Electrophoresis (1999) **20**(18): 3492-507

Exploratory Data Analysis Groupware for Qualitative and Quantitative Electrophoretic Gel Analysis Over the Internet - WebGel

Peter F. Lemkin[1], James M. Myrick[2], Yegappan Lakshmanan[3], Matthew J. Shue[4], James L. Patrick[5], Peter V. Hornbeck[6], Greggory Thornwal[7], Alan W. Partin[4]

Address correspondence to:
[1] Dr. Peter F. Lemkin, Lab. Experimental & Computational Biology, National Cancer Institute, Frederick Cancer Research and Development Center, Building 469 Room 150, Frederick MD 21702, USA. lemkin@ncifcrf.gov

[2] Centers for Disease Control and Prevention, Atlanta, GA
[3] U. Mass. Medical Center, Worcester, MA
[4] The James Buchanan Brady Urological Institute, The Johns Hopkins Hospital, Baltimore, MD
[5] The Dept. of Pathology, Medical College of Ohio, Toledo, OH
[6] PhosphoProtein Databases Inc, Baltimore, MD
[7] Science Applications International Corporation, Frederick Cancer Research and Development Center, Frederick, MD

Submitted: to *Electrophoresis*, June 4, 1999

Running title: Web-based Gel Analysis

Keywords: Web, databases, two-dimensional electrophoresis, Internet, groupware, exploratory data analysis.

Abbreviations and capitalizations:

CC - connected component, i.e. all pixels in a spot or band
CGI - Common Gateway Interface
GIF - Graphics Interchange Format
GSpot - group spot
GUI - graphical user interface
HTML - HyperText Markup Language
Java - a Web programming language
Javascript - a Web browser programming language
OD - optical density
RBP - retinol binding protein

VSpot	*- virtual spot*
Web	*- World Wide Web*
WebGel	
1DE	*- one-dimensional electrophoresis*
2DE	*- two-dimensional electrophoresis*

Abstract

Many scientists use quantitative measurements to compare the presence, and amount of various proteins and nucleotides among series of one- and two-dimensional (1DE and 2DE) electrophoretic gels. These gels are often scanned into digital image files. Gel spots are then quantified using stand-alone analysis software. However, as more research collaborations take place over the Internet, it has become useful to share intermediate quantitative data between researchers. This allows research group members to investigate their data and share their work in progress. We developed a World Wide Web group-accessible software system, WebGel, for interactively exploring qualitative and quantitative differences between electrophoretic gels. Such Internet databases are useful for publishing quantitative data and allow other researchers to explore the data with respect to their own research. Because intermediate results of one user may be shared with their collaborators using WebGel, this form of active data-sharing constitutes a groupware method for enhancing collaborative research.

Quantitative and image gel data from a stand-alone gel image processing system are copied to a database accessible on the WebGel Web server. These data are then available for analysis by the WebGel database program residing on that server. Visualization is critical for better understanding of the data. WebGel helps organize labeled gel images into montages of corresponding spots as seen in these different gels. Various views of multiple gel images, including sets of spots, normalization spots, labeled spots, segmented gels, etc. may also be displayed. These displays are active and may be used for performing database operations directly on individual protein spots by simply clicking on them. Corresponding regions between sets of gels may be visually analyzed using Flicker-comparison (*Electrophoresis* 1997;**10**(2):122-140) as one of the WebGel methods for qualitative analysis. Quantitative exploratory data analysis can be performed by comparing protein concentration values between corresponding spots for multiple samples run in separate gels. These data are then used to generate reports on statistical differences between sets of gels (e.g., between different disease states such as benign or metastatic cancers, etc.).

Using combined visual and quantitative methods, WebGel can help bridge the analysis of dissimilar gels which are difficult to analyze with stand-alone systems and can serve as a collaborative Internet tool in a groupware setting.

1 Introduction

Electrophoretic gels are crucial tools for biological, pharmaceutical and cancer research, and in many other areas of molecular biology for investigating proteins and nucleotides. Although **qualitative** differences measured in these gels are of primary importance, the ability to accurately **quantify** this data is also important. As more research groups concentrate on proteomics and protein function, they require the detection of post-translational modifications of proteins as well as the identification of these proteins and the quantitative differences seen in different disease states [1-2]. These post-translational modifications are often easily observed and quantified using two-dimensional electrophoresis (2DE). Protein identifications are typically performed using other confirmatory techniques such as protein-sequencing, mass spectrophotometry, amino-acid composition, and the use of monoclonal antibodies.

Laboratory 2DE quantitative analyses are often performed using stand-alone (not Internet-based) image processing and database software [3-11] as well as other research and commercial software systems. Some of these systems could be used for exploratory data analysis that can compare protein expression patterns of sets of proteins in large numbers of protein gels while taking into account statistical variations of the data. The stand-alone systems that have this added capability typically incorporate database manipulation, graphics, and statistical analysis as well as the image acquisition, image processing, and spot pairing needed to get the data into a form amenable to this type of exploration. There are also new gel analysis systems that are beginning to integrate their local spot data with data from Web 2D gel databases. These include Melanie-II 2.3 [6], the Carol system [12], and the Michigan leukemia database [13]. These last two systems use Java applets (i.e., a Web browser application) to help process and interface with client-server databases from the user=s Web browser.

Many of these systems can simultaneously compare all corresponding spots across a set of scanned gel images. However, the effectiveness of this procedure depends on accurate spot pairing across gels and careful normalization procedures to avoid staining intensity bias. When the gels are similar, such as with tissue culture samples, this works fairly well. When gels are dissimilar such as with human tissue samples, then the user typically has to manually landmark sets of corresponding spots between gels to adequately define the registration between gels. This may make an automated analysis difficult or impossible.

The additional effort required to build a complete composite gel database may not be necessary when one is only interested in a visual comparison to match a spot in one gel with a previously identified spot in another gel, or in the quantitative comparison of a few spots. We introduce a new method called the WebGel 2DE Web-based exploratory data analysis, which incorporates both qualitative and quantitative comparison capabilities. This paper will be an overview of WebGel. Further details are available on the Web (**http://www.lecb.ncifcrf.gov/webgel**).

Individuals in a research group may decide to share their explorations with other collaborative group members on a read-only basis while not-opening the explorations of the group to the entire Internet. Such a database might be made accessible to the research community at a later time allowing them to share the data with other researchers or to compare it with their own results. This groupware facility also improves communication between members of research teams when is not convenient or possible for them to access a single database at the same location or at the same time. Such was the rationale for the development of the groupware facility in WebGel with one of our earlier prostate cancer databases (Partin et al.) where the research group was dispersed geographically. Since most preliminary research data is proprietary until published, WebGel databases are protected using secure Internet access.

1.1 What is WebGel?

WebGel is an Internet-based, interactive, qualitative and quantitative gel database analysis system. Access to WebGel databases is restricted to collaborative groups with requisite log-in permissions. A WebGel database contains previously quantified gel data generated from a stand-alone quantitative gel analysis system. Such data can be transferred over the Internet using FTP, E-mail, etc. if it is located at a different facility than where the WebGel database is located. Once a database is set up, a user may first visually compare a few gels to find several significant spots of interest. The user can then investigate these selected spots quantitatively, using additional gels in the database, to generate statistical verification or to indicate that the change was consistent across those gels. This exploration process includes defining and manipulating lists of these spots of interest and their associated data, and analyzing these subsets of the data using both images and statistics. WebGel is not meant to be used for performing exhaustive quantitative searches over multi-gel databases, but rather to allow the user to focus on particular spot changes in gel databases where the gels are difficult to compare.

Visual comparison of corresponding spots is performed for two gels at a time using the Flicker system [14-15] which has been incorporated as a subsystem in WebGel. The user first visually aligns the gels in Flicker and then is able to compare the gels. Flicker also has image preprocessing transforms such as contrast enhancement and spatial warping (changing the geometry of one gel to the geometry of the other) which makes it easier to do the image comparison. Aligned spots that differ in intensity appear to pulse when flickered, and therefore call attention to possible significant differences that could be quantified using other WebGel tools. Spots that are missing in one of the two gels appear to blink, so qualitative differences are quite apparent. This is aided by differential-rate flickering which displays one image longer than another and is useful for scrutinizing the gel with the missing spot.

Figure 1 shows the WebGel analysis system and how the users interact with the database through their Web browsers. Project databases must be network-accessible to the Web server where WebGel resides. Generally both are on the same computer. Multiple independent projects can reside on the same server. For example, we currently have more than ten private databases on the same server. Some of these databases include: prostate nuclear matrix proteins (Partin et al);

phosphoprotein signaling pathways (Hornbeck); DNA alterations in the K-*ras* oncogene and DNA alterations in cells exposed to high linear energy transfer particles (Patrick); fetal alcohol syndrome and Rett syndrome databases (Myrick); and others in progress. The privacy of individual projects is protected through project-specific password log-ins.

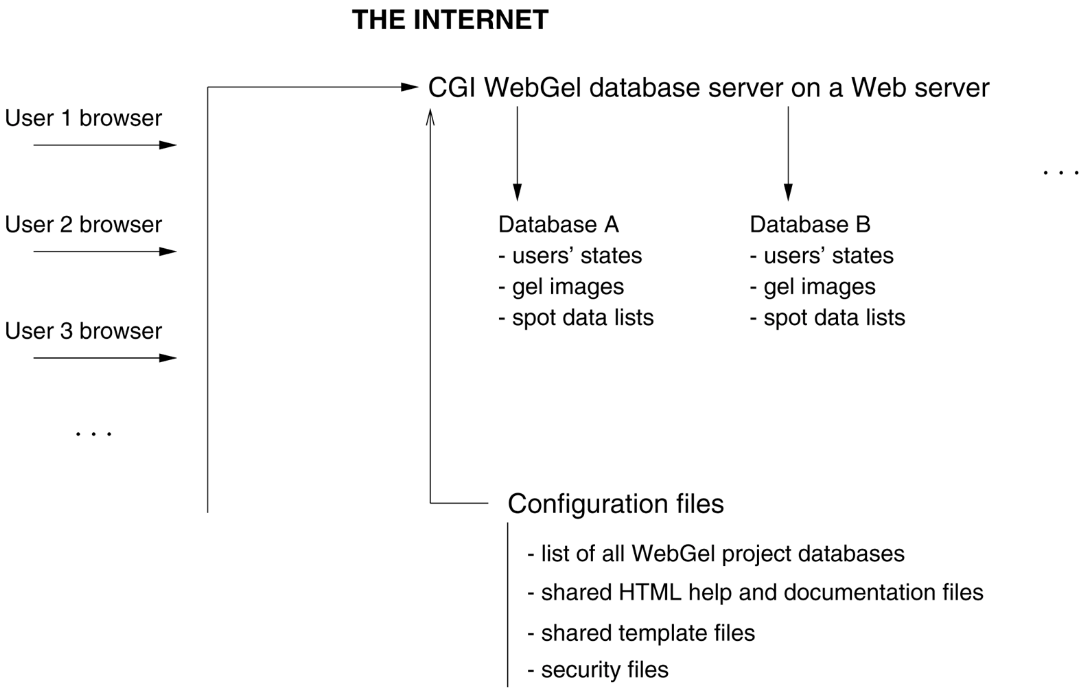


Figure 1 shows the WebGel analysis system schema and how it relates users, the Internet, and multiple project databases. One or more users may access the WebGel Web database from their Web browsers. A Javascript program running in the user=s Web browser is loaded and automatically started when the user accesses the WebGel database Web page. This program gathers and checks the user-specified information and sends it to the WebGel database when a command is requested. The URL they use specifies the particular WebGel database. The underlying WebGel configuration, template, and shared documentation files route users to the correct database. Each project database has its own data in the form of gel images, spot list files and configuration files. Within the project database, each user has data related to their particular exploration.

1.2 Spot numbering within a gel

A spot, in the case of a 2DE protein gel, represents a polypeptide. With 2DE DNA gels created using orthogonal DNA restriction enzymes, a spot is a DNA fragment [18]. WebGel currently handles spot list data from GELLAB-II [3,4], GELLAB-II+ [5], and Melanie-II-2.3 [6] systems.

We refer to a spot's unique identifier within a gel by its CC number. [CC stands for Aconnected component@ meaning all of the pixels constituting that spot.] Each spot in a gel is a CC and spots are generally numbered sequentially in a gel, starting at 1, by most of the stand-alone gel analysis systems. For example, a particular protein such as RBP (retinol binding protein) might be numbered CC 1571 in one gel but CC 1485 in another. The connected component numbers are automatically assigned by the 2D gel image spot quantification program when the spots are found and their quantitative features saved. In GELLAB-II, this spot list data is saved in a Gel Segmentation File (GSF) for each gel. A spot's quantitative features are saved in the GSF and are indexed by its CC. In GELLAB-II+ and in Melanie-II-2.3 they are saved in tab-delimited text file reports. Other stand-alone gel analysis systems call these Spot Lists, Spot List Files, Spot Reports, etc. and may store them as separate files, merged files, proprietary binary files, or in a relational database. We adopt the GSF notation for WebGel, although we can import data from these other systems.

Spot quantification and position data features in WebGel are read from these spot data files generated by stand-alone gel analysis systems. This spot data is generally not normalized between gels, although it may include integrated density corrected for background, D=, computed as *Density - Area* × *MeanBackgroundDensity*. Table 1 shows some examples of the different types of spot reports and the system-dependent spot features. Any other stand-alone spot quantification system could be used with WebGel if the data were transformed to one of these formats.

Table 1. Spot list data formats used by WebGel. This shows examples of the different formats of different spot lists obtained from the stand-alone gel analysis systems that can be read by WebGel. Any other system that produces or can be converted into one of these data formats could be used by WebGel. The minimum data required for use with WebGel is spot number, spot (x,y) position, and spot integrated density or volume. The formats include: **a**) GELLAB-II GSF (Gel Segmentation File) format, **b**) GELLAB-II+ spot report format, and **c**) BioRad Melanie-II gel AExcelList@ spot report file format.

a) NCI GELLAB-II Gel Segmentation File (GSF) spot list data

```
 CC 469 M.E.R[119:122,150:153] D.R.=[.58:.66] D/A= .613 MnB= .556
 1st MOM[120.89, 151.58] A= 12 D= 7.36 D'= .68
 Sx= .96 Sy= 1.11 Sxy= .81 V= 4.98
```

b) Scanalytics GELLAB-II+ gel segmentation spot report file

```
 Spot ID, X Loc, Y Loc, Area, Density, Mean Density, MW,     pI
 47       327    46     46    6.97737  0.151682      64.834  5.25
```

c) BioRad Melanie-II gel spot AExcelList@ spot report file

```
 Gel: ECOLI.mel
 ID VOL      %VOL     AREA     %OD      OD       X          Y
 54 0.033384 0.012056 0.643125 0.208335 0.241000 441.714294 3.904762
```

1.3 WebGel allows the user to manipulate lists of spots

A WebGel spot may be: 1) an actual spot in a spot list imported from one of these gel analysis systems; 2) a virtual spot (VSpot) indicating a position in one gel where there is no spot but which corresponds to a spot in another gel; or 3) a group spot (GSpot) defined as a set of distinct actual spots to be treated as one for quantification and visualization purposes. For example, a set of isoforms (e.g., phosphorylation or glycosylation state) of a protein can be considered to be a group-spot with the total protein concentration being the sum of its components. The individual member spots of the group spot can also be analyzed separately so interpretation of the data one way does not preclude the other.

There are three types of spot lists which may be created and manipulated for each gel. These lists include: 1) a generic list of spots called the Spot List, which is a list of all real CC numbered spots found in the segmented gel; 2) the Ratio-Normalization spot list, which is also a list of CC numbered spots used for computing the spot-ratio normalizations for each gel; and 3) the Spot-Label spot list, which is a list of spot labels which makes the assignment of text names to particular CC numbered spots.

1.4 Manipulating a spot in WebGel

A spot may be specified for use in a report by: 1) its underlying spot number for that gel (this is the CC number); 2) a spot label name that the user had assigned to a CC number in one gel, and that may also be assigned to corresponding but different CC spots in other gels; and 3) clicking on a spot in a gel image after setting the appropriate view action. Many of the WebGel commands consist of performing operations on spots or sets of spots. This allows the user to make quantitative spot measurements and comparisons on corresponding spots where the same spot label was assigned to spots in different gels.

1.5 User interaction with images through View Actions when clicking on a spot

The spot label assignment process described above may be accomplished by clicking on corresponding spots in several gel images. This process is an example of a view action. The user may then display another set of gels they wish to work on and repeat the process until all gels of interest have the same spot label assigned. Each view action can operate on all of the gels being display, although only the particular gel that user clicks on each time. Additional spot labels may be assigned to other spots by changing the default spot label prior to doing the view action. Similarly, there are other view actions that can be taken when clicking on spots in images, and these are listed in Table 3 later in the paper.

1.6 Operations on a set of corresponding spots

Once the same spot label has been assigned in different gels, the user may analyze all of these spots simultaneously. Typically one would want to normalize the density measurements for each gel prior to doing a statistical comparison. This is done by defining a set of corresponding normalization spots for each gel being compared which allows the gels to be compared quantitatively. For example, statistics on a particular spot may be generated for quantitative comparisons between gels using a *t*-test. A montage image may be generated showing the corresponding spot in the gels that have been selected. Normalized data may also be queried for particular spots.

Note that a comparison of a set of gels using a labeled spot compares that *particular* spot across the set of gels. It does not simultaneously compare *all* of the labeled spots across the set of all of the gels. That is what the stand-alone systems attempt to do, but certain gels may require considerably more work for the user because of manual editing to correct mis-pairings made by the computer. For those gels that contain messy data and are difficult to automatically quantify and pair spots, WebGel makes it easier or even possible to analyze these gels. Examples of this type of gel data include: tissue samples taken under varying conditions; some types of cell fractions; gels with large variations in isoelectric point, molecular mass, numbers of spots, or positions of spots. This is especially true when one is interested in qualitative or quantitative comparison of only a few spots.

2 Materials and methods

We now describe the computational model used with WebGel from the point of view of the user and the types of data involved. Because full documentation is available on-line (see Section 5 Software Availability), we will not go into extensive detail in this paper.

2.1 The WebGel computational model

WebGel uses a data transaction model shown in Figure 1. Computation is initially requested by the user from their Web browser. The computation is then performed at the WebGel server which accesses the gel database as needed. Results are then returned and displayed in the Web browser. Since each transaction is a complete event, the user may leave the session at any time after receiving confirmation that the transaction has occurred. Each user has a state file that tracks information necessary to reconstruct where they were in their exploration. Normally, the state is kept invisibly in the browser Web page by the associated Javascript program. However, the user can request that the state be saved in a named state file on the WebGel server. This file can be opened at a later date to recover the state. The user can review the list of saved states and view the data contained in any state, including that of other users who have granted them permission to view their data.

The WebGel method is introduced by describing a typical user session. The user first invokes a particular WebGel database in their browser by entering the URL for one of the WebGel databases and answering the top level group-user name and password security information for the collaborative group. This database-level security procedure limits database access to the collaborative group members who are allowed access.

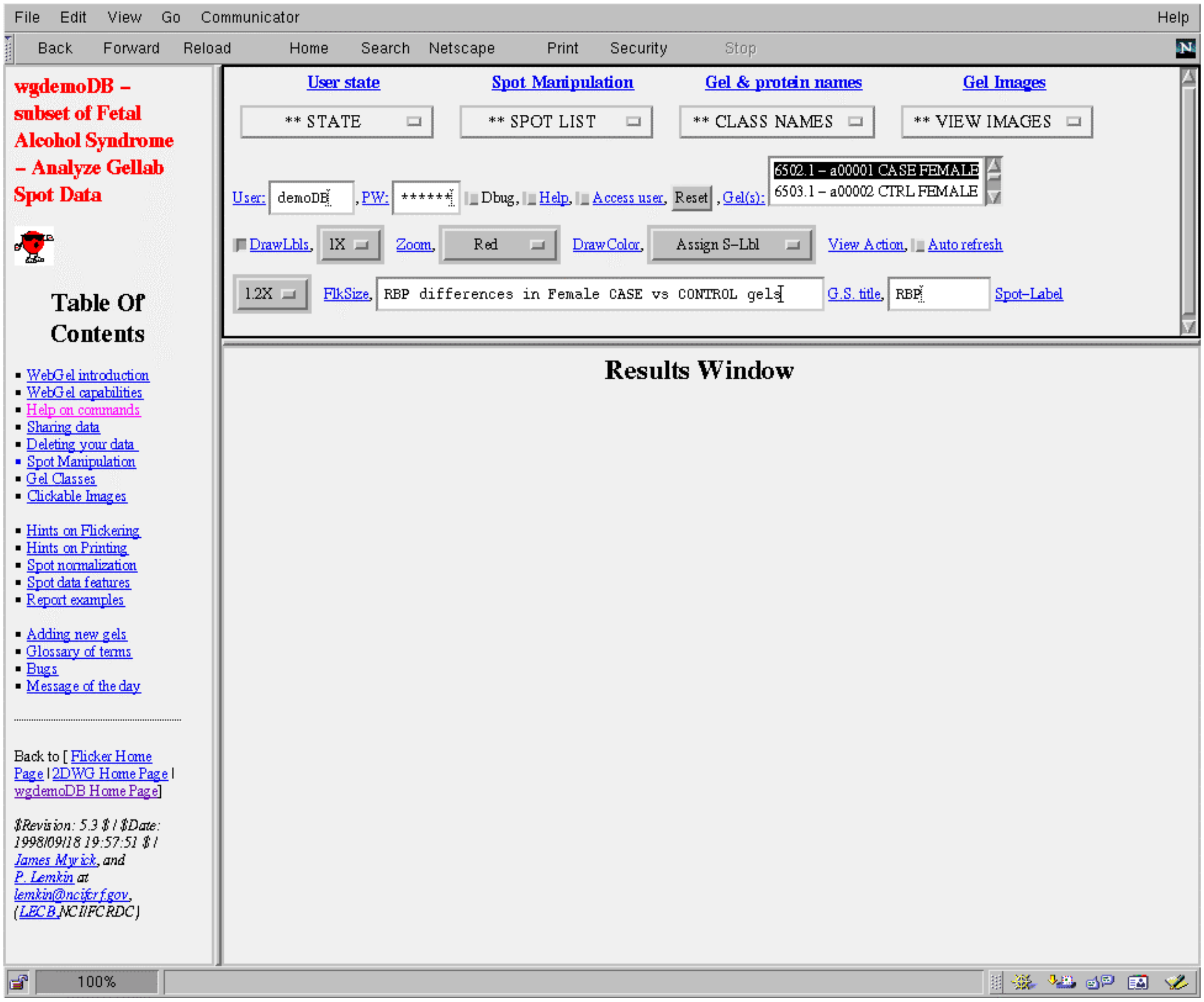
The Web page contains a Javascript program that implements this user interface and checks the consistency of the command data (additional data required by the command) that the user enters before commands are executed. If successful, it forwards the command data to the WebGel Common Gateway Interface (CGI) program running on the Web server. This CGI program is written in the C language. We currently are using WebGel on a SparcStation-20 with the Solaris 2.6 Unix operating system. However, the CGI program should be portable to other Unix systems such as Linux running Web servers such as the Apache Web server.

2.2 Screen organization of the user interface

After initially connecting with the database through their Web browser, the user sees the graphical user interface (GUI) shown in Figure 2. Because multiple researchers in a collaborative group may share the same database, a second level of personal user identification is required (note the AUser@ name and APW@ password type-in areas in the figure) and must be entered before commands are accepted. This allows each user to keep personal exploratory data separate from other members of their collaborative group. Entering a new user name and password will set up a new personal user account provided the user name was not previously used. This two level security scheme corresponds to the Unix group and individual user access methods.

The user interface for WebGel is based on a dynamic Web page consisting of three windows displayed within the user=s browser as adjacent tiled windows in a browser frame. The region on the left side of the screen is a Table of Contents Window for documentation. The region at the top is the Command Window, which consists of pull-down menus of commands, data entry for the group-user name and password, gels selection, view action, etc. Commands are accepted after the user enters their personal user name and password (User, PW) required for tracking the user=s exploration. The results for all computations, images, text reports, tables, documentation, etc. are normally presented in the large window in the lower-right center called the Results Window.

The user then selects a command from one of the four pull-down menus at the top of the Command Window. There are also several check boxes, scrollable selections, and other pull-down menus which are used to modify the commands prior to their being invoked. Additional pop-up dialog windows may appear to request additional information when needed. If the user had requested an operation (such as Display Original Images) which results in images being displayed in the Results Window, then commands may be invoked by clicking on spots in these displayed images.



wgdemoDB – subset of Fetal Alcohol Syndrome – Analyze Gellab Spot Data

Table Of Contents

- WebGel introduction
- WebGel capabilities
- Help on commands
- Sharing data
- Deleting your data
- Spot Manipulation
- Gel Classes
- Clickable Images

- Hints on Flickering
- Hints on Printing
- Spot normalization
- Spot data features
- Report examples

- Adding new gels
- Glossary of terms
- Bugs
- Message of the day

Back to [Flicker Home Page | 2DWG Home Page | wgdemoDB Home Page]

$Revision: 5.3 $ / $Date: 1998/09/18 19:57:51 $ / James Myrick, and P. Lemkin at lemkin@ncifcrf.gov, (LECB,NCI/FCRDC)

User: demoDB, PW: ******, □ Dbug, □ Help, □ Access user, Reset, Gel(s): 6502.1 – a00001 CASE FEMALE 6503.1 – a00002 CTRL FEMALE

■ DrawLbls, 1X Zoom, Red DrawColor, Assign S-Lbl View Action, □ Auto refresh

1.2X FlkSize, RBP differences in Female CASE vs CONTROL gels G.S. title, RBP Spot-Label

Results Window

Figure 2 is a screen view showing the organization of the WebGel graphical user interface when it first comes up in the user's Web browser. The region on the left side of the screen is a Table of Contents Window for documentation. The top region is the Command Window that consists of the user log-in area, pull-down menus of commands, data entry text area for the user name and password, gel state title, and spot label. There are a number of selection menus including gel selection, montage zoom size, drawing color, view action and flicker window size. There are also a number of check box options including help on individual commands, accessing other users= data, drawing labels in gel images, etc. Commands may be selected after the user enters a personal (user name, password) that tracks the user=s particular explorations. The results for all computations, images, text and tables are presented in the large window in the lower right which is called the Results Window. At times, a second pop-up Results Window may be created for use with clickable images. Additional reports may appear in a separate pop-up window (not shown). Because the user does not need to see all of the command controls once they are set, the browser windows can be resized to gain more real estate in the Results Window when viewing many images simultaneously.

2.3 Command processing

The user normally selects a command from one of the pull-down menus in the Command Window. The Command Window also contains a scrollable list of all gels in the database from which the user selects the subset of gels to be analyzed. This list is labeled AGel(s)@ in Figure 2. The Gel(s) scrollar, as well as the List Gels and the Project Documentation commands shows the gel study information, and gel image file names associated with the gels.

The Command Window contains a pull-down menu of the possible view actions to be taken when the user clicks in an active image. An active image is created by some of the commands (e.g., Show Spot Data or Assign Spot Label, etc. commands from the Gel Image menu). The user selects the view action they wish to associate with the displayed images. This is done prior to invoking the command to display those images (see Table 2). Clicking on a spot in one of the resulting gel images invokes the view action command on the spot the user clicked on or on the image as a whole, e.g., Assign Gel Class. Computed results, images, or other information is then sent back to the Web browser and displayed in the Results Window. Some view action operations will pop-up a second results window to display additional information. For example, it might show the quantified data for a particular spot, summary statistical results for a particular labeled spot in several gels, notification that a spot has been added to the normalization spot list, or the default spot label has been assigned to the spot the user clicked on. By keeping the active gel images on the screen, they can be used repeatedly with different spots in different gels. This is much more efficient than constantly reloading these images from the Web server each time.

When processing a command that requires gel data, the Web server may read the data from one or more spot list or gel image files into the memory of the CGI program for all of the specified gels. Since it is computationally expensive to read this data, it is only done when required for particular commands and only for that subset of gels which are being analyzed. Although reprocessing the gel spot list data each time puts a burden on the database server, it makes processing and creation of new databases very simple. Given the reasonably low level of computational traffic on the server for a small number of collaborative groups, we have not found this method of processing to be a problem with our shared Web server. Since computers are becoming much faster and memory much less expensive, adding additional users and databases could be handled by using a more powerful, dedicated Web server.

Table 2. List of view actions available for clickable images. The user first selects a view action option from the pull-down menu in the Control Window. This indicates the action to be taken when the user clicks on a spot in a clickable gel image after it has been displayed using either a View or Draw command from the Gel Images menu.

```
No action (the default)
Show spot data
Show Swiss-Prot ID
Add spot to Spot-List
Add spot to Ratio-Normalization list
Add spot as a member of the Group-Spot
Delete spot from Spot-List
Delete spot from Ratio-Normalization list
Delete spot Virtual-Spot at (x,y) position
Delete spot from Group-Spot
Assign spot default Spot-Label to spot
Assign Virtual-Spot to (x,y) position
Assign spot default Class name to gel
```

2.3.1 WebGel command menus

Detailed descriptions of the commands for each of the pull-down menus are described in Web pages accessed from the hypertext links above the pull-down menus. Commands are grouped by function in the four pull-down menus. Because the details of these menus can be readily investigated from WebGel when it is running, and there is on-line hypertext documentation for the commands, we will not be discussing individual commands in this paper.

The User State menu includes commands for manipulating the gel state, accessing other users, deleting the account, and manipulating notes. The Spot Manipulation menu includes commands for manipulating the Spot List, Ratio-Normalization spot list, and various operations on Spot Labels including statistical tests. The Gel & Protein Names menu includes commands for manipulating gel experimental class names and Swiss-Prot database links to proteins in a user's WebGel database. The Gel Images menu includes commands for viewing gels in various ways such as looking at the original data and segmented images, drawing images with various spot lists and spot labels, and generating montages of particular spots seen across a set of gels. It also contains commands for viewing the raw spot lists.

WebGel also has additional on-line help including a hypertext-linked glossary. The hypertext-linked user reference manual is integrated into the Table of Contents window. A useful way to begin is through the Table of Contents and the descriptions above each of the pull-down menus.

To get help on an individual command, WebGel may be put into AHelp@ mode. First select the Help check box in the Command Window. Then select a command from one of the pull-down menus. Instead of executing the command, WebGel will display information on that specific command in the Results Window. The user then clears the Help check box when they want to execute commands again.

2.4 Use of Flicker as a visual method of finding significant spots of interest

WebGel can help the user visually compare any two gel images in its database using the Web-based Flicker gel comparison system. This allows the user to easily specify the two gels to be flickered directly from the list of gels in the WebGel database rather than indirectly through the Flicker home page (**http://www.lecb.ncifcrf.gov/flicker**). Flicker is a Java applet that runs in Web browsers, and when invoked in WebGel, it appears in the Results Window. To Flicker two gels, the user first selects two gels from the list of gels and then selects the Flicker command in the Gel Images menu. The view action is integrated with Flicker when it is used with WebGel, so that clicking on a spot in one of the images executes that view action. This makes it easier to interact with the database while doing a flicker comparison. For example, when using view actions with Flicker, the user could assign a spot label for two gels that are difficult to compare.

2.5 Saving the user=s data explorations: the user and gel state subsets

The Control Window settings that a particular user specifies at any time during an exploratory data analysis session is called the current user gel state and is saved in the user=s default gel state on the WebGel server. This gel state includes a list of working gels (i.e., gels currently being analyzed), lists of spot labels and normalization spots for each gel, gel experimental class names, the subset of gels being explored, various parameters, and a title describing the data. At any point, the user may save their current gel state as a gel state subset. Any gel state subset may be opened later to become the new current user gel state. The user can then save it again under the same name or a new name after making additional changes. Any user in a collaborative group for a particular database can access another user's explorations (gel state subsets) if they were granted access by the other user. In addition to the gel state subsets, there is a user state that contains a directory of all the gel state subsets including their titles. Users may delete their user state and/or gel state subsets in future sessions. These states are saved as files in a protected database area on the Web server.

The current gel state can be restored using the Open Gel State command from any of the user's gel state subsets or from those of their collaborators in the group if they are allowed access. The User State command provides a directory of the gel state subsets previously saved. Table 3 shows an example of exploratory data saved in a gel state subset.

Table 3. Example of a user=s gel state. This shows the data in a user=s gel state which saves the state of a particular database exploration contains experimental class name, CC spot mapping data for normalization spots, spot labels and other information for each gel being investigated. Note the gels listed are those currently under investigation and are a subset of all of the gels in the WebGel database. These correspondences enable the user to compare spots across gels. The spot list data imported from stand-alone gel analysis systems is not stored in the gel state, but rather refers to CC spot numbers in those spot list files.

```
Gel-State-Title[RBP differences in Female CASE/CONTROL]05/24/1999, 04:49:07PM
#
Working_Gels[6502.1,6507.1,6510.1,6612.1]
#
Class-Name[6502.1]FEMALE CASE
Spot-List[6502.1]
Ratio-Norm-SL[6502.1]604,613,624,757,1129,1133,1034,1035,1028,1029,1013,1014
Slabel[6502.1]Alpha-1-antitrypsin=638,RBP=1571
Virtual-Spots[6502.1]
Group-Spots[6502.1]
# ---
Class-Name[6507.1]FEMALE CONTROL
Spot-List[6507.1]
Ratio-Norm-SL[6507.1]529,538,547,576,534,671,1040,1050,930,931,932,933,934,935
Slabel[6507.1]Alpha-1-antitrypsin=584,RBP=1485
Virtual-Spots[6507.1]
Group-Spots[6507.1]
# ---
Class-Name[6510.1]FEMALE CASE
Spot-List[6510.1]
Ratio-Norm-
SL[6510.1]567,575,585,618,720,1145,1156,1038,1039,1032,1033,1027,1028
Slabel[6510.1]Alpha-1-antitrypsin=619,RBP=1705
Virtual-Spots[6510.1]
Group-Spots[6510.1]
# ---
Class-Name[6612.1]FEMALE CONTROL
Spot-List[6612.1]
Ratio-Norm-SL[6612.1]589,593,604,651,718,1068,966,967,960,954,955,956,1062
Slabel[6612.1]Alpha-1-antitrypsin=597,RBP=1584
Virtual-Spots[6612.1]
Group-Spots[6612.1]
# ---
#
#
# --- Command Window GUI state --
Zoom-Montage[1]
Drawing-Color[RED]
Draw-Labels-Text[on]
ViewAction[GS]
Flicker-Canvas-Size[220]
```

2.6 Organization of a project database

Figure 3 illustrates the organizational structure of data for a specific project database in WebGel. Data within a project database is organized first by original gel data and then by exploratory data the user creates during ongoing analysis sessions. Original data includes sets of gel images and their corresponding spot lists. This is fixed except when gels are permanently added to or removed from the database. Exploratory data includes the user state and gel state files for all users in the collaborative group. This data will grow as new gel state subsets are generated and users are added.

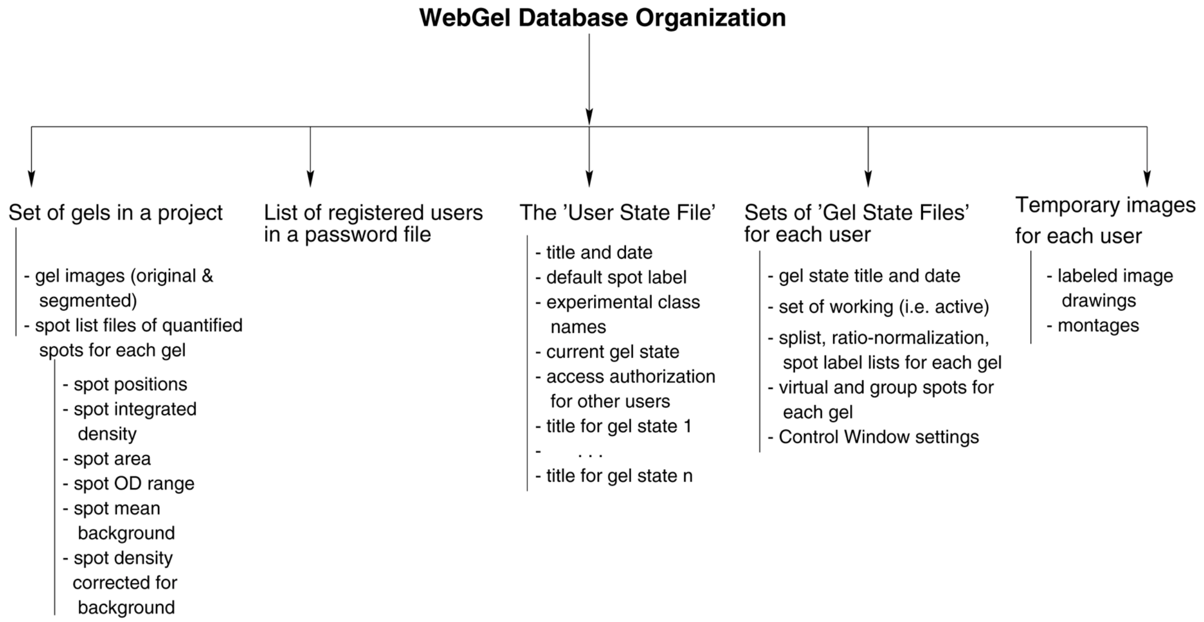

WebGel Database Organization
Set of gels in a project
- gel images (original & segmented)
- spot list files of quantified spots for each gel
- spot positions
- spot integrated density
- spot area
- spot OD range
- spot mean background
- spot density corrected for background
List of registered users in a password file
The 'User State File'
- title and date
- default spot label
- experimental class names
- current gel state
- access authorization for other users
- title for gel state 1
- . . .
- title for gel state n
Sets of 'Gel State Files' for each user
- gel state title and date
- set of working (i.e. active)
- splist, ratio-normalization, spot label lists for each gel
- virtual and group spots for each gel
- Control Window settings
Temporary images for each user
- labeled image drawings
- montages

Figure 3 shows the organizational structure of data for a specific project database in WebGel. Data is organized by the original data and by the user=s subsequent exploratory data. Original data includes sets of gel images, their corresponding spot lists, and database configuration files. Exploratory data includes the gel state and user state files.

2.7 Spot data normalization

When protein gels are stained stoichiometrically or run as autoradiographs and then scanned with a non-saturated images scanned with a calibrated scanner, the integrated density of a spot will correspond linearly within limits to the protein concentration in the gel. In order to compare corresponding spot quantification between gels, the data must be normalized. This may be done in two ways percent and ratio normalization, with the latter being preferred. Normalization depends on all gel samples having the same magnification when scanned and the gels being within the dynamic range of the stain or autoradiograph and scanner.

The percent normalization method computes the percent of total density of a spot relative to all of the other spots in the same gel. It is computed as $D_{\%}(j,g)$ for any CC spot *j* in gel *g* in equation (1) taking all spots *k* in the gel into account.

$$D_{\%}(j,g) = (100\% \, D_{=}(j,g)) \; / \sum_{\text{for all k spots}} D_{=}(k,g) \qquad (1)$$

The ratio-normalization (RN) method uses a set of relatively constant spots common to all of the gels to compute a normalization density for each gel. The set of RN spots is the set of CC spots used to compute a sum of integrated density for each gel *g*. This sum is then used in computing a normalized density Dr(j,g) for any CC spot *j* for the same gel *g* in equations (2-3). The main point is that the ratio-normalization spots are corresponding spots that are selected in each of the gels being compared. In general, the more spots in the RN list, the better the normalization. If too few spots are used, the normalization may be dependent on spots which change with the experimental conditions. Therefore, having a reasonable number of consistent normalization spots is critical to a good normalization.

$$Sr(g) = \sum_{\text{k in RN list}} D_{=}(k,g) \qquad (2)$$

$$Dr(j,g) = (100 \, D_{=}(j,g)) \; / \; Sr(g) \qquad (3)$$

2.8 Spot reports

Spot data reports may be generated for any spot or set of spots in a gel or for a set of gels. If the gels were normalized by a set of corresponding normalization spots, then WebGel will use the Dr(j,g) data in its calculations that then appear as *Dr* in the reports. Otherwise, it will use percent of total density for each gel.

There are a number of different types of spot reports. To give the flavor of how one might generate a report, we review the procedure used to generate one of these - the normalized spot data report. 1) First set the selection in the view action pull-down menu to Add to Ratio-Normalization. This will be used in step (3). 2) Then display the subset of gels to be analyzed using one of the View or Draw commands in the View Gels menu. This will generate the clickable images in the Results Window. 3) Click on the same set of spots in each gel to add them to the ratio-normalization list for these gels. This is used to compute Sr(g) shown in equation 2. 4) Next set the default spot label to a particular name (e.g., the RBP spot) using the Set Default Spot Label in the Spot Manipulation menu, and change the view action to Assign Spot Label. This will be used in the next two steps to assign the spot RBP spot label in all of the gels. 5) Regenerate a clickable map using one of the View or Draw commands. 6) Add the corresponding spots in the different gels, clicking on the same spot in each gel. 7) At this point, the user may generate a statistical report using the Test by Spot-Label command in the Spot

Manipulation menu. This report compares the data for the corresponding spot across multiple gels. Spot density will be reported as ratio normalized density.

2.9 Image displays

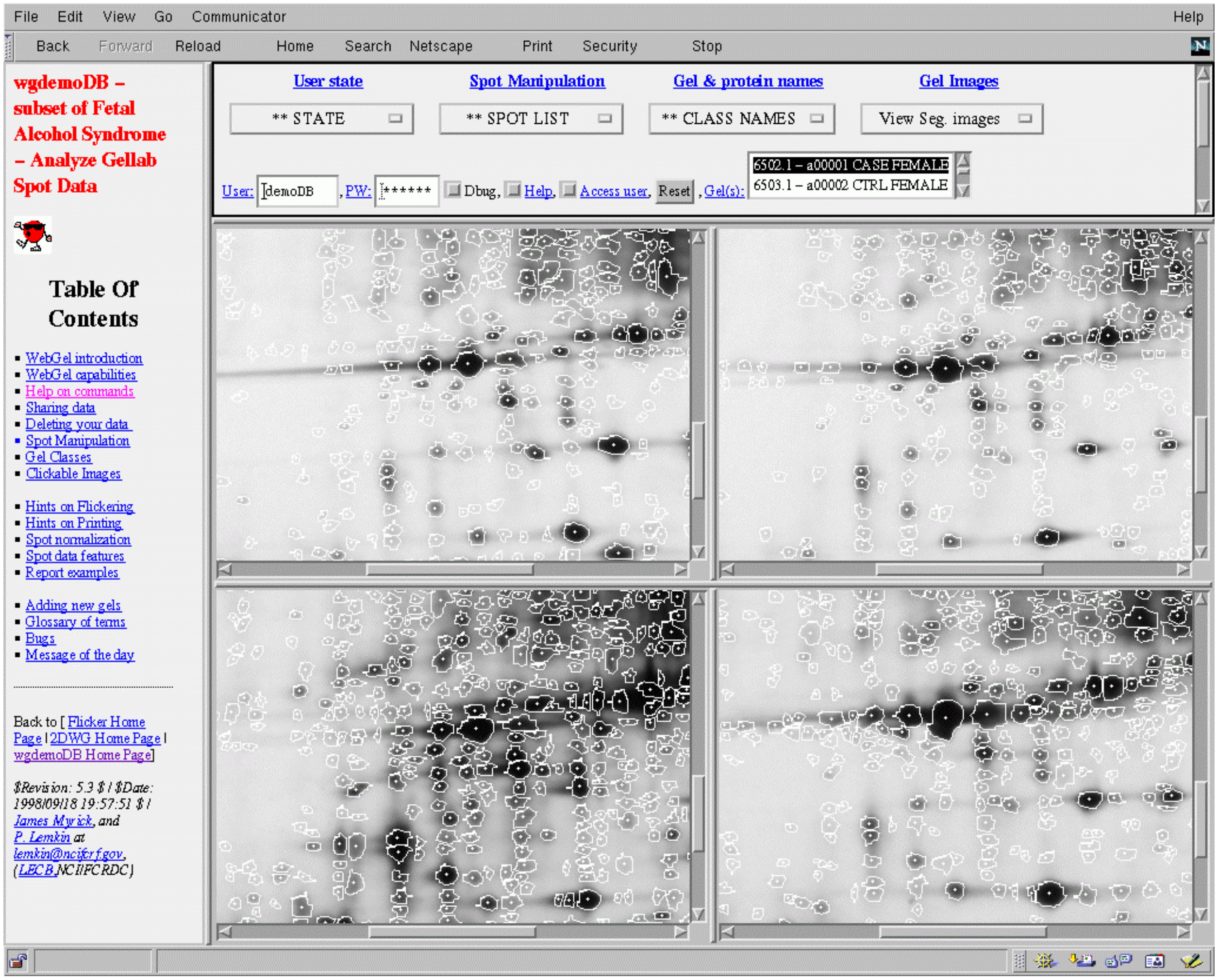
Many different types of images may be displayed including variants of the original data such as images with spot boundaries, the gel image with the spots subtracted. The latter is useful for verifying the quality of the spot segmentation. Other displays include drawings of where spot lists, ratio-normalization spots and spot labels, are located in the gels. The montage display is a static image that is a more compact form for viewing subregions around labeled spots. Montage images are subregions of a set of gels around a particular spot and are useful for verifying qualitative changes or that the correct spot is being analyzed. WebGel also allows one to track a set of spots with no labels for possible subsequent investigation. All of these displayed images may be used with the view action.

3 Results

We have applied WebGel to several databases of 2DE protein gels and of 2D DNA gels. These include a prostate cancer tissue nuclear matrix protein database (Partin et al. [19-21]); a fetal alcohol syndrome (FAS) database of serum proteins (Robinson et al. [22]); a Rett Syndrome database (Myrick); DNA alterations in the K-*ras* oncogene [18] and cell lines exposed to high linear energy transfer particles databases (Patrick et al. [18]); and a phosphoprotein database exploring different parts of the cell cycle (Hornbeck et al. [23]).

We illustrate parts of WebGel in this paper using some of the data from the FAS database [22]. Other results, reported elsewhere, include possible prostate cancer tumor markers from a nuclear matrix protein database. A protein called YL-1 with four isoforms has been found and confirmed with WebGel that increased in the aggressive forms of prostate cancer with respect to normal prostate tissue (Lakshmanan et al. [24]). We also found two proteins we call RPC-1a and RPC-1b that are reduced in the aggressive form of prostate cancer with respect to the normal prostate tissue and are reported in Lemkin et al. [25]. The ratio of these two proteins, which appear to be isoforms, changes significantly with the aggressive form of prostate cancer. Such protein changes may be useful cancer tumor markers to improve staging and treatment protocols for prostate cancer patients.

Figure 4 shows the four segmented gel images from the FAS database [22] consisting of two female Case (left) and two Control (right) samples. The segmented images show the boundaries of the spots that were found using the stand-alone spot segmentation software. Spot position, integrated density, background-corrected density, area, and other features measured for each spot are available for interrogation with WebGel.



wgdemoDB – subset of Fetal Alcohol Syndrome – Analyze Gellab Spot Data

Table Of Contents

- WebGel introduction
- WebGel capabilities
- Help on commands
- Sharing data
- Deleting your data
- Spot Manipulation
- Gel Classes
- Clickable Images

- Hints on Flickering
- Hints on Printing
- Spot normalization
- Spot data features
- Report examples

- Adding new gels
- Glossary of terms
- Bugs
- Message of the day

Back to [Flicker Home Page | 2DWG Home Page | wgdemoDB Home Page]

$Revision: 5.3 $ / $Date: 1998/09/18 19:57:51 $ / James Myrick, and P. Lemkin at lemkin@ncifcrf.gov, (LECB,NCI/FCRDC)

** STATE

** SPOT LIST

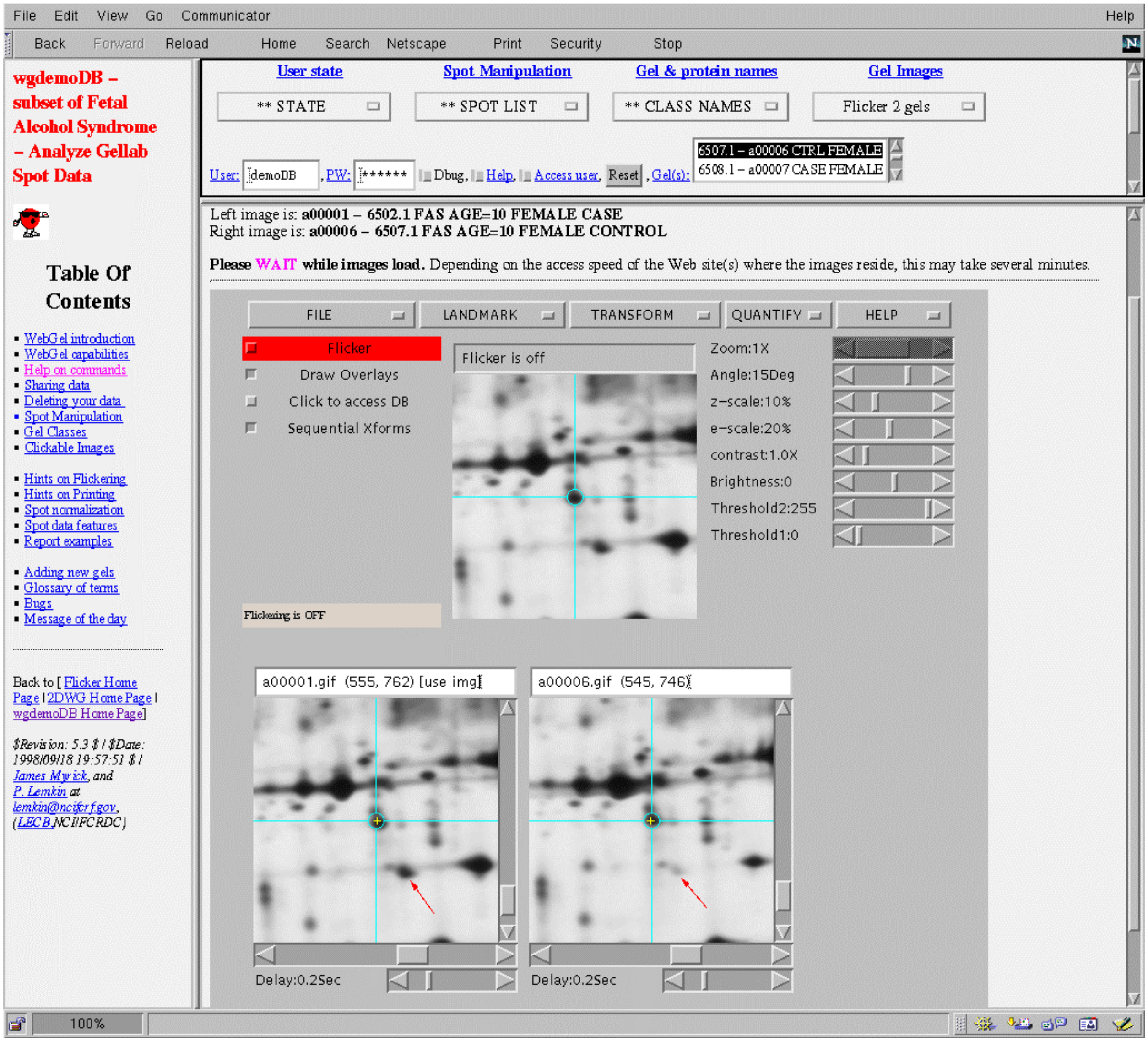
** CLASS NAMES

View Seg. images

User: demoDB , PW: ****** □ Dbug, □ Help, □ Access user, Reset , Gel(s):
6502.1 – a00001 CASE FEMALE
6503.1 – a00002 CTRL FEMALE

Figure 4 is a screen view showing four of the segmented gel images from the fetal alcohol syndrome (FAS) database [RobM95]. The two gels in the left column are Female-Case and the two on the right column are Female-Control. The stand-alone gel analysis system (in this case GELLAB-II) was used to segment and quantify of the spots in each of the original gels and generate the spot quantification lists used by WebGel.

Figure 5 shows a flicker comparison of a Case (left) and a Control (right) gel. The scrollable windows were centered on the retinol binding protein RBP which was found to be one of the marker proteins increased in fetal alcohol syndrome [22]. Also, note the spot in the lower right quadrant with the arrow. This spot is much lighter in the Control gel than in the Case gel indicating that there is probably a significant quantitative change. This spot might be a good candidate for further investigation across other gels in the database after normalizing the integrated density in those gels. In general, this method of first visually looking for differences and then quantitatively checking their validity across more gels is one of the ways WebGel might be used to suggest spots for quantitative follow up.



wgdemoDB – subset of Fetal Alcohol Syndrome – Analyze Gellab Spot Data

User: demoDB, PW: ******

Gel(s): 6507.1 – a00006 CTRL FEMALE
6508.1 – a00007 CASE FEMALE

Left image is: a00001 – 6502.1 FAS AGE=10 FEMALE CASE
Right image is: a00006 – 6507.1 FAS AGE=10 FEMALE CONTROL

Please WAIT while images load. Depending on the access speed of the Web site(s) where the images reside, this may take several minutes.

Table Of Contents

- WebGel introduction
- WebGel capabilities
- Help on commands
- Sharing data
- Deleting your data
- Spot Manipulation
- Gel Classes
- Clickable Images

- Hints on Flickering
- Hints on Printing
- Spot normalization
- Spot data features
- Report examples

- Adding new gels
- Glossary of terms
- Bugs
- Message of the day

Back to [Flicker Home Page | 2DWG Home Page | wgdemoDB Home Page]

$Revision: 5.3 $ / $Date: 1998/09/18 19:57:51 $ / James Myrick, and P. Lemkin at lemkin@ncifcrf.gov, (LECB,NCI/FCRDC)

Flicker is off

Flickering is OFF

Zoom:1X
Angle:15Deg
z-scale:10%
e-scale:20%
contrast:1.0X
Brightness:0
Threshold2:255
Threshold1:0

a00001.gif (555, 762) [use img]

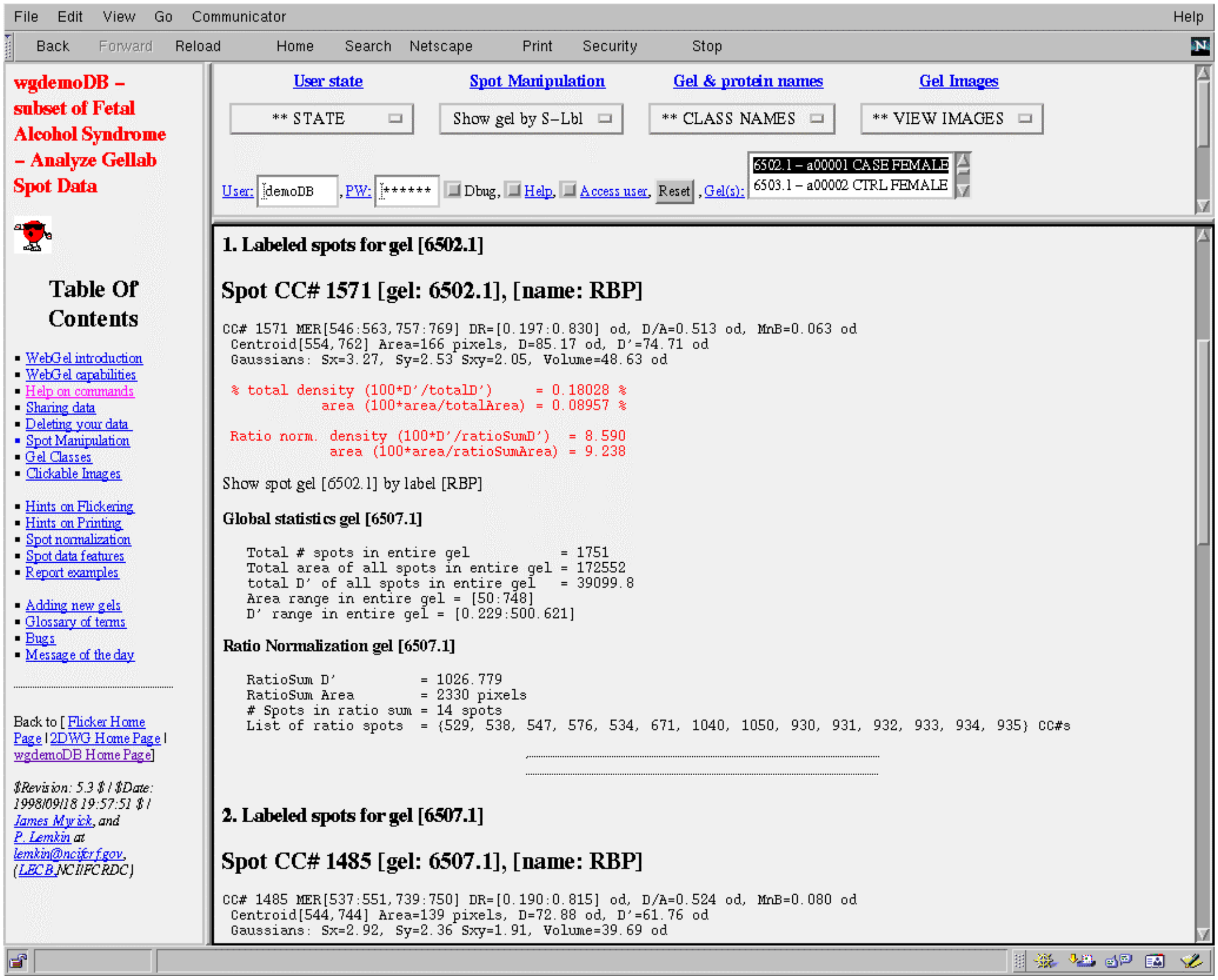
a00006.gif (545, 746)

Delay:0.2Sec

Delay:0.2Sec

Figure 5 is a screen view showing the Flicker comparison of two gels (Case and Control) from the FAS database. The gel images were scrolled so that the RBP protein was roughly centered. Flickering indicated that there was a significant qualitative difference on a spot to the lower right of the RBP (indicated by arrows). This difference could then be followed up by labeling that spot in all of the gels and generating statistics on this normalized spot or by generating a montage image.

Figure 6 shows part of a detailed spot data report for the gels for which the RBP spot label was assigned. It demonstrates features extracted from the stand-alone gel analysis system=s spot list files. Figure 7 shows a table of corresponding spots with the same RBP spot label among a set of gels where the statistical tests have been applied. Integrated density is reported as both raw density and density corrected for background (D'). The D' density is normalized as percent of total density for each gel and reported as $D_{\%}$. Because the ratio-normalization was defined for these gels, it reports the data using the ratio-normalization value, Drn, in the statistics calculations. Figure 8 shows a gel subregion montage of the RBP spot across a subset of the FAS female Case and Control samples. Because the gel panels are order left to right by decreasing normalized density, Drn, the user gets immediate visual feedback if any gels are out of ordered.


File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
User state
Spot Manipulation
Gel & protein names
Gel Images
** STATE
Show gel by S–Lbl
** CLASS NAMES
** VIEW IMAGES
User: demoDB, PW: ******
Dbug, Help, Access user, Reset, Gel(s):
6502.1 – a00001 CASE FEMALE
6503.1 – a00002 CTRL FEMALE

wgdemoDB – subset of Fetal Alcohol Syndrome – Analyze Gellab Spot Data

Table Of Contents

- WebGel introduction
- WebGel capabilities
- Help on commands
- Sharing data
- Deleting your data
- Spot Manipulation
- Gel Classes
- Clickable Images

- Hints on Flickering
- Hints on Printing
- Spot normalization
- Spot data features
- Report examples

- Adding new gels
- Glossary of terms
- Bugs
- Message of the day

Back to [Flicker Home Page | 2DWG Home Page | wgdemoDB Home Page]

$Revision: 5.3 $ / $Date: 1998/09/18 19:57:51 $ / James Myrick, and P. Lemkin at lemkin@ncifcrf.gov, (LECB NCI/FCRDC)

1. Labeled spots for gel [6502.1]

Spot CC# 1571 [gel: 6502.1], [name: RBP]

```
CC# 1571 MER[546:563,757:769] DR=[0.197:0.830] od, D/A=0.513 od, MnB=0.063 od
 Centroid[554,762] Area=166 pixels, D=85.17 od, D'=74.71 od
 Gaussians: Sx=3.27, Sy=2.53 Sxy=2.05, Volume=48.63 od

 % total density (100*D'/totalD')      = 0.18028 %
          area (100*area/totalArea) = 0.08957 %

 Ratio norm. density (100*D'/ratioSumD')  = 8.590
            area (100*area/ratioSumArea) = 9.238
```

Show spot gel [6502.1] by label [RBP]

Global statistics gel [6507.1]

```
   Total # spots in entire gel           = 1751
   Total area of all spots in entire gel = 172552
   total D' of all spots in entire gel   = 39099.8
   Area range in entire gel = [50:748]
   D' range in entire gel = [0.229:500.621]
```

Ratio Normalization gel [6507.1]

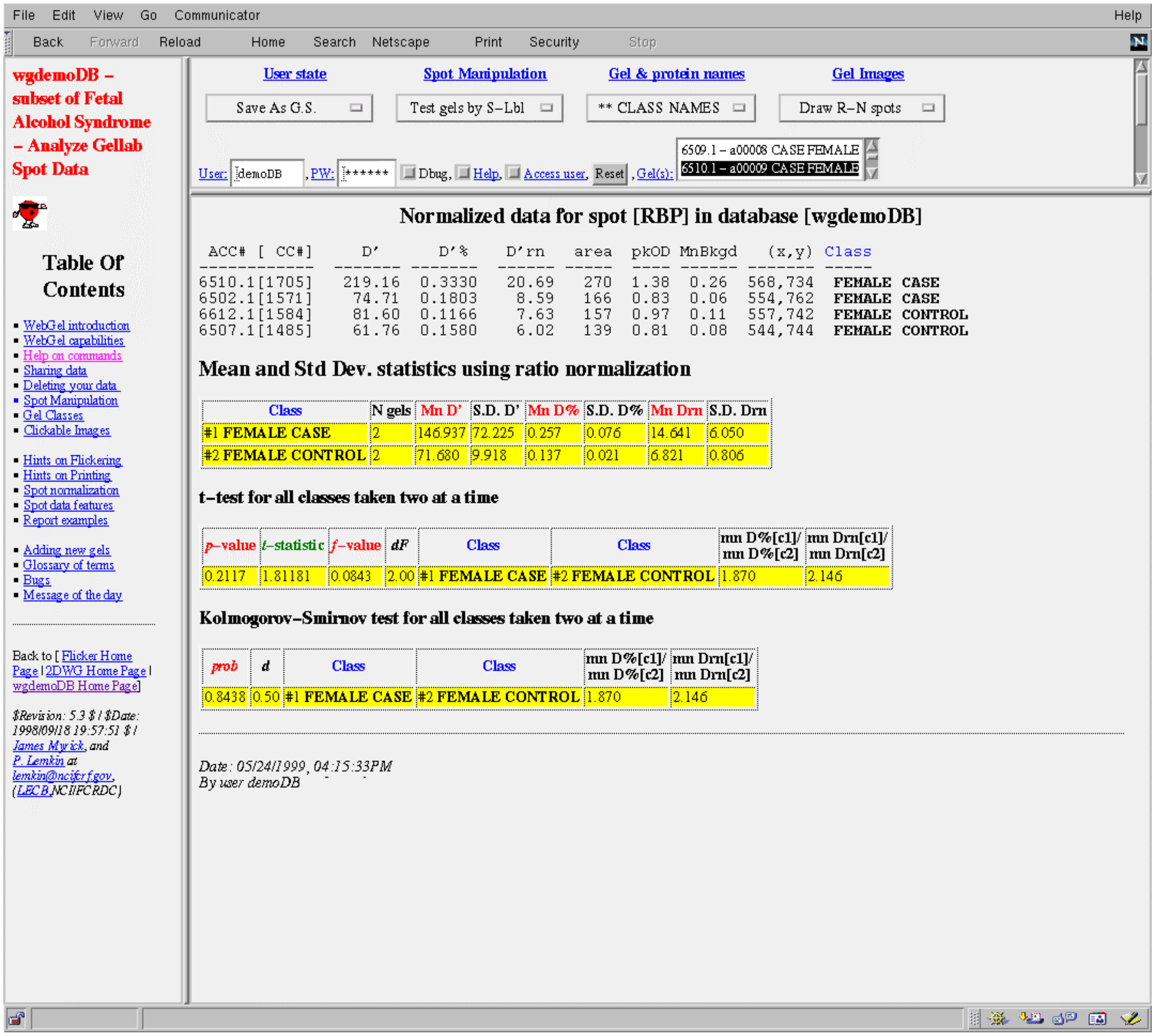
```
   RatioSum D'          = 1026.779
   RatioSum Area        = 2330 pixels
   # Spots in ratio sum = 14 spots
   List of ratio spots  = {529, 538, 547, 576, 534, 671, 1040, 1050, 930, 931, 932, 933, 934, 935} CC#s
```

2. Labeled spots for gel [6507.1]

Spot CC# 1485 [gel: 6507.1], [name: RBP]

```
CC# 1485 MER[537:551,739:750] DR=[0.190:0.815] od, D/A=0.524 od, MnB=0.080 od
 Centroid[544,744] Area=139 pixels, D=72.88 od, D'=61.76 od
 Gaussians: Sx=2.92, Sy=2.36 Sxy=1.91, Volume=39.69 od
```

Figure 6 is a screen view showing a detailed spot data report in the gels in the FAS database to which the RBP spot-label was assigned. It contains the full set of features extracted from the stand-alone gel analysis system spot list file.



User: demoDB , PW: ****** , Gel(s):
6509.1 – a00008 CASE FEMALE
6510.1 – a00009 CASE FEMALE

wgdemoDB – subset of Fetal Alcohol Syndrome – Analyze Gellab Spot Data

Table Of Contents

Back to [Flicker Home Page | 2DWG Home Page | wgdemoDB Home Page]

$Revision: 5.3 $ / $Date: 1998/09/18 19:57:51 $ / James Myrick, and P. Lemkin at lemkin@ncifcrf.gov, (LECB NCI/FCRDC)

Normalized data for spot [RBP] in database [wgdemoDB]

```
 ACC# [ CC#]      D'      D'%      D'rn   area  pkOD MnBkgd   (x,y) Class
------------  ------- -------  ------ -----  ---- ------ ------- -----
6510.1[1705]   219.16  0.3330   20.69   270  1.38   0.26 568,734  FEMALE CASE
6502.1[1571]    74.71  0.1803    8.59   166  0.83   0.06 554,762  FEMALE CASE
6612.1[1584]    81.60  0.1166    7.63   157  0.97   0.11 557,742  FEMALE CONTROL
6507.1[1485]    61.76  0.1580    6.02   139  0.81   0.08 544,744  FEMALE CONTROL
```

Mean and Std Dev. statistics using ratio normalization

Class	N gels	Mn D'	S.D. D'	Mn D%	S.D. D%	Mn Drn	S.D. Drn
#1 FEMALE CASE	2	146.937	72.225	0.257	0.076	14.641	6.050
#2 FEMALE CONTROL	2	71.680	9.918	0.137	0.021	6.821	0.806

t–test for all classes taken two at a time

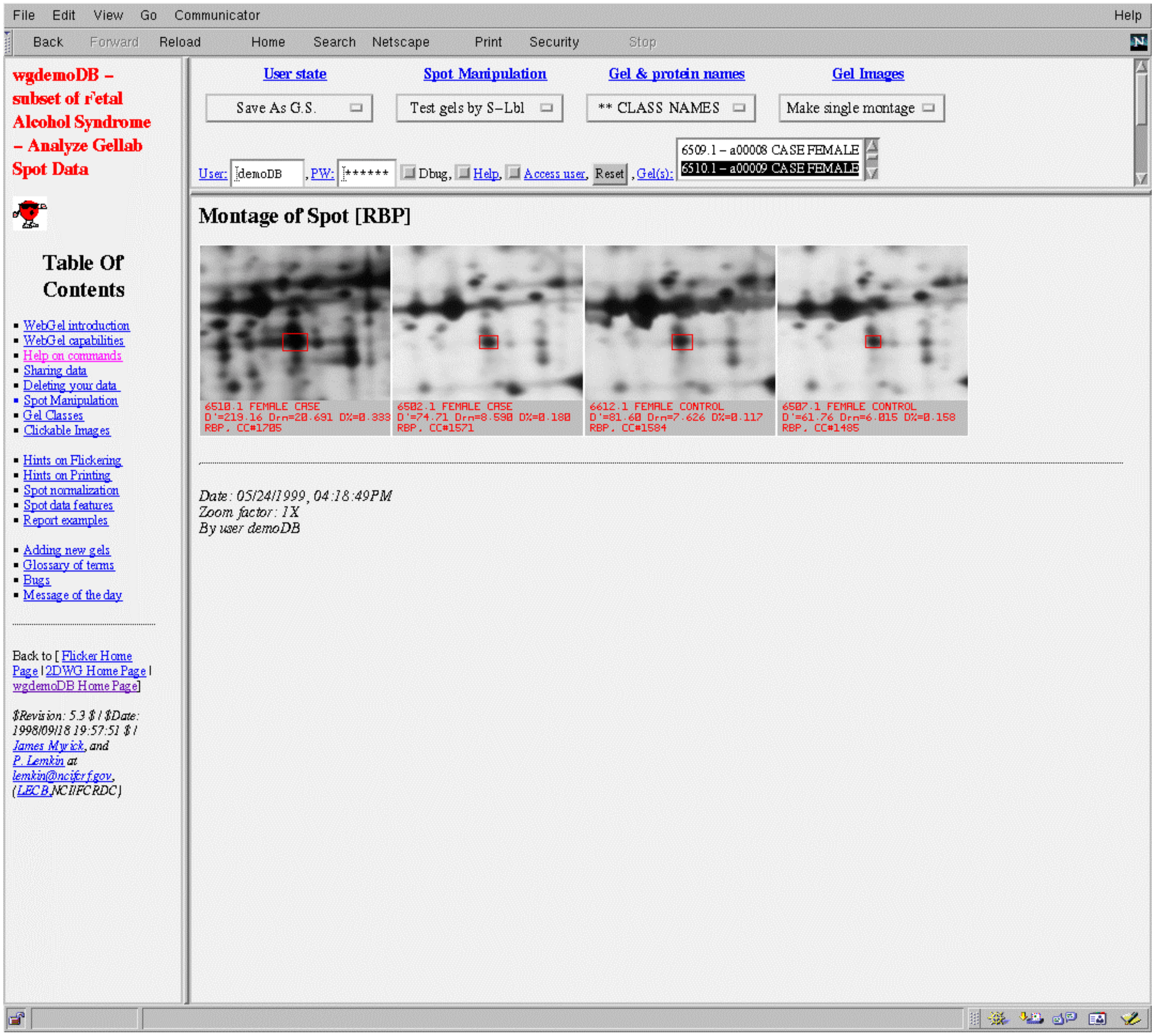
p–value	*t*–statistic	*f*–value	*dF*	Class	Class	mn D%[c1]/ mn D%[c2]	mn Drn[c1]/ mn Drn[c2]
0.2117	1.81181	0.0843	2.00	**#1 FEMALE CASE**	**#2 FEMALE CONTROL**	1.870	2.146

Kolmogorov–Smirnov test for all classes taken two at a time

prob	*d*	Class	Class	mn D%[c1]/ mn D%[c2]	mn Drn[c1]/ mn Drn[c2]
0.8438	0.50	**#1 FEMALE CASE**	**#2 FEMALE CONTROL**	1.870	2.146

Date: 05/24/1999, 04:15:33PM
By user demoDB

Figure 7 is a screen view showing a table report from the FAS database for protein RBP which was defined as a corresponding spot with the same spot label among the set of gels. The gels were all normalized using the same set of corresponding normalization spots.The statistical tests and resulting report were generated using the Test gels by Spot Label command.



wgdemoDB – subset of f'etal Alcohol Syndrome – Analyze Gellab Spot Data

Table Of Contents

- WebGel introduction
- WebGel capabilities
- Help on commands
- Sharing data
- Deleting your data
- Spot Manipulation
- Gel Classes
- Clickable Images

- Hints on Flickering
- Hints on Printing
- Spot normalization
- Spot data features
- Report examples

- Adding new gels
- Glossary of terms
- Bugs
- Message of the day

Back to [Flicker Home Page | 2DWG Home Page | wgdemoDB Home Page]

$Revision: 5.3 $ / $Date: 1998/09/18 19:57:51 $ / James Myrick, and P. Lemkin at lemkin@ncifcrf.gov, (LECB, NCI/FCRDC)

User: demoDB, PW: ******, Dbug, Help, Access user, Reset, Gel(s): 6509.1 – a00008 CASE FEMALE
6510.1 – a00009 CASE FEMALE

Montage of Spot [RBP]

```
6510.1 FEMALE CASE
D'=219.16 Drn=20.691 D%=0.333
RBP, CC#1705

6502.1 FEMALE CASE
D'=74.71 Drn=8.590 D%=0.180
RBP, CC#1571

6612.1 FEMALE CONTROL
D'=81.60 Drn=7.626 D%=0.117
RBP, CC#1584

6507.1 FEMALE CONTROL
D'=61.76 Drn=6.015 D%=0.158
RBP, CC#1485
```

Date: 05/24/1999, 04:18:49PM
Zoom factor: 1X
By user demoDB

Figure 8 is a screen view showing a montage image of the RBP protein for four of the FAS gels. The images in the montage image are sorted by normalized density and are placed with the gel with the highest concentration in the upper left hand corner and the least in the lower right-hand corner

4 Discussion

WebGel is an exploratory 2DE gel image and data analysis system that operates over the Internet using a Web browser. The distributed qualitative and quantitative analysis is performed by the user interacting with pre-processed quantitative electrophoretic gel data residing on a Web server database.

The essence of the WebGel analysis method is to first Flicker compare two gels at a time from within WebGel to visually narrow the analysis to a small set of interesting spots with robust changes. Visual analysis is useful for aligning and putatively identifying corresponding spots in difficult gels where there is a large variation in the morphology and numbers of spots. The user then uses WebGel to manually label corresponding spots between gels. This lets them use numerical data to test the statistical significance of visually detected quantifiable spot differences. These differences are then shown as either images (montages) or as reports. Spots that are missing in gels can be labeled as virtual spots for later visualization in montages to show visual verification across a set of gels from different experimental conditions. WebGel is not an effective tool for exhaustively searching an entire database because of the large number of spot and gel combinations that would have to be tested. However, major spot changes and missing spots are easily detected using Flicker and then quantified with statistical tests and documented in montage images with WebGel. This can aid in the verification of spot changes which may correlated with a disease state.

4.1 Use of gel analysis groupware in collaborative exploratory data analysis

By providing a groupware mechanism for collaborative research, WebGel helps group members share their work in progress across temporal and geographic boundaries using the Internet. Data is presented numerically in tables and in images. WebGel provides a method for quantifying small numbers of spots or bands across multiple gels. It integrates the Flicker gel comparison to visually locate or verify qualitative changes in spots that can then be quantitatively measured. It also allows the linking of known labeled proteins with entries in the Swiss-Prot database, which in turn provides links with other proteomic and genomic databases.

Although we minimize the amount of image data transported from the Web server, some image transmission is involved. Therefore, a high-bandwidth Internet connection is desirable since images are generated repeatedly. WebGel is marginally usable over a 28Kb modem, but is much more responsive with a local area network and high-speed Internet connections such as a T1-line or cable modem.

Research groups could set up their own WebGel servers to service collaborative 2DE gel analysis database projects across the Internet or behind fire walls. Running WebGel behind a firewall is important for groups who are restricted to operating only behind fire walls. The Web server requires the WebGel CGI program; 2DE image and spot list quantified data files; and configuration files pointing to this data in order to run WebGel.

Currently, WebGel can=t transfer, enter gels into the database and quantify them directly over the Internet. Future versions are being developed which will make this possible. This is not a problem for small collaborative groups because databases can be easily maintained. However, being able to expand a database so that researchers in different laboratories around the world could contribute quantified data opens up the possibility of building world-collaborative databases where meta-statistics techniques might be used on pooled data from different sources.

When multiple 2DE studies exist that include the same or sufficiently similar disease types and that have analyzed the same tissue of body fluid, increased statistical power can be obtained by pooling the studies together with WebGel. Such may be the case with some cancer studies. However, given the paucity of publicly accessible large 2DE studies now, and the unavailability of many studies for proprietary reasons, the likelihood of two or more similar data sets being available is small. Of course, merging of multiple data sets requires careful analysis using meta statistics methods which could be added to WebGel. Such a facility would require editorial oversight to maintain the quality of the database.

4.2 Advantages and disadvantages of distributed gel analysis groupware

There are a number of advantages and disadvantages to performing collaborative gel analysis over the Internet rather than solely with a stand-alone system. WebGel does not require special hardware or software to analyze gels in collaborators= laboratories B just a Web browser. This makes it much easier and faster for new collaborators to join the group since no local copy of the gel database has to be set up and continuously updated from another site. Users simply access the WebGel database with any Web browser. Then, data may be immediately shared with members of a group since the ability to share collaborative data is built into the software. Because no special local software is required for new users once the WebGel database is installed on a Web server, there is no additional cost per user seat. Since WebGel documentation is built into the Web site, no additional paper manual is required.

The groupware mechanism helps the researchers share their work. Because each database may be made secure, it is protected from access by others outside of the group. The ability to selectively share intermediate analyses with others within the group, is useful when testing hypotheses on preliminary data. To improve communication on exploratory results within the group, WebGel allows the creation and editing of both private and group notes. The latter are shared between group members and the former may be made visible to other selected group members at the user's discretion.

Using Flicker to do pair-wise visual comparison helps users wade through a large number of false-positive spot changes caused by the mis-pairings common in stand-alone 2D gel composite gel databases when analyzing sets of gels that are quite different. However, the false-negative rate (not finding actual spot differences which are significant) may be higher using Flickering since it is too time consuming to exhaustively compare all combinations of a large number of gels. If the investigator is looking for a few robust spot changes across all samples and experimental conditions, comparing a few samples and following them up across all of the gels may be an adequate preliminary analysis. The consistency of the changes found may then be verified using common spot labels and reviewing montage images and normalized tabular data for the few significant spots that are analyzed in depth.

A disadvantage of this Web-based approach is that it is currently less powerful than some of the stand-alone gel analysis database systems. However, if the numbers of spot changes to be detected and quantified are small, then this is not a serious disadvantage. Most of the stand-alone gel analysis systems can automatically pair spots between gels given a few to dozens of landmarks B but only if there is a reasonable correspondence between the gels. The lack of a spot-pairing algorithm somewhat limits WebGel=s usefulness. However, since WebGel=s primary use is to find and quantify small numbers of spots using visual Flickering, or to publish interactive quantitative data, the lack of automatic spot pairing is not a serious issue for that set of biological problems.

WebGel may also serve to give researches a convenient tool for tentative protein identification. One can Flicker-compare a series of new 2DE gel images to reference images from another database in which many proteins have been positively identified to obtain a very good candidate identification of a protein in the new gels. When the samples= matrices are the same, i.e., human sera, plasma, tissue type, etc., then the putative spot identification will be much more certain.

The data exploration techniques we have applied here to 2DE in WebGel can also be used in other biomedical databases. We are incorporating some of these techniques in a Java-based cDNA microarray database system called the Micro Array Explorer or MAExplorer and applying these techniques to gene discover in a mammary genome database.

4.3 Acknowledgments

Thanks are due to Mary Robinson (CDC, Atlanta) for use of her Fetal Alcohol Syndrome data in the WebGel demonstration database. Thanks also to Ellen Burchill and Tom Schneider for useful suggestions for improving this manuscript.

5 Software availability

We plan to make the WebGel program available for researchers to set up their own databases on their Web servers. In the meantime, we have made a demonstration WebGel database available that can be accessed from the Web. A subset of the Fetal Alcohol Syndrome database [22] can be accessed at the WebGel home page **http://www.lecb.ncifcrf.gov/webgel**.

6 References

1. Celis, J.E., Genomics and Proteomics of Cancer - paper symposium, *Electrophoresis* 1999, *20*, 221-430.

2. Williams, K.L., Genomes and proteomes: Towards a multidimensional view of biology. *Electrophoresis* 1999**,** *20*, 678-688.

3. Lipkin, L.E., Lemkin, P.F., Data base techniques for multiple PAGE (2D gel) analysis. *Clinical Chemistry* 1980, *26*, 1403-1413.

4. Lemkin, P.F., Lester, E.P., Database and search techniques for 2D gel protein data: A comparison of paradigms for exploratory data analysis and prospects for biological modeling. *Electrophoresis* 1989**,** *10* ,122-140.

5. Wu Y, Lemkin PF, Upton K (1993) A fast spot segmentation algorithm for 2D electrophoresis analysis. E*lectrophoresis* 1993**,** *14*, 1350-1356. [GELLAB-II+ **http://www.scanalytics.com/**]

6. Appel, R.D., Vargas, J.R., Palagi, P.M., Walther, D., Hochstrasser, D.F., Melanie II--a third-generation software package for analysis of two-dimensional electrophoresis images: II. Algorithms. *Electrophoresis* 1997, *18*, 2735-2748 [**http://www.expasy.ch/**]

7. Anderson, N.L., Taylor, J., Scandora, A.E., Coulter, B.P., and Anderson, N.G. The TYCHO system for computer analysis of two-dimensional gel electrophoresis patterns. *Clinical Chemistry* 1981, *27,* 1807-1820. [**http://www.lsbc.com/**]

8. Garrels, J.I., Farrar, J.T., Burwell IV, C.B., in Celis, J.E., Bravo, R. (Eds), Two-Dimensional Gel Electrophoresis of Proteins: Methods and Applications, Academic Press, NY, 1984, pp. 37-91.

9. BioImage 2DE gel analysis system. [**http://www.bioimage.com/2d.html**]

10. Phoretix 2DE gel analysis system. [**http://www.phoretix.com/**]

11. Rasband, W. NIH-Image public-domain biomedical image processing [**http://rsb.info.nih.gov/nih-image/**]

12. Pleissner, K.-P., Hoffmann, F., Kriegel, K., Wenk, C., Wegner, S., Sahlstrom, A., Oswald, H., Alt, H., Fleck,. E., Proteome data analysis and management, *Electrophoresis* 1999, *20,* 755-765. [**http://gelmatching.fu-berlin.de/**]

13. Oh, J., Hanash, S.M., Teichroew, D., Mining protein data from two-dimensional gels: Tools for systematic post-planned analysis. *Electrophoresis* 1999, *20,* 766-774.

14. Lemkin, P.F., Comparing Two-Dimensional electrophoretic gels across the Internet. *Electrophoresis* 1997, *18*, 461-470.

15. Lemkin, P.F., Comparing 2D Electrophoretic gels across Internet databases. In A2-D Protocols for Proteome Analysis@, Andrew Link (Ed), a book in Methods in Molecular Biology, Vol. 112, Humana Press, Totowa, NJ, 1997, pp 339-410.

16. Lemkin, P.F., 2DWG meta-database of 2D electrophoretic gel images on the Internet. *Electrophoresis* 1997, *18*, 2759-2773.

17. Lemkin, P.F., Thornwall, G.W., Flicker Image Comparison of 2-D Gel Images for Putative Protein Identification using the 2DWG Meta-Database. *Molecular Biotechnology* 1999, in press.

18. Patrick, J.L., Ping, G., Liu, J., Lemkin, P.F., You, M., Analysis of restriction landmark genomic scanning images of DNA from K-*ras* oncogene transformed NIH- 3T3 cells using computer assisted analysis. J. Analytical Biochem. 1999, in prep.

19. Getzenberg, R.H., Pienta, K.J., Coffey, D.S., The tissue matrix: cell dynamics and hormone action. *Endocr Rev.* 1990**,** *11* , 399-417.

20. Partin, A.W., Getzenberg, R.H., CarMichael, M.J., Vindivich, D., Yoo, J., Epstein, J.I., Coffey, D.S., Nuclear matrix protein patterns in human benign prostatic hyperplasia and prostatic cancer, Cancer Research 1993, *53*, 744-746.

21. Partin, A.W., Briggman, J.V., Subong, E.N.P., Szaro, R., Oreper, A., Wiesbrock, S., Meyer, J., Coffey, D.S., Epstein, J.I., Preliminary immunohistochemical characterization of a monoclonal antibody (PRO:4-216) prepared from human prostate cancer nuclear matrix proteins. *Urology* 1997, *50*, 800-808.

22. Robinson, M.K., Myrick, J.E., Henderson, L.O., Coles, C.D., Powell, M.K., Orr, G.A, Lemkin, P.F., Two-dimensional protein electrophoresis and multiple hypothesis testing to detect potential serum protein biomarkers in children with fetal alcohol syndrome. *Electrophoresis* 1995, *16*, 1176-1183.

23. Hornbeck, P.V., Lemkin, P.F., The Murine Lymphocytes Phosphoprotein Database. I. Protein Phosphorylation during M Phase of the Cell Cycle. 1999, in prep.

24. Lakshmanan, Y., Subong, E.N.P., Partin ,A.W., Differential Nuclear Matrix Expression in Prostate Cancers: Correlation with Pathologic Stage. J. Urology 1998, *159*, 1354-1358.

25. Lemkin, P.F., Shue, MJ.; Thornwall, G., Partin, A.W., Searching for Prostate Cancer Tumor Markers in 2D Electrophoretic Gels of Nuclear Matrix Proteins using Web-based exploratory analysis. *Urological Oncology*, 1999, in prep.