



## README: 500K Copy Number Sample Data Sets

Updated 7/26/06

### 500K Sample Data Set

The sample data set consists of 9 Tumor/Normal Pairs derived from human cancer cell lines and X Chromosome titration set (3X, 4X, and 5X). The X Chromosome titration data set contains 4 replicates of each DNA sample (total of 12 per enzyme). The total number of unique DNA samples is 21 and the total number of CEL files is 60 (30 NSP and 30 STY).

- The Tumor/Normal pairs were obtained from ATCC ([www.atcc.org](http://www.atcc.org)).
- The Chromosome X samples were obtained from Coriell Institute Cell Repository (<http://ccr.coriell.org/nigms/>).
- DNA samples were subjected to the 500K Mapping Assay according to the protocol (<http://www.affymetrix.com/products/arrays/specific/500k.affx>).
- There are 48 HapMap samples which can be used as additional Normals ([http://www.affymetrix.com/support/technical/sample\\_data/500k\\_data.affx](http://www.affymetrix.com/support/technical/sample_data/500k_data.affx)).

### Paired - Normal Samples

ID	Type	Description	Source
CRL-2325D	Normal	B lymphoblast; peripheral blood; EBV transformed	ATCC
CRL-2324D	Tumor	primary ductal carcinoma TNM stage I; grade 3	
CRL-5957D	Normal	B lymphoblast; peripheral blood; EBV transformed	
CRL-5868D	Tumor	adenocarcinoma; stage 2	
CCL-256.1D	Normal	B lymphoblast; peripheral blood; EBV transformed	
CCL-256D	Tumor	non-small cell lung carcinoma	
CRL-2319D	Normal	B lymphoblast; peripheral blood; EBV transformed	
CRL-2320D	Tumor	metastatic lymph node from ductal carcinoma TNM stage IIA; grade 3	
CRL-2362D	Normal	B lymphoblast; peripheral blood; EBV transformed	
CRL-2321D	Tumor	primary ductal carcinoma; no lymph node metastases	
CRL-2337D	Normal	B lymphoblast; peripheral blood; EBV transformed	
CRL-2336D	Tumor	primary ductal carcinoma TNM stage IIB; grade 3	
CRL-2339D	Normal	B lymphoblast; peripheral blood; EBV transformed	
CRL-2338D	Tumor	primary ductal carcinoma TNM stage IIA; grade 3	
CRL-2341D	Normal	B lymphoblast; peripheral blood; EBV transformed	
CRL-2340D	Tumor	primary ductal carcinoma TNM stage IIIA; grade 2	
CRL-2346D	Normal	B lymphoblast; peripheral blood; EBV transformed	
CRL-2314D	Tumor	primary ductal carcinoma TNM stage IIB; grade 3	

### Chromosome X Titration Set

ID	Type	Description	Source
NA04626	3X	Aneuploid Chromosome - Trisomy - 3X	Coriell Institute
NA01416	4X	Aneuploid Chromosome - Non-Trisomic - 4X	
NA06061	5X	Aneuploid Chromosome - Non-Trisomic - 5X	

### Download Information

The data consists of .CEL files in Data Transfer Tool (DTT 1.1) Archive files which can be restored in GCOS. You need to download all DTT files and then launch Data Transfer Tool to restore the

data to your GCOS system. In addition, there is a text file (Sample\_Attribute\_File.txt) which contains sample ID and disease state information about the samples.

- Download DTT 1.1 from <http://www.affymetrix.com/products/software/specific/dtt.affx>
- DTT Archive files:
  - Tumor\_Normal\_NSP.DTT (650 MB)
  - Tumor\_Normal\_NSP.D02 (650 MB)
  - Tumor\_Normal\_NSP.D03 (650 MB)
  - Tumor\_Normal\_NSP.D04 (78 MB)
  - Chromosome\_X\_NSP.DTT (650 MB)
  - Chromosome\_X\_NSP.D02 (102 MB)
  - Tumor\_Normal\_STY.DTT (650 MB)
  - Tumor\_Normal\_STY.D02 (650 MB)
  - Tumor\_Normal\_STY.D03 (650 MB)
  - Tumor\_Normal\_STY.D04 (78 MB)
  - Chromosome\_X\_STY.DTT (650 MB)
  - Chromosome\_X\_STY.D02 (102 MB)

Note: The .D0x notation indicates that the DTT archive file was broken up into multiple files due to the large size. You need to download all DTT files (e.g. Chromosome\_X\_NSP.DTT and Chromosome\_X\_NSP.D02) and then launch Data Transfer Tool to restore the chromosome X NSP data to your GCOS system.

### **Release Notes**

When analyzing tumor samples using the Mapping 500K Array Set, it is important to note that depending on the extent of chromosomal rearrangements present in the sample, it is possible that deletions or amplifications may introduce modifications to:

- **Overall call rates and heterozygosity:** The genotype calling algorithm used in the Mapping 500K product expects heterozygous genotype calls to see an equal balance in the signals for each allele. In the presence of copy number change the signal balance for heterozygotes is typically skewed and the genotyping algorithm starts to find it more difficult to accurately identify heterozygous genotypes. In the case of less extreme imbalance true heterozygotes tend to be called N and in the case of more extreme imbalance true heterozygotes can be called homozygote. There are symptoms which can be observed when this happens. Overall call rate and heterozygosity rate declines.
- **Gender calls:** The gender estimate is based on chromosome X SNPs, so if there are amplifications or deletions in the X chromosome it can cause errors in the gender call. For example if much of one X chromosome is deleted in a female the gender call may be male.
- **Common SNP calls:** There are 50 SNPs tiled on both the Nsp and the Sty arrays, for the purpose of verifying sample matches. If there are any changes in chromosome copy number in the vicinity of these SNPs it is possible that the changes may induce mismatches among the common SNPs.

### **Additional Information and Support**

Please refer to the complete CNAT manual for additional information. For support information, please contact your FAS or local support team <http://www.affymetrix.com/support/index.affx>  
United States / Canada: 888-DNA-CHIP (888-362-2447)  
Europe: +44 (0) 1628 552550  
Japan: +81 3-5730-8222