Genome-Wide Human SNP Array 6.0 (Affymetrix)

Clear Cell Sarcoma of the Kidney (CCSK) – Copy Number

Nucleic acid labeling, hybridization, and array scanning were performed on 11 CCSKs according to the manufacturer's protocol for the Affymetrix 6.0 SNP array (Affymetrix, Santa Clara, CA, USA) and processed with the Affymetrix Genotyping Console (GTC) 4.0 software. Reference normalization was performed as described by Pounds et al[1]. Circular binary segmentation (CBS) was performed using DNAcopy from BioConductor. Segmented regions of autosomal chromosomes containing at least 8 markers in which the log2 value was > +0.5 or < -0.5 were considered regions of gain or loss, respectively. For the other 2 CCSK samples, copy number was assessed by using relative coverage generated by whole genome sequencing.

Specifically:

*Protocol performed at Ann and Robert H. Lurie Children's Hospital and St. Jude's Children's Research Hospital.

DNA was extracted from normal kidney, tumor, or blood samples at Nationwide Children's BioPathology Center (BPC) by using the standard BPC protocol. Pico green analysis was performed to verify concentration of gDNA. Spectrophotometry was performed to verify DNA purity and gel electrophoresis was performed to verify DNA quality. Tumor and corresponding normal specimens (blood and/or normal kidney) were supplied to St. Jude Children's Research Hospital on 96-well plates allowing for the inclusion of two controls.

Nucleic acid labeling, hybridization and array scanning protocols were performed according to the Affymetrix manufacturer's protocol for the AffyMetrix 6.0 SNP array at St Jude's Children's Research Hospital.

Data were provided by St Jude's Children's Research Hospital in the Affymetrix CEL file format (Level 1 data) and the CEL files were processed using AffyMetrix Genotyping Console (GTC) 4.0 software to generate corresponding Birdseed .chp and .txt files (Level 2 data) by using the Birdseed v2 algorithm with the default parameters. Several quality control parameters were used:

- Contrast QC (quality control): The average contrast QC was 1.83 for all samples, which is above the minimal of 1.7 recommended by AffyMetrix. Less than 10% of samples had a Contrast QC <0.4, and those samples with contrast QC <0.4 were deemed acceptable based on their heterozygosity values and Birdseed call rates.
- DNA gender check: Samples were classified into genders using AffyMetrix Genotyping Console software; no inconsistencies were noted. Only 0.03% of all samples could not be classified according to gender ("unknown"); all of these samples were tumor samples in which the gender of the corresponding normal sample was called correctly.
- Sample Call Rate: AffyMetrix GTC 4.0 software was used to check the calling rate of constitutional DNA samples and all samples had calling rates greater than the cut-off of

>95.5% (range, 94.1–99.5%; mean, 97.9%). Furthermore, the calling rate of tumor samples ranged from 93.4–97.3% (mean, 97.4%).

- DNA Autosomal Heterozygosity rate: The percentage of heterozygous SNPs among all measured SNPs was determined per sample using AffyMetrix GTC 4.0 software. The heterozygosity rates of normal samples ranged from 24–32%, which is within normal limits. This rate, which is expected to be lower for tumors compared to normals, ranged from 15–32% in our tumor samples.
- Normalization: The reference normalization procedure utilized for our data normalization relies on an algorithm developed at St. Jude that utilizes a diploid chromosome for each sample to guide data normalization, as described[1]. In the first step, the CEL files (Level 1 data) and Birdseed.txt files (Level 2 dta) are read into dChip and model-based expression analysis (MBEI) is performed to generate probe level summarization values for each individual probe. This results in a file containing two columns for each individual sample: (1) the summarized probe value and (2) the genotype call. This file containing un-normalized data is exported from dChip as a text file and imported into R for reference normalization according to Pounds et al[1]. This algorithm requires two input files: (1) the dChip output file described above and (2) a text file defining each SNP on the AffyMetrix 6.0 chip according to chromosome and location. The reference chromosome for each sample was selected by using Nexus 6.0 software. The reference normalization algorithm provides an output text file containing two columns for each sample: (1) the normalized probe value and (2) the genotype call.

Circular binary segmentation (CBS) was then applied to the output files in order to obtain segmented copy number information. This was performed in R using the DNAcopy BioConductor package. First, the log (base=2) of the ratios of each tumor sample's signal values over the signal values of the corresponding normal samples was calculated. After detecting outliers and smoothing the log ratio signal data, CBS was applied to segment the data into regions of estimated equal copy number. CBS was performed using default parameters including “nperm = 10,000”, “alpha=0.01”,”undo.splits=sdundo”, and “undo.SD=1”. This algorithm resulted in a segmented file for each tumor sample relative to the corresponding normal sample.

References

1. Pounds S, *et al.* (2009). Reference alignment of SNP microarray signals for copy number analysis of tumors. *Bioinformatics* **25**, 315-21 (PMID: 19052058)