

The Cancer Genome Atlas Pilot Project
Stanford University School of Medicine
Normalized Data Description

Platform: Illumina Infinium 550K SNP Array

Normalized Data

The Illumina Beadstudio software will be used to generate normalized intensity values for each allele of every SNP, the logR (log of total intensity, summed over both alleles) and B allele frequency values for each SNP, as well the differences in logR and B allele frequency between each pair of tumor and normal samples. Such pairwise differences will be the basis of inferring copy number changes.

We will also provide additional normalized logR and B allele frequency data files that result from our custom normalization procedures. We developed these procedures to correct for additional sources of noise or bias, such as sample-specific, bead pool-specific, and SNP-specific effects in the intensity data that have not been adequately removed by Illumina's genotyping software.