Visualizing Set Concordance with Permutation Matrix and Fan Diagram

1 Permutation Matrix

an overview of relationships among sets and elements

- each row represents a set
- each column represents an element



improved permutation matrix shows similarity information without overlaps and provides better support for showing the names of elements.

aggregates elements with the same-membership into a single column

- saving a significant amount of screen space
- clear overview in a compact form



3 power analysis results with 7643 genes

aggregation drastically reduced the number of columns from **7643** to **7**

2 Fan Diagram

a new diagram to represent set relationships without inconsistencies in Venn diagrams



- no missing valid intersections
- no invalid intersection areas
- accurate size coding of all possible areas
- support for fundamental set operations
- two or three sets only

Bohyoung Kim, Bongshin Lee, and Jinwook Seo

Motivation:

While some tools are available to help users perform set concordance analysis, there are no scalable tools to provide users with comprehensive overviews and dynamic filtering and details-on-demand, which we believe play an important role in helping users integrate various perspectives.



ConSet with 16 sets (or gene ontology terms) and 32 elements (or genes)

Solution: ConSet

a novel visualization tool to effectively support important tasks for set concordance analysis of a large number of sets and elements, featuring: improved permutation matrix visualization

- fan diagram for consistent visualization of set relationships
- support for the information visualization mantra

	_ D × _ B ×
2) [16] of [16] sets V {31} of {31} elements	Select sets to visualize Select All
MembershipDegree of Aggregation (log10)[20] extracellular[18] extracellular space[8] extracellular matrix[5] structural molecule activity[6] cell adhesion[6] regulation of cellular process[6] regulation of biological process[5] cell growth[3] cytokine activity[4] death[2] S phase of mitotic cell cycle[3] carbohydrate binding[6] protein catabolism[6] proteolysis and peptidolysis[5] peptidase activity	set name size extracellular 20 extracellular space 18 extracellular matrix 8 structural molecule activity 5 cell adhesion 6 Add Delete Clear extracellular matrix structural molecule activity cell adhesion Fairy Diagram Venn Diagram Fairy Diagram Venn Diagram Image: Colored structural molecule activity cell adhesion Image: Colored structural molecule activity cell adhesion Image: Control Image: Control
element COL4A1 COL3A1	_U {extracellular}, {extracellular space}, {extracel _U {extracellular}, {extracellular matrix}, {structura
2 7 1	
	NUM //



when users hover over an area, the information about that area is shown as a tooltip and the names of elements are shown in the elements list.

3 Filtering Sets/Elements

enable users to filter sets/elements using dynamic

		[21] of [21] sets { 163 } of { 163 } elements	
E. J. C. Gender St. C. Linger, Jack 11, 1924 (n. 1924 (n. 1944 (n.			
		 Membership Degree of Aggregation (log10) [59] cellular macromolecule metabolism [56] cellular protein metabolism [25] protein modification [25] biopolymer modification [35] biopolymer metabolism [25] transferase activity [21] adenyl nucleotide binding [21] ATP binding [30] purine nucleotide binding [18] biosynthesis [24] negative regulation of cellular physiological process [25] cell proliferation [36] regulation of metabolism [37] response to stress [25] response to other organism [25] response to biotic stimulus [39] defense response [38] immune response [22] transmembrane receptor activity 	
	(a) (10 Company (10) (10) (10) (10) (10) (10) (10) (10)] of [21] sets 3 } of { 163 } elements	
HOLE COLECT & COLECT & COLECT AND HOLE COLE AND HOLE OF MARKING HOME ON AND HOLE OF AND A COLECT AND A COLECT A	POP PLS PTMS ST OST XRCCA HUM		
	🚓 🖶 🏶 🏶 🏶 Membership Degree of Aggregatio	n (log10)	
	[59] cellular macrom	olecule metabolism	
	[57] protein metabolis [56] cellular protein n	sm netabolism	
	[40] response to bioti	c stimulus	
	[39] defense respons [38] immune respons	e e	
	[37] response to stre	SS	
	[36] regulation of metabolism [35] biopolymer metabolism		
	[30] purine nucleotide	e binding	
it sets whose cardina	lity is loss than	30 from (a)	
it sets whose cardina		50 monn (a)	
18/22 122 122	() [40]-6[24]4-		
MAN DAN D. U U (C.A) U L. U.	(2) (2) $(10$ $[10]$ $[21]$ sets (2) (24) of $\{163\}$ elements		
SETRERA HUNHUMHUMA HUND HUMMANNA	And How		
	Membership		
	Degree of Aggregation (log10)		
	[30] purine nucleotide binding		
	[35] biopolymer metabolism		
	[36] regulation of metabolism		
	[38] immune response		
	[39] defense response		
	[40] response to biotic stimulus		
	[56] cellular protein metabolism		
	[57] protein metabolism		
	Taal centriar macromolecule metabolism		

(c) filtering out elements that belong to less than 5 sets from (b)

applies the Rank-by-Feature framework to diagrams of

shows the top 10 diagrams ranked by some criteria • ranking criteria: "intersection size" / "overlap metric" Fan diagram or Venn diagram

<u> </u>		element	belongs to
		TNC_D	{extracellular}, {extracellular space}, {extracell
		COL4A1_U	{extracellular}, {extracellular space}, {extracell
₩ <u>_</u>] [TNFRSF11B_D	{extracellular}, {extracellular space}, {extracell
		NTN4_U	{extracellular}, {extracellular space}, {extracell
[6 elements]		LUM_D	{extracellular}, {extracellular space}, {extracell
<u> </u>		TIMP1_D	{extracellular}, {extracellular space}, {extracell
{extracellular}			
AND /outrocollular oppool			
AND	2		
{extracellular matrix}		•	Image: Second se
		<u>, </u>	

