Technical Brief

VennDIS: A JavaFX-based Venn and Euler Diagram Software to Generate Publication Quality Figures

Vladimir Ignatchenko[1], Alexandr Ignatchenko[1], Ankit Sinha[1,2], Paul C. Boutros[2,3,4] and Thomas Kislinger[1,2].

[1]Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada.

[2]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

[3]Informatics and Bio-computing Program, Ontario Institute for Cancer Research, Toronto, ON, Canada

[4]Department of Pharmacology & Toxicology, University of Toronto, Toronto, ON, Canada

Correspondence: Thomas Kislinger, PhD

Princess Margaret Cancer Center

TMDT, Room 9-807

101 College Street, ON M5G 1L7

Canada

Tel. 416-581-7627

E-mail: thomas.kislinger@utoronto.ca

Received: 10-Jul-2014; Revised: 20-Oct-2014; Accepted: 17-Dec-2014

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/pmic.201400320.

Venn diagrams are graphical representations of the relationships among multiple sets of objects and are often used to illustrate similarities and differences amongst genomic and proteomic datasets. All currently existing tools for producing Venn diagrams evince one of two traits; they require expertise in specific statistical software packages (such as R), or lack the flexibility required to produce publication-quality figures. We describe a simple tool that addresses both shortcomings, Venn Diagram Interactive Software (VennDIS), a JavaFX-based solution for producing highly customizable, publication-quality Venn and Euler diagrams of up to 5 sets. The strengths of VennDIS are its simple graphical user interface and its large array of customization options, including the ability to modify attributes such as font, style and position of the labels, background color, size of the circle/ellipse and outline color. It is platform independent and provides real-time visualization of figure modifications. The created figures can be saved as XML files for future modification or exported as high-resolution images for direct use in publications.

Keywords:

Venn Diagram / Euler Diagram / Data visualization software

Venn diagrams [1] are used extensively in proteomics and genomics to visualize intersections and complements across multiple experiments or datasets. Euler diagrams are an alternative way to display relationships amongst biological data, where empty intersections and complements are not displayed and subsets are positioned inside the superset. Currently existing software packages or online tools can be broadly divided into two categories. First, there are R-based tools such as the VennDiagram [2], gplots, limma [3-5], gplots:venn [6] and venneuler [7] packages. These tools provide broad capabilities, but require users with knowledge of the R programming language [8]. A number of web-based tools have also been developed, including VENNY [9], BioVenn [10] VennMaster [11], Google chart [12], GeneVenn [13], Venn Diagram Plotter [14] and eulerAPE [15]. These are user-friendlier, but have limited customization options to produce publication quality figures (see **Supplemental Table 1** for a detailed comparison). We introduce a new software package, VennDIS (Venn Diagram Interactive Software), which produces highly customizable, publication quality Venn and Euler figures with a simple and intuitive graphical user interface.

VennDIS is a platform-independent open source software written in Java, using the JavaFX software platform, which is included within the standard JDK 1.8 and JRE 1.8 bundles (Java Runtime Environment; http://www.oracle.com/technetwork/java/). Java was chosen as it is open-source; enabling developers to design, create and deploy rich client applications that can be deployed across a wide variety of operating systems such as Windows, Mac OS X and Linux. VennDIS is simple to install and easy to use. The interface was designed to create intuitive access to a wide range of customization features, such as background color and transparency, font size and style, circle/ellipse size and border color, style and weight. Parameter modifications are instantly presented in the diagram preview panel. In the case of 2-way Venn and Euler diagrams the circle size and the overlap is area proportional (*i.e.* size of the circles and overlap are calculated proportional to the input list). Following visualization of each diagram users have the option to manually change the circle size and overlap to enhance visualization of diagrams with largely different circle sizes. For 3-way Venn and Euler diagrams the graphics are size-proportional but not overlap-proportional. As described above, the user can manually change circle sizes. For more complex 4- and 5-way Venn diagrams the ellipse size is kept constant. VennDIS supports several types of data input, such as simple lists, excel style tables, and pre-calculated values for overlaps and differences.

Since VennDIS was created with proteomics data in mind, MaxQuant output tables [16] can be uploaded and the appropriate columns selected for graphical display. Based on the user input, VennDIS analyzes the data and instantaneously offers the user suitable diagram options. Created diagrams can be saved in XML format for future modification, exported as high-resolution bitmap images in a variety of formats (*i.e.* SVG, PNG, JPG or BMP). The current version of VennDIS supports four different types of Venn diagrams shown in **Figure 1A** and a total of 58 Euler diagrams, of which 12 are shown on **Figure 1B**. Initiating VennDIS takes users to the main window, which contains a toolbar (top left corner). Here, users can start the creation of a new diagram, open a previously saved diagram or save and export generated Venn or Euler diagrams. On the right hand side of the start window a "help button" displays a brief step-by-step user guide of the main software functions (see **Supplemental Document 1** for more details). By clicking on the "new" Venn or Euler diagram button, a data input window will appear. Currently VennDIS supports three different ways to import data, which can be selected from a right-hand menu in the input window by selecting the desired icon (**Figure 2A-C**). The default input for VennDIS is "Lists" (**Figure 2A**). This method enables user to submit lists of various identifiers for one or more datasets by copying and pasting the respective identifiers into the individual dataset boxes. The identifiers are matched and appropriate diagram options are displayed as icons on the lower left-hand corner. Individual datasets can either be named at this stage or later in the main diagram interface. Another type of data input are "Tables", where tab delimited text files can be uploaded or copy/pasted from the clipboard, selected through a tab at the top left-hand corner of the input window. One specific proteomics centric option that we have implemented is the direct import of MaxQuant tables, followed by the selection of specific columns for graphical display (**Figure 2B**). Imported tables are considered binary data and any value within a sample column considers the matching identifier present, while zero or absent values consider the identifier not present (**Figure 2B**). Finally, a value-based data input option is also available (**Figure 2C**). Here, users can manually add pre-calculated numerical values for the intersections and complements of chosen Venn or Euler diagrams. Regardless of the chosen data input format, the most likely variations of resulting Venn or Euler diagrams are instantly calculated and displayed as icons on the bottom left-hand corner of the input window (**Figure 2A-C**). The diagram option button returns the user to the main window of VennDIS. Here users have additional customizations options prior to saving or exporting the diagram as an image.

In the main window, which is shown in **Figure 3A**, a default version of the selected Venn or Euler diagram is displayed on the right-hand side. Located on the left-hand side of the main window are a variety of customization tabs. Through these tabs users can modify a wide range of diagram properties. Customization options for the B&O (Background and Overlaps) tab are shown on the left-hand side of the main window (**Figure 3A**) and include background color, transparency, font size and style and color of the values displayed in the Venn/Euler diagrams. Slightly different sets of customization options are available through the "Ellipse" tabs (**Figure 3B**). These include size, color, transparency and z-order of ellipses, fill and stroke, as well as font size, style and color of the labels. After the desired customization is complete the Venn or Euler diagram can be saved as XML format for future modifications. Bitmap images of customized diagrams with different user adjustable levels of resolution can be generated though the "Export" tab in the main dialog window (**Figure 3C**). After selection of the required resolution the diagram can be saved in three different image formats (PNG, JPG or BMP) or as Scalable Vector Graphics (SVG) for future modification in a vector graphics editor.

In summary, although a wide array of tools exist that enable users to display the differences and overlaps of genomics and proteomics data in the form of Venn or Euler diagrams, these tools require either knowledge of specific programming languages or lack display and/or customization options [2-7, 9-15]. To overcome these limitations we have developed VennDIS, a user-friendly program with an intuitive interface. Currently VennDIS supports 62 types of Venn and Euler diagrams (see **Supplemental Table 2** for details). VennDIS runs on most modern operating systems and is open source, enabling developers to add additional functionalities if required. The tool enables a variety of user-friendly customizations, including input types and saving options. Importantly, once data has been entered, VennDIS automatically determines the most suitable Venn or Euler diagrams while automatically adjusting the ellipse and circle sizes. Real-time feedback on graphical changes is given in the main diagram window. VennDIS (including the source code) is available for download from our lab web page (http://kislingerlab.uhnres.utoronto.ca/projects/VennDIS_v1.0.1.zip) and requires installation of Java Runtime Environment 1.8 or higher, which is freely available from the Oracle website (http://www.oracle.com/technetwork/java).

Acknowledgements

This study was funded by the Canadian Institutes of Health Research (CIHR) to T.K. and P.C.B. (MOP-114896), Global Leadership Round in Genomics and Life Sciences to TK (GL2-01012) as well as the Ontario Institute for Cancer Research to P.C.B. through funding provided by the Government of Ontario. A.I. was supported in parts by the Princess Margaret Cancer Center Head and Neck Translational Research Program. T.K. is supported by the Canadian Research Chair program. P.C.B. is supported by a Terry Fox Research Institute New Investigator Award and a CIHR New Investigator Award. We gratefully acknowledge the philanthropic support from Joe's Team. Support is also provided from the Campbell

Family Institute for Cancer Research and the Ministry of Health and Long-term Planning.

Competing financial interests

The authors declare no competing financial interests.

References

[1] Venn, J., On the diagrammatic and mechanical repre- sentation of propositions and reasonings. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 1880, *9*, 1-18.

[2] Chen, H., Boutros, P. C., VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC bioinformatics* 2011, *12*, 35.

[3] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., *et al.*, Bioconductor: open software development for computational biology and bioinformatics. *Genome biology* 2004, *5*, R80.

[4] Smyth, G. K., Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions using R and Bioconductor, Statistics for biology and Health, Springer, New York* 2005, 397-420.

[5] Smyth, G. K., Yang, Y. H., Speed, T., Statistical issues in cDNA microarray data analysis. *Methods Mol Biol* 2003, *224*, 111-136.

[6] Moeller, S., Warnes, G. R., gplots:venn. *http://cran.r-project.org/*.

[7] Wilkinson, L., Exact and Approximate Area-Proportional Circular Venn and Euler Diagrams. *IEEE Transactions on Visualization & Computer Graphics* 2012, *18*, 321-331.

[8] Dessau, R. B., Pipper, C. B., ["R"--project for statistical computing]. *Ugeskrift for laeger* 2008, *170*, 328-330.

[9] Oliveros, J. C., VENNY. An interactive tool for comparing lists with Venn Diagrams. *http://bioinfogp.cnb.csic.es/tools/venny/index.html*. 2007.

[10] Hulsen, T., de Vlieg, J., Alkema, W., BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics* 2008, *9*, 488.

[11] Kestler, H. A., Muller, A., Kraus, J. M., Buchholz, M., *et al.*, VennMaster: area-proportional Euler diagrams for functional GO analysis of microarrays. *BMC bioinformatics* 2008, *9*, 67.

[12] Google Chart. *https://developers.google.com/chart/*.

[13] Pirooznia, M., Nagarajan, V., Deng, Y., GeneVenn - A web application for comparing gene lists using Venn diagrams. *Bioinformation* 2007, *1*, 420-422.

[14] Venn Diagram Plotter. *http://omics.pnl.gov/software/venn-diagram-plotter*.

[15] Micallef, L., Rodgers, P., eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PloS one* 2014, *9*, e101717.

[16] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature biotechnology* 2008, *26*, 1367-1372.

Figure 1. A) Examples of customized two, three, four and five way Venn diagrams. **B**) Examples of 12 out of 58 Euler diagrams available in VennDIS.

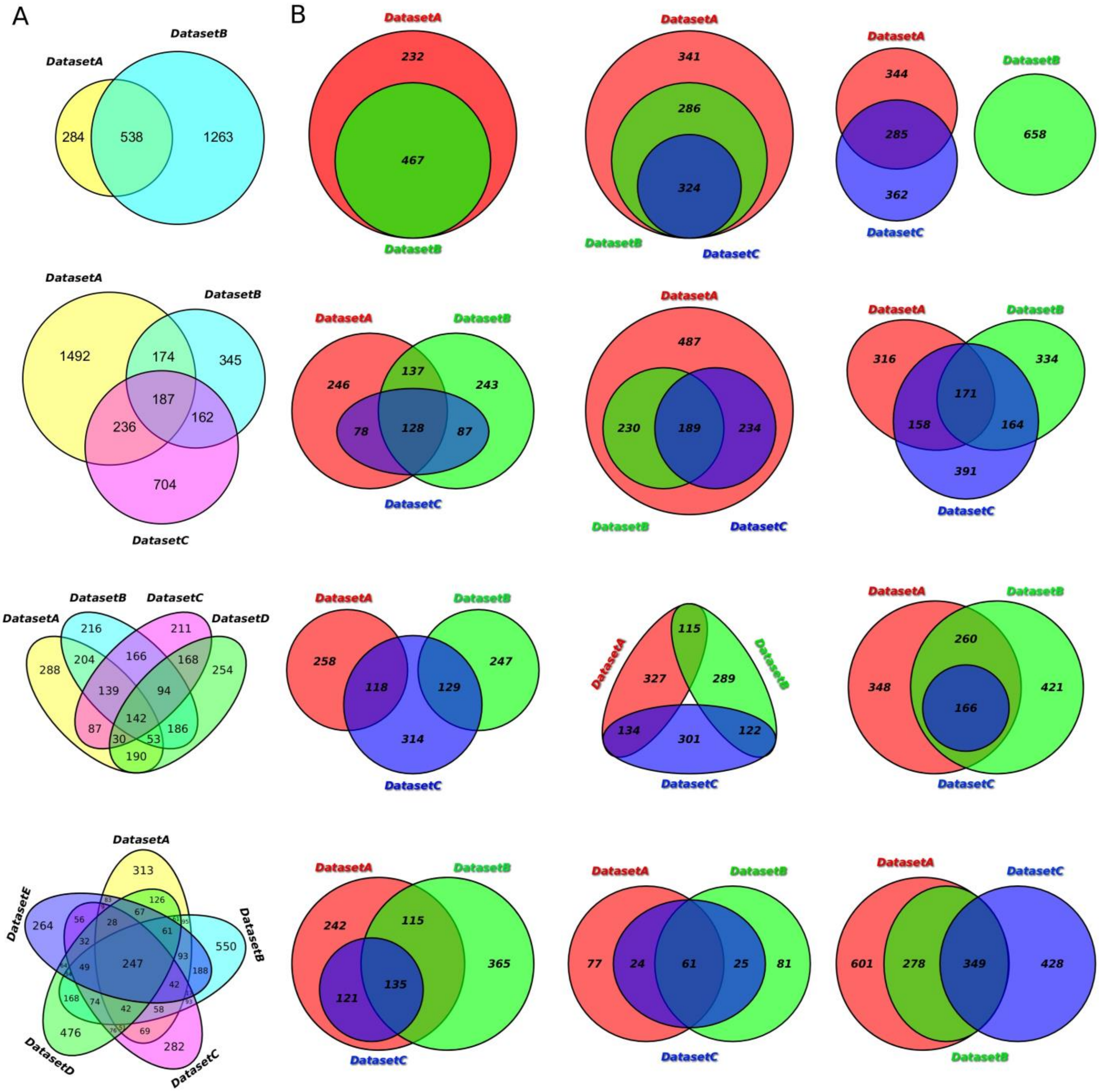

Figure 2. A) Data input in the form of a "list". **B**) Data input in the form of a "table" (a MaxQuant table is shown as example). **C**) Data input in the form of pre-calculated "values".

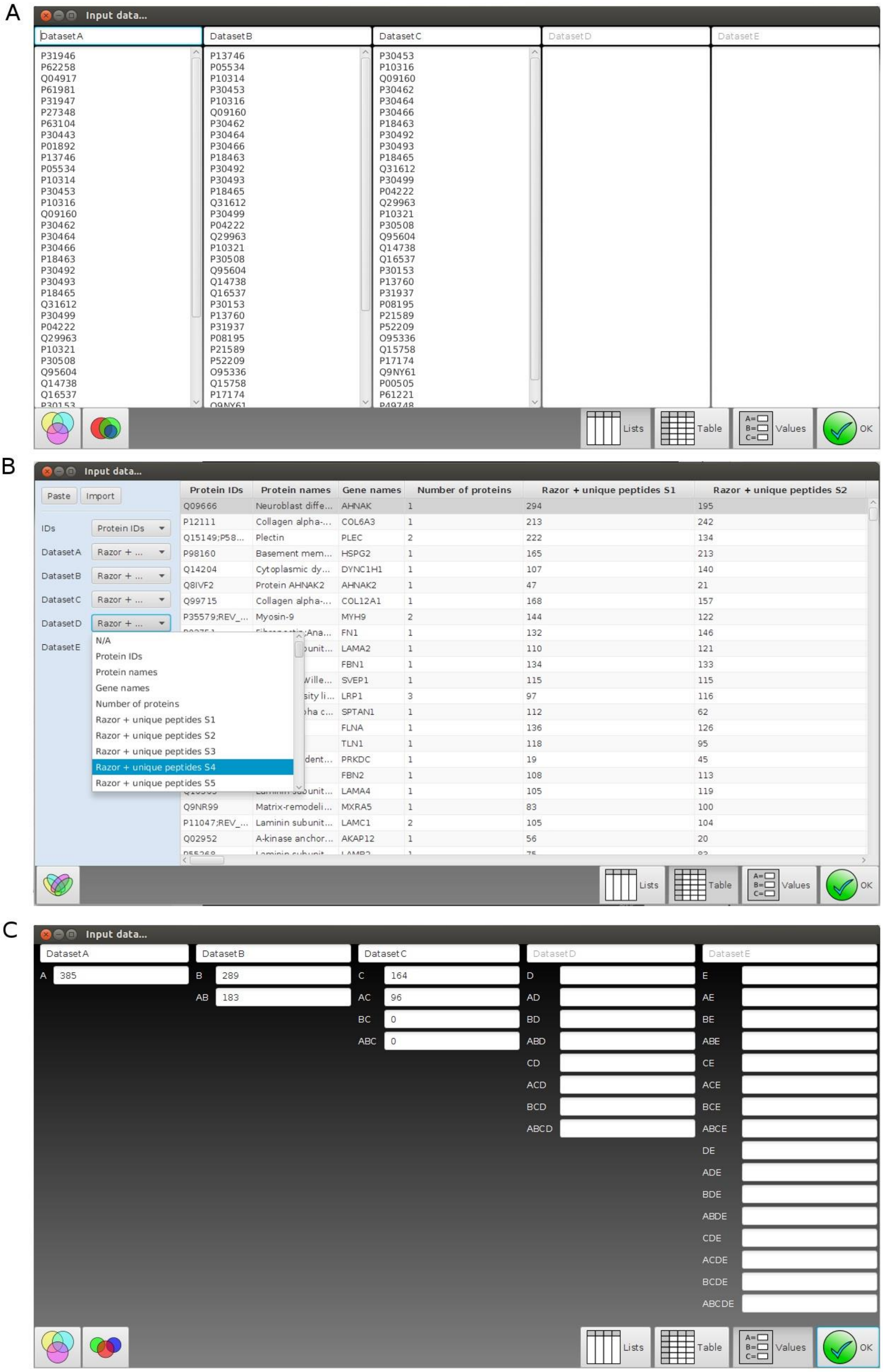

Figure 3. **A**) VennDIS main window with customization panel for B&O (Background and Overlaps) tab on the right and rendered diagram preview on the left. **B**) Customization panel for “Ellipse” tabs. **C**) Export image window for customized diagram.