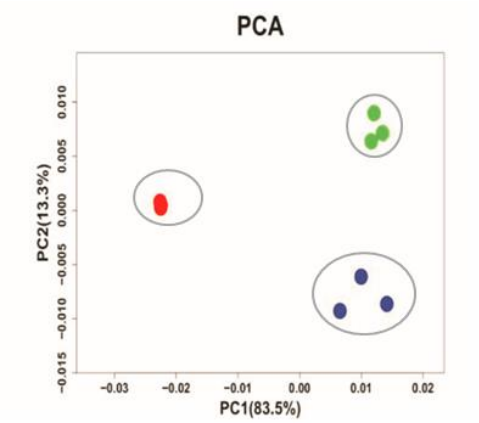
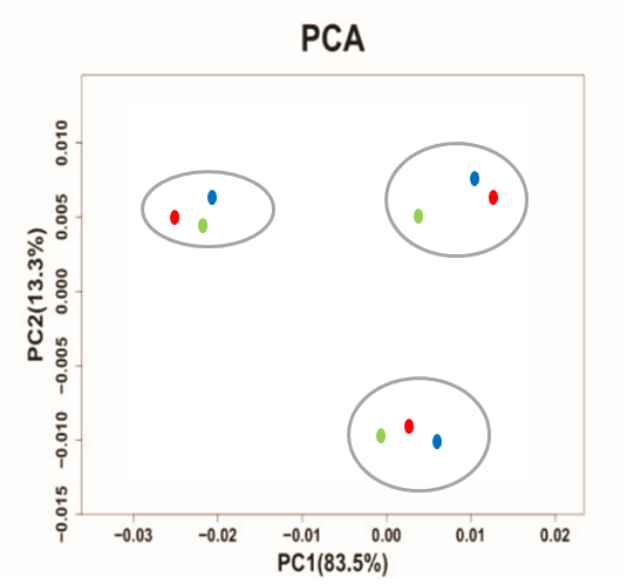
RNA 测序作为一种很灵敏的 RNA 表达量定量技术，很多因素会对实验精度产生影响。批次效应一种常见的误差来源。我们先用一个案例来解释批次效应可能对结果产生的影响。

这是一例研究某种处理对大鼠细胞系 RNA 表达影响的案例。红、蓝、绿分别代表三种处理条件，每种处理三个生物学重复（红：对照；蓝：处理 1；绿：处理 2）。理论上，9 个样本预期的 PCA 聚类结果应该如下图，每种处理方式聚为一类。



但实际的 PCA 聚类结果却如下图。在每一个聚类中，分别含有各一个不同处理的样本。结合老师的实验背景，我们发现由于实验仪器的限制，整个实验的 9 个样本是分三个批次完成的，每个批次各含有不同处理的样本一个（即每次处理各含一个红，蓝，绿处理）。由于实验批次间存在差异（本次实验，三个批次的细胞系传代数存在差异），导致最终 PCA 聚类结果中，样本的表达模式按照批次聚类，而不是实验处理聚类。即这样的实验处理，无形中扩大了处理组内的差异（噪音），处理组间的差异显著程度降低，导致某些与处理相关的差异基因无法被有效检测。



这类由实验条件不一致导致的批次效应在高通量测序项目中是普遍存在的。样本处理的背景条件不同，RNA 样本提取、建库、测序的条件不同（实验方法，操作人员等因素）等，都会引入批次效应，最后导致与实验无关的误差被放大，降低了实验准确度。类似的，如果从网络数据库（例如 GEO 数据库）下载数据，我们也会发现不同 project 的数据间也存在明显的批次效应。

虽然批次效应对 RNA-seq 的影响要小于表达谱芯片，而且批次效应通过数据预处理可以被部分校正（例如以上的大鼠细胞系项目，通过 paired samples t-test 可以部分校正）。但这样的干扰，依然是我们要尽可能要避免的。因此，在 RNA-seq 等对实验处理比较敏感的实验设计中，我们建议老师通过对实验的预先设计和控制，尽可能将与实验处理无关的背景条件控制在一个水平，减少批次效应对结果的影响。