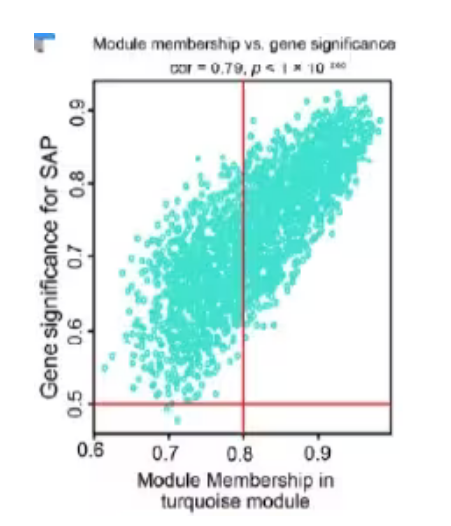
Module-trait-realtionship是模块与表型特征关联图谱，也是WGCNA中我们最关注的结果。横坐标是表型性状（trait），上图中显示有两个表型性状，一个是Control，另一个是COVID-19。纵坐标是对应模块，用每个模块的特征基因（eigengene）来表示这个模块（ps：回顾一下，eigengene代表的是每个模块的第一主成分）。在上图中有10个模块，分别用不同颜色表示。横坐标是表型性状和纵坐标是每个模块的相关性以及p-value值，黄的越深，越正相关；蓝色越深，越负相关。根据上图可以看出分析得到的10个模块中，MEpink（r = 0.57，P = 4E−20）、MEtan（r = 0.33，P = 6E−07）与COVID-19的相关性较高。

**4. MM-GS相关性图**

模块与指定性状的相关性我们已经分析得到了，但是一个模块通常有很多基因。我们如何找到最相关的基因呢？WGCNA通过GS和MM给出了一个合理的解决办法。

首先我们将模块中的每个基因与eigengenes进行相关性分析，得到的结果就是module membership（MM）。当结果越接近于0，则我们认为该基因与其所在的模块越不相关。而结果越接近于1或者-1，则我们认为该基因与模块基因高度正或负相关。此外，MM与模块内的连通性是高度相关的，所以高连通性的hub genes倾向于有更高的MM值。

模块内的基因，我们不仅想知道它与模块的相关性，还想知道它与模块对应的性状的相关性。于是WGCNA定义了一个新的变量GS (gene significance)，即计算模块内gene 与对应性状的显著性。



MM-GS相关性散点图代表的是MM与GS的相关性，从上图我们可以看到，GS和MM相关性很高（cor=0.79，p<1 × 10−200），这表明与性状高度相关的基因，通常也是该性状对应模块内比较重要的基因。因此当我们选择感兴趣的重要基因时，推荐选取散点图右上角部分的基因。