

Venn diagrams in bioinformatics

Anqiang Jia[†], Ling Xu[†] and Yi Wang

Corresponding author: Yi Wang, State Key Laboratory of Silkworm Genome Biology, Southwest University, Chongqing 400715, China; Biological Science Research Center, Southwest University, Chongqing 400715, China. Tel.: +86 2368251683, Fax: +86 2368251128; E-mail: yiwang28@swu.edu.cn

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

Abstract

Venn diagrams are widely used tools for graphical depiction of the unions, intersections and distinctions among multiple datasets, and a large number of programs have been developed to generate Venn diagrams for applications in various research areas. However, a comprehensive review comparing these tools has not been previously performed. In this review, we collect Venn diagram generators (i.e. tools for visualizing the relationships of input lists within a Venn diagram) and Venn diagram application tools (i.e. tools for analyzing the relationships between biological data and visualizing them in a Venn diagram) to compare their functional capacity as follows: ability to generate high-quality diagrams; maximum datasets handled by each program; input data formats; output diagram styles and image output formats. We also evaluate the picture beautification parameters of the Venn diagram generators in terms of the graphical layout and briefly describe the functional characteristics of the most popular Venn diagram application tools. Finally, we discuss the challenges in improving Venn diagram application tools and provide a perspective on Venn diagram applications in bioinformatics. Our aim is to assist users in selecting suitable tools for analyzing and visualizing user-defined datasets.

Key words: Venn diagrams; visualization; generator; application

Background

Massive amounts of data are frequently generated with technology improvement, which leads to increasingly complex data relationships. Understanding the relationships between different datasets is vitally important for data analysis, and the use of suitable visualization methods is especially important for clarifying established relationships. One of the simplest and most effective ways to illustrate these relationships is by generating a Venn diagram, which was first proposed in 1880 by John Venn [1]. A typical Venn diagram uses overlapping circles to illustrate the intersections, differences and unions between datasets. For example, one of the most basic Venn diagrams has three circles [2, 3] and is used to visualize the overlaps between three datasets (Figure 1). Intersections between sets are represented in the overlapping portions of the circles, differences are shown in the

non-overlapping portions of the circles and unions are represented by the data in both or all three circles.

In bioinformatics, Venn diagrams enable researchers to quickly observe the information between experiments, conditions and groups to explore the relationships between datasets [4–8]. With the advancement of sequencing technology, such as genomics, transcriptomics and proteomics have been widely used in enormous biological studies to unearth core gene sets. After obtaining the gene list from different omics, Venn diagrams are often used to display the shared or unique genes among these gene lists [9–11]. To collect the most widely used Venn diagram tools to date, we used ‘Venn diagram’ as the keyword to search in the Google Scholar engine, and a total of 149 000 results were retrieved as of 23rd November 2020. More than 75% of these results are related to biology based on the search

Anqiang Jia is a PhD student in the Biological Science Research Center at Southwest University, Chongqing. He majors in genome biology, multi-omics and single-cell transcriptome analysis.

Ling Xu is a Post-Doctoral Researcher at the University of California, Berkeley. Her research focuses on discovering mechanisms underlying the interactions between sorghum and its microbiome under drought.

Yi Wang is a Professor in the Biological Science Research Center at Southwest University, Chongqing. His expertise includes genome biology, database development and next-generation sequence analysis.

Submitted: 9 December 2020; **Received (in revised form):** 4 March 2021

© The Author(s) 2021. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

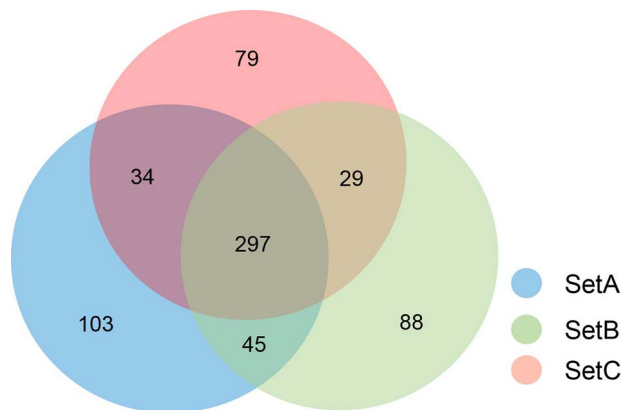


Figure 1. Typical Venn diagram. It consists of three regular circles, each of which represents a dataset. This diagram was drawn using PowerPoint.

using the keyword ‘Venn diagram biology’. To date, many tools have been developed to utilize different styles of Venn diagrams, such as jvenn [12], VennDiagram [13], VennPainter [14] and UpSetR [15]. Additionally, some application tools have been developed to show or analyze biological data based on the Venn diagrams. For example, NetVenn [16] uses network diagrams and Venn diagrams to visualize the network relationships of shared genes between different gene lists. OrthoVenn [17] analyzes orthologous genes between different genomes and displays them with a Venn diagram.

Venn diagrams have brought great convenience to some scientists in diverse fields. For example, the classic Venn diagrams were used to visualize the orthologous clusters of the four plants and seven animals [18–19], as well as to visualize the overlapping genes from the differential expressed genes (DEGs) of the transcriptome with four samples [20]. Area-proportional Venn diagrams (Euler diagram) can show the size relationships among datasets. Gerbracht et al. [21] used the diagram to display the intersection genes and set relationships between four DEGs lists. To elucidate the intersections and unions between many datasets, Bailey et al. [22] used the matrix-based diagram tool UpSetR [15] to plot highlights the intersections in 10 samples. In addition, some Venn diagram application tools are not inferior. For instance, the OrthoVenn [17] tool was used to identify shared and unique gene clusters of four insects in a Venn diagram [23]. Zhang et al. [24] utilized the application tool VennPlex [25] to visualize the difference, upregulated and downregulated genes in a Venn diagram. Overall, the Venn diagram is not only used to visualize the intersections and unions of several target lists but also to analyze, filter and visualize complex biological data. These findings clearly show that the Venn diagram has immeasurable application prospects in bioinformatics. Hence, selecting the optimal tool is important for users.

Several previous studies compared the types of Venn diagrams, but they did not indicate which tool has the best performance for drawing graphics with different layouts. The ‘performance’ represents the ability to generate high-quality pictures, the maximum number of datasets processed, the parameters of picture beautification, etc. For instance, Chapman et al. [26] compared and evaluated the running time and error rate of four diagram types, including Venn diagrams, Euler diagrams with shading, Euler diagrams without shading and linear diagrams, and they clarified that linear diagrams represented using parallel horizontal line segments and line overlaps represented using

intersections are most effective for representing sets. Alsalkh et al. [27] summarized some Venn diagram outputs and provided guidance for choosing appropriate graphic types for different visualization requirements. However, a comprehensive comparison of Venn diagram generators and their application tools in bioinformatics is lacking. Here, we reviewed the most widely used Venn diagram generators as well as Venn diagram related application tools by comparing the performance of these generators and the functional characteristics of each application tool in bioinformatics. We selected several Venn diagram generator tools that performed relatively well in all aspects and reviewed the prospects for future development directions of the generators and application tools. Our aim is to provide guidance to readers in choosing optimal Venn diagram application tools for particular research needs.

Venn diagram generators

Collection of Venn diagram generators

Based on our investigation, 32 Venn diagram generators were collected for the following section (Tables 1–5). To compare their performance, we present tools that can draw multiple layouts separately in each section and use test data drawn from partially modified RainBio mammary datasets [28]. Some of these generators were excluded from using the above dataset, for example, eulerGlyphs [29], BranchingSets [30] and Bubble Sets [31], because they do not support user-defined datasets. We also excluded the AggreSet [32] tool due to incompatibility with our dataset. In terms of graphical layouts, these Venn diagram generators were categorized into six groups (Figure 2):

- (1) Classic Venn diagrams, which have smooth curves that form regular or symmetrical diagrams [14];
- (2) Edwards Venn diagrams, which consist of rounded squares, circles and ‘cogwheel’ curves [33];
- (3) Euler diagrams, which use circles or ellipses of different sizes to show the relationship between sets [34];
- (4) Venn and Euler diagram variants, which include irregular or polygonal graphics [27, 28];
- (5) Aggregation or matrix-based diagrams, which are composed of matrixes or graphics and arranged side by side [27] and
- (6) Other technical diagrams, which include linear, node-link and overlay diagrams [31, 35, 36].

In general, 14 of the 32 tools supported > 2 layouts, and most generators were designed to be used in a single application environment, either web platforms, the R environment, localized interfaces, the command line (Windows and Linux) or JavaScript. However, three of them (nVenn [37], Intervene [38] and eulerAPE [39]) supported more than two environments. In addition, most of the Venn diagram tools could support 2–6 datasets, while aggregation and matrix-based technology tools were more suitable for > 6 datasets.

Classic Venn diagrams

Classic Venn diagrams are mainly composed of regular circles or ellipses that compare the compositions of several datasets. Table 1 shows 16 of the 32 tools that can draw classic Venn diagrams: 9 are based on web platforms, 5 are based on the R and 2 are localized interface tools. In the following sections, we present the tools based on their working environment.

Table 1. Basic features of 16 tools used to generate classic Venn diagrams

Tools	Application	Sets	Input data format	Output image format	Key parameters	Text output	Reference
jvenn	Web	2–5	Uploaded lists, pasted values	PNG, SVG	Diagram color, font size and family	CSV	[12]
VennDiagramWeb	Web	2–5	Uploaded lists	PNG, SVG, TIFF	Diagram color, font size, border style, width and color	TXT	[40]
InteractiVenn	Web	2–5	Uploaded lists, pasted values	PNG, SVG	Diagram color, font size and opacity	TXT	[41]
Intervene	Web	2–4	Uploaded lists	PNG, SVG, PDF, TIFF	Label size, font size, border width, type and color	Online table	[38]
EVENN	Web	2–6	Uploaded lists, pasted values	PNG, SVG	Diagram color, font size and family	CSV	http://www.ehbio.com/test/venn/#/
Multiple List Comparator	Web	2–4	Uploaded lists, pasted values	PNG, SVG	Diagram color	TXT	http://www.molbio.tools.com/listcompare.html
Calculate and draw custom Venn diagrams	Web	2–5	Uploaded lists, pasted values	PNG, SVG	No	TXT	http://bioinformatics.psb.ugent.be/webtools/Venn/
Pangloss Venn diagram generator	Web	2–3	Pasted values	No	No	Online table	http://www.pangloss.com/seidel/Protocols/venn.cgi
Venny	Web	2–4	Pasted values	No	Font size and family, label size and family, border width and percentage	No	https://bioinfogp.cnb.csic.es/tools/venny/
VennDiagram	R	2–5	Lists	TIFF	Diagram color, font size and family, label size and family, border width and style	R text	[13]
Vennerable	R	2–4	Lists	R graphics	No	R text	R package
eVenn	R	2–4	Lists	PNG, SVG	Diagram color and font size	TXT	R package
venn	R	2–7	Lists	R graphics	Diagram color, font size and opacity	R text	R package
colorfulVennPlot	R	2–4	Lists	R graphics	Diagram color	R text	R package
VennPainter	Localized interface	2–5	Uploaded lists	SVG	Diagram color and load color palette	TXT	[14]
VennDIS	Localized interface	2–5	Uploaded lists, pasted values, table	PNG, SVG, BMP, JPG	Diagram color, font size, family and rotation, border color, style and size	TXT	[42]

The user-friendly interfaces of web tools enable users to obtain high-quality pictures through interactive real-time selection of parameters online. Among the nine web tools we listed (Table 1), InteractiVenn [41], jvenn [12], EVENN and VennDiagramWeb [40] have comprehensive parameters to beautify the pictures (Table 1) for the 2–5 datasets. All of them support user-defined lists as input by which users can obtain high-quality images for publication by modifying the diagram color and label font with just a few mouse clicks and generate figures in vector or bitmap format. Moreover, jvenn [12] uses a classic Venn diagram and histogram to display datasets (Figure 3A). EVENN is consistent with the jvenn diagram, and InteractiVenn [41] can merge the subsets of different datasets (Figure 3B). Furthermore, Intervene [38], Calculate and draw custom Venn diagrams also support 2–5 datasets, and Intervene (Supplementary Figure S1A) cannot customize the diagram color while Calculate and draw

custom Venn diagrams (Supplementary Figure S1B) does not have adjustable parameters for modifying pictures. Although Multiple List Comparator (Figure 3E) and Venny (Figure 3F) support 2–4 datasets with ellipses, Multiple List Comparator cannot modify label attributes, and Venny does not support output vector diagrams.

Among the R packages, VennDiagram [13] supports 2–5 datasets and has many parameters that allow users to customize the diagrams (Figure 3C), such as the font size, label style, border style and size, diagram color and rotation. However, eVenn, venn and colorfulVennPlot do not allow the modification of these parameters to generate high-quality pictures. eVenn uses circles to display up to 4 datasets (Figure 3G), venn can process up to seven samples at present (Figure 3H) and colorfulVennPlot can draw colorful Venn diagrams with up to 4 datasets (Supplementary Figure S1D).

Table 2. Basic features of eight tools used to generate Edwards' Venn diagrams

Tools	Application	Sets	Input data format	Output image format	Key parameters	Text output	Reference
jvenn	Web	2–6	Uploaded lists, pasted values	PNG, SVG	Font size and family and border color	CSV	[12]
InteractiVenn	Web	6	Uploaded lists, pasted values	PNG, SVG	Diagram color, font size and opacity	TXT	[41]
Calculate and draw custom Venn diagrams	Web	2–6	Uploaded lists, pasted values	PNG, SVG	No	TXT	http://bioinformatics.psb.ugent.be/webtools/Venn/
Intervene	Web	2–5	Uploaded lists	PNG, SVG, PDF, TIFF	Label size, font size, border width, type and color	Online table	[38]
EVenn	Web	2–6	Uploaded lists, pasted values	PNG, SVG	Font size and family and border color	CSV	http://www.ehbio.com/test/venn/#/
Vennerable	R	2–9	Lists	R graphics	No	R text	R package
Vennature	Localized interface	2–6	Lists	PPT, EMF	No	No	[46]
VennPainter	Localized interface	2–6	Uploaded lists	SVG	Diagram color, loadable color palette	TXT	[14]

Table 3. Basic features of eight tools used to generate Euler diagrams

Tools	Application	Sets	Input data format	Output image format	Key parameters	Text output	Reference
VennDiagramWeb	Web	2–3	Uploaded lists	PNG, SVG, TIFF	Diagram color, font size, border style, width and color	TXT	[40]
EVenn	Web	2–6	Uploaded lists, pasted values	PDF	Diagram color, font size, label size, legend position and border style	No	http://www.ehbio.com/test/venn/#/
VennDiagram	R	2–3	Lists	TIFF	Diagram color, font size and family, label size and family, border width and style	R text	[13]
Vennerable	R	2–3	Lists	R graphics	No	R text	R package
eulerr	R	2–6	Values	R graphics	Diagram color and area-proportional	No	R package
venneuler	R	2–4	Values	R graphics	Diagram color	No	R package
Venn Diagram Plotter	Localized interface	2–3	Pasted values	PNG, SVG, BMP	Diagram colors, area size and rotation	No	https://omics.pnl.gov/software/venn-diagram-plotter
VennDIS	Localized interface	2–3	Uploaded lists, pasted values, tables	PNG, SVG, BMP, JPG	Diagram color, font size, family and rotation, border color, style and size	TXT	[42]

In the localized interface tools, both VennDIS [42] and VennPainter [14] can produce high-quality pictures for 2–5 datasets. VennDIS has more customizable parameters than VennPainter for beautifying the picture. For example, VennDIS can modify the background color, font size, style and rotation, border style and size (Figure 3D, Table 1).

Classic Venn diagrams are frequently used to visualize the shared and unique genes between 2–5 datasets. To elucidate the shared and unique homologous genes, Jarvis et al. [43] used InteractiVenn [41] to visualize the shared and unique orthologous gene clusters between four highly nutritious crops. The

jvenn [12] was used to visualize the shared and unique gene families among four metazoans [44]. Furthermore, Melo et al. [45] also identified the shared and unique proteins from five cancer exosomes with a Venn diagram by VennDiagram [13].

Edwards' Venn diagrams

Edwards' Venn diagrams were first proposed by Edwards [33] and are produced by projecting a circular Venn diagram [46] or a three-dimensional ball onto a spherical surface (Supplementary Figure S2A), which are consisted of rounded squares, circles

Table 4. Basic features of 10 tools used to generate Venn and Euler diagram variants

Tools	Application	Sets	Input data format	Output image format	Key parameters	Text output	Reference
jvenn	Web	6	Uploaded lists, pasted values	PNG, SVG	Diagram color, font size and family	CSV	[12]
Intervene	Web	2–5	Uploaded lists	PNG, SVG, PDF, TIFF	Label size, font size, border width, type and color	Online table	[38]
Pangloss Venn diagram generator	Web	4	Pasted values	No	No	Online table	http://www.pangloss.com/seidel/Protocols/venn.cgi
EVenn	Web	>12	Pasted values	PDF	Diagram color and transparency	No	http://www.ehbio.com/test/venn/#/
Vennerable	R	2–4	Lists	R graphics	No	R text	R package
VennPainter	Localized interface	5–8	Uploaded lists	SVG	Diagram color and loadable color palette	TXT	[14]
nVenn	Web	2–6	Pasted values	PNG, SVG	Diagram color, font size and border width	Online table	[37]
	R		Lists	R graphics	Diagram color, opacity and border width	R text	
	Command line (Windows)		Lists	SVG	No	TXT	
eulerAPE	Localized interface	2–3	Uploaded lists, Pasted values	PNG, SVG	Area-proportional	others	[39]
eulerForce	Localized interface	2–5	Pasted values	PNG	Force-directed layout	No	[51]
eulerGlyphs	Localized interface	2–3	No	PNG	Area-proportional and glyph representations	No	[29]

and ‘cogwheel’ curves (Supplementary Figure S2B and 2C), and helps users to further visualize the overlaps between datasets. To date, the Edwards’ Venn diagram has been integrated into multiple tools. For instance, InteractiVenn [41], jvenn [12], EVenn and VennPainter [14] can generate publication-quality Edwards’ Venn diagrams (Table 2); and InteractiVenn, jvenn and EVenn maintain the same interactive capabilities as those used for classic Venn diagrams. In terms of the graphical interface, InteractiVenn [41] can only display Edwards’ Venn diagrams for 6 datasets and supports changing the color of different areas (Figure 4A), while jvenn (Figure 4B) [12] and EVenn (with graph the same as jvenn) can change the color of the borders. The localized interface tool VennPainter can draw an Edwards’ Venn diagrams of 2–6 datasets with filled area colors (Table 2 and Figure 4C), although it cannot change the label style or font size. In addition, Calculate and draw custom Venn diagrams (Supplementary Figure S3A) and Vennture ([46]; Supplementary Figure S3D) can display Edwards’ Venn diagrams for up to 6 datasets, although their graphics cannot be modified interactively. Intervene [38] can display ‘battle’ (Supplementary Figure S3B) and ‘common’ (Supplementary Figure S3C) Edwards’ Venn diagrams with up to 5 datasets, as well as customizable label and font attributes, but the color of the filled area cannot be modified. The R package Vennerable can display the largest number of datasets at 9 using the ‘battle’ Edwards’ diagram (Figure 4D); it has deficiencies in terms of custom parameter settings and visualization of intersections and unions.

Edwards’ Venn diagrams are also used to visualize the intersections and unions between 2–6 samples. For instance, the Edwards’ Venn diagrams in the Vennture [46] software was used to visualize the shared and unique miRNAs of the six differentially expressed lists [47]. Georgiadis et al. [48] used the VennPainter [14] tool-generated Edwards’ Venn diagram to illustrate the overlapping CpG sites between six DNA methylation samples.

Euler diagrams

Euler diagrams use circles or ellipses of different scales or sizes to clarify relationships between sets, such as inclusions, exclusions and intersections [34]. In addition to drawing classic Venn diagrams, VennDiagramWeb [40], VennDIS [42], VennDiagram [13] and EVenn also draw Euler diagrams. The output picture can be configured with customizable parameters, including the diagram color, font size, labels, border style, width, color, etc. (Table 3). VennDiagramWeb [40] and EVenn can show accurate proportional relations regardless of the presence of intersections (Figure 5A, Supplementary Figure S4A) between sets. VennDiagram [13] is similar to VennDiagramWeb [40], while VennDIS [42] cannot display the sizes proportionally (Figure 5B). Furthermore, EVenn and Eulerr use ellipses (Figure 5C) and circles (Figure 5D) to produce Euler diagrams with up to 6 datasets, although the accurate proportions cannot be displayed. Vennerable (Supplementary Figure S4B), venneuler (Supplementary Figure S4C) and

Table 5. Basic features of 11 tools used to generate aggregate, matrix and others diagrams

Tools	Application	Sets	Input data format	Output image format	Key parameters	Text output	Reference
RainBio	Web	2–15	Uploaded lists	No	Displays the size of main intersections and similarities between sets and interactions	Online table	[28]
AggreSet	Web	≥ 50	Uploaded lists	No	Selection, filtering, comparison and analysis of set relations	No	[32]
Multiple List Comparator	Web	≥ 50	Uploaded lists, pasted values	PNG, SVG	Paired heatmap (infinite), diagram color and labels	TXT	http://www.molbio.tools.com/listcompare.html
UpSet	Web	2–30	Uploaded lists	No	Aggregation by degrees, sets, deviation and overlaps, diagram color and other settings	No	[54]
EVenn	Web	> 6	Pasted values	PDF (UpSet), SVG (Network)	Network has nodes font, edges smooth and physics repulsion parameters	No	http://www.ehbio.com/test/venn/#/
SuperExactTestR		> 7	Lists	R graphics	Statistical distributions and exact probability, sort and color settings	R text	[55]
Intervene	Web	2–30	Uploaded lists	PNG, SVG, PDF, TIFF	Diagram color, plot width and height, point size and label size	Online table	[38]
UpSetR	R	2–30	Lists, values	R graphics	Shows intersections of sets with bar chart statistics, sort, color and other	No	[15]
BranchingSets	Web	2–10	No	No	Node contraction and expansion and color modification	JavaScript	[30]
Linear Diagram Generator	Web	≥ 7	Pasted values	SVG	Background using Gray bar chart or line style	Web	https://www.cs.kent.ac.uk/people/staff/pjr/linear/index.html
Bubble Sets	Web	No	No	No	Intersect using bubble diagram	JavaScript	[31]

Venn Diagram Plotter (Supplementary Figure S4D) can also display Euler diagrams of up to 3 datasets, but they have almost no customizable parameters for beautifying images in real-time (Table 3).

Euler diagrams are popular because of their ability to show proportional relationships, while they demonstrate accuracy for 2–3 datasets. For example, Bárcena et al. [49] used the VennDiagram [13] package to generate proportional Euler diagrams to visualize the shared chip peaks between two methyltransferases. The shared and host-specific microbiome of the rhizosphere and root were also discovered and visualized using the Euler diagrams [50] in the VennDIS tool [42].

Venn and Euler diagram variants

Several tools are available that produce irregular diagrams and variants of the Venn/Euler diagrams (Table 4), and they can process more datasets than tools that display classic Venn diagrams and Euler diagrams. The nVenn [37] is available in a web platform, the R and Linux environments, which can draw quasiproportional Euler diagrams of up to 6 datasets. One thing to be note is: the term ‘quasiproportional’ was described by

literature Pérez-Silva et al. [37] (nVenn). The nVenn inserts different sizes of circles in each subset area to represent the proportional relationships between 6 datasets (Figure 6A), which picture the size of each sub area in a clearer way as compared with EVenn (Figure 5C) and eulerr (Figure 5D). The disadvantage is that nVenn [37] runs slower than VennPainter [14] and jvenn [12]. In addition, the VennPainter can display a new nested Venn diagram with 5–8 sets (Figure 6B) and provide text output results for up to 31 datasets. The jvenn uses triangles and histograms to display 6 datasets (Figure 6C) and has a user-friendly interface. In addition to displaying pictures similar to jvenn [12], EVenn can also depict flower plots (Figure 6D). Intervene [38] and venerable provide a variety of Venn and Euler layouts, such as squares (Figure 6D), triangles (Figure 6E) and ChowRuskey diagrams (Figure 6F), although they have few customizable parameters (Table 4). Moreover, eulerGlyphs [29] treats the set elements as points and places them into a graph (Figure 6H), but eulerGlyphs does not support user-defined data.

The quasiproportional Euler diagram and the nested Venn diagram were used to show multiple datasets. To name a few, the nVenn [37] tool was used to visualize the overlapping DEGs in a quasiproportional Euler diagram across the five

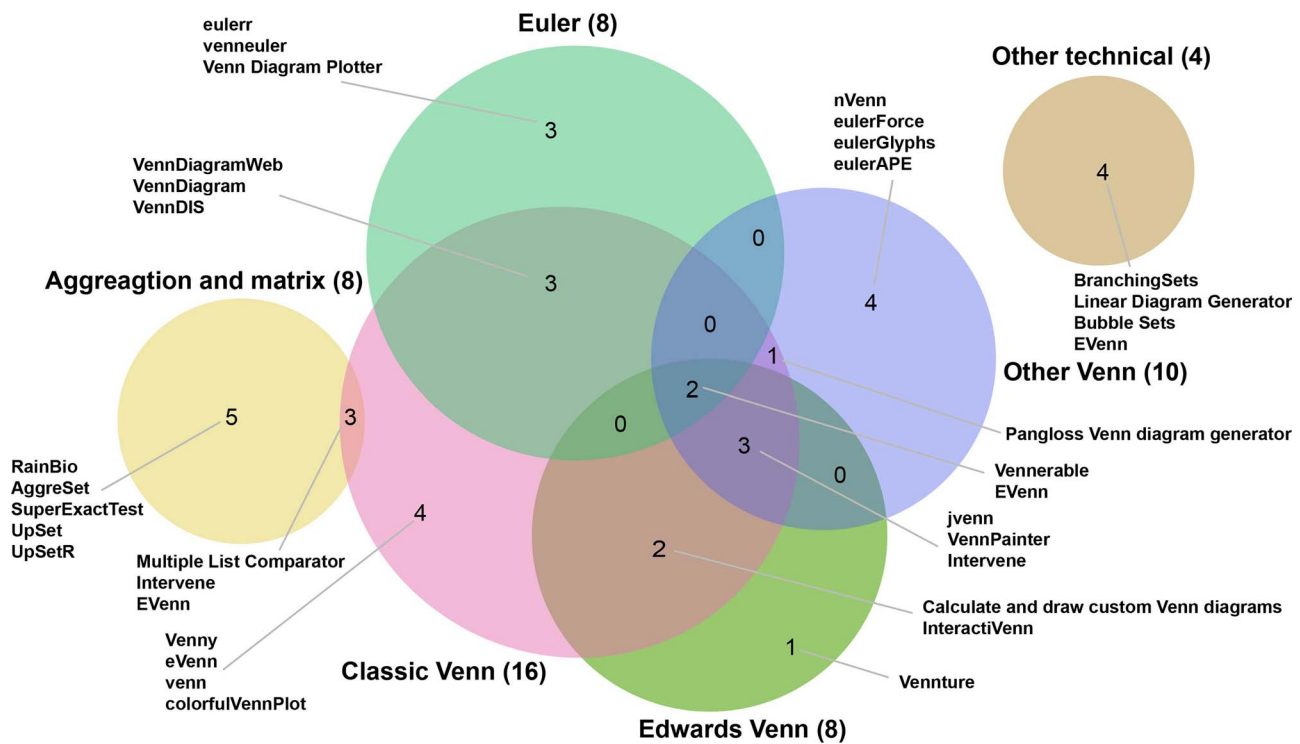


Figure 2. Thirty-two tools that can draw Venn diagrams with different layouts. The numbers in parentheses represent the number of tools that can draw the corresponding layouts, and the intersections represent tools that can draw multiple layouts. Intervene and EVenn appear more than twice in the picture due to the inability to further simplify the layout. The diagram was drawn using PowerPoint.

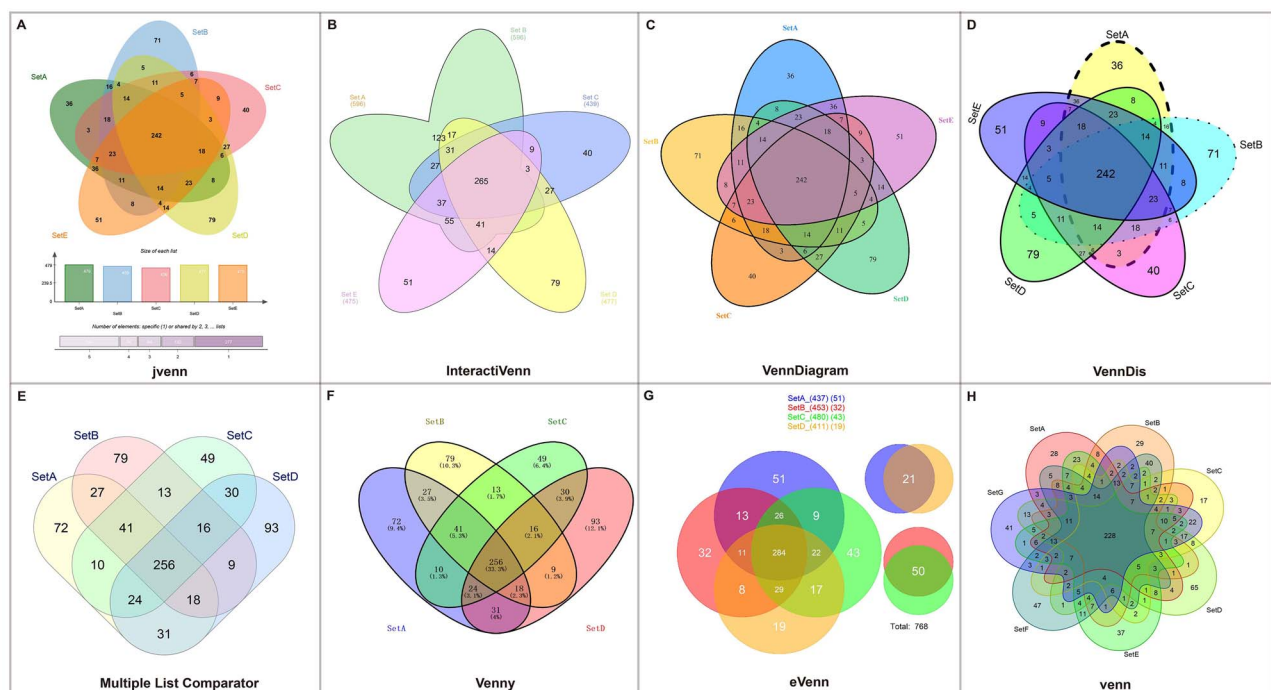


Figure 3. Classic Venn diagrams. The names of the tool used to create the diagram are listed at the bottom of the panels (similarly for the remaining images). (A-D) are classic Venn diagrams with 5 datasets; (E-G) are classic Venn diagrams with 4 datasets and (H) is a classic Venn diagram with 7 datasets.

samples (four mutant neurons and one patient brain) in humans, and it revealed that a substantial number of downregulated

genes were shared in two mutant neurons [52]. Tong et al. [53] used the VennPainter [14] tool to visualize the shard and

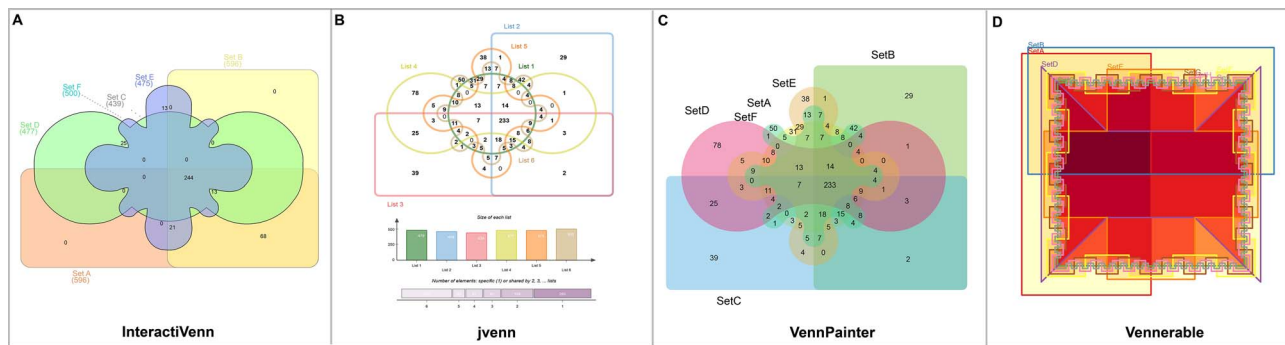


Figure 4. Edwards' Venn diagrams. (A–C) Edwards' Venn diagrams with 6 datasets; (D) 'Battle' Edwards' Venn diagrams with 9 datasets.

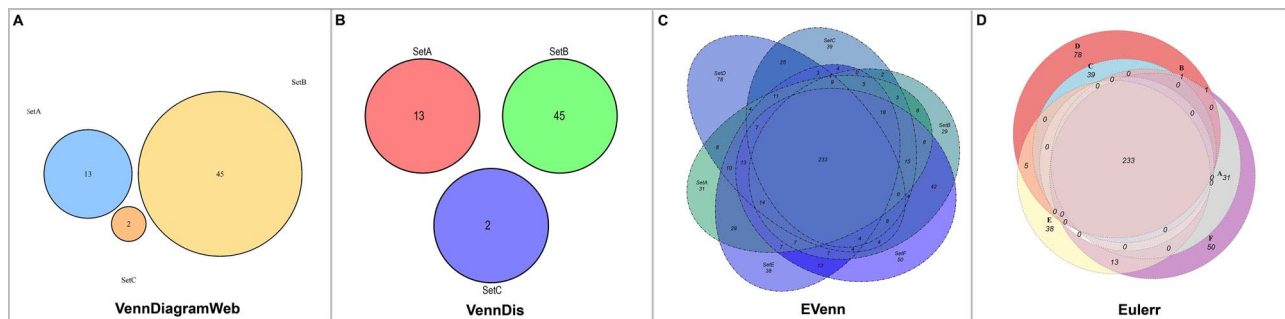


Figure 5. Euler diagrams. (A) proportional Euler diagram without intersections; (B) Euler diagrams without intersections; (C and D) elliptical and circular Euler diagrams, respectively, with 6 sets.

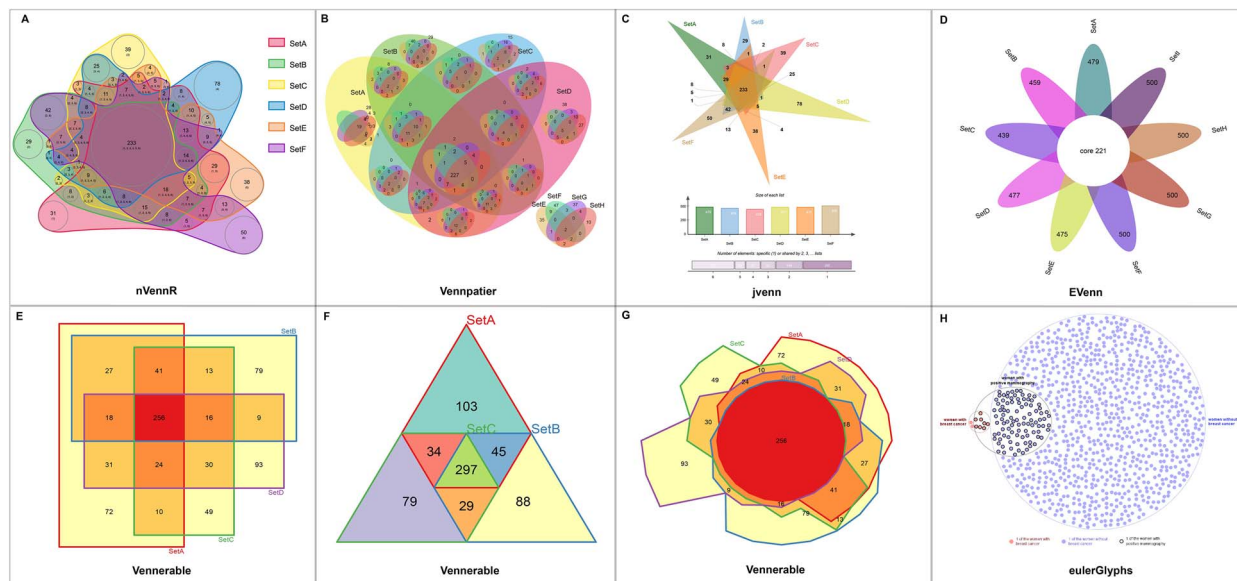


Figure 6. Venn and Euler diagram variants. (A) Quasiproportional Euler diagram with 6 datasets in nVennR (R version of nVenn); (B) nested Venn diagrams with 8 datasets; (C) triangular Venn diagrams with 6 datasets; (D) flower plot with 8 datasets; (E) square diagram with 4 datasets; (F) triangular diagram with 3 datasets; (G) ChowRuskey diagrams with 4 datasets and (H) set element-based Venn diagram (each point represents an element).

unique gene families among eight species with a nested Venn diagram.

Aggregate or matrix-based diagrams

Although Venn and Euler diagrams are the most popular visualization methods, the development of 'omics' data has led to visualization requirements that Venn and Euler diagrams cannot

meet, such as presenting the intersection between dozens of datasets or more. Aggregate and matrix diagrams can be used to compare large numbers of datasets (Table 5). The R package UpSetR [15] is one of the most widely used tools (google scholar citations: >600 by 27 February 2021) for visualizing a large number of datasets, and it has a concise and simplified interface. Furthermore, UpSetR can optimize the picture to use multiple parameters and mark intersections of interest using

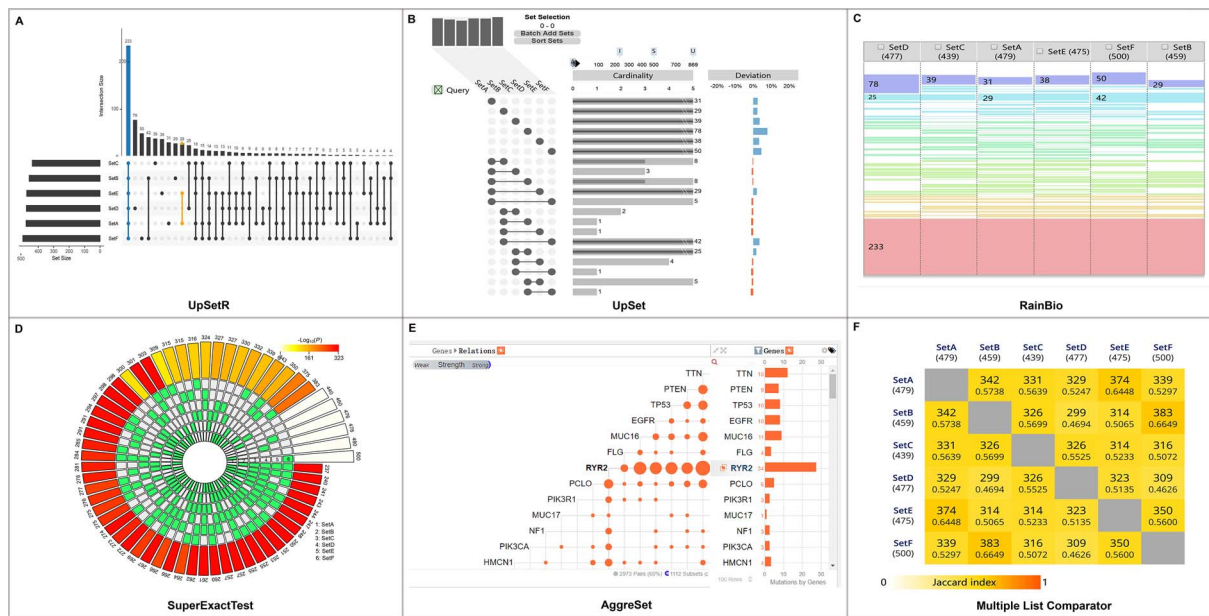


Figure 7. Aggregate and matrix-based diagrams. (A and B) UpSet diagrams with 6 datasets; (C) proportional rainbow boxes with 6 datasets; (D) multiple-set intersections and statistical circle diagrams with 6 datasets; and (E, F) pairwise comparison diagrams with multiple datasets.

special symbols or colors (Figure 7A). One thing to note is that UpSetR [15] also has a JavaScript version. The web tools UpSet [54], RainBio [28], EVenn, AggreSet [32] and Multiple List Comparator can display multiple datasets and allow users to view the intersection interactively. UpSet (Figure 7B), EVenn and RainBio (Figure 7C) are able to show the intersections up to 30, 30 and 15 datasets, respectively. EVenn is similar to UpSetR [15]. AggreSet (Figure 7E) and Multiple List Comparator (Figure 7F) perform pairwise comparisons between paired datasets. However, only EVenn can generate vector graphics. Moreover, several tools are based on aggregation and matrix techniques. For instance, SuperExactTest (Figure 7D) [55] uses a multilayer circular or two-dimensional matrix to display multiple datasets, and it provides the statistical calculations for multiple-set intersections. PowerSet [56] visualizes the intersection of different sets and displays them in treemaps. Radial Sets [57] analyzes overlapping patterns between datasets using Chord diagrams. ConSet [58] uses a permutation matrix to display relationships among multiple sets and compares 2–3 sets with Fan diagrams.

Taken together, the UpSetR [15] plot is the most popular for displaying multiple datasets compared with other aggregate or matrix-based tools. For example, the shared and unique marker genes from five endothelial cell samples were found using UpSetR [15] plot [60]. The eight experiment groups were plotted in a UpSetR [15] plot to uncover the common RNA-binding proteins [61]. Ellrott et al. [62] displayed the shared and unique Variant counts of 16 samples using the UpSetR [15] plot.

Other technical diagrams

To provide a comprehensive evaluation of dataset visualization techniques, we also tested some other technical diagram tools, such as linear, node-link and overlay diagrams (Table 5). Linear Diagram Generator [59] can display user data by generating a linear diagram composed of line segments when the user enters a list of intersection and union values (Figure 8A). However, the Linear Diagram Generator only has a simple web version

at present. EVenn can display the relationship between the different elements of datasets in a node-link diagram (Figure 8B), and it can modify the node color and smooth edges, among other actions. To explore the relationships between the genes of different networks or pathways, BranchingSets [30], Bubble Sets [31] and BiSet [63] were designed to view the relationship between elements based on Node-link or Overlay technology. BranchingSets adopts node-link diagrams to show the intersections between different pathways, and genes from the same pathway are displayed in the same color. Nodes with overlapping colors indicate that the genes are involved in multiple pathways (Figure 8C). Bubble (Figure 8D) and kelp diagrams [64, 65] are other options to show the relationships between sets.

Venn diagram application tools

Collection of Venn diagram application tools

Venn diagram application tools have been used in numerous biological fields. To understand the development trends of these application tools, we collected a total of 13 tools. Table 6 lists their application platforms, data input formats, supported maximum datasets, output formats and functions. Nine of these tools use a web platform interface, while the remaining ones use a local interface. According to the functional features, we divided these tools into five categories. In the following sections, we compare and visualize the tools in each category.

- (1) Gene networks that compare and analyze network relationships between sets, e.g. NetVenn [16], DiVenn [66], Venn-diaNet [67] and Netsets.js [68];
- (2) Set visualization and database linking to view information, e.g. OrthoVenn2 [69], GeneVenn [70] and BioVenn [71];
- (3) Software output file visualization that input the results data from certain software and display overlapping and nonoverlapping elements within the Venn diagram, e.g. VennMaster [72], VennPlex [25], VennBLAST [73] and BedSect [74];

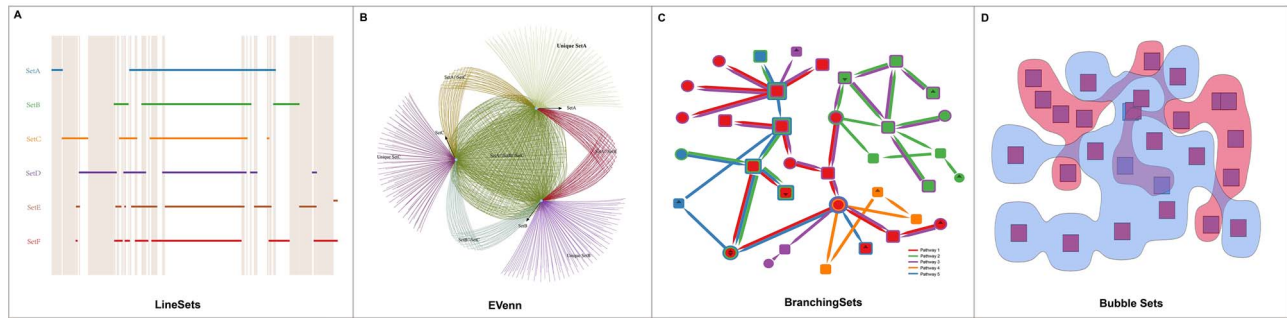


Figure 8. Linear, node-link and overlay diagrams. (A) Linear diagram with 6 datasets, where sets are arranged in parallel and areas of overlapping lines represent intersections; (B) network diagram with 3 datasets, where the three tiny lightblue nodes represent Sets A, B and C, respectively (labeled as SetA, SetB and SetC), the biggest darkgreen nodes in the middle represent the shared intersections between 3 sets (labeled as $\text{SetA} \cap \text{SetB} \cap \text{SetC}$), the top nodes of the three sectors represent the unique elements of each set (labeled as Unique SetA, Unique SetB and Unique SetC), and the remaining nodes represent the intersections between every 2 sets (labeled as $\text{SetA} \cap \text{SetB}$, $\text{SetA} \cap \text{SetC}$ and $\text{SetB} \cap \text{SetC}$); (C) node-link pathway diagram with 5 datasets, where the nodes connected by the same color represent a pathway, the overlap of lines or nodes represent intersections and (D) bubble diagram with 2 sets, the diagram mainly displays the spatial relationship between datasets, where bubbles connected by the same color represent a dataset.

- (4) Microbiome analysis tools that discover and visualize core microbiomes, e.g. MetaCoMET [75];
- (5) Other tools, e.g. DynaVenn [76] and some toolkits [77, 78].

The network relationships between genes are crucial for understanding gene functions [89]. Accordingly, they are important for mining shared and unique networks or genes from multiple gene lists. Here, we investigated four web tools that have Venn diagram functions for analyzing gene network relationships. In NetVenn [16], gene lists can be uploaded or pasted into the corresponding area as input for up to 4 sets, then the gene network relationship of each list is analyzed from the men-tha database [79] (the database archives protein interaction networks for many species, such as human, mouse and Arabidopsis), and overlapping or non-overlapping elements are displayed in a Venn diagram while gene interactions are displayed in a network diagram (Figure 9A). NetVenn also has a user-friendly interaction, such as highlighting genes in a subset and viewing the Gene Ontology (GO) terms for genes by clicking the mouse [16]. As such, JavaScript framework Netsets.js [68] can also display gene network relationships using the Venn and network diagrams, which enables a link to the web for use or connection to the Cytoscape application by Netsets-Cyapp. Particularly, the tool takes interaction lists as inputs and displays them in jvenn [12] and the interaction network diagram (Figure 9D) for up to 6 datasets. Venn-diaNet [67] supports up to three gene lists and displays genes as points in the Venn diagram (Figure 9B), and it allows users to zoom in to view the relationships of the gene network and screen core genes by ranking. DiVenn [66] takes the DEGs of the transcriptome as input files, displays relationships in a network interaction diagram for up to 8 datasets, and allows users to interactively visualize gene pathways and ontologies with an online table (Figure 9C).

To view the usage of these tools, we searched some cases to clarify their functions. For instance, NetVenn [16] was used to compare the autophagy networks in human [80]. The DiVenn [66] tool was used to visualize the DEGs from the six different stages of wheat in a node-link network [81], which was also used to plot the protein genes identified by mass spectrometry from two stress conditions in a node-link network plot [82].

Set visualization and database linking to view information

Exploring the relationships among different sets of genes and linking them to databases has been a long-term research interest

[17]. Numerous tools have been developed to compare different sets of genes. For instance, OrthoVenn2 [69], GeneVenn [70] and BioVenn [71] can visualize the intersections and unions of gene datasets when inputting multiple gene sequences or lists and provide annotation information online for each gene or links to biological databases for viewing. In addition, OrthoVenn2 can analyze orthologous clusters for up to 12 protein sequence lists and displays shared or unique orthologous clusters via jvenn ([12]; 2–6 datasets) or ClusterVenn-based (2–8 datasets) diagrams (Figure 10A; [69]). Furthermore, it also provides online information or links for querying the GO enrichment results of each cluster. BioVenn (Figure 10B) and GeneVenn (Supplementary Figure S5A) display the overlaps and nonoverlaps of gene lists in area-proportional and classic Venn diagrams. GeneVenn can provide gene information by linking to NCBI's Entrez Nucleotide database [70], and BioVenn is able to query the biological information of identifiers by linking them to public databases when lists contain Affymetrix and EntrezGene identifiers [71].

For these tools, OrthoVenn2 [69] was frequently used to identify orthologous clusters between species, while GeneVenn [70] and BioVenn [71] were mainly used to visualize the intersections between several datasets. For example, Kim et al. [83] and Rodrigues et al. [84] used OrthoVenn2 [69] to visualize the distribution of orthologous genes among five *Streptomyces venezuelae* strains and six *Candida* spp. related species, respectively. GeneVenn [70] was used to display the shared and unique genes between several gene sets [85, 86]. In addition, the proportional Venn diagrams of BioVenn [71] were also used to display the intersections and unions between samples [87, 88].

Software output file visualization

To obtain shared or unique identifiers from multiple datasets, basic programming skills are required to extract identifiers from the output files of certain software tools. The local interface tools VennMaster [72], VennPlex [25] and VennBLAST [73] can visualize the data from the output of software programs or multi-omics analysis. VennMaster is integrated into the GoMiner [90] application, which has the ability to display an area-proportional Euler diagram to represent GO categories with flagged genes (Figure 11A; [72]). VennPlex assesses DEGs lists from multi-omics analysis and displays their relationships through a Venn diagram (Figure 11B; [25]). VennBLAST is allowed to extract the output of blast alignments [91] and display it in a Venn diagram format (Figure 11C; [73]). The web tool BedSect [74] takes a bed

Table 6. List of basic features for 13 Venn diagram application tools

Tools	Application	Input	Sets	Output	Function	Reference	Category
NetVenn	Web	Uploaded/pasted gene lists	2–4	SVG, PNG, PDF	Uses 2–4 gene lists as input to display a Venn diagram, predicts the interaction network of each gene list, combines Venn diagrams to highlight overlapping genes and provides gene annotations	[16]	Gene networks
DiVenn	Web	Uploaded gene lists	2–8	SVG, PNG	Accepts up- and downregulation information of gene lists and displays them in a network diagram	[66]	
Venn-diaNet	Web	Uploaded gene lists	2–3	TXT	Accepts multiple gene lists to predict the interaction network, displays it in a Venn diagram and can select seed genes based on scores from subsets	[67]	
NetSets.js	JavaScript	Pasted gene lists	2–6	SVG, PNG, TXT	Support input of 2–6 interaction lists and displays overlaps within a Venn diagram and interaction relationships within a network diagram	[68]	
OrthoVenn2	Web, Command line (Linux)	Protein sequences	2–12	SVG, PNG	The online version supports the input of genome sequences of 2–12 species, uses jvenn or clustervenn to display orthologous genes between them and provides gene GO enrichment information	[69]	Set visualization and database linking to view information
GeneVenn	Web	Uploaded/pasted gene lists	2–3	No	Supports display 2–3 gene lists in a Venn diagram, and the elements of each subset can be linked to NCBI to query biological information	[70]	
BioVenn	Web	Uploaded/pasted gene lists	2–3	SVG, PNG	Supports a 2–3-sample area-proportional Venn diagram, and the elements of each subset can be linked to public databases to query biological information	[71]	
VennMaster	Localized interface	GO result files	>6	SVG, JPG	Visualizes GO analysis results in an area-proportional Euler diagram	[72]	Software output file visualization
VennPlex	Localized interface	Uploaded gene lists	2–4	BMP, JPG, PNG, GIF	Accepts 2–4 gene lists with the up- and downregulation information and displays them in a Venn diagram	[25]	
VennBLAST	Localized interface	Blast result files	2–3	TXT	Visualizes 2–3 blast results in a Venn diagram	[73]	
BedSect	Web	Bed files	>6	PNG, PDF, TXT	Compares multiple bed files and displays subsets with an UpSet diagram or correlation heatmap	[74]	
MetaCoMET	Web	Biom files	2–6	SVG, PNG, PDF	Supports input BIOM format files and displays the shared core microbiome of 2–6 samples in a Venn diagram	[75]	Microbiome analysis
DynaVenn	Web	Uploaded gene lists	2–3	SVG	Compares 2–3 gene lists with dynamic Venn diagrams and computes the p-value for the overlapping elements	[76]	Others

file as input and displays the results with an UpSet diagram or correlation heatmap (Supplementary Figure S5B). Although these tools support the input of data in unique formats, challenges remain, such as a lack of customizable picture attributes and user-friendly interactions for localization interface tools.

Identification of the overlaps of output results of some types of software is convenient for using existing tools. VennMaster [72] was used to display significant DEGs in different GO categories [92], as well as to plot the overlapping parts in a proportional Venn diagram [93]. Additionally, the tblastx alignments of

top hit sequences from the comparison of *Cancer borealis* and *Homarus* transcriptomes were obtained [94] by VennBLAST [73].

Microbiome analysis

Identifying core microbiomes from multiple experimental conditions helps researchers further understand microbial environments and functions [95, 96]. The operational taxonomic unit (OTU) is a taxonomic unit that distinguishes microbiome groups, and some software tools can store OTUs under multiple

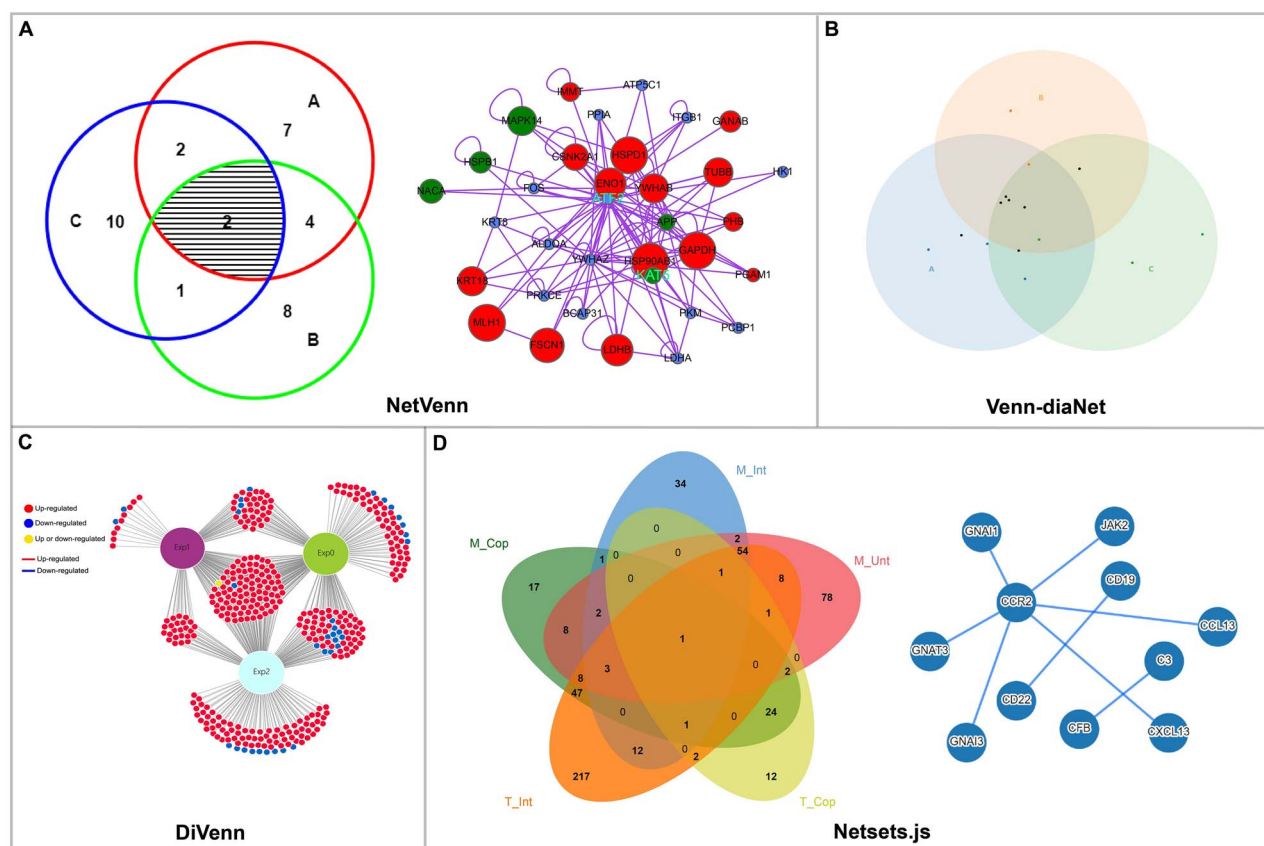


Figure 9. Gene network and Venn diagrams. (A) Venn and network diagram with three gene lists, in which the genes in the green fluorescent font in the network diagram represent the two genes at the intersection of the shaded part in the Venn diagram. (B) Venn diagram with three gene lists, in which each point represents a gene; the magnification shows the interaction relationships between genes. (C) Network diagram with three gene lists. (D) Venn diagrams with five interaction pair lists and network diagram in a subset.

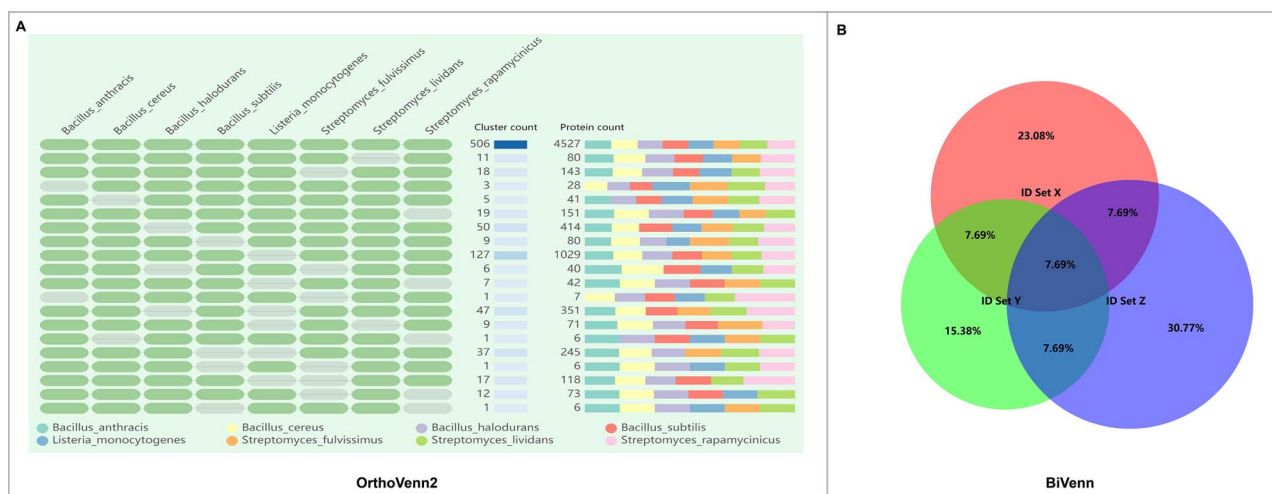


Figure 10. ClusterVenn and area-proportional Venn diagrams. (A) ClusterVenn diagram with protein sequences of eight genomes. (B) Area-proportional Venn diagram with 3 datasets.

experimental conditions as a common format file (Biological Observation Matrix Data, BIOM) for downstream microbiome analysis [97–99]. Venn diagrams have been used in multiple

studies to visualize core microbiomes [10, 100, 102]. MetaCoMET [75] takes a BIOM file as input and enables the discovery and visualization of the core microbiomes while providing a user-

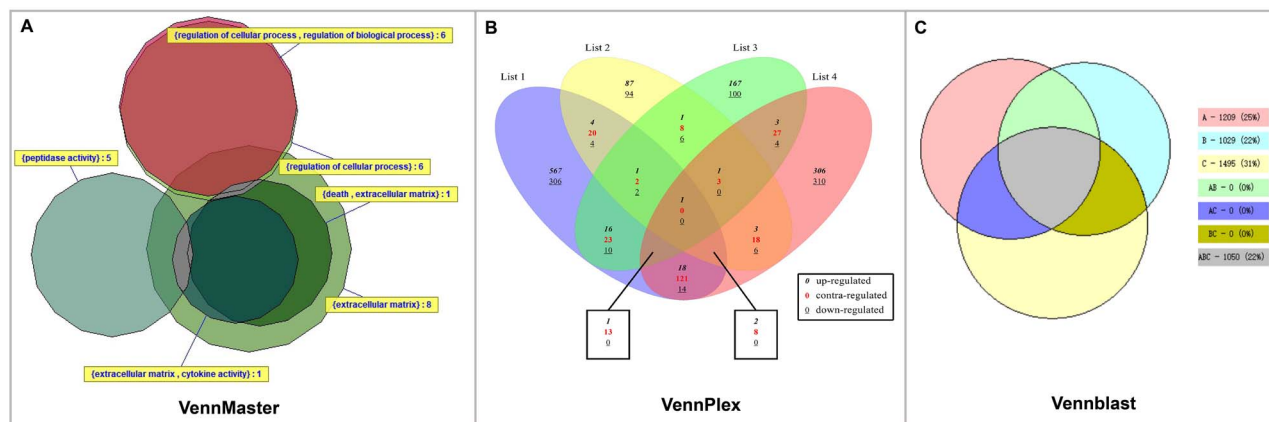


Figure 11. Area-proportional Euler and Venn diagrams. (A) Area-proportional Euler diagram with six GO categories; (B) Venn diagram with four DEGs lists, in which the three numbers in each subset represent up, contra and downregulated genes from top to bottom; and (C) Venn diagram with 3 blast alignment datasets.

friendly and interactive graphical interface for analyzing each subset defined by the union or disjunction of groups within the Venn diagram (Supplementary Figure S5C).

Multiple attempts have shown MetaCoMET [75] has been widely utilized to discover core microbiomes with a Venn diagram based on input BIOM format data. For example, the core microbiomes were exposed in Venn diagrams from five experimental sets [103] with the help of MetaCoMET [75]. Toro-Londono et al. [104] depicted the OTU distribution of the gut microbiota in the four experimental groups of children using MetaCoMET [75].

Other Venn diagram application tools

In addition to the application tools we described above, we also retrieved some Venn diagram application tools that function from other aspects or are integrated into some toolkits. For example, DynaVenn [76] uses two to three gene lists as an input to visualize their intersection and compute the p-value for overlapping elements while displaying dynamic Venn diagrams (Supplementary Figure S5D). Moreover, Venn diagrams have also been integrated into toolkits, such as limma [105], gplots [106] and TBtools [107], and they are also involved in some bioinformatics pipelines [77, 78, 99].

Conclusions and future outlook

In this review, we aim to provide guidelines for Venn diagram generator selection. Generally, more favored generators have a high capacity for more datasets, allowing for customization and beautification of the graphics, accepting simple input formats, supporting high-quality output pictures and assisting user-friendly interactions. Our review indicated that the web tool EVenn can draw high-quality graphics with multiple layouts (classic and Edwards' Venn, Euler, aggregate and node-link diagrams), which are supported by multiple parameters. The InteractiVenn [41], VennDiagram [13] and VennDIS [42] tools have more comprehensive functions for drawing classic Venn diagrams with 2–5 datasets. The jvenn [12] and VennPainter [14] can depict the classic and Edwards' Venn diagrams with 2–5 and 2–6 datasets, respectively. The VennPainter supports 5–8 datasets with nested Venn diagrams. Among the Euler diagram tools, nVenn [37] and VennDiagramWeb [40] provide more accurate displays of set relationships. Among the aggregate and matrix-based technical diagram tools, UpSetR [15] has more

parameters to optimize the display via simplified and concise pictures.

Taken together, most of classic and Edwards' Venn diagrams are highlighted in the visualization of 5 and 6 datasets, respectively. The Euler diagrams are outstanding in showing the proportional relationship of 2–3 datasets. In order to show more datasets, some recent studies have developed some Venn and Euler Variant tools that can generate the quasiproportional Euler diagrams of 6 datasets [37] and the nested Venn diagrams of 8 datasets [14]. One exception is the matrix diagram tool UpSetR [15] can process up to 30 datasets. Conversely, there are also some shortcomings of the Venn diagram generators. For instance, (i) most of the generators support a few datasets due to layout difficulties, because there are $2^n - 1$ subsets of n datasets that need to be laid out; (ii) a substantial number of generators center on the data, ignoring the potential relationships and other attributes between the elements of datasets; (iii) an extremely large number of generators makes it difficult to quickly observe certain overlaps for 5–6 datasets when interactions are lacking and (iv) none of the Euler diagram tools can generate quasiproportional Euler diagrams using circles or ellipses for more than 3 datasets. Based on the above issues, many potential challenges remain for developing more comprehensive Venn diagram tools. Therefore, future development of Venn and Euler diagram tools is warranted for a larger number of datasets, such as algebraic construction [108], nested [109], 3D [110] and multiple graphics cross-diagrams, among others. Furthermore, user-friendly interactions, aesthetically pleasing graphics and accurately defined area proportions must be optimized and expanded. For matrix-based, aggregation and other technology diagrams, tools with user-friendly interactions and aesthetically pleasing interfaces must be developed, and more effective parameters must be customizable to optimize the graphics.

Some Venn diagram application tools support input biological data with a special format (e.g. gene sequences, and output result files from some software), as well as performing the follow-up calculations, filtering and visualization [68, 69, 72, 75], whereas others accept gene lists and calculate the overlaps between gene sets [16, 67]. However, these tools (e.g. VennMaster [72], VennPlex [25] and VennBLAST [73]) have single functions and poor output picture quality. We suggest that future studies of Venn diagram application tools should be based on a new perspective that analyzes the intersection between multiple biological data, especially the original datasets of multiple omics

directly used as input, and display the core genes in the figure. The function of an ideal Venn diagram application tool should include but not be limited to (i) permitting the input of user-defined multiple omics data; (ii) automatically screening core genes and visualizing them within a Venn diagram; (iii) combining the Venn diagram and the network diagram to show a higher resolution of gene interaction relationships; (iv) integrating Venn diagrams into various bioinformatics pipelines or platforms and (v) allowing visualization of the output files from more software programs and providing user-friendly interactions and aesthetically pleasing interfaces. More importantly, users should be able to obtain customized and high-quality pictures for publication.

Key Points

- Venn diagram generators and Venn diagram application tools play an indelible role in the screening of core elements between datasets and can effectively help users further understand and explore potentially shared characteristics of their data.
- We conducted a comprehensive comparison of 32 Venn diagram generators in terms of data input, image output, layouts, graphical interfaces and functions. In addition, we compared 13 Venn diagram application tools in different fields of bioinformatics.
- Venn diagram generators are highlighted in terms of the visualization of intersections and unions between sets with multiple layouts, and different layouts support different numbers of datasets. Venn diagram application tools are also outstanding in visualizing information from the multiple fields of bioinformatics. We recommend a number of high-performance tools for user reference in each Venn diagram generator layout category. We also discuss the challenges and future directions for Venn diagram generators and application tools to satisfy the growing demand in the omics era.

Supplementary Data

Supplementary data are available online at [https://academic.oup.com/bib](https://academic.oup.com/bib/article/22/5/bbab108/6220174).

Acknowledgments

We thank Guoqing Zhang and Hailong Guo for their help with the collection of the Venn Diagram generators. We also thank Zhaoyuan Wei and Fang Lu for their help with the use of the Venn Diagram applications.

Funding

This work was supported by the National Natural Science Foundation of China (31871330), the Scientific Research Starting Foundation of Southwest University (SWU118103), the Fundamental Research Funds for the Central Universities (XDJK2019TJ003) and the Chongqing Municipal Training Program of Innovation and Entrepreneurship for Undergraduates (S201910635003).

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Venn JI. On the diagrammatic and mechanical representation of propositions and reasonings. *London, Edinburgh, Dublin Philos Mag J Sci* 1880;10:1–18.
2. Chow S, Rodgers P. *Constructing Area-Proportional Venn and Euler Diagrams with Three Circles*. Paris: Euler Diagrams, 2005.
3. Poythress VS, Sun HSJTP. A method to construct convex, connected Venn diagrams for any finite number of sets. *Pentagon* 1972;31:80–2.
4. Edwards J, Johnson C, Santos-Medellín C, et al. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A* 2015;112:E911–20.
5. Jin S, Zong Y, Gao Q, et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* 2019;364:292–5.
6. Lake BB, Ai R, Kaeser GE, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* 2016;352:1586–90.
7. Marks H, Kalkan T, Menafra R, et al. The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 2012;149:590–604.
8. Probst AV, Almouzni G. Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. *Trends Genet* 2011;27:177–85.
9. Huang AC, Jiang T, Liu YX, et al. A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science* 2019;364:eaau6389.
10. Chen MX, Lu CC, Sun PC, et al. Comprehensive transcriptome and proteome analyses reveal a novel sodium chloride responsive gene network in maize seed tissues during germination. *Plant Cell Environ* 2020;44:88–101.
11. Jacob F, Pather SR, Huang, et al. Human pluripotent stem cell-derived neural cells and brain organoids reveal SARS-CoV-2 neurotropism predominates in choroid plexus epithelium. *Cell Stem Cell* 2020;27:937, e9–50.
12. Bardou P, Mariette JM, Escudié F, et al. Jvenn: an interactive Venn diagram viewer. *BMC Bioinf* 2014;15:1–7.
13. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinf* 2011;12:35.
14. Lin G, Chai J, Yuan S, et al. VennPainter: a tool for the comparison and identification of candidate genes based on Venn diagrams. *PLoS One* 2016;11:e0154315.
15. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017;33:2938–40.
16. Wang Y, Thilmony R, Gu YQ. NetVenn: an integrated network analysis web platform for gene lists. *Nucleic Acids Res* 2014;42:W161–6.
17. Wang Y, Coleman-Derr D, Chen G, et al. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 2015;43:W78–84.
18. Sierro N, Battey JN, Ouadi S, et al. Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol* 2013;14:R60.
19. Zhou X, Wang B, Pan Q, et al. Whole-genome sequencing of the snub-nosed monkey provides insights into folivory and evolutionary history. *Nat Genet* 2014;46:1303–10.

20. Tschopp P, Sherratt E, Sanger TJ, et al. A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature* 2014;**516**:391–4.
21. Gerbracht JV, Volker B, Thiago BB, et al. CASC3 promotes transcriptome-wide activation of nonsense-mediated decay by the exon junction complex. *Nucleic Acids Res* 2020;**48**:8626–44.
22. Bailey MH, Tokheim C, Porta-Pardo E, et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* 2018;**173**:371–85.
23. Barreto FS, Watson ET, Lima TG, et al. Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. *Nat Ecol Evol* 2018;**2**:1250–7.
24. Zhang H, Gan H, Wang Z, et al. RPA interacts with HIRA and regulates H3.3 deposition at gene regulatory elements in mammalian cells. *Mol Cell* 2017;**65**:272–84.
25. Cai H, Chen H, Yi T, et al. VennPlex—a novel Venn diagram program for comparing and visualizing datasets with differentially regulated datapoints. *PLoS One* 2013;**8**:e53388.
26. Chapman P, Stapleton G, Rodgers P et al. Visualizing sets: an empirical comparison of diagram types. In: *International Conference on Theory and Application of Diagrams*. Berlin, Heidelberg 2014; pp 146–60.
27. Alsallakh B, Micallef L, Aigner W, et al. The state-of-the-art of set visualization. *Comput Graph Forum* 2016;**35**:234–60.
28. Lamy JB, RainBio TR. Proportional visualization of large sets in biology. *IEEE Trans Vis Comput Graph* 2019;**3285**–98.
29. Micallef L, Dragicevic P, Fekete J. Assessing the effect of visualizations on Bayesian reasoning through crowdsourcing. *IEEE Trans Vis Comput Graph* 2012;**18**:2536–45.
30. Paduano F, Etemadpour R, Forbes, et al. *BranchingSets. Proceedings of the 9th International Symposium on Visual Information Communication and Interaction - VINCI '16*. In: Dallas TX USA, 2016, 9–16.
31. Collins C, Penn G, Carpendale S. Bubble sets: revealing set relations with isocontours over existing visualizations. *IEEE Trans Vis Comput Graph* 2009;**15**:1009–16.
32. Yalcin MA, Elmquist N, Bederson BB. AggreSet: rich and scalable set exploration using visualizations of element aggregations. *IEEE Trans Vis Comput Graph* 2016;**22**:688–97.
33. Edwards AWF. *Cogwheels of the Mind: the Story of Venn Diagrams*. JHU Press, Baltimore and London, 2004.
34. Baron MEA. Note on the historical development of logic diagrams: Leibniz, Euler and Venn. *Math Gazette* 1969;**53**:113.
35. Rao S, Mahoney JV. *Analyzing an Image Showing a Node-Link Structure*. Google Patents, 1996;5:522–022.
36. Aggarwal A, Guibas LJ, Saxe J, et al. A linear-time algorithm for computing the Voronoi diagram of a convex polygon. *Discrete Comput Geom* 1989;**4**:591–604.
37. Pérez-Silva JG, Araujo-Voces M, Quesada V, et al. nVenn: generalized, quasi-proportional Venn and Euler diagrams. *Bioinformatics* 2018;**34**:2322–4.
38. Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinf* 2017;**18**:287.
39. Micallef L, Rodgers P. eulerAPE: drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* 2014;**e101717**:9.
40. Lam F, Lalansingh CM, Babaran HE, et al. VennDiagramWeb: a web application for the generation of highly customizable Venn and Euler diagrams. *BMC Bioinf* 2016;**17**:401.
41. Heberle H, Meirelles GV, da Silva FR, et al. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinf* 2015;**16**:169.
42. Ignatchenko V, Ignatchenko A, Sinha A, et al. VennDIS: a JavaFX-based Venn and Euler diagram software to generate publication quality figures. *Proteomics* 2015;**15**:1239–44.
43. Jarvis DE, Ho YS, Lightfoot DJ, et al. The genome of *Chenopodium quinoa*. *Nature* 2017;**542**:307–12.
44. Luo YJ, Takeuchi T, Koyanagi R, et al. The Lingula genome provides insights into brachiopod evolution and the origin of phosphate biomineralization. *Nat Commun* 2015;**6**:1–10.
45. Melo SA, Luecke LB, Kahlert C, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature* 2015;**523**:177–82.
46. Martin B, Chadwick W, Yi T, et al. VENNTURE—a novel Venn diagram investigational tool for multiple pharmacological dataset analysis. *PLoS One* 2012;**7**:e36911.
47. Xing L, Zhang D, Zhao C, et al. Shoot bending promotes flower bud formation by miRNA-mediated regulation in apple (*Malus domestica* Borkh.). *Plant Biotechnol J* 2016;**14**:749–70.
48. Georgiadis P, Gavriil M, Rantakokko P, et al. DNA methylation profiling implicates exposure to PCBs in the pathogenesis of B-cell chronic lymphocytic leukemia. *Environ Int* 2019;**126**:24–36.
49. Bárcena C, Quirós PM, Durand S, et al. Methionine restriction extends lifespan in progeroid mice and alters lipid and bile acid metabolism. *Cell Rep* 2018;**24**:2392–403.
50. Samad A, Trognitz F, Compant S, et al. Shared and host-specific microbiome diversity and functioning of grapevine and accompanying weed plants. *Environ Microbiol* 2017;**19**:1407–24.
51. Micallef L, Rodgers P. eulerForce: force-directed layout for Euler diagrams. *J Visual Lang Comput* 2014;**25**:924–34.
52. Caldwell AB, Liu Q, Schroth GP, et al. Dedifferentiation and neuronal repression define familial Alzheimer's disease. *Sci Adv* 2020;**6**:eaba5933.
53. Tong GX, Xu W, Zhang YQ, et al. De novo assembly and characterization of the Hucho taimen transcriptome. *Ecol Evol* 2018;**8**:1271–85.
54. Lex A, Gehlenborg N, Strobel H, et al. UpSet: visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 2014;**20**:1983–92.
55. Wang M, Zhao Y, Efficient Test ZB. Visualization of multi-set intersections. *Sci Rep* 2015;**5**:16923.
56. Alsallakh B, Ren L. PowerSet: a comprehensive visualization of set intersections. *IEEE Trans Vis Comput Graph* 2017;**23**:361–70.
57. Alsallakh B, Aigner W, Miksch S, et al. Radial sets: interactive visual analysis of large overlapping sets. *IEEE Trans Vis Comput Graph* 2013;**19**:2496–505.
58. Kim B, Lee B, Seo J. Visualizing set concordance with permutation matrices and fan diagrams. *Interact Comput* 2007;**19**:630–43.
59. Rodgers P, Stapleton G, Chapman P. Visualizing sets with linear diagrams. *ACM Trans Comput-Hum Interact* 2015;**22**:1–39.
60. Kalucka J, de Rooij L, Goveia J, et al. Single-cell transcriptome atlas of murine endothelial cells. *Cell* 2020;**180**:764–79.
61. Perez-Perri JI, Rogell B, Schwarzl T, et al. Discovery of RNA-binding proteins and characterization of their dynamic responses by enhanced RNA interactome capture. *Nat Commun* 2018;**9**:1–13.
62. Ellrott K, Bailey MH, Saksena G, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst* 2018;**6**:271–81.

63. Sun M, Mi P, North C, et al. BiSet: semantic edge bundling with Biclusters for Sensemaking. *IEEE Trans Vis Comput Graph* 2016;**22**:310–9.
64. Dinkla K, van Kreveld MJ, Speckmann B, et al. Kelp diagrams: point set membership visualization. *Comput Graphics Forum* 2012;**31**:875–84.
65. Meulemans W, Riche NH, Speckmann B, et al. KelpFusion: a hybrid set visualization technique. *IEEE Trans Vis Comput Graph* 2013;**19**:1846–58.
66. Sun L, Dong S, Ge Y, et al. DiVenn: an interactive and integrated web-based visualization tool for comparing gene lists. *Front Genet* 2019;**10**:421.
67. Hur B, Kang D, Lee S, et al. Venn-diaNet: venn diagram based network propagation analysis framework for comparing multiple biological experiments. *BMC Bioinf* 2019;**20**:667.
68. Nagpal S, Kuntaland BK, Mande SS. NetSets.js: a JavaScript framework for compositional assessment and comparison of biological networks through Venn integrated network diagrams. *Bioinformatics* 2020;btaa723.
69. Xu L, Dong Z, Fang L, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 2019;**47**:W52–8.
70. Pirooznia M, Nagarajan V, Deng Y. GeneVenn - a web application for comparing gene lists using Venn diagrams. *Bioinformatics* 2007;**1**:420–2.
71. 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.
72. Kestler HA, Muller A, Kraus JM, et al. VennMaster: area-proportional Euler diagrams for functional GO analysis of microarrays. *BMC Bioinf* 2008;**9**:67.
73. Zahavi T, Stelzer G, Strauss L, et al. VennBLAST-whole transcriptome comparison and visualization tool. *Genomics* 2015;**105**:131–6.
74. Mishra GP, Ghosh A, Jha A, et al. BedSect: an integrated web server application to perform intersection, visualization, and functional annotation of genomic regions from multiple datasets. *Front Genet* 2020;**11**:3.
75. Wang Y, Xu L, Gu YQ, et al. MetaCoMET: a web platform for discovery and visualization of the core microbiome. *Bioinformatics* 2016;**32**:3469–70.
76. Amand J, Fehlmann T, Backes C, et al. DynaVenn: web-based computation of the most significant overlap between ordered sets. *BMC Bioinf* 2019;**20**:743.
77. Blom J, Kreis J, Spänig S, et al. EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 2016;**44**:W22–8.
78. Chong J, Soufan O, Li C, et al. MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* 2018;**46**:W486–94.
79. Calderone A, Castagnoli L, Cesareni G. mentha: a resource for browsing integrated protein-interaction networks. *Nat Methods* 2013;**253**:690–1.
80. Altmann C, Hardt S, Fischer C, et al. Progranulin overexpression in sensory neurons attenuates neuropathic pain in mice: role of autophagy. *Neurobiol Dis* 2016;**96**:294–311.
81. Ray P, Guo Y, Chi MH, et al. Serendipita bescii promotes winter wheat growth and modulates the host root transcriptome under phosphorus and nitrogen starvation. *Environ Microbiol* 2020.
82. Pant BD, Oh S, Lee HK, et al. Antagonistic regulation by CPN60A and CLPC1 of TRXL1 that regulates MDH activity leading to plant disease resistance and thermotolerance - ScienceDirect. *Cell Rep* 2020;**33**:108512.
83. Kim W, Lee N, Hwang S, et al. Comparative genomics determines strain-dependent secondary metabolite production in *Streptomyces venezuelae* strains. *Biomolecules* 2020;**10**:864.
84. Rodrigues LS, Gazara RK, Passarelli-Araujo H, et al. First genome sequences of two multidrug-resistant *Candida haemulonii* var. *vulnera* isolates from Pediatric patients with Candidemia. *Front Microbiol* 2020;**11**:1535.
85. Ruiz N. Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in *Escherichia coli*. *Proc Natl Acad Sci U S A* 2008;**105**:15553–7.
86. Giannakis N, Sansbury BE, Patsalos A, et al. Dynamic changes to lipid mediators support transitions among macrophage subtypes during muscle regeneration. *Nat Immunol* 2019;**20**:626–36.
87. Spiller KJ, Restrepo CR, Khan T, et al. Microglia-mediated recovery from ALS-relevant motor neuron degeneration in a mouse model of TDP-43 proteinopathy. *Nat Neurosci* 2018;**21**:329–40.
88. Leidal AM, Huang HH, Marsh T, et al. The LC3-conjugation machinery specifies the loading of RNA-binding proteins into extracellular vesicles. *Nat Cell Biol* 2020;**22**:187–99.
89. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;**47**:D607–13.
90. Zeeberg BR, Feng W, Wang G, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol* 2003;**4**:R28.
91. McGinnis SD, TlJNAR M. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 2004;**32**:20–5.
92. Zhang Y, Zhang S, Han S, et al. Transcriptome profiling and in silico analysis of somatic embryos in Japanese larch (*Larix leptolepis*). *Plant Cell Rep* 2012;**31**:1637–57.
93. Supek F, Miñana B, Valcárcel J, et al. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* 2014;**156**:1324–35.
94. Northcutt AJ, Lett KM, Garcia VB, et al. Deep sequencing of transcriptomes from the nervous systems of two decapod crustaceans to characterize genes important for neural circuit function and modulation. *BMC Genomics* 2016;**17**:868.
95. Turnbaugh PJ, Ley RE, Hamady M, et al. The human microbiome project. *Nature* 2007;**449**:804–10.
96. Hamady M, Knight R. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res* 2009;**19**:1141–52.
97. Daniel MD, Clemente JC, Justin K, et al. The biological observation matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* 2012;**1**:1–6.
98. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;**75**:7537–41.
99. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.
100. Shade A, Handelsman JJEM. Beyond the Venn diagram: the hunt for a core microbiome. *Environ Microbiol* 2012;**14**:4–12.

101. Kokou F, Sasson G, Friedman J, et al. Core gut microbial communities are maintained by beneficial interactions and strain variability in fish. *Nat Microbiol* 2019;**4**:2456–65.
102. Zhang J, Liu Y, Zhang N, et al. NRT1.1B is associated with root microbiota composition and nitrogen use in field-grown rice. *Nat Biotechnol* 2019;**37**:676–84.
103. Saha S, Jeon BH, Kurade MB, et al. Microbial acclimatization to lipidic-waste facilitates the efficacy of acidogenic fermentation. *Chem Eng J* 2019;**358**:188–96.
104. Toro-Londono MA, Bedoya-Urrego K, Garcia-Montoya GM, et al. Intestinal parasitic infection alters bacterial gut microbiota in children. *PeerJ* 2019;**7**:e6200.
105. Smyth GK. Limma: linear models for microarray data. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 2005; pp: 397–420.
106. Warnes GR, Bolker B, Bonebakker L, et al. gplots: various R programming tools for plotting data. *R Package Version* 2009;**2**:1.
107. Chen C, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 2020;**13**:1194–202.
108. Bannier A, Bodin N. A new drawing for simple Venn diagrams based on algebraic construction. *J Comput Geom* 2017;**8**:153–73.
109. Radcliffe NJ. Nested Venn Diagrams. *Stochastic Solutions Com*, 2010.
110. Rodgers P, Flower J, Stapleton G. Introducing 3D Venn and Euler diagrams. In: *Proceedings of the 3rd International Workshop on Euler Diagrams*, Canterbury, UK 2012; p:92–106.