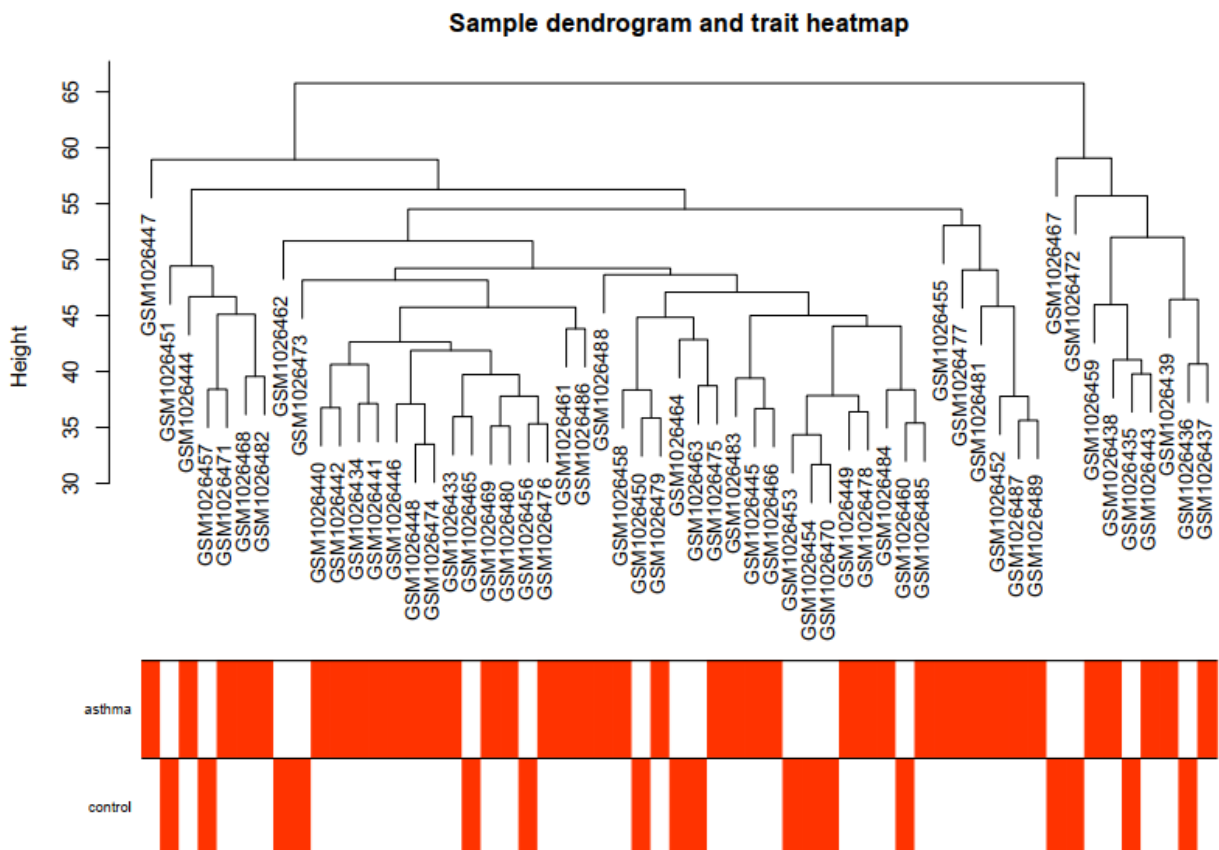


WCGNA 分析

WGCNA (weighted gene co-expression network analysis, 权重基因共表达网络分析) 是一种分析多个样本基因表达模式的分析方法, 可将表达模式相似的基因进行聚类, 并分析模块与特定性状或表型之间的关联关系, 因此在疾病以及其他性状与基因关联分析等方面的研究中被广泛应用。其算法首先假定基因网络服从无尺度分布, 并定义基因共表达相关矩阵、基因网络形成的邻接函数, 然后计算不同节点的相异系数, 并据此构建分层聚类树, 该聚类树的不同分支代表不同的基因模块(module), 模块内基因共表达程度高, 而分属不同模块的基因共表达程度低。基于模块与性状之间相关性分析, 可以得知哪些基因可能与所关注的性状相关。

分析软件: R, WGCNA

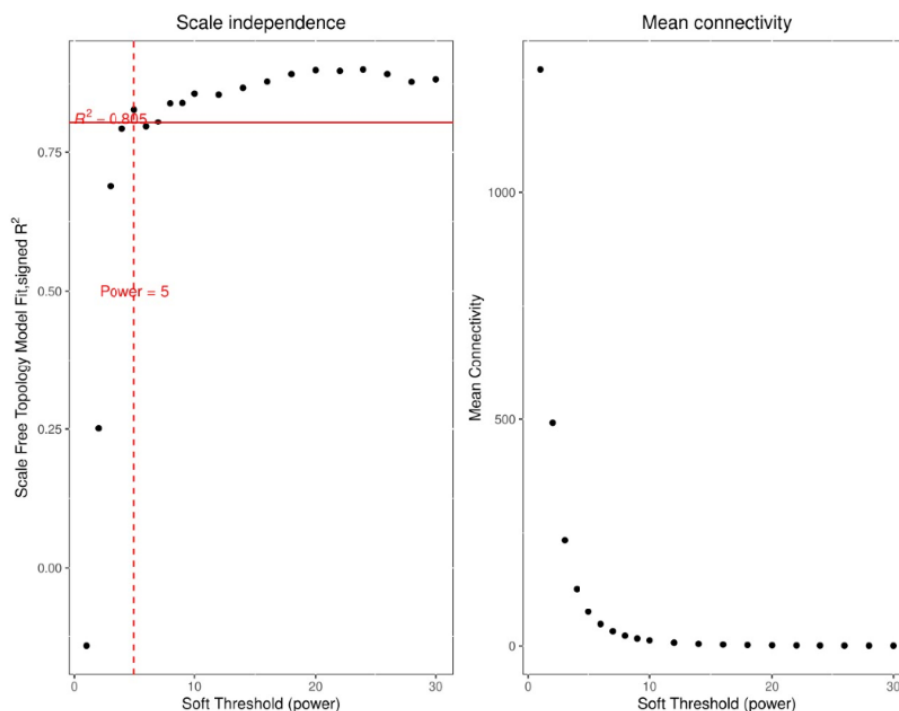
1. 样本聚类树



每个小分支对应的 GSM 加数字, 代表具体的一个样本, 分支越近得样本表明其关系越接近。图片下方为该项目所关注的分组对应的样本情况。

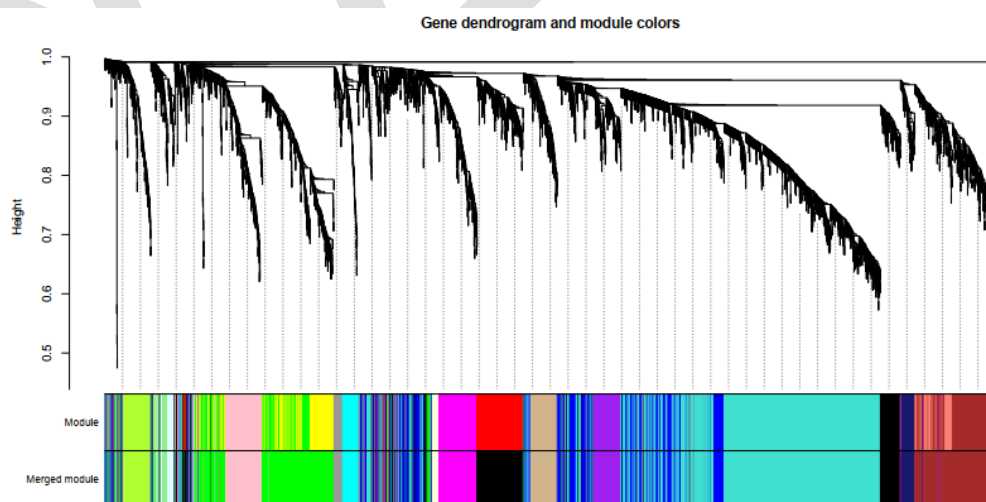
但是所得结果一般与所关注的性状分组并不是完全一致的, 所以参考意义不大, 一般不放进文章中。

2. 筛选合适的阈值



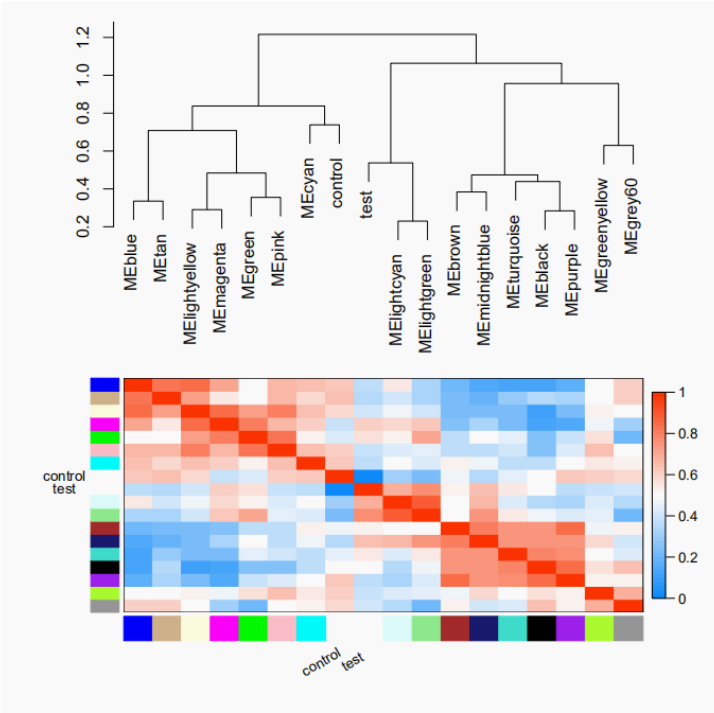
选一个合适的值,使得构建的网络更符合无标度网络特征,即使得 WGCNA 分析的结果更为可靠。一般 power 在 30 以内,要存在一个 power 值可以使无标度网络图谱结构 R^2 达到 0.8 (相关性系数的平方越高,说明该网络越逼近无网络尺度的分布),平均连接度在 100 以内,否则则说明部分样品与其他样品差别太大。这张图一般是放在文章中的。

3. 基因模块的可视化



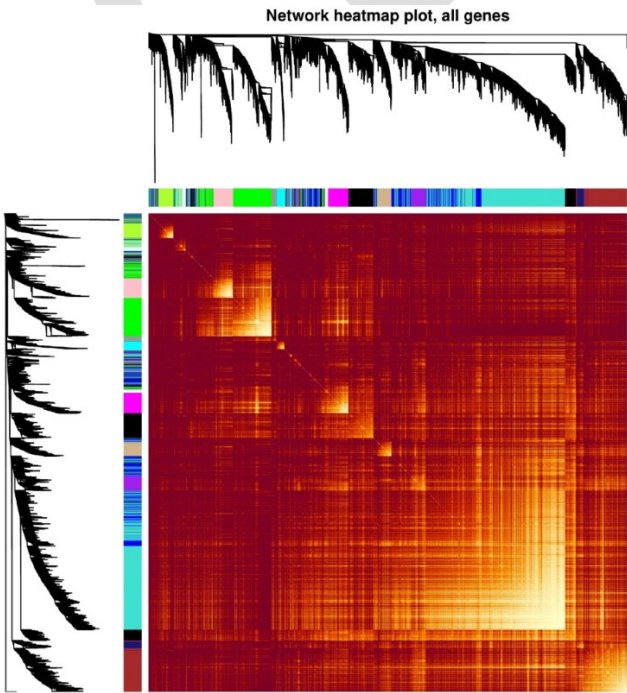
根据基因间表达量的相关性构建聚类树,并划分模块。如果某些基因在一个生理过程或不同组织中总是具有类似的表达变化,那么这些基因在功能上可能相关,可以把他们定义为一个模块(module)。对于上半部分的树图,纵向距离代表两个节点间(基因间)的距离,横向距离无意义。一般放在文章里。

4. 模块间相关性热图



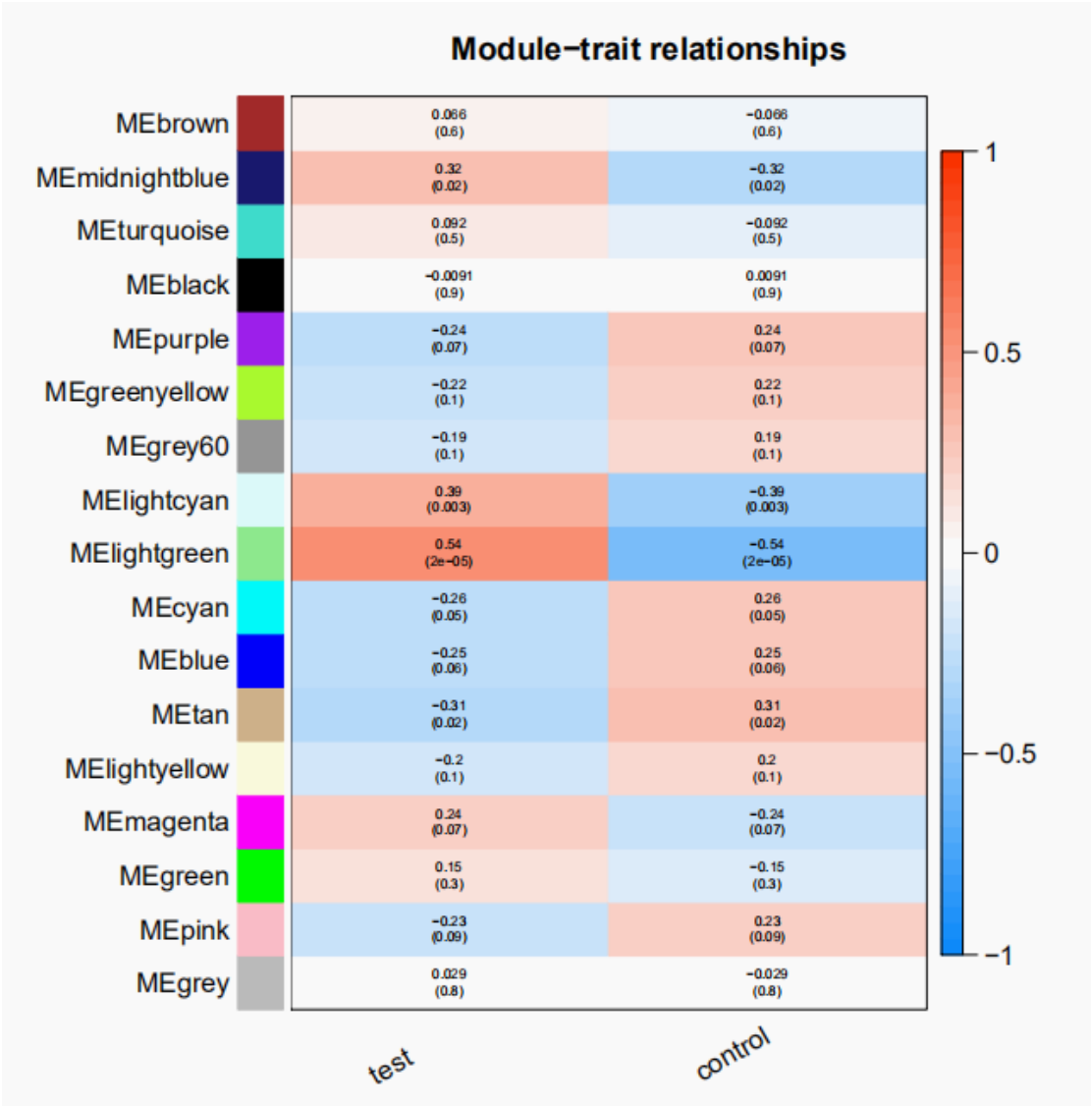
用来展示模块之间关系的远近。图可分为两部分，上部分根据模块特征值（eigengene）对模块进行聚类，纵坐标代表节点的相异程度。下半部分图形中每一行和列代表一个模块，方块颜色越红，相关性越强；方块颜色越蓝，负相关性越强。一般不放在文章里。

5. 模块基因聚类热图



基于层次聚类的结果，绘制层次聚类树状图，每一树状图代表一个模块，每一个分支代表一个基因，每个点的颜色越浅（红→黄→白）代表行和列对应的两个基因间的连通性越强。一般不放在文章里。

6. 模块与性状的相关性图



如果某模块与样品的相关性显著高于其它模块，说明这一个模块可能与该样品存在最强的关联关系。横坐标为性状，纵坐标为模块，每个格子的数字代表模块与样品的相关性，该数值越接近 1，表示模块与样品正相关性越强；越接近-1，表示模块与样品负相关性越强。括号里的数字代表显著性 P value，该数值越小，表示显著性越强。这张图是必放在文章里的。