Kit Curtius, 4 May 2017

Bioinformatics pipeline for low pass whole genome sequencing (LP-WGS) copy number analysis

Trim adapter read-through (skewer)

BASH Script

Quality check of sequencing (FastQC)

Align to hg19 (BWA-MEM)

Convert SAM to unsorted BAM,
filter out low mapping quality scores
(MAPQ value < 37, Samtools)

sort BAM/ Mark duplicates (Picard),
remove duplicates (Samtools),
index final BAM (Picard)

R Script (QDNAseq 1.10)

Quality check of BAM file (BamQC),
Calculate coverage statistics (GATK)

Arrange BAM raw reads in 500 kbp bins,
filter, correct for GC content & mappability,
normalize, smooth ouliters

Segment copy number profile (DNAcopy)
by circular binary segmentation (CBS)

Call copy number aberrations
using a six state mixture model (CGHcall)

Calculate % of bins with abnormal
copy number across the binned genome