

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

**Platform:** Illumina Infinium 550K SNP Array

**Raw Data**

Among the raw data files that we will provide are the IDAT files, which are the binary data files produced by the Illumina scanner, one for each color channel of each sample, that contain the average intensity data for each SNP averaged over >20 beads. These files can be read by the Illumina BeadStudio analysis software to produce all the other data files.

We will provide all the genotype calls for each sample, as well as a cluster file that defines the genotype cluster positions for each SNP that we generated from the TCGA samples. We will also export from the Illumina Beadstudio software the raw intensity values and genotyping quality scores.