BMC Bioinformatics

Poster presentation

A permutation-based method to identify loss-of-heterozygosity using paired genotype microarray data

Stan Pounds*[1], Cheng Cheng[1], Wenjian Yang[2], Arzu Onar[1], Christine Hartford[3], Susana Raimondi[4] and Mary Relling[2]

Address: [1]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA, [2]Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN 38105, USA, [3]Department of Pediatrics, University of Chicago, Chicago, IL 60637, USA and [4]Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA

Email: Stan Pounds* - Stanley.pounds@stjude.org

* Corresponding author

This abstract is available from: http://www.biomedcentral.com/1471-2105/9/S7/P12

Background
SNP genotyping microarrays may be used to detect regions of loss-of-heterozygosity (LOH). Genotype array data are collected for tumor tissue and germline tissue samples from each subject. For each subject, an initial call of LOH or non-LOH is generated for each marker via straightforward comparison of the genotype call across each tissue sample pair [1]. The genotype calls are generated with some error. Therefore, statistical models are used to analyze the pattern of LOH calls to infer regions of LOH for each subject [1].

Materials and methods
We propose call-based segmentation analysis (CBSA) as a permutation-based method to infer regions of LOH from this type of data. Chromosome endpoints and the positions of markers with initial LOH calls are used to divide the genomes of study subjects into a series of distinct segments that are indexed by subject and location. The size of each segment is measured by the number of non-LOH calls it contains.

CBSA performs a permutation test to determine whether a segment has significantly fewer non-LOH calls than expected by chance. Permuting the assignment of initial LOH calls to subject and genomic position generates an empirical null distribution of segment size for computing p-values. In practice, p-values may be computed with a very accurate analytical approximation of the permutation distribution [2].

Next, the false discovery rate (FDR) is estimated with a robust method [3]. Finally, each segment defined by the observed positions of LOH calls has a size, p-value, and FDR estimate associated with it. Each segment with an FDR estimate below a selected threshold is inferred to be a segment of LOH. Mathematical proofs establish that the FDR estimate is conservative, i.e., the estimated FDR is expected to be greater than the actual FDR [3].

Results
In our study of LOH in secondary leukemia [4], we applied CBSA with an estimated FDR of 10%. CBSA showed similar or greater sensitivity than dChip SNP [1] to detect LOH on each chromosome with one-copy loss according to cytogenetics [5]. Additionally, CBSA was robust against poor quality. After exclusion of two subjects with poor quality data, CBSA inferences were concordant with original CBSA inferences for the remaining eleven subjects at 99.6% of all markers.

Conclusion
CBSA is a practically useful method for detecting LOH. CBSA is conceptually simple, computationally efficient,

statistically sound, and robust. Furthermore, CBSA may be a more powerful method than dChip SNP for some studies.

References

1. Lin M, Wei L-J, Seller WR, Lieberfarb M, Wong WH, Li C: **dChip-SNP: significance curve and clustering of SNP-array based loss-of-heterozygosity data.** *Bioinformatics* 2004, **20:**1233-1240.
2. Pyke R: **Spacings.** *J Roy Stat Soc B* 1965, **27:**395-449.
3. Pounds S, Cheng C: **Robust estimation of the false discovery rate.** *Bioinformatics* 2006, **22:**1979-1987.
4. Hartford C, Yang W, Cheng C, Fan Y, Liu W, Trevino L, Pounds S, Neale G, Raimondi SC, Bogni A, Dolan ME, Pui C-H, Relling MV: **Genome scan implicates adhesion biological pathways in secondary leukemia.** *Leukemia* 2007, **21:**2128-2136.
5. Raimondi SC, Mathew S, Pui C-H: **Cytogenetics as a diagnostic aid for childhood hematologic disorders: conventional cytogenetic techniques, fluorescence in situ hybridization, and comparative genomic hybridization.** In *Tumor Marker Protocols. Methods in Molecular Medicine* Edited by: Hanausek M, Walaszek Z. Totowa, NJ: Humana Press; 1998:209-227.