

The Cancer Genome Atlas Pilot Project
Memorial Sloan Kettering Cancer Center
Normalized Data Description

Platform: Agilent 244K Array

Normalized data

Although the Feature Extraction program from Agilent does normalize the two channels we have found that this simple normalization misses an effect related to the local GC content near the probe region on the genome. These genomic-based artifacts can cause problems in the downstream analysis so we have implemented a more complex normalization procedure, which removes this effect. After our GC-normalization we take the LogRatio (which we convert to the usual log base 2 convention) for each sample in a given batch and create one tab delimited text file, which has each sample in a different column. The initial columns of the file contain the probe ids, chr, and genomic location of the probe and a pseudo-genome wide coordinate to facilitate graphing and analysis of the data. Finally we will filter out any samples that fail the QC metrics that have been established. Level 2 will only contain samples that pass QC.