Supplementary Material for:

A novel SNP analysis method to detect copy number alterations with an unbiased reference signal directly from tumor samples

Definitions

Log2 ratio: if I1 is an intensity signal from array 1 and I2 is an intensity signal from array 2 then log2 = log2(I1/I2).

Standard deviation (SD): the square root of a variance; the variance is a sum of squared differences between the signal value at a given point and a signal mean.

Unpaired data set: the tumor (test) and normal samples were collected from different subjects. In this case the raw copy number (CN) for each tumor (test) sample is computed using the (log2) ratio between a given tumor (test) sample intensity and the average intensity of all normal samples.

Paired data set: the tumor (test) and normal samples were acquired from the same subject. In this case the raw CN for each tumor (test) sample is computed using the (log2) ratio between given tumor (test) sample intensity and the normal sample intensity from the same subject.

Segmentation: assigning a constant value to the subset of adjacent signal data points based on some data points with similar criteria. The assigned value could be a mean or median (CBS, GADA) or the copy number (HMM).

Affymetrix 250K Sty Assay for Formalin Fixed, Paraffin Embedded Tissue (FFPET) Samples

In most cases, genomic DNA samples extracted from formalin fixed, paraffin embedded tissues (FFPET) are more fragmented than genomic DNA extracted from fresh frozen tissue counterparts (Lyons-Weiler, et al., 2008). Therefore, we use the Sty assay to process FFPET samples, as the 250K Sty assay preferentially cuts a greater amount of shorter fragment length SNP's when compared to the 250K Nsp assay. The number of PCR reactions was increased since FFPET samples consistently under-perform when compared to frozen tissue with regard to PCR yield (not shown). We routinely run 4 replicates of each FFPET sample.

DNA was extracted from FFPET samples as previously described (Lyons-Weiler, et al., 2008). One microgram of whole genomic DNA was digested according to a modified restriction digest protocol (Lyons-Weiler, et al., 2008) with Sty I and restriction digest buffer 3 (New England Biolabs, MA) to begin the Affymetrix 250K assay. In short, an excess of both the Sty I enzyme and buffer are used to digest the excess genomic DNA. Upon DNA digestion, the sample is concentrated to 20 µl. The digested DNA is then ligated according to the manufacturer's instructions (Affymetrix 500K assay protocol). The digested, ligated DNA is then diluted and each sample undergoes PCR using 4 PCR reactions instead of 3 per sample (Lyons-Weiler, et al., 2008). PCR purification was carried out using Agencourt AmpPure beads (Beckman Coulter) (Affymetrix® Cytogenetics Copy Number Assay User Guide). Seventy to ninety micrograms of labeled, fragmented DNA was then loaded onto the Affymetrix Gene Chip 250K Sty arrays. Samples were hybridized according to the manufacturers instructions for 16 hours at 48°C with 60rpm rotation in the Gene Chip hybridization oven (Affymetrix). Quality Control metrics were generated for each array using the Affymetrix -AGCC software.
Paired tumor/normal samples

The paired tumor/normal data set used here was obtained from Affymetrix:
http://www.affymetrix.com/support/technical/sample_data/copy_number_data.affx

These are:
(CRL-2324D, CRL-2325D);
(CRL-5868D, CRL-5957D);
(CCL-256D, CCL-256.1D);
(CRL-2320D, CRL-2319D);
(CRL-2321D, CRL-2362D);
(CRL-2336D, CRL-2337D);
(CRL-2338D, CRL-2339D);
(CRL-2340D, CRL-2341D);
(CRL-2314D, CRL-2346D).

References Cited in the Supplement

Supplemental Figure Legends

Figure S1: Heat Maps - in-lab FFPE normals reference
The heatmap plots for each chromosome represent the segmentation results for the tumor set using the in-lab FFPE normals as a reference set. The red regions denote amplifications and the green regions denote deletions. The maximum red/green brightness corresponds to amplification = 2.5 (red) and deletion = 1.5 (green). The low threshold value was selected to take into account the contamination of samples by normal tissue, which caused the reduction of signal values corresponding to the copy numbers 1 and 3.

Figure S2: Heat Maps - VN reference based on in-lab FFPE normals data set
The heatmap plots for each chromosome represent the segmentation results for the tumor set using the VN algorithm and the FFPE normals data set as a template. The red regions denote amplifications and the green regions denote deletions. The maximum red/green brightness corresponds to amplification = 2.5 (red) and deletion = 1.5 (green). The low threshold value was selected to take into account the contamination of samples by normal tissue, which caused the reduction of signal values corresponding to the copy numbers 1 and 3.

Figure S3: Heat Maps - VN reference based on HapMap data set
The heatmap plots for each chromosome represent the segmentation results for the tumor set using the VN algorithm and the HapMap data set as a template. The red regions denote amplifications and the green regions denote deletions. The maximum red/green brightness corresponds to amplification = 2.5 (red) and deletion = 1.5 (green). The low threshold value was selected to take into account the contamination of samples by normal tissue, which caused the reduction of signal values corresponding to the copy numbers 1 and 3.

Figure S4: Segmentation plots - in-lab FFPE normals based reference
The segmentation plots represent the segmentation results displaying only the regions that fall outside the predefined threshold range. The figures represent results for the tumor set using the in-lab FFPE normals as a reference set. The red lines denote the amplifications and the green lines denote the deletion. The threshold for amplification/deletion corresponds to a copy number of 2.4 / 1.6. The low threshold value was selected to take into account the contamination of samples by normal tissue, which caused the reduction of signal values corresponding to the copy numbers 1 and 3.

Figure S5: Segmentation plots - VN reference based on in-lab FFPE normals data set
The segmentation plots represent the segmentation results displaying only the regions that fall outside the predefined threshold range. The figures represent results for tumor set processing using the VN algorithm and the FFPE normals data set as a template. The red lines denote the amplifications and the green lines denote the deletion. The threshold for amplification/deletion corresponds to a copy number of 2.4 / 1.6. The low threshold value was selected to take into account the contamination of samples by normal tissue, which caused the reduction of signal values corresponding to the copy numbers 1 and 3.

Figure S6: Segmentation plots - VN reference based on HapMap data set
The segmentation plots represent the segmentation results displaying only the regions that fall outside the predefined threshold range. The figures represent results for tumor set processing using the VN algorithm and the HapMap data set as a template. The red lines denote the amplifications and the green lines denote the deletion. The threshold for amplification/deletion corresponds to a copy number of 2.4 / 1.6. The low threshold value was selected to take into account the contamination of samples by normal tissue, which caused the reduction of signal values corresponding to the copy numbers 1 and 3.
Figure S7: Comparison of raw CN and segmentation results (Affymetrix 6.0 array) for a renal carcinoma sample on Chr.2 utilizing three different sources of the reference signal.
A: HapMap based normal reference set.
B: In-lab normal reference.
C: VN reference signal using HapMap as a template.
VN (on FFPE normal)/based reference