中文

DNA 提取

利用组织样品 DNA 提取试剂盒，按照试剂盒使用说明书步骤，提取样品的基因组 DNA，采用 Nanodrop ND2000 检测 DNA 的纯度(OD260／280 比值>1．80)，采用 Qubit 2．0 精确定量 DNA 浓度，利用 1％的琼脂糖凝胶电泳检测提取 DNA 的质量，分析 DNA 降解程度以及是否存在 RNA 或蛋白污染。

文库构建

样品检测合格后，加入一定比例的阴性对照(1ambda DNA)，首先使用 Covans S220 将基因组 DNA 随机打断至 200—300bp；对打断后的 DNA 片段进行末端修复、加 A 尾，并连接上所有胞嘧啶均经过甲基化修饰的测序接头；随后进行 Bisulfite 处理(采用 EZ DNAMethylatlon Gold K1t，Zymo Research)，经过处理，未发生甲基化的 C 变成 U(PCR 扩增后变为 T)，而甲基化的 C 保持不变，最后进行 PCR 扩增，最终的 DNA 文库。文库构建流程图如下：

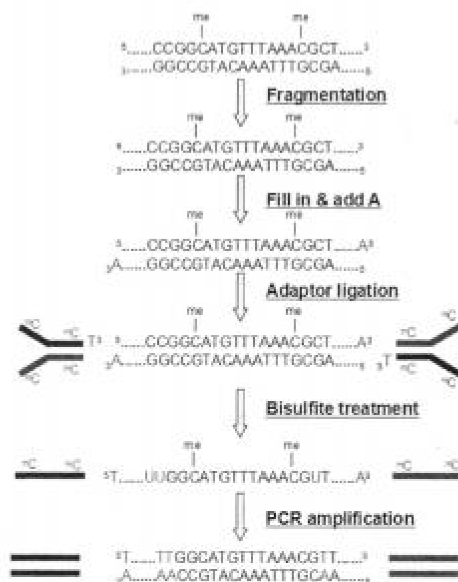

图 3-1 文库构建流程图

文库构建完成后，先使用 Qubit2．0 进行初步定量，稀释文库至 1ng／gL，随后使用 Agilent 2100 对文库的插入片段长度进行检测，符合预期后，使用 qPCR 方法对文库的有效浓度进行准确定量(文库有效浓度>2nM)，以保证文库质量。

库检合格后，把不同文库按照有效浓度及目标下机数据量的需求 pooling 后进行 Illumina Hiseq 2500 测序。测序的基本原理是边合成边测序(Sequencing by Synthesis)。在测序的 flow cell 中加入四种荧光标记的 dNTP、DNA 聚合酶以及接头引物进行扩增，在每一个测序簇延伸互补链

时，每加入一个被荧光标记的 dNTP 就能释放出相对应的荧光，测序仪通过捕获荧光信号，并通过计算机软件将光信号转化为测序峰，从而获得待测片段的序列信息。

数据预处理

随着高通量测序技术的快速发展，通过测序可以获得海量的数据信息，如何从得到的数据中获取合格的数据是信息分析的基础。因此对下机数据进行质量控制(QC)是数据分析的首项内容。FastQC 是目前常用的数据质量评估软件。采用 FastQC 对下机后的数据(raw reads)质量进行基本的统计。

数据过滤的主要目的是去除低质量的数据，保证 clean data 的质量。我们采用 trim 的方式截去测序数据的测序接头和低质量片段，从而较高效地利用测序数据，后续分析都基于 clean data。在进行原始数据 Trimming 时我们使用 Tnmmomatic 软件(Bolger et al 2014)，其数据处理的步骤具体如下：

(1)截去低质量 reads，使用滑动窗口的方式，4 个碱基为一个窗口，若该窗口的平均碱基质量值低于 1 5，则从该处截去 reads，参数选择：SLIDINGWINDOW：4：1 5；

(2)截去 reads 首尾质量低于 3 或者含 N(N 表示无法确定碱基信息)的 reads，参数选择‘LEADING．3，TRAILING．3；

(3)截去接头污染的 reads，使用两种模式去除接头：1．sample alignment mode：seed 与接头序列比对分值达到 7(约 12bp)；2．pahndrome mode：当 readl 和 read2 的重叠区碱基评分大于 30 时，截去 seed 部分序列。参数选择：ILLUMINACLIP：adaptm．fa：2：30：7：1：true；

(4)舍弃修剪后短于 36nt 的 reads；

(5)舍弃不能形成 paired 的 reads。

差异分析

本研究采用 DSS(Feng et al 2014，Park and Wu 2016)软件分析鉴定 DMS&DMR。对于差异甲基化区域分析考虑因素：(1)甲基化位点的距离相关性(spatml correlation)：适当的使用临近的甲基化位点的信息可以提高每个位点甲基化水平的评估，提高 DMR 的鉴定；使用 smoothing 的方式能够鉴定相对较长的 DMRs。(2)位点测序深度(read depth ofthe sites)：位点的测序深度可以提高精确的信息，提高 DMR 鉴定统计值。(3)生物学重复异质性(the variance among biological replicates)：忽略生物学重复的差异会导致假阳性，DSS 软件基于 beta-binomial distribution，考虑到生物学重复问的异质性；对于无生物学重复样本分析时，DSS 使用 smoothing 后临近的位

点作为生物学重复位点进行分析。

英文

Genomic DNA extraction

Frozen liver tissues were quickly pulverized with BioPulverizer (Bio Spec Products Inc.) on dry ice. DNA was extracted with Allprep DNA/RNA/Protein Mini Kit (Qiagen) according to the manufacturer's protocol. DNA was further purified by phenol–chloroform extraction followed by ethanol precipitation.

Whole-genome bisulfite sequencing library preparation

DNA (1 µg) was spiked with 1 ng unmethylated lamda DNA (Promega), fragmented (average size; 300 bp), end-repaired, A-tailed, and adapter-ligated using Truseq DNA sample prep kit (Illumina) according to the manufacturer's protocol. Adapter-ligated DNA was gel isolated, (2% agarose gel, DNA ranged from 400 to 500 bp) and recovered using a QIAquick gel extraction kit (Qiagen). After clean up with AMPure XP beads, bisulfite conversion was performed using EpiTect Bisulfite kit (Qiagen) with the following thermal cycles, 95 °C 5 min, 60 °C 25 min, 95 °C 5 min, 60 °C 85 min, 95 °C 5 min, 60 °C 175 min, 95 °C 5 min, 60 °C 180 min. After clean up with AMPure XP beads, bisulfite converted DNA was amplified with PfuTurbo Cx Hotstart DNA Polymerase with following thermal cycles, 95 °C 5 min, 98 °C 30 s, 12 cycles of (98 °C 10 s, 65 °C 30 s, 72 °C 30 s), 72 °C 5 min. DNA was cleaned with AMpure XP beads, and stored at −30 °C until use. Sequencing was performed in HiSeq2000 using PE100 base format.

Whole-genome bisulfite sequencing data processing

General quality control checks were performed with FastQC v0.8.0 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Each dataset was filtered for average base quality score (>20).Filtered datasets were aligned to a reference genome using Bismark v0.7.8 (parameters-X 10000 --non_bs_mm -n 2 -l 50 -e 70 --chunkmbs 1024)41, using Bowtie v0.12.842 as the underlying alignment tool. The reference genome index contained the genome sequence of enterobacteria phage λ (NC_001416.1) in addition to all chromosomes of the mm9 assembly (NCBI 37). Mappings for all datasets generated from the same library were merged, and duplicates removed

via the Bismark deduplication tool (deduplicate_bismark_alignment_output.pl). Mapped reads were then separated by genome (mm9 or phage λ) and by source strand (plus or minus). The first four and last one base of each read2 in all read pairs was clipped due to positional methylation bias, and any redundant mapped bases due to overlapping mates from the same read pair were trimmed to avoid bias in quantification of methylation status. Finally, the alignments for multiple libraries from the same animal were merged. Read pairs mapped to phage λ were used as a QC assessment to confirm that the observed bisulfite conversion rate was >99%. Read pairs mapped to the mm9 reference genome were used for downstream analysis

DMR detection

Using the DSS R package v2.15.0, DMCs were identified by DSS with the callDML function (default parameters), and DMRs were identified with the callDMR function (all other parameters default). DMC and DMR calls were also made via Metilene v.0.2-6 (all other parameters default), with a p value threshold of **0.05 and mean methylation difference of 0.2** for DMCs, with a p value threshold of **0.05 and mean methylation difference of 0.1** for DMRs. All DMR calls from both tools were subject to additional filters, as described below. DMRs were required to contain at least 3 validated CpG sites and have a minimum length for 50bp .