BIOINFORMATICS APPLICATIONS NOTE Vol. 23 no. 12 2007, pages 1550–1552
doi:10.1093/bioinformatics/btm122

Genome analysis

SNPTools: a software tool for visualization and analysis of microarray data

Frank J. Sørensen[1], Claus L. Andersen[2] and Carsten Wiuf[1,2,*]

[1]Bioinformatics Research Center, University of Aarhus, Høegh Guldbergs Gade 10, Building 1090, DK-8000 Aarhus C and [2]Molecular Diagnostic Laboratory, Aarhus University Hospital, Brendstrupgaardsvej 100, DK-8200 Aarhus N, Denmark

Received on January 11, 2007; revised on February 28, 2007; accepted on March 20, 2007

Advance Access publication March 24, 2007

Associate Editor: Chris Stoeckert

ABSTRACT

Summary: We have created a software tool, SNPTools, for analysis and visualization of microarray data, mainly SNP array data. The software can analyse and find differences in intensity levels between groups of arrays and identify segments of SNPs (genes, clones), where the intensity levels differ significantly between the groups. In addition, SNPTools can show jointly loss-of-heterozygosity (LOH) data (derived from genotypes) and intensity data for paired samples of tumour and normal arrays. The output graphs can be manipulated in various ways to modify and adjust the layout. A wizard allows options and parameters to be changed easily and graphs replotted. All output can be saved in various formats, and also re-opened in SNPTools for further analysis. For explorative use, SNPTools allows various genome information to be loaded onto the graphs.

Availability: The software, example data sets and tutorials are freely available from http://www.birc.au.dk/snptools

Contact: wiuf@birc.au.dk

1 INTRODUCTION

Analysis and visualization of microarray data are important for proper interpretation of data. In particular, visualization of array intensities in relation to their chromosomal positions is important for investigating copy number variation in the genome (e.g. SNP and comparative genomic hybridization (CGH) array intensities) and for identifying genes that are regulated by the same molecular mechanisms because of physical proximity (e.g. gene and miRNA array intensities). These and other similar issues have been the focus of attention in many research publications, e.g. Baehrecke *et al.* (2004) and Lin *et al.* (2004).

In a number of articles, we have developed and applied simple methods for comparing groups of array intensities and visualization of these in relation to their chromosomal positions, Andersen *et al.* (2007), Lamy *et al.* (2006), Tørring *et al.* (2007) and Zieger *et al.* (2005). Each group is defined by pathological, clinical or molecular characteristics with relevance for the sampled biological specimens. Based on the developed methods, we have created a software tool, SNPTools, for analysis and visualization of microarray data. In particular, the software is developed to handle Affymetrix SNP arrays, but most of the features have broader applicability and apply to gene expression arrays, miRNA expression arrays and CGH arrays as well. In the following, we use unit as short for SNP, gene, clone, etc. The software has the following features,

- Creation of data sets that combine intensity values, physical positions of units and array information, e.g. clinical, molecular and pathological information, for further analysis.
- Plotting of arrays intensities, one array at the time, on chromosomal maps.
- Plotting of group mean/median intensities for two or more groups of arrays on chromosomal maps, and testing for differences between the groups. Test probabilities are shown on the map, colour coded or as bars.
- Identification of segments of units that differ in intensity level between groups according to some definition of segment.
- Joint plotting of loss-of-heterozygosity (LOH) data and SNP intensities for paired arrays of normal and tumour tissues. LOH is derived from the genotypes of the normal and the tumour sample.
- Genome browser showing genes in the RefSeq database (http://www.ncbi.nlm.nih.gov/RefSeq/), validated miRNA from the miRBase (http://microrna.sanger.ac.uk/sequences/) or copy number variants (CNV; http://www.sanger.ac.uk/humgen/cnv/) with weblinks to genome web browsers providing detailed information about the genes, miRNAs or CNVs.
- Extensive help facilities to help the user with all aspects of the program.

An example screenshot is shown in Figure 1. SNPTools is a wizard—on each page the user is presented with a series of options and parameters that allow the user to control the analysis and the output. Among the options are whether the data should be (high level) normalized to facilitate comparison

*To whom correspondence should be addressed.

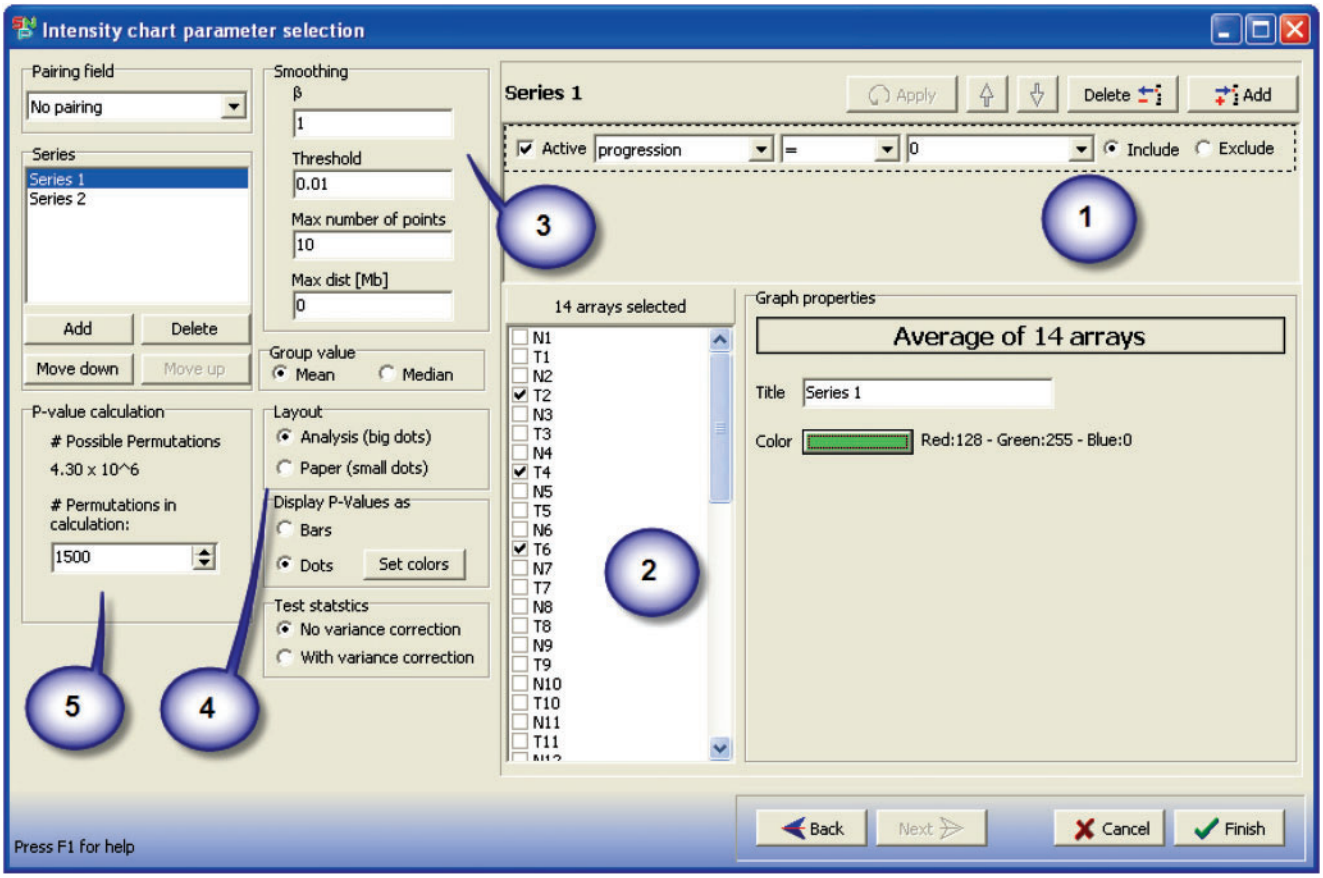

Fig. 1. An example page, where groups of arrays are defined, either using logical constraints (1), or manually (2) is shown. Also parameters for smoothing intensity values are controlled here (3), as well as layout parameters (4) and the number of permutations used for testing for group differences (5).

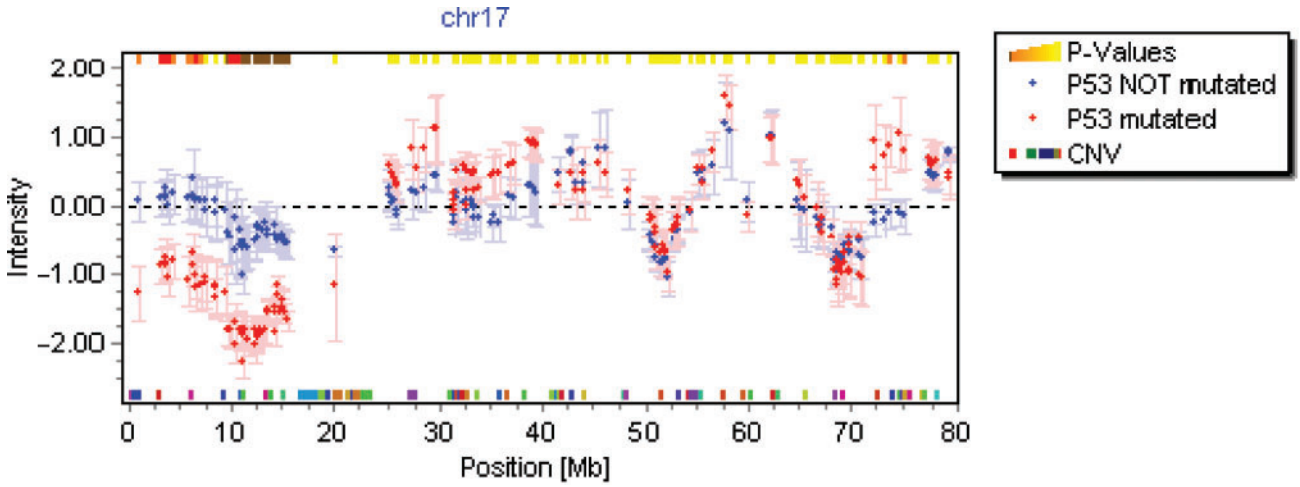

Fig. 2. An example output from SNPTools showing two groups of SNP arrays that differ in copy number in several locations on chromosome 17 is shown. The two groups comprise tumour samples only (mutations in the TP53 gene versus no mutations), normalized with reference to a collection of samples from normal tissue—hence zero represent two copies. Top bar shows *P*-values, yellow indicates that the two groups are not significantly different at 5% level; other colours indicate significant differences at various levels. Also shown are error bars for each group and SNP (±1 SD). Bottom bar shows known CNVs according to the Sanger CNV database.

of different units, smoothed to reduce noise in the intensities or whether missing data should be imputed. Groups of arrays can be selected easily, either by manual selection or by logical constraints, as illustrated in Figure 1. The wizard allows the user to go back and change parameters and options, then to replot without having to go through the whole range of pages again.

An example output plot is shown in Figure 2. The curves and other information that are shown in the plot can be manipulated and changed in various ways to make the output appear according to the user's preferences. Output can be saved and re-opened in SNPTools, saved to pdf, csv and other formats for further processing.

SNPTools come with several example data sets and tutorials that can guide the user through the initial phase of using the program.

ACKNOWLEDGEMENTS

We thank Karsten Zieger for testing the software. C.W. is supported by the Danish Cancer Society, F.J.S. by the Aarhus University Research Foundation and by the Danish Research Council. C.L.A. is supported by the Danish Research Council, and the John and Birthe Meyer Foundation.

Conflict of Interest: none declared.

REFERENCES

Andersen,C.L. *et al*. (2007) Frequent occurrence of uniparental disomy in colorectal cancer. *Carcinogenesis*, **28**, 38–48.

Baehrecke,E.H. *et al*. (2004) Visualization and analysis of microarray and gene ontology data with treemaps. *BMC Bioinformatics*, **5**, 84.

Lamy,P. *et al*. (2006) Are microRNAs located in cancer specific regions in the genome? *Br. J. Cancer*, **95**, 1415–1418.

Lin,M. *et al*. (2004) dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics*, **20**, 1233–1240.

Tørring,N. *et al*. (2007) Genome-wide analysis of allelic imbalance in prostate cancer using the Affymetrix 50K SNP mapping array. *Br. J. Cancer*, **96**, 499–506.

Zieger,K. *et al*. (2005) Role of activating fibroblast growth factor receptor 3 mutations in the development of bladder tumors. *Clin. Cancer Res.*, **11**, 7709–7719.