EVpedia 2.0

User manual of

an improved web portal

for the systematic analyses of extracellular vesicles

July 2017

lab. *of* INTERCELLULAR COMMUNICATION NETWORK

Contents

Chapter 1. Introduction

Extracellular vesicles (EVs) are spherical bilayered proteolipids with an average diameter of 20-1,000 nm. EVs contain various bioactive molecules, such as proteins, genetic materials, and lipids. They are secreted into the extracellular milieu either constitutively or in a regulated manner. Their secretion is evolutionarily conserved from prokaryotes to eukaryotes. In addition, EVs have been named with various terms: (i) Eukaryote-derived EVs: exosomes, microvesicles, ectosomes, membrane particles, exosome-like vesicles, and tolerosomes; (ii) Bacteria-derived EVs: outer membrane vesicles (Gram-negative bacteria) and membrane vesicles (Gram-positive bacteria); (iii) Archaea-derived EVs: membrane vesicles.

EVs have been suggested to play important roles in pathophysiological functions. For instance, tumor cell-derived EVs play various roles in tumor progression involving immune modulation, angiogenesis, invasion, and metastasis. In addition, there are studies suggesting the importance of EVs for the diagnosis of human diseases, including cancer and cardiovascular disease. However, it has been difficult to study EVs, due to the complexity of their components. To solve this problem, many high-throughput analyses have been performed on both prokaryotic and eukaryotic EVs: (i) Proteomes: mass-spectrometry-based studies; (ii) Transcriptomes: microarray- or next-generation sequencing-based studies; (iii) Lipidomes: chromatography-based studies. Until now, these studies have resulted in identification of over 590,000 EV-associated molecules (proteins, mRNAs, miRNAs, and lipids).

However, there had been no databases which catalog EV components derived from various types of prokaryotes and eukaryotes. In addition, there had been no systematic analytical tools which could (i) compare EV datasets by ortholog identification; (ii) perform Gene Ontology enrichment analyses; and (iii) conduct network analyses. These systematic analyses on EV components could provide new insights into the pathophysiological functions of EVs as well as EV biogenesis. To meet the needs of an integrated database and systematic analytical tools for EV components, we presented EVpedia 1.0 in 2012.

After launching, we improved several aspects of EVpedia, and now we present EVpedia 2.0. The following aspects are what we improved in EVpedia 2.0:

1. Coverage expansion
2. Automatic and frequent updates
3. Personalization and user survey
4. Quantitative analysis on EV-associated RNAs.

In addition, EVpedia 2.0 also provides the lists of publications on EV studies. This free web-based database could serve as a fundamental repository to stimulate the studies on EVs.

Chapter 2. Database contents

2.1 Size and contents

	All	Eukaryotes	Prokaryotes
Publications			
Articles	12,384	11,146	1,238
Principal investigators	6,402	5,752	713
Proteomes			
Studies	332	259	73
Datasets	685	556	129
Proteins	439,606	408,459	31,147
Transcriptomes			
mRNA			
Studies	20	20	0
Datasets	34	34	0
mRNAs	92,050	92,050	0
miRNA			
Studies	33	33	0
Datasets	131	131	0
miRNAs	57,592	57,592	0
Lipidomes			
Studies	52	47	5
Datasets	94	85	9
Lipids	3,156	2,971	185
Metabolomes			
Studies	6	6	0
Datasets	13	13	0
Metabolites	466	466	0

2.2 Available analyses

	Protein	mRNA	miRNA	Lipid	Metabolite
Experiment	O	O	O	O	O
Browse	O	O	O	O	O
Analysis					
Sequence search	O	O	O	X	X
Set analysis	O	O	O	O	O
GO enrichment analysis	O	O	O	X	X
Network analysis	O	O	O	X	X
Top 100 EV markers	O	O	O	O	O
My EVpedia	O	O	O	O	O

Red: Added or improved analyses in EVpedia 2.0 compared to original one.

Chapter 3. Connecting to EVpedia

How to connect to EVpedia

- URL: http://evpedia.info/
- All of the menus except “Home” requires “sign in” to use. You could create your own account for free. In addition, you could keep yourself signed in EVpedia.
- The account information you provide will be kept confidential and used only for the development of EVpedia.

Systems requirements for best performance

- Operating system: Microsoft Windows 7
- Browser: Google Chrome
- Resolution: 1,920 x 1,080

Compatible systems

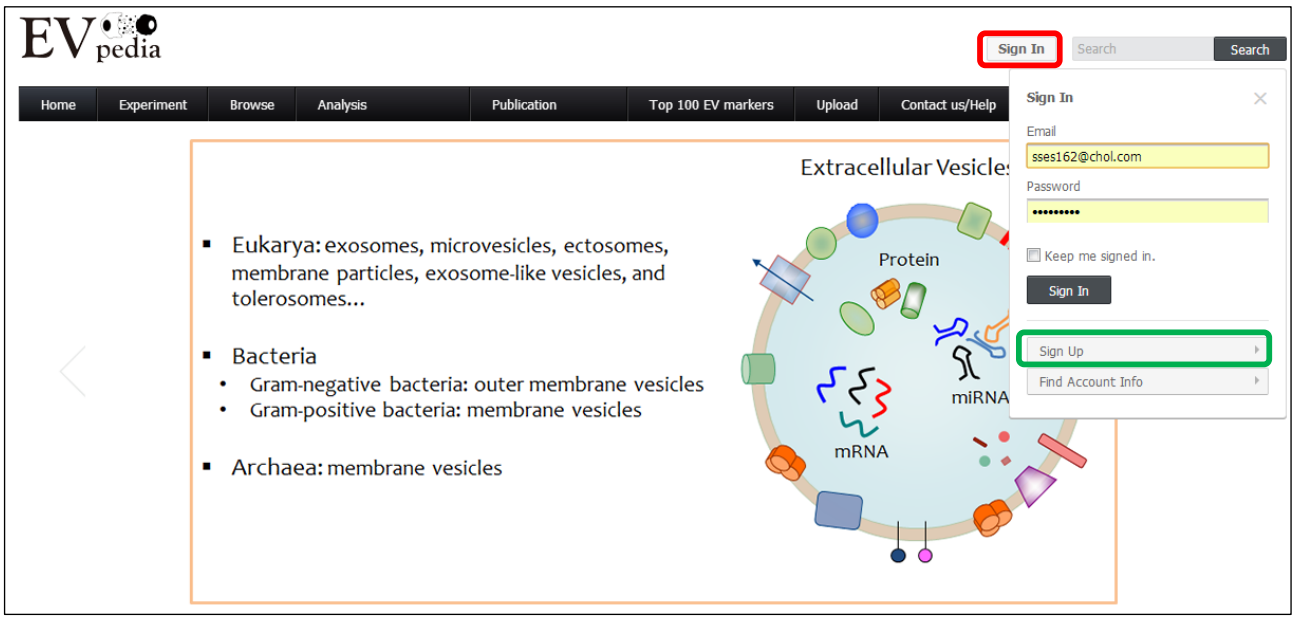
- OS: MS Windows XP/7 and Apple OS X for PC; Google Android and Apple iOS for cell phone
- Browser: Google Chrome, MS Internet Explorer (**no less than version 11**), Apple Safari, and Mozilla Firefox

Chapter 4. Exploring EVpedia

4.1 Home

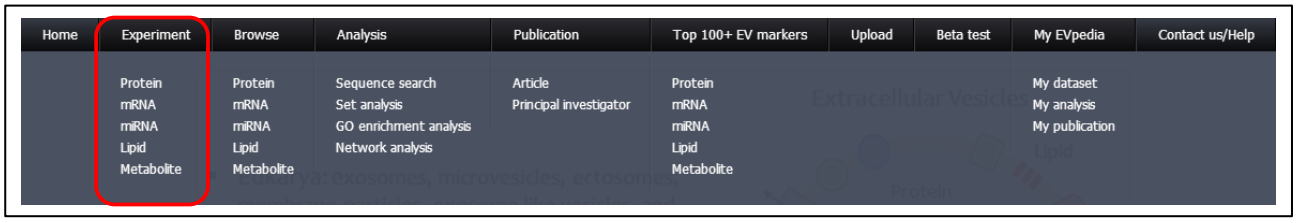
"Home" menu briefly introduces EVpedia with summary figures (slide shows), brief descriptions, references, and notices. In this menu, the user should "sign in". Note that all of the menus except "Home" requires "sign in" to use. You could create your own account for free. In addition, you could keep yourself signed in EVpedia. The account information you provide will be kept confidential and used only for development of EVpedia.

To sign in, please click the "sign in" button in the upper right corner of EVpedia (red box). If you first visit EVpedia, please click the "sign up" button to register your account (green box). Detailed license is displayed in the following page.

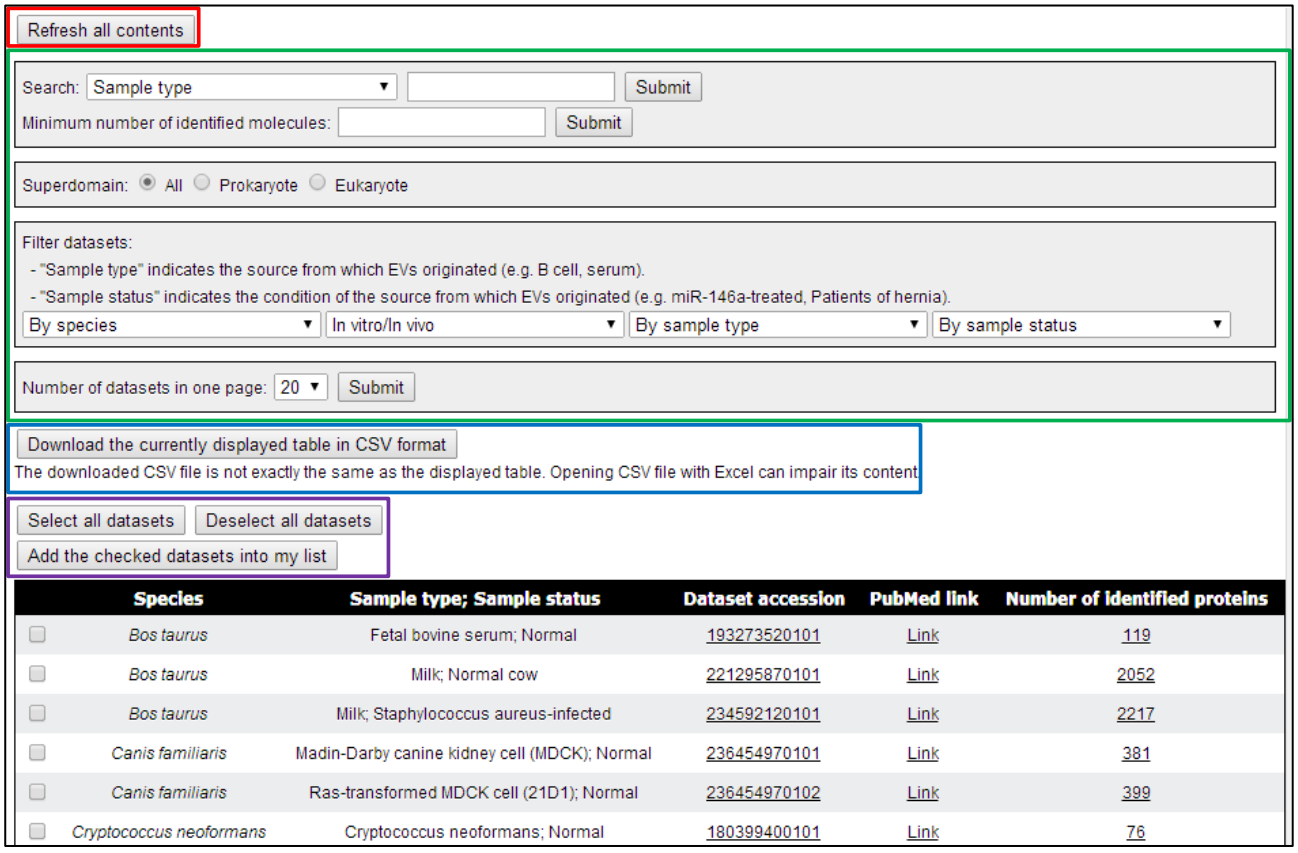


4.2 Experiment

“Experiment” menu shows the list of protein/mRNA/miRNA/lipid datasets in EVpedia and their properties (e.g. species, sample type, sample status, number of identified molecules…). In this manual, we will briefly show the functions of “Experiment” menu with protein datasets. You could use the menu with mRNA/miRNA/lipid datasets, similarly as you do with protein datasets.

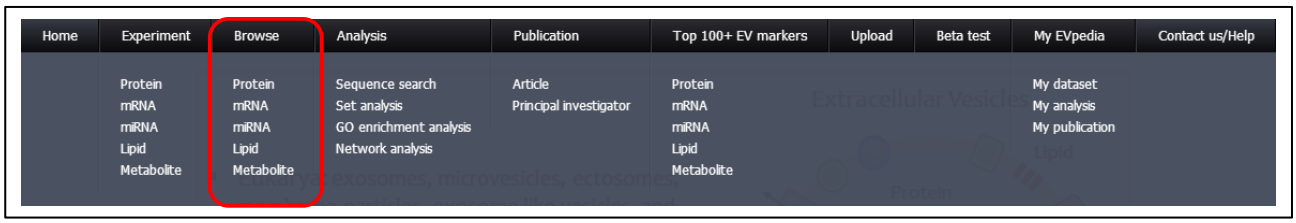


If you click the “Experiment-Protein” menu, you will see the screen captured below. First, if you click the “Refresh all contents” button (red box), all the filters and searches will be cancelled and you will see the initialized screen. Second, you could set your own filters and searches to narrow down the datasets (green box); for example, datasets acquired from *Homo sapiens*. After you set your filters and searches, the content of filters will be also changed so that you could set additional filters within the filtered results. Third, by clicking the “Download the currently displayed table in CSV format” button (blue box), you could download the current table. Although the extension of the downloaded file is “txt”, you could open it with Microsoft Excel for better views. Fourth, by clicking the “Add the checked datasets into my list” button (purple box), you could save the checked datasets in “My EVpedia – My dataset” menu and look again whenever you want. In addition, if you click the underlined items in the table, you could see the detailed information for each item; for example, if you click “Link” in the “PubMed link” column, you will be directly moved to NCBI PubMed site for the corresponding articles.

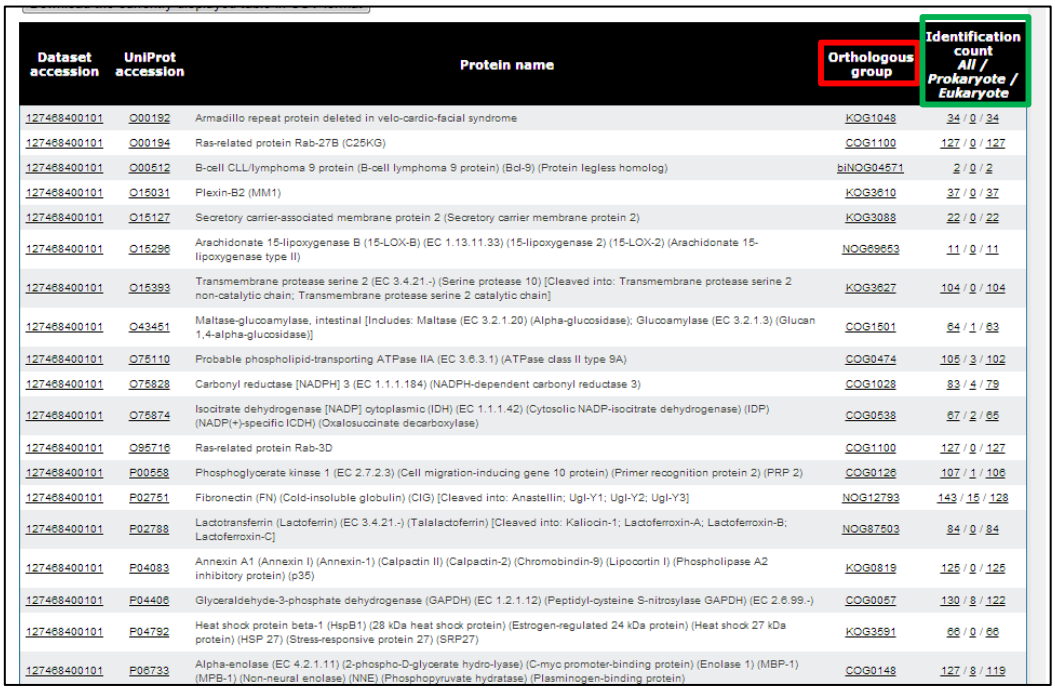


4.3 Browse

“Browse” menu provides the list of EV components identified by high-throughput analyses. Since the overall composition of screen is similar to that of “Experiment” menu, we will only discuss about the differences.



First, in the “Browse - Protein” menu, you could see the column of “orthologous group” (red box) and “identification count” (green box). An orthologous group indicates a group of proteins with similar sequences (Paralogs and orthologs; *Science*. **278**(5338):631-637, 1997.). We catalogued EV proteins as orthologous groups. In addition, we defined identification counts as the number of datasets that contains the corresponding protein or the member of the orthologous group. Therefore, if a protein has a higher identification count than others, the protein is more likely to be identified in EVs. We also applied the concepts of orthologous groups and identification count to mRNA and miRNA. However, due to the lack of a unified database, we could not catalogue lipids as orthologous groups.



Dataset accession	UniProt accession	Protein name	Orthologous group	Identification count All / Prokaryote / Eukaryote
127468400101	O00192	Armadillo repeat protein deleted in velo-cardio-facial syndrome	KOG1048	34 / 0 / 34
127468400101	O00194	Ras-related protein Rab-27B (C25KG)	COG1100	127 / 0 / 127
127468400101	O00512	B-cell CLL/lymphoma 9 protein (B-cell lymphoma 9 protein) (Bcl-9) (Protein legless homolog)	biNOG04571	2 / 0 / 2
127468400101	O15031	Plexin-B2 (MM1)	KOG3610	37 / 0 / 37
127468400101	O15127	Secretory carrier-associated membrane protein 2 (Secretory carrier membrane protein 2)	KOG3088	22 / 0 / 22
127468400101	O15296	Arachidonate 15-lipoxygenase B (15-LOX-B) (EC 1.13.11.33) (15-lipoxygenase 2) (15-LOX-2) (Arachidonate 15-lipoxygenase type II)	NOG69653	11 / 0 / 11
127468400101	O15393	Transmembrane protease serine 2 (EC 3.4.21.-) (Serine protease 10) [Cleaved into: Transmembrane protease serine 2 non-catalytic chain; Transmembrane protease serine 2 catalytic chain]	KOG3627	104 / 0 / 104
127468400101	O43451	Maltase-glucoamylase, intestinal [Includes: Maltase (EC 3.2.1.20) (Alpha-glucosidase); Glucoamylase (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)]	COG1501	64 / 1 / 63
127468400101	O75110	Probable phospholipid-transporting ATPase IIA (EC 3.6.3.1) (ATPase class II type 9A)	COG0474	105 / 3 / 102
127468400101	O75828	Carbonyl reductase [NADPH] 3 (EC 1.1.1.184) (NADPH-dependent carbonyl reductase 3)	COG1028	83 / 4 / 79
127468400101	O75874	Isocitrate dehydrogenase [NADP] cytoplasmic (IDH) (EC 1.1.1.42) (Cytosolic NADP-isocitrate dehydrogenase) (IDP) (NADP(+)-specific ICDH) (Oxalosuccinate decarboxylase)	COG0538	67 / 2 / 65
127468400101	O95716	Ras-related protein Rab-3D	COG1100	127 / 0 / 127
127468400101	P00558	Phosphoglycerate kinase 1 (EC 2.7.2.3) (Cell migration-inducing gene 10 protein) (Primer recognition protein 2) (PRP 2)	COG0126	107 / 1 / 106
127468400101	P02751	Fibronectin (FN) (Cold-insoluble globulin) (CIG) [Cleaved into: Anastellin; Ugl-Y1; Ugl-Y2; Ugl-Y3]	NOG12793	143 / 15 / 128
127468400101	P02788	Lactotransferrin (Lactoferrin) (EC 3.4.21.-) (Talalactoferrin) [Cleaved into: Kaliocin-1; Lactoferroxin-A; Lactoferroxin-B; Lactoferroxin-C]	NOG87503	84 / 0 / 84
127468400101	P04083	Annexin A1 (Annexin I) (Annexin-1) (Calpactin II) (Calpactin-2) (Chromobindin-9) (Lipocortin I) (Phospholipase A2 inhibitory protein) (p35)	KOG0819	125 / 0 / 125
127468400101	P04406	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (EC 1.2.1.12) (Peptidyl-cysteine S-nitrosylase GAPDH) (EC 2.6.99.-)	COG0057	130 / 8 / 122
127468400101	P04792	Heat shock protein beta-1 (HspB1) (28 kDa heat shock protein) (Estrogen-regulated 24 kDa protein) (Heat shock 27 kDa protein) (HSP 27) (Stress-responsive protein 27) (SRP27)	KOG3591	66 / 0 / 66
127468400101	P06733	Alpha-enolase (EC 4.2.1.11) (2-phospho-D-glycerate hydro-lyase) (C-myc promoter-binding protein) (Enolase 1) (MBP-1) (MPB-1) (Non-neural enolase) (NNE) (Phosphopyruvate hydratase) (Plasminogen-binding protein)	COG0148	127 / 8 / 119

In addition, we normalized mRNA/miRNA datasets and provide the statistical values for mRNAs/miRNAs. Using the Gaussian mixture modeling, we plotted the absent and present distributions for each datasets and calculated false positive rate (FPR) and true positive rate (TPR) for each mRNA/miRNA; FPR of mRNA ABC means the ratio of false positive mRNAs with the higher intensity than ABC, in other words, probability that an absent mRNA accidently have higher intensity than ABC; TPR of mRNA ABC indicates the ratio of true positive mRNAs with higher intensity than ABC, in other words, the percentile of mRNA ABC among the present mRNAs. For example, the FPR and TPR of Q8K194 (red box) is 0.000154 and 0.496; FPR of 0.000154 the probability that an absent mRNA accidently have higher intensity than Q9K194 is only 0.000154, which refers this mRNA is highly likely to be present in EVs; TPR of 0.496 means Q9K194 ranks at top ~50% among present mRNAs, which indicates this mRNA has intermediate intensity. Note that, as default option, we defined EV mRNA/miRNA as the ones with FPR lower than 0.05 and TPR lower than 0.5.

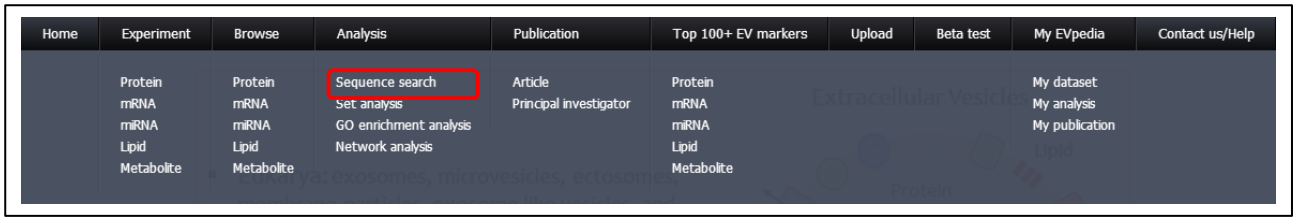
Dataset accession	UniProt accession	mRNA name	FPR	TPR	Orthologous group	Identification count *All / Prokaryote / Eukaryote* (FPR<0.05,TPR<0.5)
174861130201	P62830	60S ribosomal protein L23	0	0.094622381	COG0093	16 / 0 / 16
174861130201	Q8K194	U4/U6.U5 small nuclear ribonucleoprotein 27 kDa protein (U4/U6.U5 snRNP 27 kDa protein) (U4/U6.U5-27K) (U4/U6.U5 tri-snRNP-associated protein 3)	0.000154	0.496	KOG3263	2 / 0 / 2
174861130201	Q6PDX6	E3 ubiquitin-protein ligase Rnf220 (EC 6.3.2.-) (RING finger protein 220)	5.4e-06	0.446	ppiNOG03264	9 / 0 / 9
174861130201	Q9Z1Q2	Abhydrolase domain-containing protein 16A (EC 3.-.-.-) (HLA-B-associated transcript 5)	6.98e-06	0.45	KOG1553	6 / 0 / 6
174861130201	Q3TBX5	MCG133576, isoform CRA_a	1.37e-06	0.428	KOG2898	5 / 0 / 5
174861130201	Q6NZF1	Zinc finger CCCH domain-containing protein 11A	0	0.216133244	KOG4791	10 / 0 / 10
174861130201	Q91WG4	Elongator complex protein 2 (ELP2) (STAT3-interacting protein 1) (StIP1)	2.35e-05	0.467	COG2319	28 / 0 / 28
174861130201	P47740	Fatty aldehyde dehydrogenase (EC 1.2.1.3) (Aldehyde dehydrogenase 3) (Aldehyde dehydrogenase family 3 member A2)	0.000152	0.496	COG1012	23 / 0 / 23
174861130201	Q9R0Q4	Mortality factor 4-like protein 2 (MORF-related gene X protein) (Sid 393) (Transcription factor-like protein MRGX)	9.78e-11	0.326	KOG3001	18 / 0 / 18
174861130201	P63260	Actin, cytoplasmic 2 (Gamma-actin) [Cleaved into: Actin, cytoplasmic 2, N-terminally processed]	0	0.116546819	COG5277	25 / 0 / 25
174861130201	P61957	Small ubiquitin-related modifier 2 (SUMO-2) (SMT3 homolog 2) (Sentrin-2) (Ubiquitin-like protein SMT3A) (Smt3A)	0	0.060650916	KOG1769	17 / 0 / 17
174861130201	P42932	T-complex protein 1 subunit theta (TCP-1-theta) (CCT-theta)	8.81e-06	0.453	COG0459	23 / 0 / 23
174861130201	Q61142	Spindlin-1 (30000 Mr metaphase complex) (SSEC P)	1.81e-13	0.276	NOG40069	7 / 0 / 7
174861130201	Q61142	Spindlin-1 (30000 Mr metaphase complex) (SSEC P)	2.98e-09	0.358	NOG40069	7 / 0 / 7
		E3 ubiquitin-protein ligase UHRF1 (EC 6.3.2.-) (Nuclear protein 95) (Nuclear zinc				

4.4 Analysis

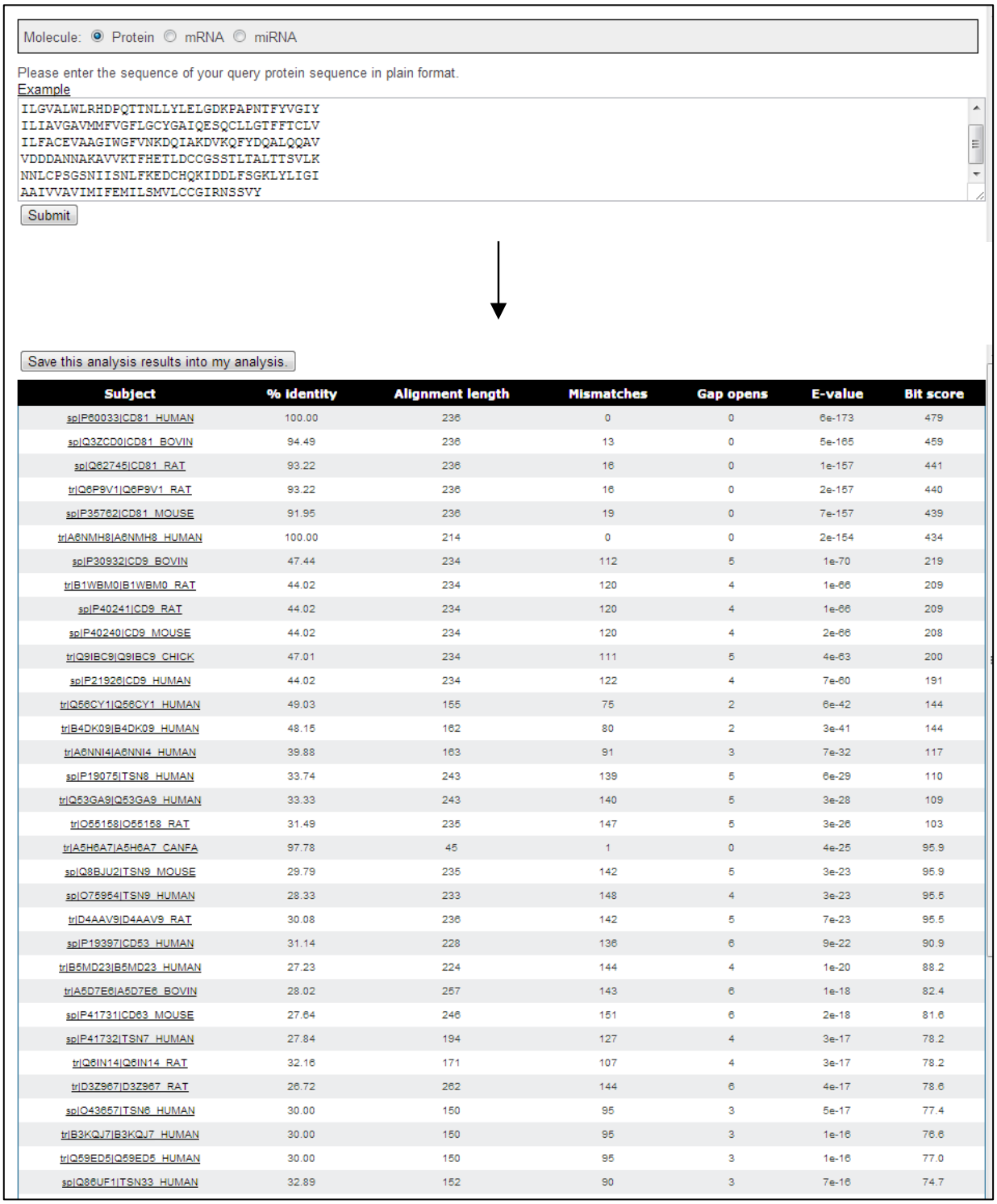
EVpedia provides variety of bioinformatic analyses to take a deeper look into high-throughput datasets; these includes “Sequence search”, “Set analysis”, “Gene Ontology enrichement analysis”, and “Network analysis”.

4.4.1 Sequence search

“Sequence search” (red box) helps to find the sequence of molecules in interest, out of the catalogued molecules in EVpedia. You could use this menu with protein, mRNA, or miRNA.



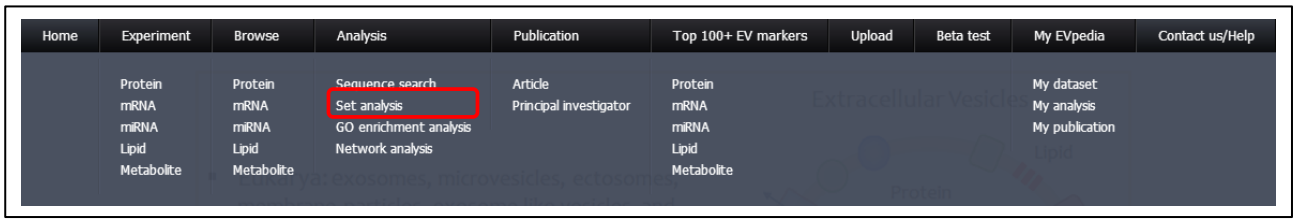
If you enter the sequence of protein, mRNA, or miRNA in the plain format, the EVpedia will search similar EV components with NCBI BLAST (*Nucleic Acids Res*. **36**(Web Server issue):W5-W9, 2008). The screenshot below shows an example of CD81 protein. Note that all the result of EVpedia analyses could be saved in My EVpedia – My analysis.



Subject	% identity	Alignment length	Mismatches	Gap opens	E-value	Bit score
sp|P60033|CD81_HUMAN	100.00	236	0	0	6e-173	479
sp|Q3ZCD0|CD81_BOVIN	94.49	236	13	0	5e-165	459
sp|Q62745|CD81_RAT	93.22	236	16	0	1e-157	441
tr|Q6P9V1|Q6P9V1_RAT	93.22	236	16	0	2e-157	440
sp|P35762|CD81_MOUSE	91.95	236	19	0	7e-157	439
tr|A6NMH8|A6NMH8_HUMAN	100.00	214	0	0	2e-154	434
sp|P30932|CD9_BOVIN	47.44	234	112	5	1e-70	219
tr|B1WBM0|B1WBM0_RAT	44.02	234	120	4	1e-66	209
sp|P40241|CD9_RAT	44.02	234	120	4	1e-66	209
sp|P40240|CD9_MOUSE	44.02	234	120	4	2e-66	208
tr|Q9IBC9|Q9IBC9_CHICK	47.01	234	111	5	4e-63	200
sp|P21926|CD9_HUMAN	44.02	234	122	4	7e-60	191
tr|Q56CY1|Q56CY1_HUMAN	49.03	155	75	2	6e-42	144
tr|B4DK09|B4DK09_HUMAN	48.15	162	80	2	3e-41	144
tr|A6NNI4|A6NNI4_HUMAN	39.88	163	91	3	7e-32	117
sp|P19075|TSN8_HUMAN	33.74	243	139	5	6e-29	110
tr|Q53GA9|Q53GA9_HUMAN	33.33	243	140	5	3e-28	109
tr|O55158|O55158_RAT	31.49	235	147	5	3e-26	103
tr|A5H6A7|A5H6A7_CANFA	97.78	45	1	0	4e-25	95.9
sp|Q8BJU2|TSN9_MOUSE	29.79	235	142	5	3e-23	95.9
sp|O75954|TSN9_HUMAN	28.33	233	148	4	3e-23	95.5
tr|D4AAV9|D4AAV9_RAT	30.08	236	142	5	7e-23	95.5
sp|P19397|CD53_HUMAN	31.14	228	136	6	9e-22	90.9
tr|B5MD23|B5MD23_HUMAN	27.23	224	144	4	1e-20	88.2
tr|A5D7E6|A5D7E6_BOVIN	28.02	257	143	6	1e-18	82.4
sp|P41731|CD63_MOUSE	27.64	246	151	6	2e-18	81.6
sp|P41732|TSN7_HUMAN	27.84	194	127	4	3e-17	78.2
tr|Q6IN14|Q6IN14_RAT	32.16	171	107	4	3e-17	78.2
tr|D3Z967|D3Z967_RAT	26.72	262	144	6	4e-17	78.6
sp|O43657|TSN6_HUMAN	30.00	150	95	3	5e-17	77.4
tr|B3KQJ7|B3KQJ7_HUMAN	30.00	150	95	3	1e-16	76.6
tr|Q59ED5|Q59ED5_HUMAN	30.00	150	95	3	1e-16	77.0
sp|Q86UF1|TSN33_HUMAN	32.89	152	90	3	7e-16	74.7

4.4.2 Set analysis

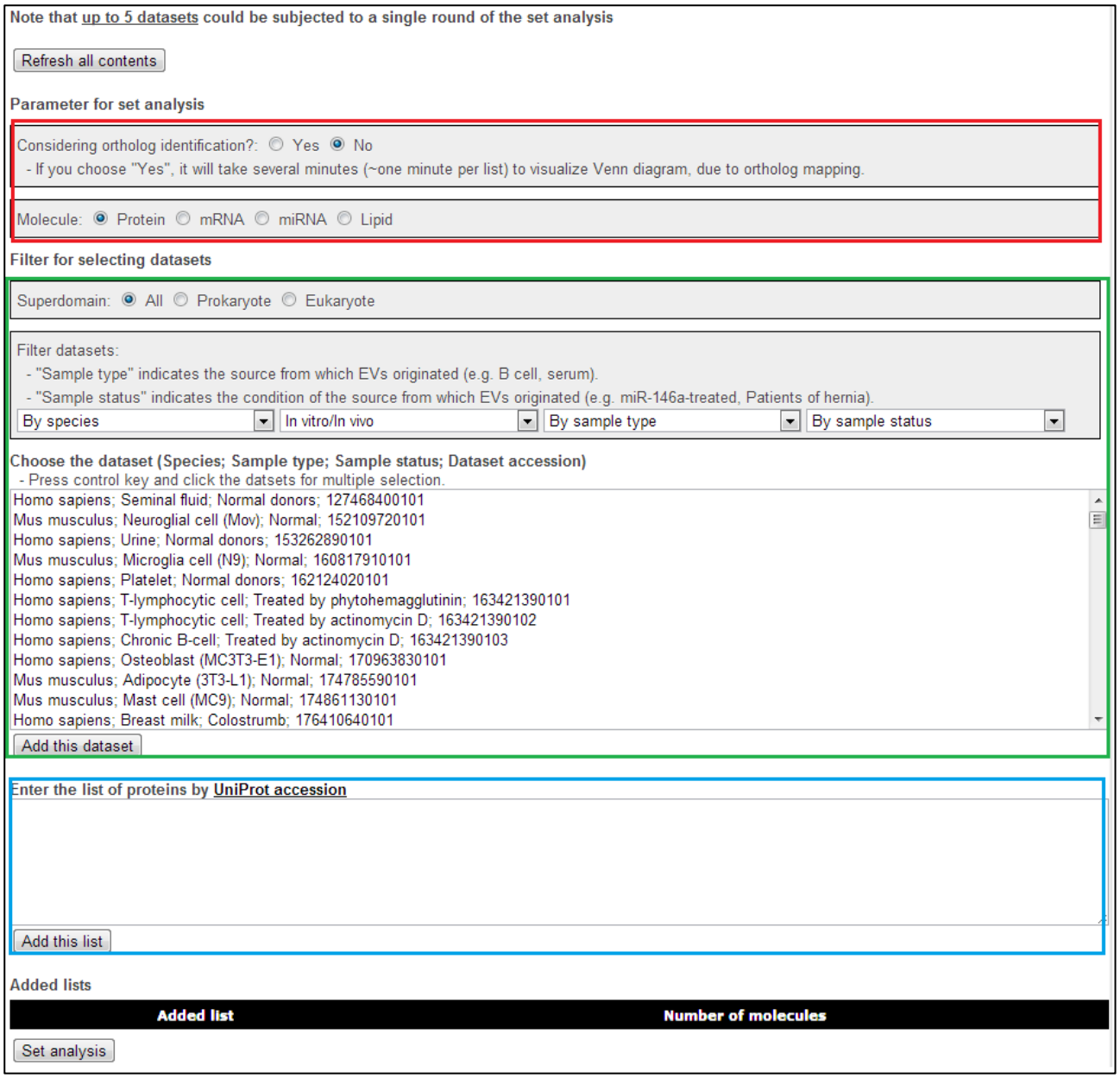
“Set analysis” (red box) helps to compare EV components acquired from different sources. You could compare EV components identified from different studies/species/statuses. Since comparing components among different sets is frequently required after performing high-throughput analyses, “Set analysis” provides a convenient solution for such use. Besides comparing EV components, you could also compare your own lists of molecules in “Set analysis”. Note that up to 5 datasets could be compared in a single round of “Set analysis”, due to the limit of depicting a Venn diagram in the two-dimensional space.



There are two parameters for “Set analysis” (red box): (i) Whether you consider ortholog identification or not; and (ii) The type of molecules. The default setting is not considering ortholog identification. However, if you want to consider ortholog identification, you could do it by simply choosing “Yes”. If you choose “Yes”, it will take several minutes (~ one minute per list) to visualize a Venn diagram, due to ortholog mapping. In addition, you could perform “Set analysis” either with proteins, mRNAs, miRNAs, or lipids.

In addition, you could select datasets (green box) by filtering datasets by four parameters (Species, *in vitro*/*in vivo*, sample type, and sample status). When you choose one or more certain parameter(s), the list of datasets is automatically changed, to show the list satisfying the chosen parameter(s). Then you could choose datasets in interest and add the dataset(s). Alternatively, you could enter the list of molecules by yourself (blue box).

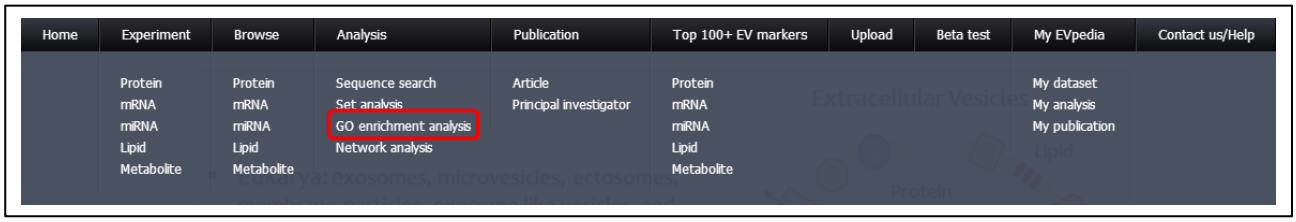
After adding the datasets or new lists, you could execute “Set analysis” by clicking the “Set analysis” button.


Note that up to 5 datasets could be subjected to a single round of the set analysis
Refresh all contents
Parameter for set analysis
Considering ortholog identification?: Yes No
- If you choose "Yes", it will take several minutes (~one minute per list) to visualize Venn diagram, due to ortholog mapping.
Molecule: Protein mRNA miRNA Lipid
Filter for selecting datasets
Superdomain: All Prokaryote Eukaryote
Filter datasets:
- "Sample type" indicates the source from which EVs originated (e.g. B cell, serum).
- "Sample status" indicates the condition of the source from which EVs originated (e.g. miR-146a-treated, Patients of hernia).
By species
In vitro/In vivo
By sample type
By sample status
Choose the dataset (Species; Sample type; Sample status; Dataset accession)
- Press control key and click the datsets for multiple selection.
Homo sapiens; Seminal fluid; Normal donors; 127468400101
Mus musculus; Neuroglial cell (Mov); Normal; 152109720101
Homo sapiens; Urine; Normal donors; 153262890101
Mus musculus; Microglia cell (N9); Normal; 160817910101
Homo sapiens; Platelet; Normal donors; 162124020101
Homo sapiens; T-lymphocytic cell; Treated by phytohemagglutinin; 163421390101
Homo sapiens; T-lymphocytic cell; Treated by actinomycin D; 163421390102
Homo sapiens; Chronic B-cell; Treated by actinomycin D; 163421390103
Homo sapiens; Osteoblast (MC3T3-E1); Normal; 170963830101
Mus musculus; Adipocyte (3T3-L1); Normal; 174785590101
Mus musculus; Mast cell (MC9); Normal; 174861130101
Homo sapiens; Breast milk; Colostrumb; 176410640101
Add this dataset
Enter the list of proteins by UniProt accession
Add this list
Added lists
Added list
Number of molecules
Set analysis

The following screenshot shows the result of comparing EVs from the seminal fluid and urine of normal human donors. If you click the number in the Venn diagram (red box), you could see the list of corresponding elements in the lower box (green box).

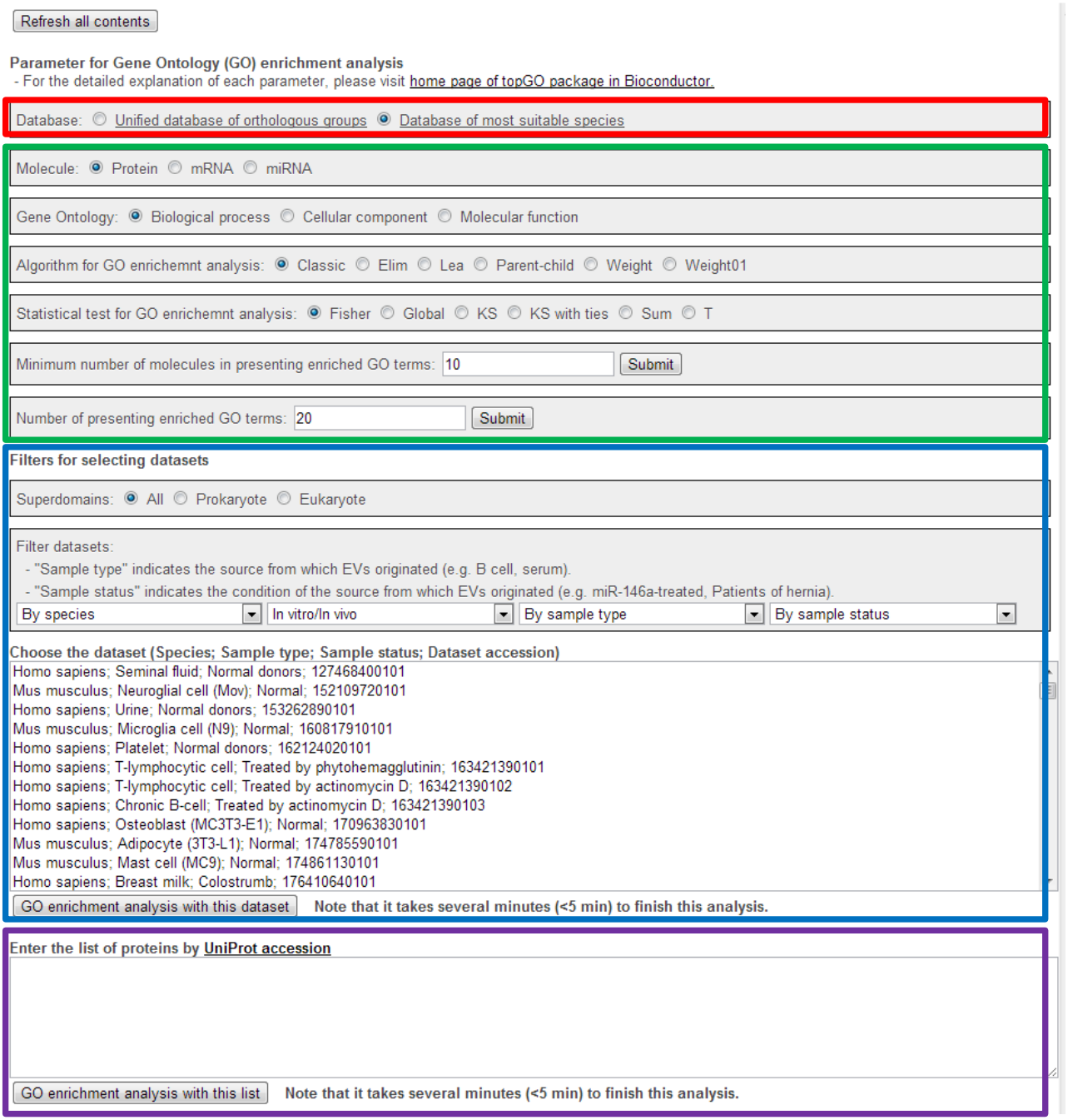
4.4.3 Gene Ontology enrichment analysis

One of the most important purposes of performing high-throughput analyses is building new hypotheses from the identified lists of molecules. However, it is quite laborious to look up the lists one by one. If we categorize the list of molecules with certain criteria, it would be even easier to deduce new hypotheses. Gene Ontology (GO) could provide the criteria, namely GO terms which represents the properties of gene products. There are three types of GO terms: (i) Biological process; (ii) Cellular component; and (iii) Molecular function. “GO enrichment analysis” (red box) provides a tool to find the enriched GO terms in a certain list of molecules.



In EVpedia, “GO enrichment analysis” could be performed using two types of databases (red box): (i) Database of most suitable species (default): a species-specific database; and (ii) Unified database of orthologous group: a newly annotated database of the functions of orthologous groups. In addition, EVpedia offers detailed options for Gene Ontology enrichment analysis (green box; please see the homepage of topGO package in Bioconductor for more information). Note that proteins, mRNAs, and miRNAs could be subject to the analysis, and all the three types of GO terms (Biological process, cellular component, and molecular function) could be displayed.

As in “Set analysis”, you could select datasets (blue box) by filtering datasets by four parameters (Species, *in vitro*/*in vivo*, sample type, and sample status). When you choose one or more certain parameter(s), the list of datasets is automatically changed, to show the list satisfying the chosen parameter(s). Then you could choose datasets in interest and add the dataset(s). Alternatively, you could enter the list of molecules by yourself (purple box). Note that a single round of the analysis could be performed with one dataset, and it takes several minutes (< 5 min) to complete the analysis.


Refresh all contents
Parameter for Gene Ontology (GO) enrichment analysis
- For the detailed explanation of each parameter, please visit home page of topGO package in Bioconductor.
Database: Unified database of orthologous groups Database of most suitable species
Molecule: Protein mRNA miRNA
Gene Ontology: Biological process Cellular component Molecular function
Algorithm for GO enrichemnt analysis: Classic Elim Lea Parent-child Weight Weight01
Statistical test for GO enrichemnt analysis: Fisher Global KS KS with ties Sum T
Minimum number of molecules in presenting enriched GO terms: 10 Submit
Number of presenting enriched GO terms: 20 Submit
Filters for selecting datasets
Superdomains: All Prokaryote Eukaryote
Filter datasets:
- "Sample type" indicates the source from which EVs originated (e.g. B cell, serum).
- "Sample status" indicates the condition of the source from which EVs originated (e.g. miR-146a-treated, Patients of hernia).
By species
In vitro/In vivo
By sample type
By sample status
Choose the dataset (Species; Sample type; Sample status; Dataset accession)
Homo sapiens; Seminal fluid; Normal donors; 127468400101
Mus musculus; Neuroglial cell (Mov); Normal; 152109720101
Homo sapiens; Urine; Normal donors; 153262890101
Mus musculus; Microglia cell (N9); Normal; 160817910101
Homo sapiens; Platelet; Normal donors; 162124020101
Homo sapiens; T-lymphocytic cell; Treated by phytohemagglutinin; 163421390101
Homo sapiens; T-lymphocytic cell; Treated by actinomycin D; 163421390102
Homo sapiens; Chronic B-cell; Treated by actinomycin D; 163421390103
Homo sapiens; Osteoblast (MC3T3-E1); Normal; 170963830101
Mus musculus; Adipocyte (3T3-L1); Normal; 174785590101
Mus musculus; Mast cell (MC9); Normal; 174861130101
Homo sapiens; Breast milk; Colostrumb; 176410640101
GO enrichment analysis with this dataset
Note that it takes several minutes (<5 min) to finish this analysis.
Enter the list of proteins by UniProt accession
GO enrichment analysis with this list
Note that it takes several minutes (<5 min) to finish this analysis.

The following is the result of “Gene Ontology enrichment analysis” using the proteome of mouse mast cell (MC9)-derived EVs. You could save the analysis results in “My EVpedia – My analysis” by clicking “Save this analysis results into my analysis” after performing a single round of the analysis.

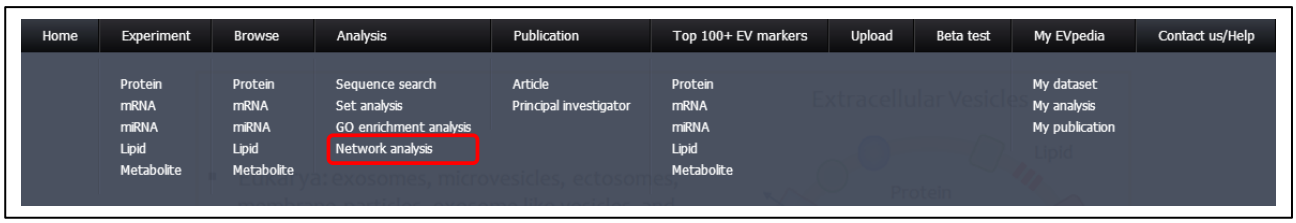
Save this analysis results into my analysis.

This table was sorted in ascending order of *P*-value.

GO ID	GO term	Number of annotated molecules	Number of annotated moleculess in this list	Expected number of annotated molecules in random list	*P*-value
GO:0034622	cellular macromolecular complex assembly	1157	48	4.81	< 1e-30
GO:0044085	cellular component biogenesis	3984	78	16.57	< 1e-30
GO:0006412	translation	1871	56	7.78	< 1e-30
GO:0034621	cellular macromolecular complex subunit ...	1367	49	5.68	< 1e-30
GO:0065003	macromolecular complex assembly	2093	57	8.7	7.1e-30
GO:0043933	macromolecular complex subunit organizat...	2326	58	9.67	1.9e-28
GO:0022607	cellular component assembly	3678	68	15.29	6.1e-26
GO:0071844	cellular component assembly at cellular ...	2832	60	11.78	9.1e-26
GO:0044267	cellular protein metabolic process	9501	101	39.51	1.2e-20
GO:0019538	protein metabolic process	11533	111	47.96	1.2e-19
GO:0071841	cellular component organization or bioge...	8710	94	36.22	2.0e-19
GO:0071840	cellular component organization or bioge...	10559	104	43.91	6.8e-19
GO:0044238	primary metabolic process	25504	176	106.05	1.0e-18
GO:0009987	cellular process	41945	232	174.42	1.6e-17
GO:0051258	protein polymerization	379	21	1.58	1.8e-17
GO:0044237	cellular metabolic process	24689	169	102.66	5.3e-17
GO:0008152	metabolic process	29250	186	121.63	2.8e-16
GO:0006334	nucleosome assembly	208	16	0.86	8.0e-16
GO:0044260	cellular macromolecule metabolic process	18630	139	77.47	1.1e-15

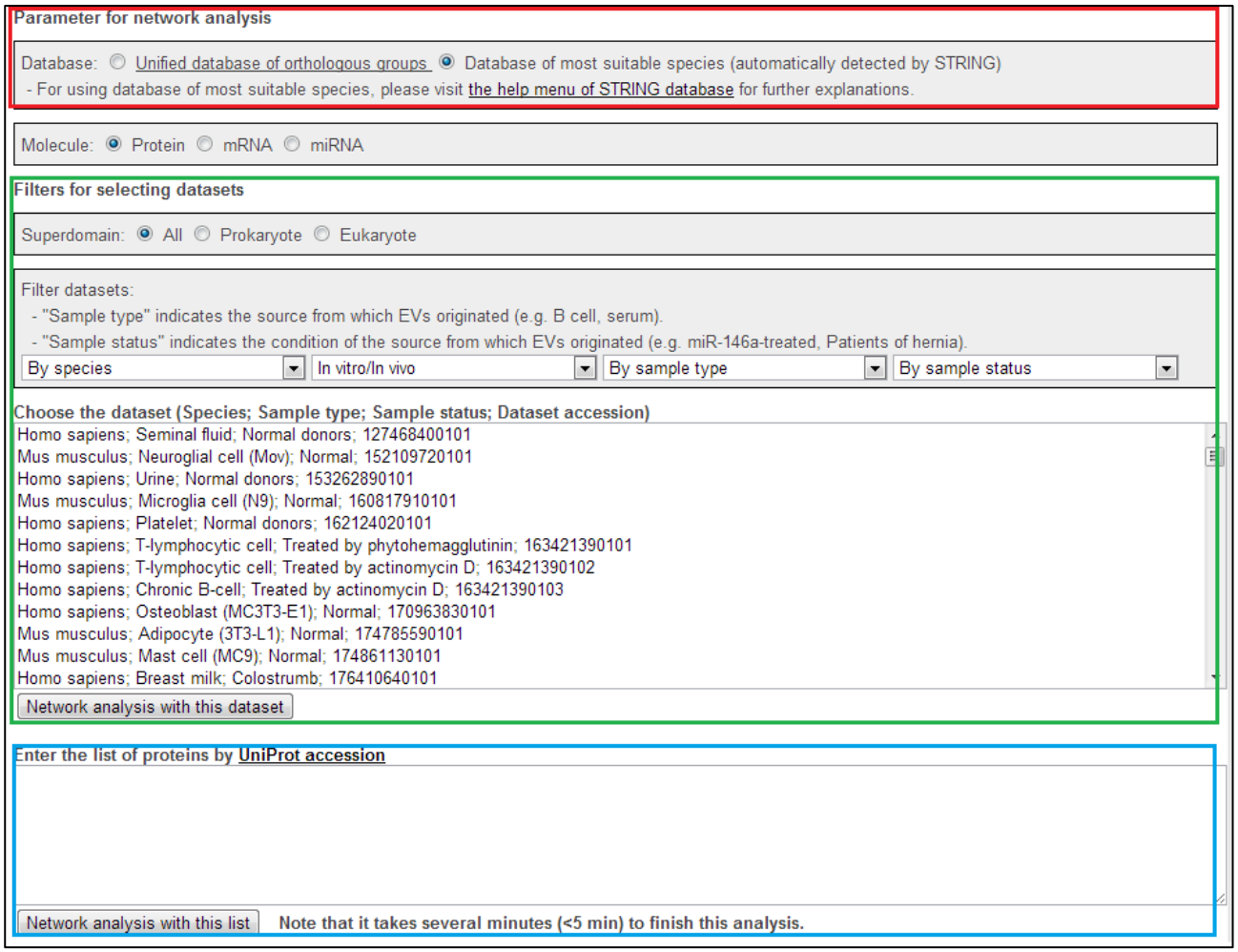
4.4.4 Network analysis

“Network analysis” (red box) helps to visualize the interactions among the list of EV-associated molecules.

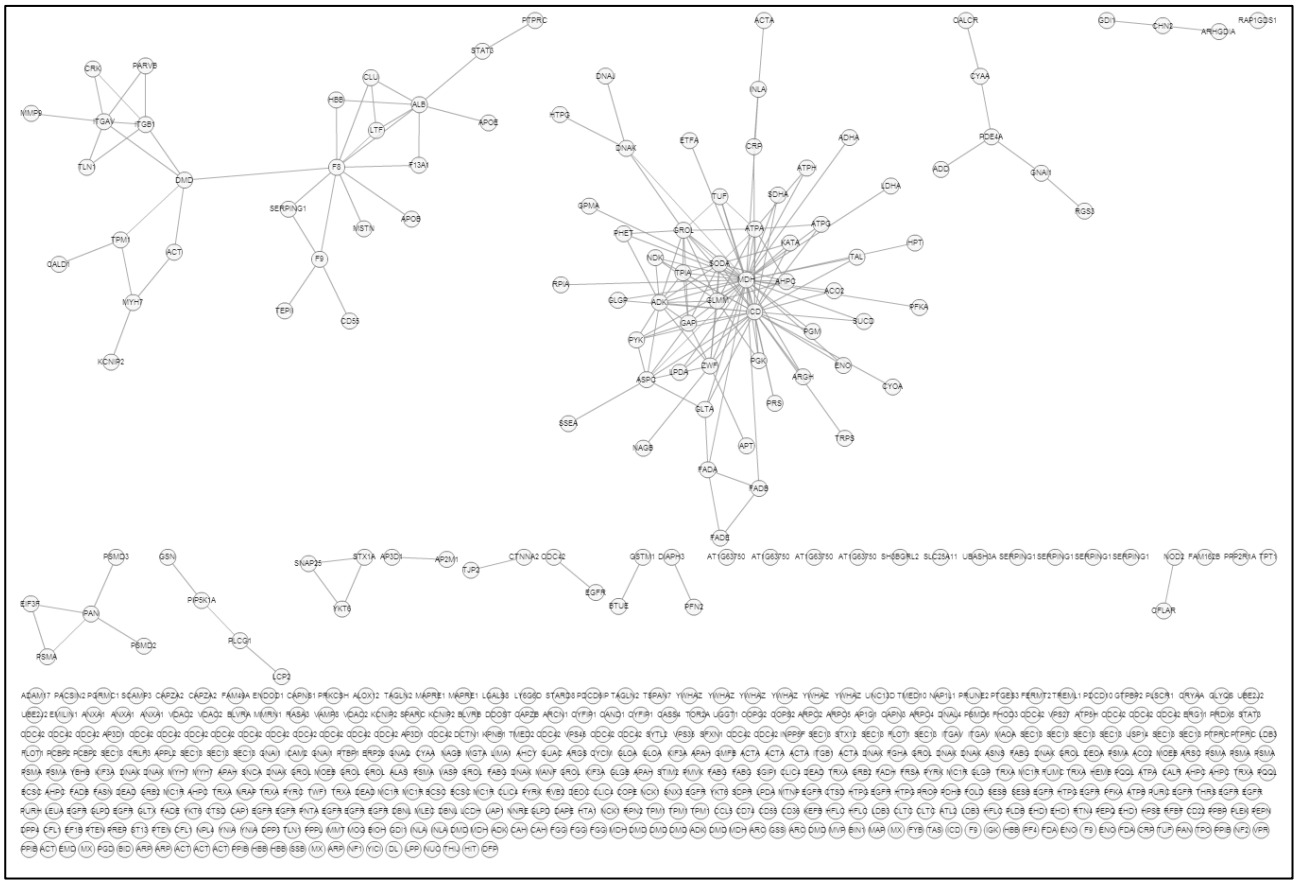


In EVpedia, “Network analysis” could be performed using two types of databases (red box): (i) Database of most suitable species (default): a species-specific database, which is automatically detected by STRING (Please visit the help menu of STRING database for further explanations); and (ii) Unified database of orthologous group: a newly annotated database of the interactions of orthologous groups.

As in “Set analysis” and “GO enrichment analysis”, you could select datasets (green box) by filtering datasets by four parameters (Species, *in vitro*/*in vivo*, sample type, and sample status). When you choose one or more certain parameter(s), the list of datasets is automatically changed, to show the list satisfying the chosen parameter(s). Then you could choose datasets in interest and add the dataset(s). Alternatively, you could enter the list of molecules by yourself (blue box). Note that a single round of the analysis could be performed with one dataset, and it takes several minutes (< 5 min) to complete the analysis.



The following is the result of “Network analysis” of the proteome of human platelet-derived EVs.



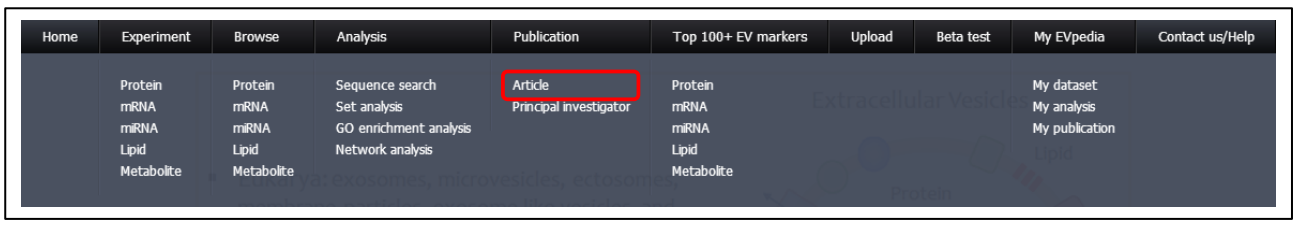
4.5 Publication

It is important for start-up researchers to review the EV-related articles and researchers before conducting their own research. To collect the possible candidate papers related to EVs, we used NCBI PubMed (http://www.ncbi.nlm.nih.gov/pubmed) for text-mining solution. For publications on prokaryote or eukaryote-derived EVs, we employed argosome*, "blebbing vesicle", "blebbing vesicles", "budding vesicle", "budding vesicles", dexosome*, ectosome*, "extracellular vesicle", "extracellular vesicles", exosome*, exovesicle*, "matrix vesicle", "matrix vesicles", microparticle*, microvesicle*, "membrane particle", "membrane particles", "membrane vesicle", "membrane vesicles", nanovesicle*, oncosome*, "outer membrane bleb", "outer membrane blebs", prostasome*, "shedding vesicle", "shedding vesicles", tolerosome*' as the search parameters. All the search outputs were manually reviewed to verify whether they are related to EVs; for example, studies about exosomes of RNA degradation activity were ruled out. More detailed information such as bibliographies, authors, and abstracts was excerpted from NCBI PubMed by means of the automatized Python code (Python version 2.7.3).

EVpedia also provide the list of major principal investigators related to EV studies. We investigated all the downloadable EV papers (approximately 85%), and matched the corresponding authors for each paper.

4.5.1 Article

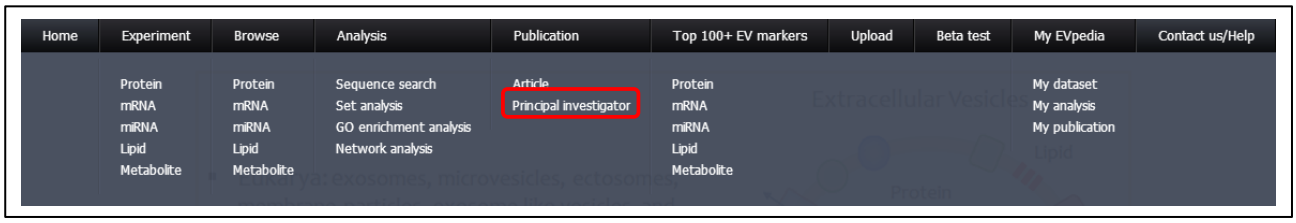
The list of articles and the related analysis results are deposited in "Publication – Article" menu (red box).



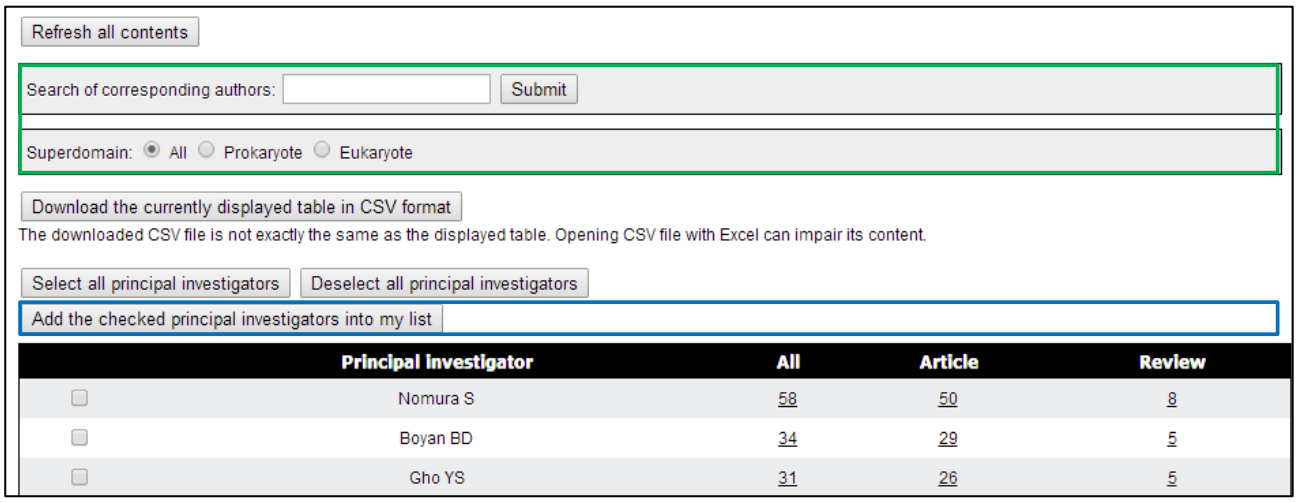
The publication trend and the word cloud (red box) could be shown by clicking the links in the upper panel. In addition, you could select articles (green box) by filtering articles by (i) Article type: article/review, article, review; (ii) EV type; and (iii) searching terms. When you choose one or more certain parameter(s), the list of articles is automatically changed, to show the list satisfying the chosen parameter(s). Note that you could save articles in interest in “My EVpedia-My publication” by checking the articles and clicking “Add the checked articles into my list” (blue box).

4.5.2 Principal investigator

The list of principal investigators studying on EVs are arranged in “Publication – Principal investigator” menu (red box).

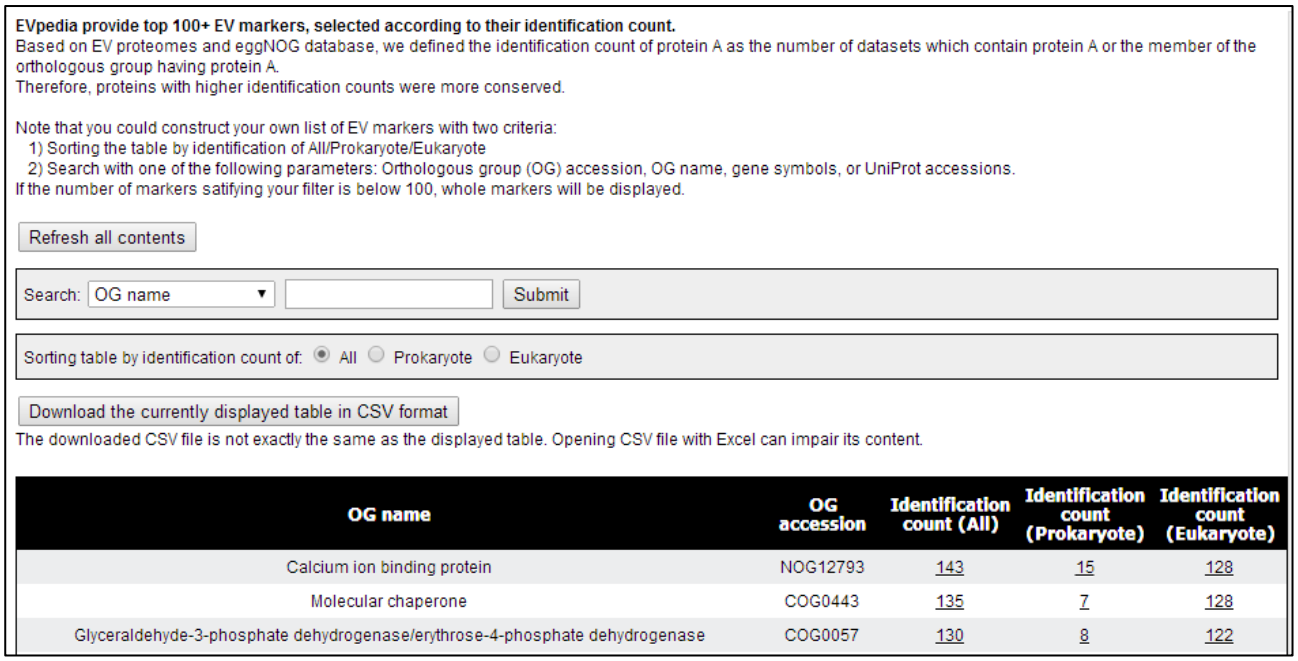


You could select principal investigators (green box) by filtering datasets by (i) Superdomain: All, Prokaryote, Eukaryote; and (ii) searching corresponding authors. When you choose one or more certain parameter(s), the list of principal investigators is automatically changed, to show the list satisfying the chosen parameter(s). Note that you could save the list of principal investigators in interest in “My EVpedia-My publication” by checking the articles and clicking “Add the checked principal investigators into my list” (blue box).



4.6 Top 100+ EV markers

EV markers could be defined as molecules identified in large number of datasets (large identification counts). EVpedia provide top 100+ EV markers for proteins, mRNAs, miRNAs, and lipids, selected according to their identification counts. Note that you could construct your own list of EV markers with two criteria: (i) Sorting the table by identification counts of All/Prokaryote/Eukaryote; and (ii) search with one of the following parameters: Orthologous group (OG) accession, OG name, gene symbols, or UniProt accessions. If the number of markers satisfying your filter is below 100, whole markers will be displayed.



EVpedia provide top 100+ EV markers, selected according to their identification count.
Based on EV proteomes and eggNOG database, we defined the identification count of protein A as the number of datasets which contain protein A or the member of the orthologous group having protein A.
Therefore, proteins with higher identification counts were more conserved.

Note that you could construct your own list of EV markers with two criteria:
1) Sorting the table by identification of All/Prokaryote/Eukaryote
2) Search with one of the following parameters: Orthologous group (OG) accession, OG name, gene symbols, or UniProt accessions.
If the number of markers satifying your filter is below 100, whole markers will be displayed.

Refresh all contents

Search: OG name Submit

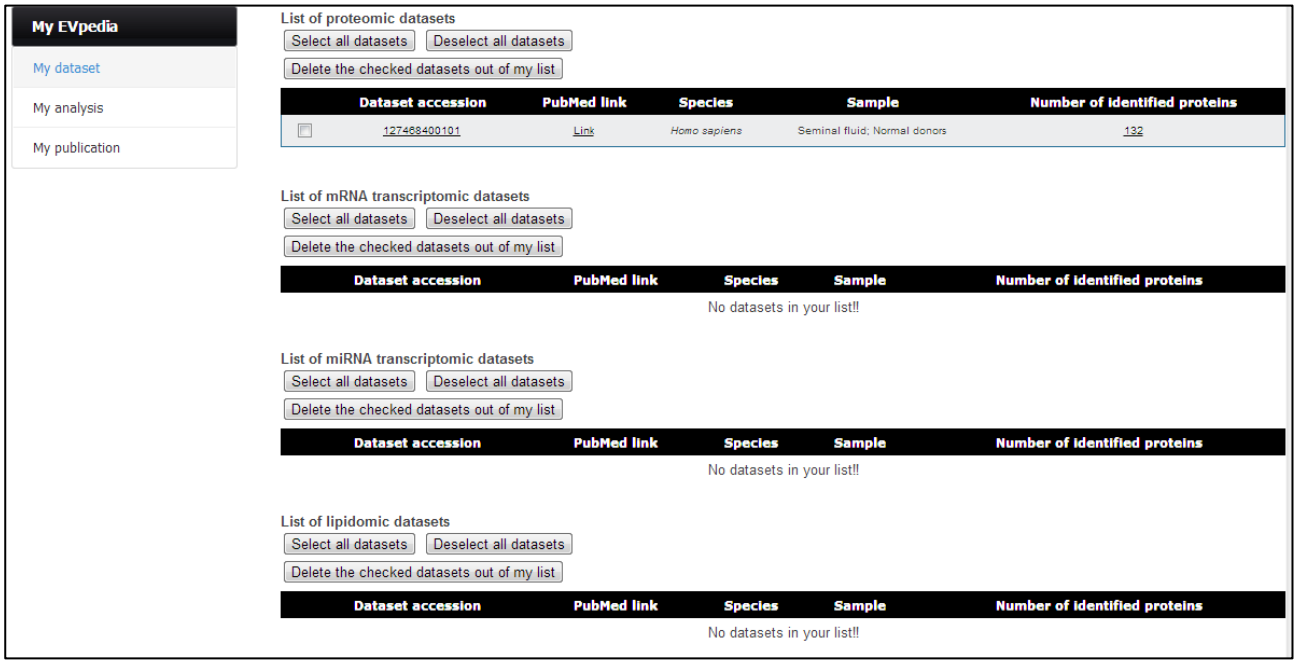
Sorting table by identification count of: All Prokaryote Eukaryote

Download the currently displayed table in CSV format
The downloaded CSV file is not exactly the same as the displayed table. Opening CSV file with Excel can impair its content.

OG name	OG accession	Identification count (All)	Identification count (Prokaryote)	Identification count (Eukaryote)
Calcium ion binding protein	NOG12793	143	15	128
Molecular chaperone	COG0443	135	7	128
Glyceraldehyde-3-phosphate dehydrogenase/erythrose-4-phosphate dehydrogenase	COG0057	130	8	122

4.7 My EVpedia

As mentioned previously, you could save the datasets, analysis results, and publication information (articles/principal investigators) into “My EVpedia”. The saved datasets, analysis results, and publication information could be shown in “My dataset”, “My analysis”, and “My publication”, respectively. Note that the contents of “My EVpedia” could be initialized after the notice via e-mail, for maintenance.



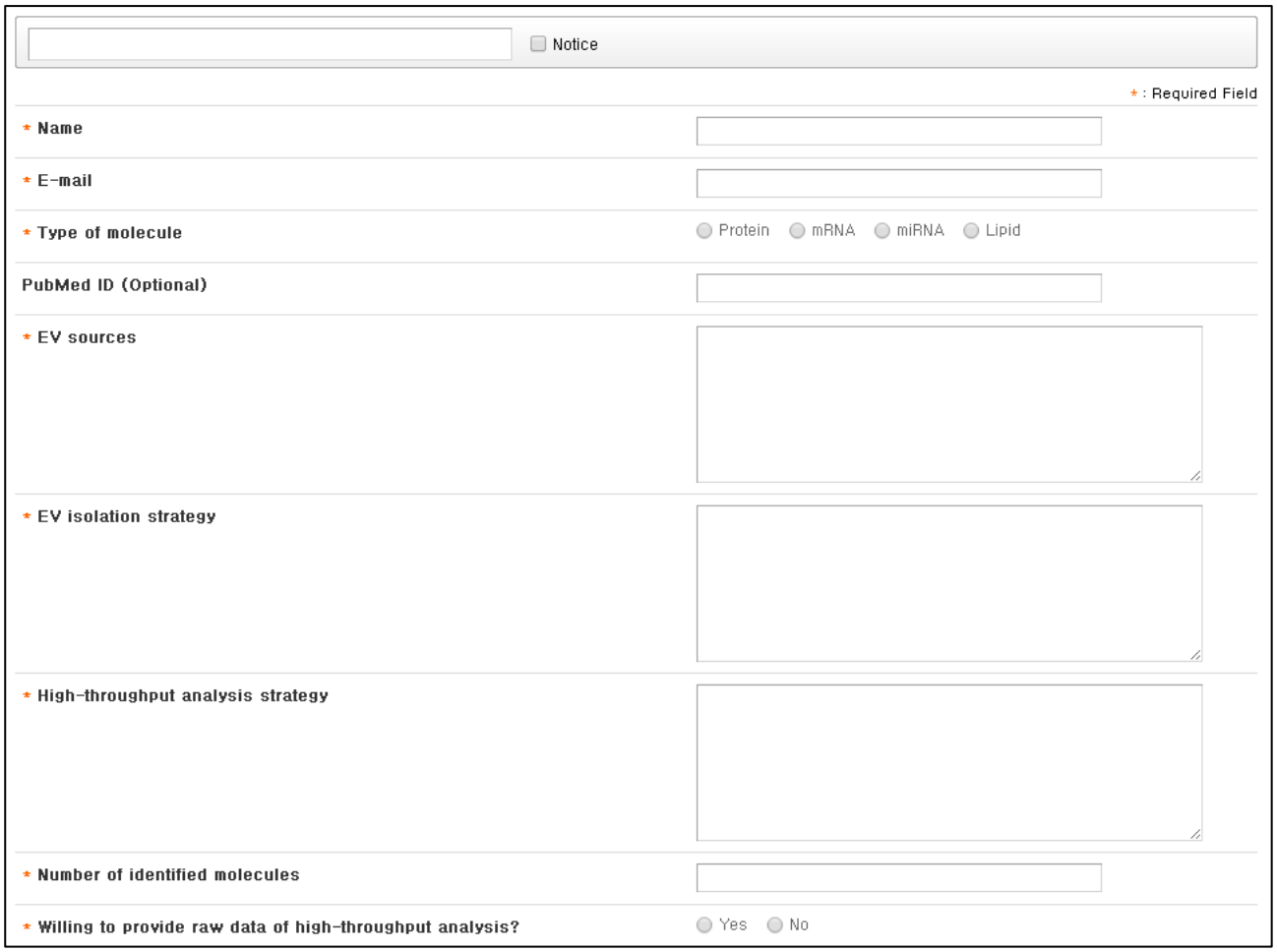
Chapter 5. Participating in EVpedia

You could help us improve EVpedia by:

- Uploading the published or unpublished protein/mRNA/miRNA/lipid dataset(s) of EVs (It could be made private until the final publication.)
- Informing us with any EV-related publication(s) which we have not addressed in EVpedia

5.1 Upload

You could upload your own datasets to EVpedia via “Upload” menu. Click “Write” button in “Upload” menu. Then you could write articles regarding your own datasets.



Required fields:

1. Name
2. E-mail
3. Type of molecules
4. EV isolation strategy
5. High-throughput analysis strategy
6. Whether you are willing to provide raw file for further analysis
7. List of molecules you identified (by attaching files in tsv or Excel format)

You could write any comments and upload your raw files and/or list of molecules you identified after you fill the required fields.

5.2 User forum

Please feel free to write your questions and/or opinions on EVpedia. You could write your questions/opinions in “Beta test” menu. All your comments could not be read except by the administrator of EVpedia. After the end of the beta test, this menu will be changed to “User forum” menu.

Chapter 6. Contact information and references

Contact information

- Jaewook Lee (Ph.D.)
  Post-doctoral researcher
  Lab. of Intercellular Communication Network
  Department of Life Sciences, Pohang University of Science and Technology (POSTECH), Pohang, Gyeongbuk 37673, Republic of Korea
  E-mail: jaewook8@postech.ac.kr
  Tel: 82-54-279-8611
  Fax: 82-54-279-8609
- Dae-Kyum Kim (Ph.D.)
  Alumni of Lab. of Intercellular Communication Network
  E-mail: sses162@postech.ac.kr
- Professor Yong Song Gho
  Laboratory head
  Lab. of Intercellular Communication Network
  Department of Life Sciences, POSTECH, Pohang, Gyeongbuk 37673, Republic of Korea
  E-mail: ysgho@postech.ac.kr
  Tel: 82-54-279-8611
  Fax: 82-54-279-8609

References

- Kim DK, Lee J, Kim SR, Choi DS, Yoon YJ, Kim JH, Go G, Nhung D, Hong K, Jang SC, Kim SH, Park KS, Kim OY, Park HT, Seo JH, Aikawa E, Baj-Krzyworzeka M, Van Balkom BW, Belting M, Blanc L, Bond V, Bongiovanni A, Borràs FE, Buée L, Buzás EI, Cheng L, Clayton A, Cocucci E, Dela Cruz CS, Desiderio DM, Di Vizio D, Ekström K, Falcon-Perez JM, Gardiner C, Giebel B, Greening DW, Gross JC, Gupta D, Hendrix A, Hill AF, Hill MM, Hoen EN, Hwang DW, Inal J, Jagannadham MV, Jayachandran M, Jee YK, Jørgensen M, Kim KP, Kim YK, Kislinger T, Lässer C, Lee DS, Lee H, Van Leeuwen J, Lener T, Liu ML, Lötvall J, Marcilla A, Mathivanan S, Möller A, Morhayim J, Mullier F, Nazarenko I, Nieuwland R, Nunes DN, Pang K, Park J, Patel T, Pocsfalvi G, Del Portillo H, Putz U, Ramirez MI, Rodrigues ML, Roh TY, Royo F, Sahoo S, Schiffelers R, Sharma S, Siljander P, Simpson RJ, Soekmadji C, Stahl P, Stensballe A, Stępień E, Tahara H, Trummer A, Valadi H, Vella LJ, Wai SN, Witwer K, Yáñez-Mó M, Youn H, Zeidler R, Gho YS. "EVpedia: A community web portal for extracellular vesicles research." Bioinformatics. 2015 Mar 15;31(6):933-9. [PubMed]
- Kim DK, Kang B, Kim OY, Choi DS, Lee J, Kim SR, Go G, Yoon YJ, Kim JH, Jang SC, Park KS, Choi EJ, Kim KP, Desiderio DM, Kim YK, Lotvall J, Hwang D, and Gho YS. "EVpedia: an integrated database of high-throughput data for systemic analyses of extracellular vesicles." *J Extracell Vesicles*. **2**: 20384, 2013. [PubMed]
- Choi DS, Kim DK, Kim YK, and Gho YS. "Proteomics, transcriptomics, and lipidomics of exosomes and ectosomes." *Proteomics*. **13**(10-11):1554-1571, 2013. [PubMed]
- Choi DS, Kim DK, Kim YK, and Gho YS. "Proteomics of extracellular vesicles: Exosomes and ectosomes." *Mass Spectrometry Reviews.* 2015 Jul-Aug;34(4):474-90 [PubMed]

Thank you for using

EVpedia