

Software

Open Access

Visualization and analysis of microarray and gene ontology data with treemaps

Eric H Baehrecke*¹, Niem Dang², Ketan Babaria² and Ben Shneiderman²

Address: ¹Center for Biosystems Research, University of Maryland Biotechnology Institute, College Park, Maryland 20742, USA and ²Department of Computer Science and Human-Computer Interaction Laboratory, University of Maryland, College Park, Maryland 20742, USA

Email: Eric H Baehrecke* - baehreck@umbi.umd.edu; Niem Dang - niem.dang@towersperrin.com; Ketan Babaria - ketanb@gmail.com; Ben Shneiderman - ben@cs.umd.edu

* Corresponding author

Published: 28 June 2004

Received: 31 January 2004

BMC Bioinformatics 2004, 5:84 doi:10.1186/1471-2105-5-84

Accepted: 28 June 2004

This article is available from: <http://www.biomedcentral.com/1471-2105/5/84>

© 2004 Baehrecke et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The increasing complexity of genomic data presents several challenges for biologists. Limited computer monitor views of data complexity and the dynamic nature of data in the midst of discovery increase the challenge of integrating experimental results with information resources. The use of Gene Ontology enables researchers to summarize results of quantitative analyses in this framework, but the limitations of typical browser presentation restrict data access.

Results: Here we describe extensions to the treemap design to visualize and query genome data. Treemaps are a space-filling visualization technique for hierarchical structures that show attributes of leaf nodes by size and color-coding. Treemaps enable users to rapidly compare sizes of nodes and sub-trees, and we use Gene Ontology categories, levels of RNA, and other quantitative attributes of DNA microarray experiments as examples. Our implementation of treemaps, Treemap 4.0, allows user-defined filtering to focus on the data of greatest interest, and these queried files can be exported for secondary analyses. Links to model system web pages from Treemap 4.0 enable users access to details about specific genes without leaving the query platform.

Conclusions: Treemaps allow users to view and query the data from an experiment on a single computer monitor screen. Treemap 4.0 can be used to visualize various genome data, and is particularly useful for revealing patterns and details within complex data sets.

Background

Genome sequencing has presented biologists with new challenges in data analysis. This advance coupled with the advent of methods to empirically analyze whole genome changes in RNA levels, protein levels, and protein activities [1-5], presents difficulties in visualizing summaries of data while obtaining meaningful details. The use of colored mosaics and hierarchical clustering to query relative RNA levels revolutionized DNA microarray analyses [6], and has been the prominent mechanism for assessing this data. While continued development of this data anal-

ysis platform is useful, these methods limit the ability to simultaneously visualize multiple data attributes including the analysis of qualitative information about either gene families or biological function and quantitative information such as RNA level and p-value simultaneously.

The Gene Ontology (GO) consortium has established a vocabulary that provides a hierarchical structure for the analysis of genome data [7,8]. GO provides a classification of gene products into molecular functions, biological

processes, and cellular components. Therefore, GO classification is particularly useful for getting overviews of data such as the percentage of genes transcribed within each category or node, and also provides a rapid mechanism for researchers to classify genes that are often given non-descript numerical names during genome annotation. The dynamic nature of GO data, which is updated weekly for active genome projects, however, challenges researchers to be vigilant in the analysis and re-analysis of data. Ideally, researchers would be able to obtain information about both qualitative attributes such as GO category, and quantitative attributes such as RNA level for an entire experiment, and query these data sets for details without losing the overview of the entire data structure.

Several computational approaches have been developed to visualize and query microarray data including Spotfire [9] and Genespring [10]. While both of these platforms are capable of analyzing both qualitative and quantitative data, neither provides an ideal platform to visualize multiple attributes simultaneously while allowing dynamic queries of data in the context of the GO classification. Further, limited mechanisms exist for merging quantitative attributes such as RNA level with GO categories. Several programs have been developed to edit, browse, and facilitate studies of GO [8]. Among these applications, FatiGO [11], GoMiner [12], MAPPFinder [13], and GoSurfer [14] provide useful platforms for the analysis of microarray data in the context of the GO hierarchy, but their use of typical windows-style browsers and tree diagrams lacking quantitative data limits the ability to rapidly see patterns and obtain details on demand.

Treemaps were developed to facilitate visualization of both hierarchical and quantitative information [15,16]. This technique has been used to visualize and query several forms of data including the stock market [17] and electronic product catalogs [18]. While treemaps have been recognized as a strategy that can be used to visualize clusters and GO annotation data [19], previous use of this strategy has been limited to static approaches of pre-selected data and have failed to take advantage of several important strengths of Treemap 4.0.

Here we extend Treemap 4.0 to visualize and query microarray data in the GO framework, and use studies of programmed cell death during development [20] as an example. We present the advantages of the use of size and color to represent attributes of the genome. By using code that merges up-to-date GO assignments with various user-defined attributes such as RNA level, p-value, and others, we utilize treemaps to get overviews of data, and to query for details. Treemap 4.0 provides users with rapid responses to queries (usually a fraction of a second), and the results can be saved and exported such that they can

be analyzed using other software. A link between Treemap 4.0 and organism-specific web sites enables users to get details about specific genes as needed. Thus, Treemap 4.0 fills a critical void for genome researchers who want to integrate and query GO information with various quantitative data.

Results and Discussion

Treemaps enable visual overviews of complex genome data with details on demand

We have used treemaps to visualize DNA microarray data that examine changes in RNA levels during steroid-triggered programmed cell death in *Drosophila* [20]. RNA was extracted from salivary glands dissected from animals that were staged 6 and 12 hours following puparium formation; before and after the rise in steroid that triggers cell death. Three independent salivary gland RNA samples were collected from each stage and used to hybridize *Drosophila* Affymetrix oligonucleotide Genechips. These Genechips contain 13,197 unique gene transcripts, and 2,876 gene transcripts were consistently detected in all 3 samples of either 6- or 12-hour salivary glands.

Treemap 4.0 was used to analyze the 2,876 gene transcripts that were consistently detected in dying salivary glands. The first step in this analysis involved parsing the 2,876 gene transcripts in the GO framework. This was accomplished using a Perl script that assigns each gene transcript a GO path in the molecular function, biological process, and cellular component categories, and then writes an output file for use in Treemap 4.0 (software and demonstration data available at HCIL [21]). An advantage of this parser is that it enables the analysis of data in the context of up-to-date GO classification, but it also has disadvantages because of limitations associated with Perl. Perl is an interpreted language that runs slowly on large data sets, and 5 minutes were required to parse and write the data to a file with 2,876 genes containing 11,638 GO assignments. The positive aspect of Perl is that comments are provided internally so that competent programmers can optimize this parser. While implementation of this parser in a language such as C++ would be faster and is a future objective, this may provide an impediment to biologists that use Perl because of user friendly attributes. The second limitation of this parser is that users must define their own column headings and types after the treemap file is written, as the column definition and data type must match for Treemap 4.0 to read the file, and this will differ between users and data files. This is one of many possible parser improvements that are future objectives including accounting for missing values, offering extended error checking with error messages, and providing weighting functions. However, our primary objective was to implement and test the utility of Treemap 4.0 for dynamic queries of microarray data in the context of up-to-date GO

classification, and this parser enabled us to accomplish this goal and will allow others to use treemaps to query genome data.

Treemaps allow the visualization of large data sets in the context of the GO classification. While a typical browser screen displays a list of 50 or less rows of characters (such as genes) with 10 or less columns of either quantitative or qualitative values (such as RNA level, p-value, GO classification, etc...), a treemap can display large amounts of information on a single screen (Figure 1A). Although Treemap 4.0 performance prevented the analysis of unprocessed DNA microarray data from 13,197 genes representing nearly the entire fly genome in the context of GO classification, it did perform well in the analysis of the 2,876 gene transcripts that are consistently detected in dying salivary glands [20] with 11,638 GO assignments. The Treemap 4.0 display is divided into three regions: (1) the data display and query window on the left, (2) the details of selected node on the top right, and (3) the control panel on the bottom right. The Treemap 4.0 data display and query window uses area to convey quantitative information. For example, users can immediately see that more area and, therefore, more nodes (gene assignments within GO) exist in the biological process category, and that less area/fewer nodes exist in the molecular function, cellular component, and unlisted term (annotated genes without GO assignments) categories (Figure 1A). Closer examination of the biological process category allows the conclusion that physiological process GO assignments are greater than either cell process or development GO assignments based on area. Selection of a category, such as the entire Gene Ontology, highlights this area in yellow, displays detailed information about the selected area in the "details of selected node" window at the top right, and enables the user to see that 11,638 gene nodes exist in this window (Figure 1A). This illustrates one of the greatest strengths of treemaps, as they provide an overview of the data while allowing details-on-demand with rapid updates.

Treemap users gain a meaningful overview of the data (Figure 1A), and also have access to detailed information about single genes without leaving this initial display. For example, steroids are bound by receptors encoding nucleic acid binding proteins that activate RNA transcription by binding to DNA. Therefore, it is interesting to determine which genes are in the nucleic acid binding category within the molecular function GO. By selecting a single box within this category, a pop-up window indicates that the gene encoding Myb is transcribed in dying salivary glands (Figure 1B). In addition, the details of selected node provides immediate quantitative and qualitative information about Myb in this experiment including the average RNA values at 6 and 12 hours, that this

gene decreased 3.0556-fold in RNA level, that the p-value of a t-test was 0.0073, several gene annotation identifiers, and the GO path for Myb (Figure 1B). The ability to extract details about overviews of data does not only depend on the user judging area, however, as the selection of either a category or a gene (and the details of the selected node displayed) are determined by the user. Of the 11,638 gene nodes in this GO (Figure 1A), selection of the biological process category displays that 5,529 genes are assigned to this GO category (Figure 2A). Within seconds of selection of other categories, it was determined that 2,726 gene nodes are assigned to the molecular function GO category, 2,019 gene nodes are assigned in the cellular component GO category, and 1,364 unlisted terms or unknown genes are transcribed during salivary gland cell death (Figure 2B,2C,2D).

Tools that facilitate visualization and queries of genome data

Size and color are two attributes that can be used to display quantitative differences in data using treemaps. Users have the flexibility to assign labels, size and color to different gene attributes in the "legend" section of the control panel in the bottom right region of the display (Figure 3A). In this and all examples to follow, the "label" has been assigned to the fixed symbol category which is the common gene name [22], "size" has been assigned to absolute average fold change in RNA level between the 6- and 12-hour time points (absolute values enable the display of negative values), and "color" has been assigned to the p-value of a t-test. To assist in visualizing significant p-values, the nodes have been displayed in two distinct categories by selecting "user defined bins" from a scroll down window in the control panel (Figure 3A). Significant values between 0 and 0.05 were colored from red to yellow while insignificant values greater than 0.05 were colored black. Therefore, users can rapidly see that large boxes represent genes with large changes in RNA level during steroid triggered cell death (Figure 3A). In addition, while all non-black boxes (genes) have significant p-values, the more red boxes are genes with the most significant p-value.

Treemap 4.0 enables users to zoom on details by double-clicking on an area of interest, and this results in a rapid update of the area selected (a right mouse click enables the user to zoom out one level). For example, programmed cell death involves the degradation of cell components, and this is known to require the activity of peptidase (protease) enzymes. By double-clicking the molecular function GO category that contains 2,726 gene nodes, and then selecting enzymes, the user can immediately see that enzymes are one of the largest categories within molecular function with 1,186 gene nodes (Figure 3A). Furthermore, by double-clicking and zooming on the

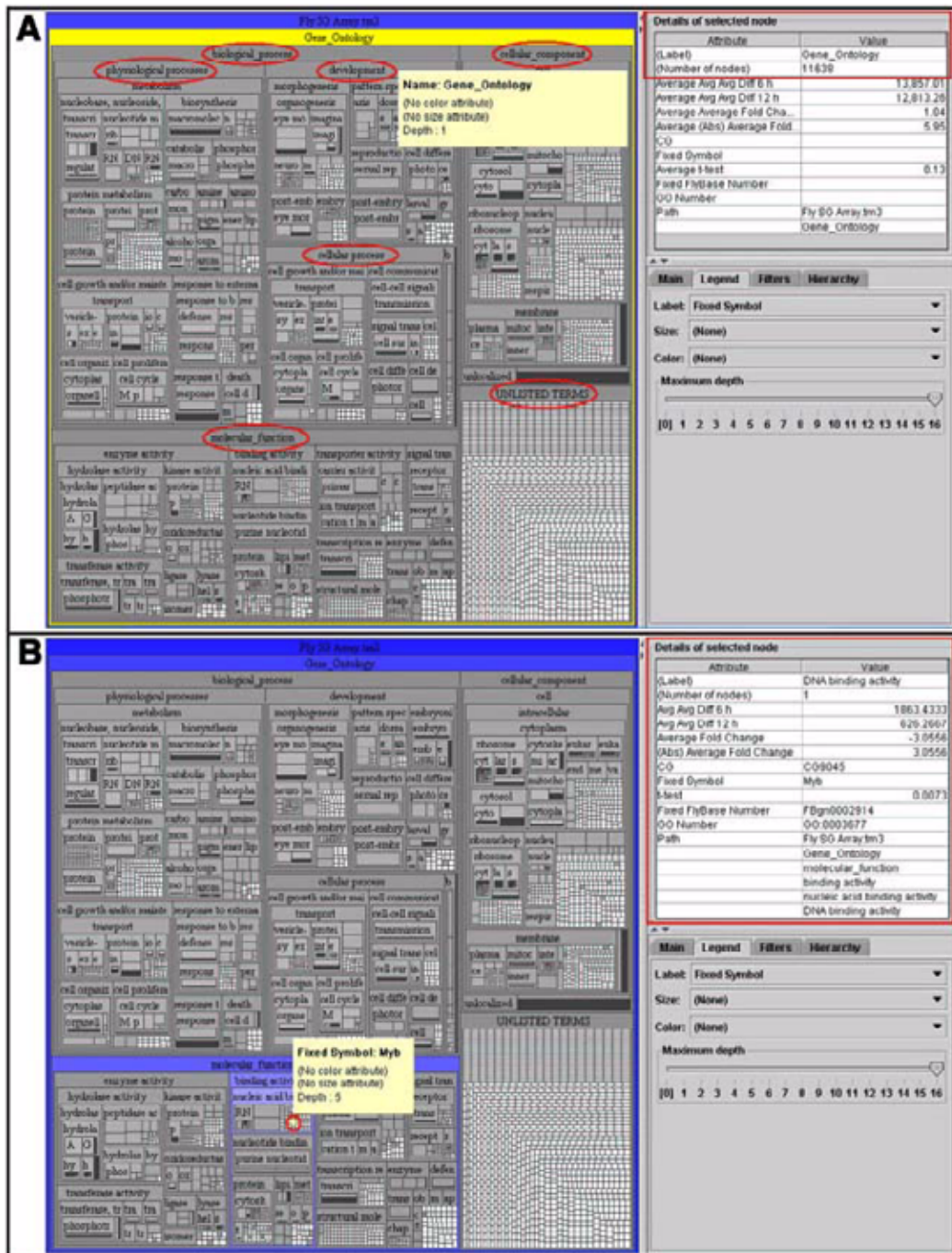


Figure 1

Treemap allows users to visualize data from an entire DNA microarray experiment in the context of the GO hierarchy on a single screen, and to access details about any gene on demand. (A) The Treemap data display and query window on the left uses area to convey quantitative information. A larger number of gene nodes exist in the biological process category of the GO (red circle), and that less area and fewer gene nodes exist in the molecular function (red circle), cellular component (red circle), and unlisted term (annotated genes without GO assignments) (red circle) categories. Selection of the entire Gene Ontology highlights this boxed area in yellow, and displays detailed information about the selected area in the "Details of selected node" window at the top right (red box). (B) Treemap allows access to data details without leaving the overview of the data. Selection of a single gene node box enables the user to see that the gene encoding Myb is transcribed in dying salivary glands (red circle), and the "Details of selected node" provides immediate quantitative and qualitative information about Myb in this experiment (red box).

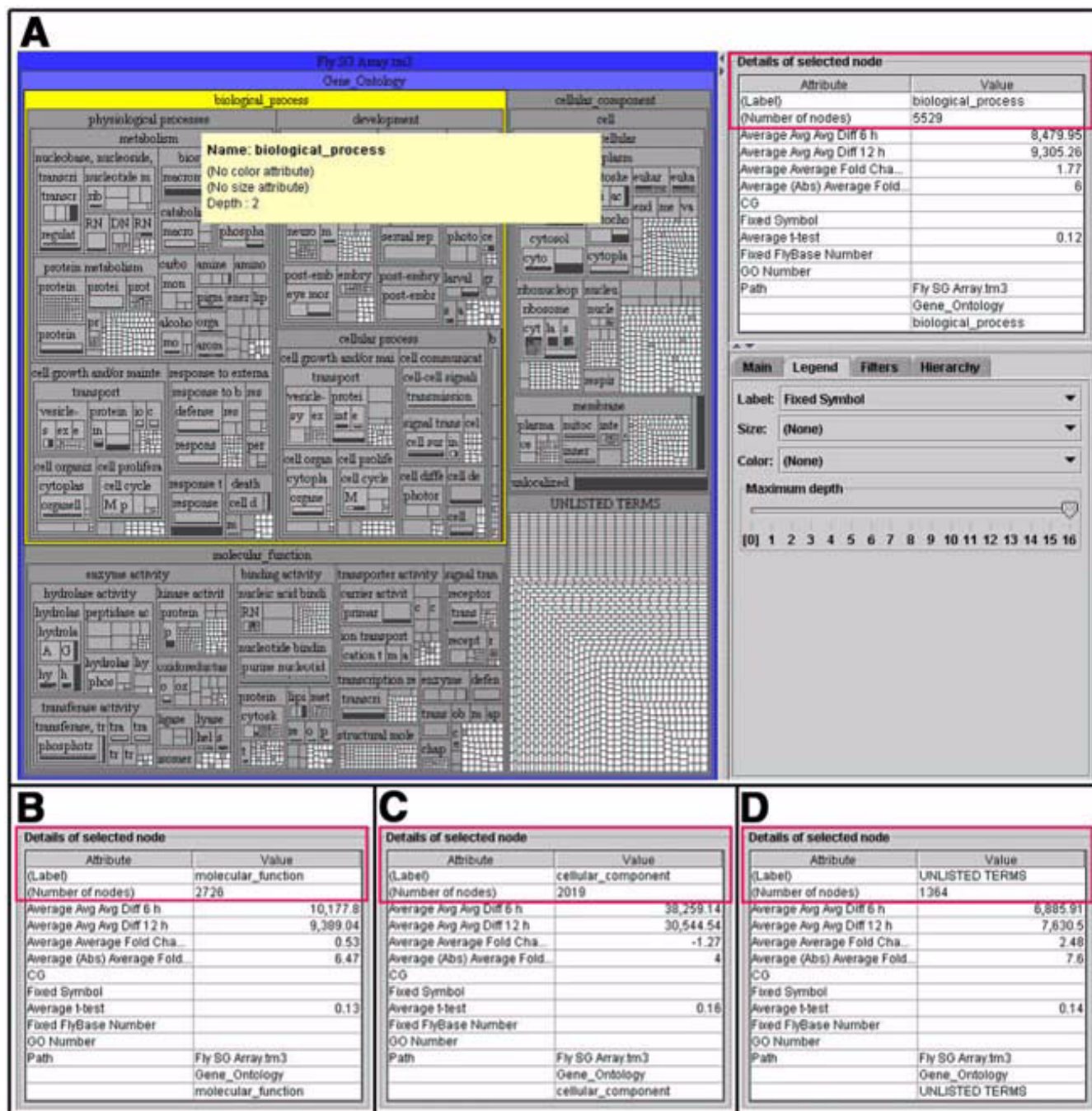


Figure 2

Over-views of genome data can be rapidly obtained using Treemap. (A) Selection of the biological process region displays that 5,529 genes are assigned to this GO category in the "Details of selected node" window at the top right (red box). (B-D) 2,726 gene nodes are assigned to the molecular function GO category, 2,019 gene nodes are assigned in the cellular component GO category, and 1,364 unlisted terms or unknown genes are transcribed during salivary gland cell death (red boxes).

peptidase activity category and then selecting cysteine-type peptidase within that category, the user can immedi-

ately see in the "Details of selected node" window that 23 gene nodes are present, and that most of the boxes are

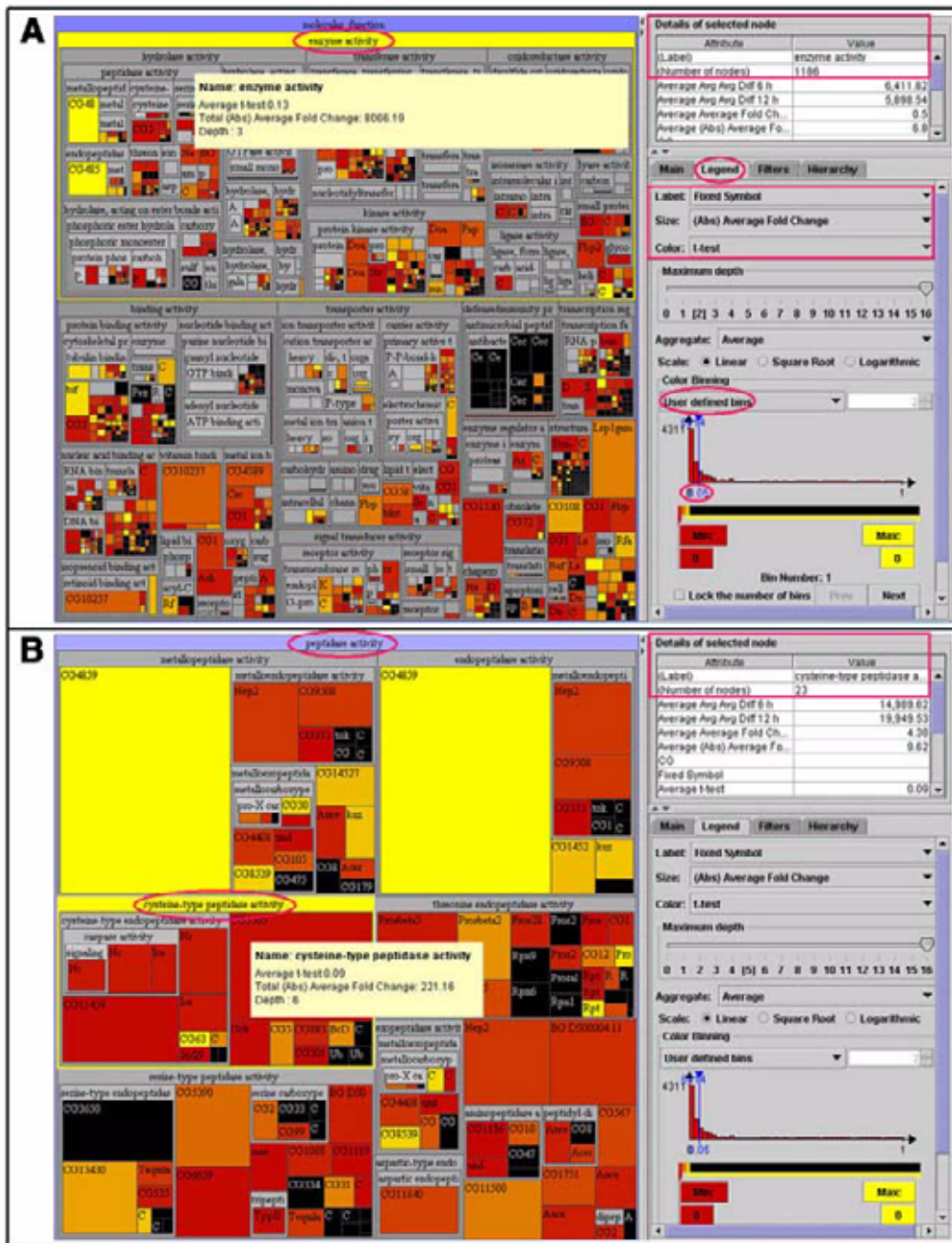


Figure 3

Size and color provide users with a rapid mechanism to evaluate data, and zooming allows access to details about genes of interest. (A) The "legend" section (red circle) of the control panel in the bottom right region of the display allows users to assign "label", "size", and "color" to qualitative or quantitative data (red box). Label has been assigned to the fixed symbol category (common gene name), size has been assigned to absolute average fold change in RNA level, and color has been assigned to the p-value of a t-test. "User defined bins" (red circle) were selected from a scroll down window in the control panel to assist with rapid identification of significant p values. Significant values between 0 and 0.05 were colored from red to yellow while insignificant values greater than 0.05 were colored black. Large boxes represent genes with large changes in RNA level. (B) Treemap users can zoom on details by double-clicking on an area of interest. For example, programmed cell death involves the degradation cells components, and this is known to require the activity of peptidase (protease) enzymes. By double-clicking and zooming on the peptidase activity category (red circle) and selecting cysteine-type peptidase (red circle and outlined in yellow), the user immediately obtains details including that 23 gene nodes are present (red box in details of selected node), that most of the boxes are large and red, and that the Caspases Nc (Dronc) and Ice (Drice) are present.

large and red. Closer examination reveals the presence of Nc (Dronc) and Ice (Drice) (Figure 3B), which encode two cysteine-type peptidases that are commonly known as Caspases, and these are critical regulators of programmed cell death in higher animals including *Drosophila* and humans [23]. Focused evaluation of the proteases that are present in dying cells, such as Caspases, provides one rapid mechanism to determine if either expected or predicted genes are transcribed in dying cells. The association of non-Caspase proteases that may not have been previously implicated in programmed cell death, however, enables users to test if these genes are involved in cell death by conducting additional biological experiments.

Treemap 4.0 provides power and flexibility by allowing users to query microarray data available in the context of the entire GO classification with little loss of time. Programmed cell death is an important cellular process and is represented by only 48 gene nodes within the 5,529 gene nodes of the biological process GO classification of this microarray data (Figure 4A). This result could indicate that a small number of cell death genes are required in dying cells.

Alternatively, salivary gland cell death may either be regulated at the level of protein, or perhaps cell death genes may not have been classified within this category of the GO because they may not have been identified to function in this process at this time. The Caspase co-factor Ark (Dark/Dapaf-1/Hac-1) is easily recognized upon closer examination of the genes within the cell death category because of its large box 50.8333 fold-change in RNA level and red 0.0036 p-value (Figure 4B). Several other interesting cell death genes including the Caspases Nc (Dronc) and Ice (Drice), the Bcl-2 family member Buffy, the inhibitor of apoptosis th (Thread/Diap1), the CD36 family member crq, and the serine/threonine kinase Akt1 are present with significant red to orange p-values. In contrast, the presence of black insignificant p-values for the known cell death regulators rpr (Reaper), W (Wrinkled/Hid) and Eip93F (E93) are surprising (Figure 4B). By selecting either rpr, W, or Eip93F, it was rapidly determined that these insignificant p-values must be due to biological variation in RNA levels in cells or different salivary glands, as all three of these genes are detected at reasonable levels 12 hours after puparium formation (Figures 4C,4D,4E). It is not surprising to have such variation in the RNA levels for these 3 genes, as these RNAs are transcribed in a very transient manner in dying salivary glands [24-27].

While queries for genes in GO categories that are known to be implicated in a process such as nucleic acid binding, peptidases, and cell death are useful, the search for important genes with unknown function (unlisted terms in GO)

may provide the greatest advances in biology. To facilitate meaningful queries for important genes, users can use filters to select genes based on specific quantitative attributes. If the genes that increase following the rise in hormone that triggers salivary gland cell death are of greatest interest, users can go to the "filters" window of the control panel and move the "average fold change" slider to include values greater than +2 (Figure 5A). The display is rapidly updated such that all of the negative values and values less than 2 turn grey (Figure 5A). If filtered values are not of interest, they can be immediately removed by selecting the "hide filtered" button in the control panel, and this removes all of the grey boxes (Figure 5B). A second approach to separating quantitative attributes is available by implementing user-defined bins in the "hierarchy" window of the control panel (Figure 6A). To accomplish this task, users must first select the default hierarchy and "remove" it. Quantitative attributes of the hierarchy can then be added, such as absolute average fold change, and "user-defined bins" can be used to separate values from 0 to 2 (Figure 6A). A second gene attribute, such as p-value of a t-test, can be added and user-defined bins enable the separation of values greater than 0.05 (Figure 6B). Thus, the use of filtering and hierarchy enables users to rapidly sort data based on meaningful quantitative attributes.

Genome researchers require rapid access to details about genes such as map position within the genome, nucleotide and protein sequence, and literature published to name a few examples. Treemap 4.0 was adapted to contain a direct link to organism-specific websites within the main window of the lower right control panel. By enabling this link, users who select a gene node while holding the control key will be directly connected to the Flybase page for the gene of interest. While we have only implemented this for use in Flybase at this time, we intend to make this function flexible and enable all GO organism databases to be selected in the future. Finally, Treemap 4.0 is not meant to serve as a stand-alone tool for genome analysis, and any queried file that is selected while holding the control key will be saved to a tab-delimited file that can then be used in other software such as hierarchical clustering.

Conclusions

Treemaps is a space-filling visualization technique that facilitates exploration of hierarchical structures. While treemaps have been widely used in the evaluation of business performance, the stock market, inventory management, and product catalogs, it has not been widely used in genome research. By presenting attributes of genes by size (RNA level) and color-coding (p-value), we demonstrate how treemaps can facilitate queries of DNA microarray data in the context of Gene Ontology categories. User-

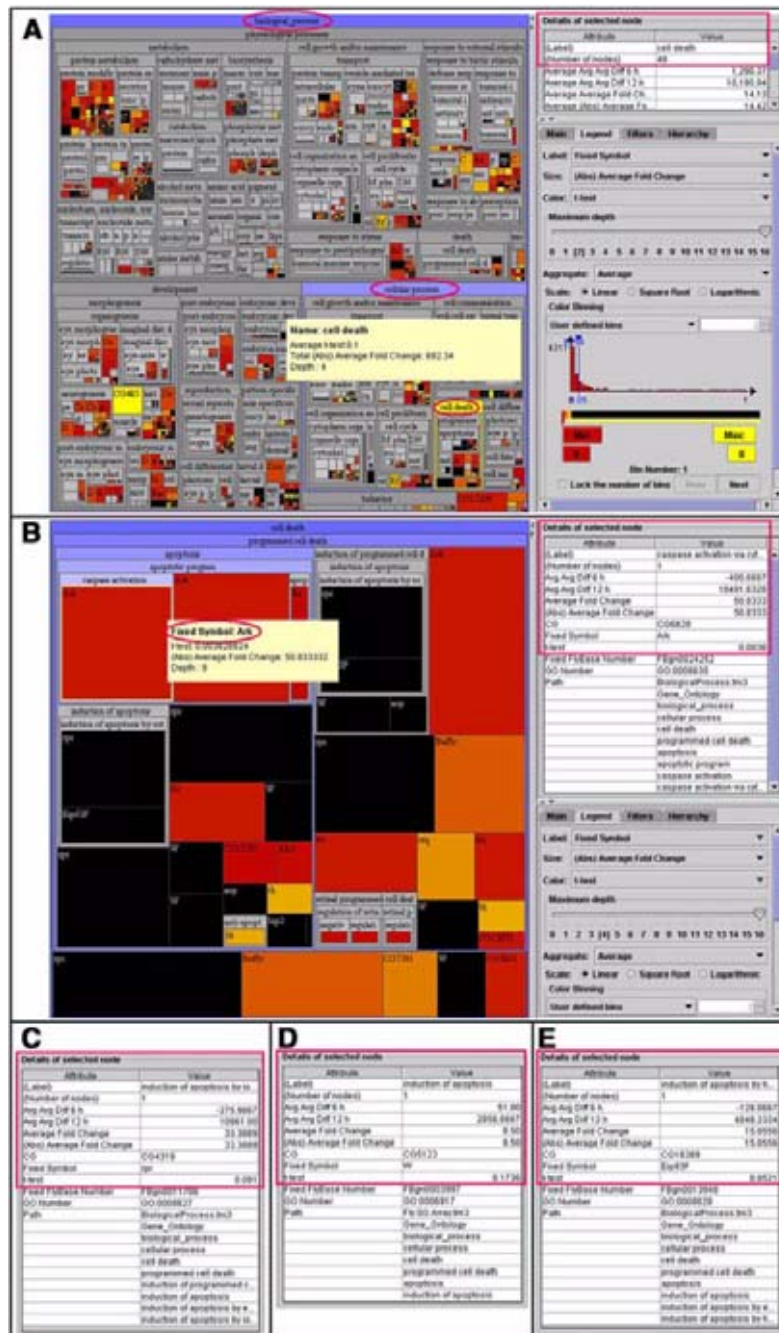


Figure 4

Treemaps allows users to query data in the context of the entire GO classification with little loss of time. (A) The user rapidly determined that programmed cell death is represented by 48 gene nodes within the 5,529 gene nodes of the biological process GO classification (red circles in the data display and query window, and red box in the details of selected node). (B) Zooming enabled the identification of the Caspase co-factor Ark (Dark/Dapaf-1/Hac-1) as a possible significant factor in salivary gland cell death because of its large box 50.8333 fold-change in RNA level and red 0.0036 p-value (red circle in the data display and query window, and red box in details of selected node). Several other interesting cell death genes including the Caspases Nc (Dronc) and Ice (Drice), the Bcl-2 family member Buffy, the IAP th (Thread/Diap1), the CD36 family member crq, and the serine/threonine kinase Akt1 are present with significant red to orange p-values. (C-D) The black insignificant p-values for rpr (Reaper), W (Wrinkled/Hid) and Eip93F (E93) were rapidly assessed in the details of selected node, and the fact that all three of these genes are detected at extremely elevated levels 12 hours after puparium formation suggests that these p-values must be due to biological variation in RNA levels in either different cells or different salivary glands (red boxes).

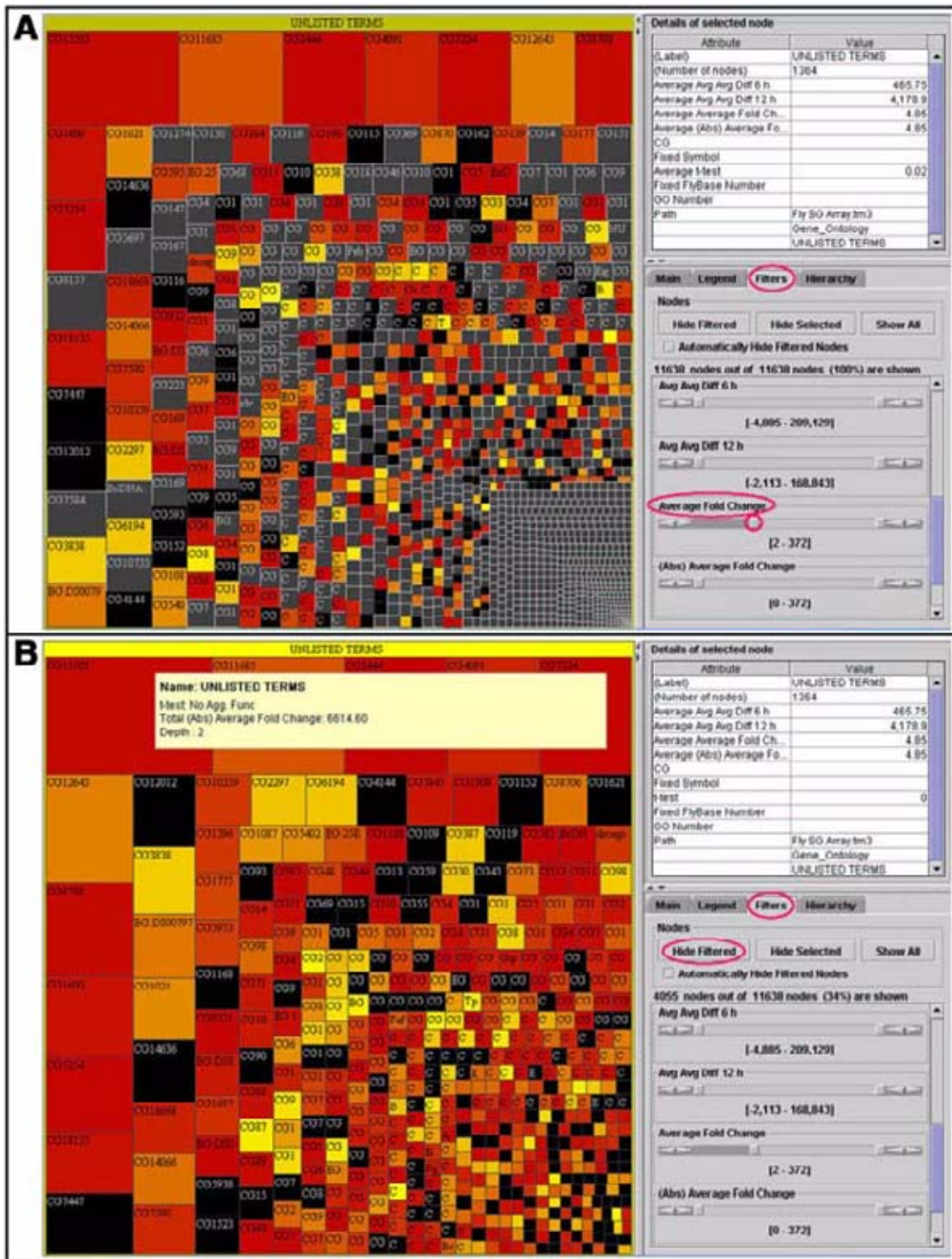


Figure 5
 Filters allow Treemap users to rapidly identify genes of interest based on quantitative attributes. (A) Filters can be used to select genes based on specific quantitative attributes in Treemap. Users can apply "filters" in the control panel (red circle). In this example, the "average fold-change" slider was moved to include values greater than 2 (red circles), and the Treemap display was updated so that all values less than 2 turn grey. (B) Filtered values that are not of interest can be removed by selecting the "hide filtered" button in the "filters" control panel (red circles), and this removes grey boxes.

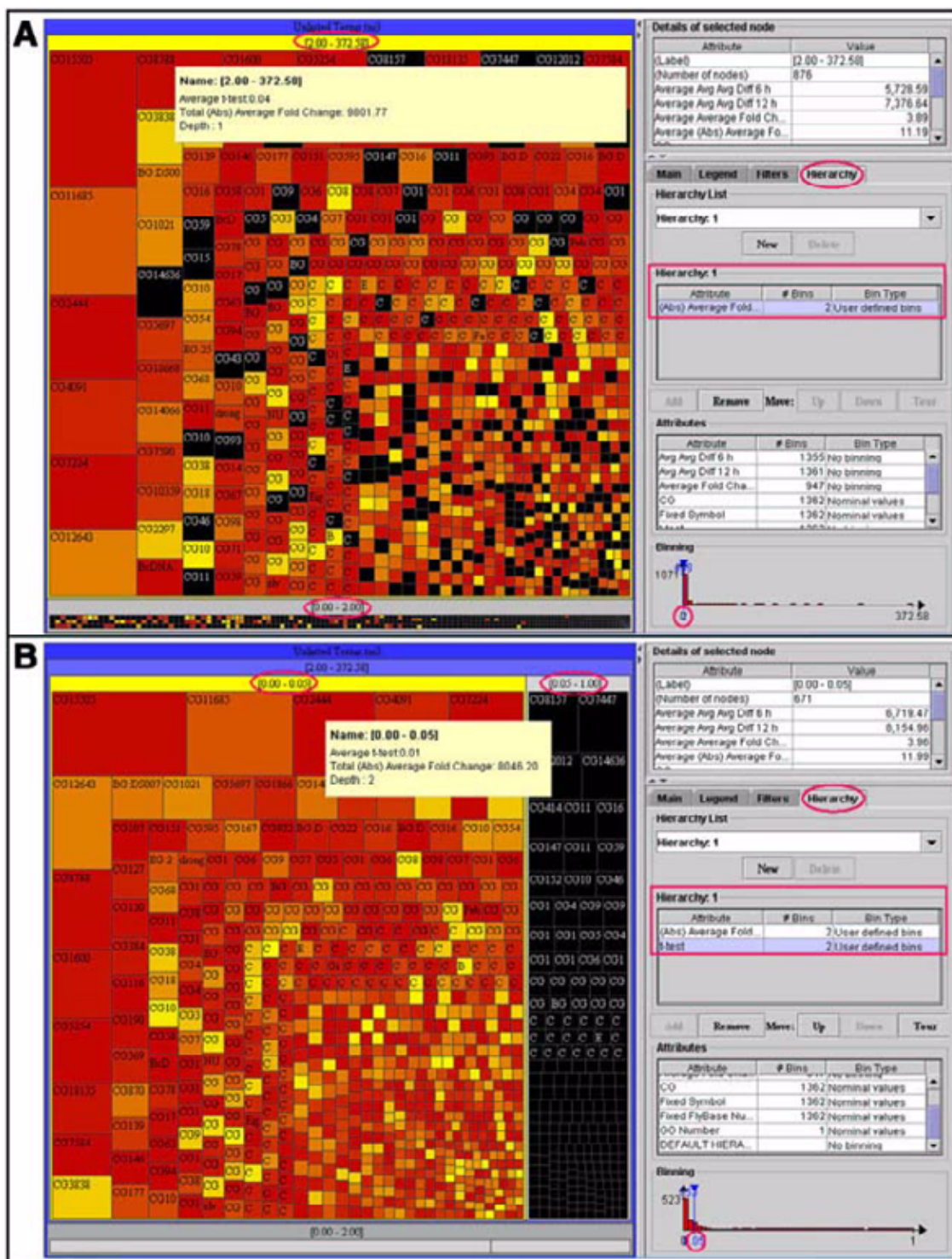


Figure 6
 Genes can be displayed in distinct categories based on quantitative attributes in Treemap. (A) Quantitative attributes, such as "absolute average fold change" (red box), can be added in the "hierarchy" window of the control panel (red circle), and values that are less and greater than 2 can be separated by applying user defined bins (red circles in the data display and query window). (B) A second quantitative attribute can be added in the "hierarchy" window of the control panel (red circle), such as p-value of a t-test (red box), and user defined bins enable the separation of values that are less and greater than 0.05 (red circles in the data display and query window).

defined filtering within Treemap 4.0 allows the selection of data of greatest interest that can then be exported for secondary analyses with other software. For simplicity, we have only used Affymetrix DNA microarray data in the examples provided. Treemaps can be used to visualize all forms of DNA microarray data, however, and this platform should also enable researchers to query for patterns within a variety of complex genome data sets. For example, Treemap can also be used to query other qualitative and quantitative attributes of genomes including map position, degree of conservation between species, size of a gene family, extent of alternative splicing, and protein levels. In addition, genome annotation may be facilitated by treemap presentations, as it allows users to see complex information in a single display. For example, a single gene is represented many times within the GO, and this information may not be apparent in either large tables or genome profiling summary statistics. Comprehensive GO assignments are intended to provide researchers with maximum information, but such results can also be misleading as the same gene may appear many times within a single category (see Figures 3,4 for examples). Although treemaps allows users to see the number of nodes (genes) in each category, presentation of the complete data allows informed decisions and critical evaluation of GO assignments.

Methods

Data

DNA microarray data were taken from previously described experiments aimed at understanding how RNA levels change during steroid-triggered programmed cell death [20], and the data is available at <http://www.umbi.umd.edu/~cbr/baehrecke/data.htm>. RNA was extracted from salivary glands dissected from wild-type Canton S *Drosophila melanogaster* that were staged 6 and 12 hours following puparium formation. Three independent salivary gland RNA samples were collected from each stage, converted into double-stranded cDNA, and used to synthesize biotinylated cRNA [28]. Affymetrix *Drosophila* Genome Arrays were hybridized, washed, scanned, and analyzed with Affymetrix Micro Array Suite 4.0. Genes were excluded from this study if their hybridization signal was not consistently detected above background (determined by algorithms in Affymetrix Micro Array Suite 4.0) in all 3 replicates of a single time point within an experiment. The fold change values were averaged across the 3 replicates, and p-values were determined by conducting a t-test using average difference values (determined by algorithms in Affymetrix Micro Array Suite 4.0).

Software

Treemap 4.0 (run-treemap.bat) was implemented in Java 2 (J2SE v. 1.4.1 or later) and the parser (run-parser.bat) was written in Perl. Treemap and the parser can be run on

both Windows and Macintosh computers, but Macintosh requires a recent version of system 10 to utilize a suitable version of Java. This software is freely available for non-commercial use (e.g. academic use, evaluation by a single individual in a company, etc.) and can be licensed for commercial use [21]. All data files are parsed and processed into a single tab delimited text file. When started, Treemaps will first load all input data into main memory and subsequently process any necessary computations. For data with less than 20 thousand nodes with each node containing 10 or fewer attributes, the memory requirement is moderate (less than 256 MB on a 32-Bit processor at 700 MHz). The most time consuming step is the initial parsing of the microarray data, the FlyBase gene association flat file, and the file derived from the molecular function, biological process, and cellular component categories of the GO. While the Perl scripts may need at least five minutes to finish parsing all input data and format the tab delimited format, Treemap 4.0 will need only seconds to display the application.

Authors' contributions

EHB and BS conceived the use of treemaps for analyses of genome data. ND and KB developed the parser and modified Treemap 4.0 for use in genomic studies. EHB analyzed the data. EHB and BS wrote the paper. All authors read and approved the manuscript.

Acknowledgements

We thank Catherine Plaisant, Harry Hochheiser, Steve Mount, Jinwook Seo, Emily Clough, Alvaro Godinez, Louisa Wu, and members of the Baehrecke laboratory for helpful comments and discussions. Studies of genome regulation of cell death are supported by NIH grant GM59136 to E.H.B..

References

1. Schena M, et al.: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**:467-470.
2. Link AJ, et al.: **Direct analysis of protein complexes using mass spectrometry.** *Nat Biotechnol* 1999, **17**:676-682.
3. Lockhart DJ, et al.: **Expression monitoring by hybridization to high-density oligonucleotide arrays.** *Nat Biotechnol* 1996, **14**:1675-1680.
4. MacBeath G, Schreiber SL: **Printing proteins as microarrays for high-throughput function determination.** *Science* 2000, **289**:1760-1763.
5. Zhu H, et al.: **Global analysis of protein activities using proteome chips.** *Science* 2001, **293**:2101-2105.
6. Eisen MB, et al.: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863-14868.
7. Consortium GO: **Creating the gene ontology resource: design and implementation.** *Genome Res* 2001, **11**:1425-1433.
8. **GeneOntology** [<http://www.geneontology.org/doc/GO.doc.html>]
9. **Spotfire** [<http://www.spotfire.com/>]
10. **Genespring** [<http://www.silicongenetics.com/cgi/SiG.cgi/index.smf>]
11. Al-Shahrour F, Diaz-Uriarte R, Dopazo J: **FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes.** *Bioinformatics* 2004, **20**:578-580.
12. Zeeberg BR, et al.: **GoMiner: a resource for biological interpretation of genomic and proteomic data.** *Genome Biol* 2003, **4**:R28.
13. Doniger SW, et al.: **MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data.** *Genome Biol* 2003, **4**:R7.
14. **GoSurfer** [<http://biosun1.harvard.edu/complab/gosurfer/>]

15. Shneiderman B: **Tree visualization with tree-maps: A 2-dimensional space filling approach.** *ACM Transactions on Graphics* 1992, **11**:92-99.
16. Bederson B, Shneiderman B, Wattenberg M: **Ordered and quantum treemaps: Making effective use of 2D space to display hierarchies.** *ACM Transactions on Graphics* 2002, **21**:833-854.
17. **Smartmoney** [<http://www.smartmoney.com/marketmap>]
18. **Hivegroup** [<http://www.hivegroup.com>]
19. McConnell P, Johnson K, Lin S: **Applications of Tree-Maps to hierarchical biological data.** *Bioinformatics* 2002, **18**:1278-1279.
20. Lee C-Y, et al.: **Genome-wide analyses of steroid- and radiation-triggered programmed cell death in *Drosophila*.** *Curr Biol* 2003, **13**:350-357.
21. **HCIL** [<http://www.cs.umd.edu/hcil/treemap/TMDemolicense.shtml>]
22. **Flybase** [<http://flybase.bio.indiana.edu/>]
23. Shi Y: **Mechanisms of Caspase Activation and Inhibition during Apoptosis.** *Mol Cell* 2002, **9**:459-470.
24. Baehrecke EH, Thummel CS: **The *Drosophila* E93 gene from the 93F early puff displays stage- and tissue-specific regulation by 20-hydroxyecdysone.** *Dev Biol* 1995, **171**:85-97.
25. Jiang C, Baehrecke EH, Thummel CS: **Steroid regulated programmed cell death during *Drosophila* metamorphosis.** *Development* 1997, **124**:4673-4683.
26. Jiang C, et al.: **A steroid-triggered transcriptional hierarchy controls salivary gland cell death during *Drosophila* metamorphosis.** *Mol Cell* 2000, **5**:445-455.
27. Lee C-Y, et al.: **E93 directs steroid-triggered programmed cell death in *Drosophila*.** *Mol Cell* 2000, **6**:433-443.
28. **Affymetrix** [<http://www.affymetrix.com/index.affx>]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

