

测序说明

1. MGI平台一次上机2张芯片，1张芯片4条lane(L01/L02/L03/L04)，每条lane产出360-380M reads。
2. MGI的barcode必须成套使用，目前有501-508,509-516共两套，测序时，每条lane必须保证至少有一套barcode，可以多套一起使用，lane与lane之间的barcode可以重合，二者不关联。例如，如果一条lane测一个样本，则该样本的barcode可以是501-508或者509-516，如果一条lane测两个样本，则该lane的barcode组合可以是501-504和505-508(即加起来是一套barcode)。依此类推，一条lane最多可以测16个样本。
3. 拿到下机信息时，每个样本都有对应的芯片号、lane号以及barcode号，需要根据这些信息来找样本的下机数据。每个芯片-lane-barcode的组合下都会有一批数据，如果某样本对应的是501-508共8个barcode，则需要把这8个barcode的数据合并在一起，作为该样本的数据进行分析。如果某样本对应的是2个lane号8个barcode号，则需要把2条lane下8个barcode的数据(共16批数据)合并到一起作为该样本的数据进行分析，依此类推。每个样本对应的芯片号、lane号和barcode可能都不是单一的，原则是将涉及到该样本所有数据都合并。
4. 所有的下机数据都是按照barcode拆分好的fq文件,每个barcode有reads1和reads2两个fq文件。fq文件以gzip的格式存放，文件名为“芯片号_Lane_barcode_1.fq.gz”和“芯片号_Lane_barcode_2.fq.gz”

示例

送样形式	样本编号	样本名称	待测数据量	芯片号	Lane 号	barcode	下机数据量 bases/G
DNA	D1803012-TH	Xinjiang	60/150	V100007345	L01	501-508	58.698
DNA	D1804662-JQ	Bt-Drone-2	15	V100007345	L01	509	18.267
DNA	D1804420-ZJ-2	LX5	912	V100007345	L02	509-516	75.525
DNA	D1804420-ZJ-1	LX5	912	V100007359	L03/L04	509-516	41.156
DNA	D1810146A-GQ	D26	1	V100007345	L02	501	0.518
DNA	D1804260-XH	E1	150	V100007345	L03/L04	501-508	151.880

数据交付、释放的时候项目管理会提供上面的表格（可能会有差异）。以表格最下方的“E1”这个样本为例，可以从表格中获取到的信息有：

- 该数据是测序的芯片号是V100007345
- 该数据测了两条lane，分别是L03和L04
- 样本上机的时候选用的barcode是501-508这8个barcode

根据上述信息可以推测，这个样品在这一批测序中总共可以产生32个文件，文件名列

表:

V100007345_L03_501_1.fq.gz
V100007345_L03_501_2.fq.gz
V100007345_L03_502_1.fq.gz
V100007345_L03_502_2.fq.gz
V100007345_L03_503_1.fq.gz
V100007345_L03_503_2.fq.gz
V100007345_L03_504_1.fq.gz
V100007345_L03_504_2.fq.gz
V100007345_L03_505_1.fq.gz
V100007345_L03_505_2.fq.gz
V100007345_L03_506_1.fq.gz
V100007345_L03_506_2.fq.gz
V100007345_L03_507_1.fq.gz
V100007345_L03_507_2.fq.gz
V100007345_L03_508_1.fq.gz
V100007345_L03_508_2.fq.gz
V100007345_L04_501_1.fq.gz
V100007345_L04_501_2.fq.gz
V100007345_L04_502_1.fq.gz
V100007345_L04_502_2.fq.gz
V100007345_L04_503_1.fq.gz
V100007345_L04_503_2.fq.gz
V100007345_L04_504_1.fq.gz
V100007345_L04_504_2.fq.gz
V100007345_L04_505_1.fq.gz
V100007345_L04_505_2.fq.gz
V100007345_L04_506_1.fq.gz
V100007345_L04_506_2.fq.gz
V100007345_L04_507_1.fq.gz
V100007345_L04_507_2.fq.gz
V100007345_L04_508_1.fq.gz
V100007345_L04_508_2.fq.gz

确定是同一个样品的数据，在分析的时候可以合并后处理。