Taylor & Francis
Taylor & Francis Group

SHORT COMMUNICATION

🔓 OPEN ACCESS   Check for updates

A novel community driven software for functional enrichment analysis of extracellular vesicles data

Mohashin Pathan [a], Shivakumar Keerthikumar [a], David Chisanga [b], Riccardo Alessandro[c], Ching-Seng Ang[d], Philip Askenase[e], Arsen O. Batagov[f], Alberto Benito-Martin[g], Giovanni Camussi[h], Aled Clayton[i], Federica Collino[h,j], Dolores Di Vizio[k], Juan Manuel Falcon-Perez[l], Pedro Fonseca [m], Pamali Fonseka [a], Simona Fontana [c], Yong Song Gho[n], An Hendrix[o], Esther Nolte-'t Hoen[p], Nunzio Iraci [q,r], Kenneth Kastaniegaard [s], Thomas Kislinger[t], Joanna Kowal [u], Igor V. Kurochkin[f], Tommaso Leonardi [r,v], Yaxuan Liang[w], Alicia Llorente[x,y], Taral R. Lunavat[z], Sayantan Maji[aa], Francesca Monteleone[c], Anders Øverbye [x,y], Theocharis Panaretakis[m], Tushar Patel[aa], Héctor Peinado [g,bb], Stefano Pluchino[r], Simona Principe [t], Goran Ronquist[cc], Felix Royo[l], Susmita Sahoo[w], Cristiana Spinelli [k], Allan Stensballe[s], Clotilde Théry[u], Martijn J.C. van Herwijnen[p], Marca Wauben[p], Joanne L. Welton [dd], Kening Zhao[a] and Suresh Mathivanan[a]

aDepartment of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Australia; bDepartment of Computer Science and Information Technology, La Trobe University, Melbourne, Australia; cDepartment of Biopathology and Medical Biotechnologies, University of Palermo, Palermo, Italy; dThe Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, Australia; eSection of Allergy and Clinical Immunology, Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA; fBioinformatics Institute, A*STAR, Singapore, Singapore; gChildren's Cancer and Blood Foundation Laboratories, Departments of Pediatrics, Cell and Developmental Biology, Drukier Institute for Children's Health, Meyer Cancer Center, Weill Cornell Medical College, New York, NY, USA; hDepartment of Medical Science, University of Torino, Medical School, Torino, Italy; iDivision of Cancer and Genetics, School of Medicine, Cardiff University, Velindre Cancer Centre, Cardiff, UK; jCarlos Chagas Filho Institute of Biophysics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil; kDivision of Cancer Biology and Therapeutics, Departments of Surgery, Biomedical Sciences and Pathology and Laboratory Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; lExosomes Lab, Metabolomics Unit, CIC bioGUNE, CIBERehd. Bizkaia Technological Park, Derio 48160, Bikaia and IKERBASQUE Research Foundation, Bilbao, SPAIN; mDepartment of Oncology-Pathology, Cancer Centrum Karolinska, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden; nDepartment of Life Sciences, Pohang University of Science and Technology, Pohang, Republic of Korea; oLaboratory of Experimental Cancer Research, Department of Radiation Oncology and Experimental Cancer Research, Ghent University, Ghent, Belgium; pDepartment of Biochemistry and Cell Biology, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands; qDepartment of Biomedical and Biotechnological Sciences (BIOMETEC), University of Catania, Catania, Italy; rDepartment of Clinical Neurosciences, Wellcome Trust-Medical Research Council Stem Cell Institute and National Institute for Health Research Biomedical Research Centre, University of Cambridge, Cambridge, UK; sLaboratory for Medical Mass Spectrometry, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark; tPrincess Margaret Cancer Centre, Toronto, Canada; uInstitut Curie, PSL Research University, INSERM U932, Paris, France; vEuropean Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK; wCardiovascular Research Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA; xDepartment of Molecular Cell Biology, Institute for Cancer Research, Oslo University Hospital-The Norwegian Radium Hospital, Oslo, Norway; yCentre for Cancer Biomedicine, Faculty of Medicine, University of Oslo, Oslo, Norway; zKrefting Research Centre, Institute of Medicine at the Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; aaDepartment of Transplantation, Mayo Clinic, Jacksonville, FL, USA; bbMicroenvironment and Metastasis Group, Department of Molecular Oncology, Spanish National Cancer Research Center (CNIO), Madrid, Spain; ccDepartment of Medical Sciences, Uppsala University, Uppsala, Sweden; ddDepartment of Biomedical Sciences, Cardiff School of Health Sciences, Cardiff Metropolitan University, Cardiff, UK

ABSTRACT

Bioinformatics tools are imperative for the in depth analysis of heterogeneous high-throughput data. Most of the software tools are developed by specific laboratories or groups or companies wherein they are designed to perform the required analysis for the group. However, such software tools may fail to capture "what the community needs in a tool". Here, we describe a novel community-driven approach to build a comprehensive functional enrichment analysis tool. Using the existing FunRich tool as a template, we invited researchers to request additional features and/or changes. Remarkably, with the enthusiastic participation of the community, we were able to implement 90% of the requested features. FunRich enables plugin for extracellular vesicles wherein users can download and analyse data from Vesiclepedia database. By involving researchers early through community needs software development, we believe that comprehensive analysis tools can be developed in various scientific disciplines.

CONTACT Suresh Mathivanan ✉ S.Mathivanan@latrobe.edu.au 🖅 Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Bundoora, Victoria 3086, Australia

ⓘ The supplemental material for this article can be accessed here.

Advances in high-throughput techniques including next generation sequencing, RNA sequencing and proteomics have spurred enormous volume of data [1–3]. Currently, it is amenable to characterise the genome, transcriptome, metabolome and proteome of an organism in a robust manner. These technological developments reduced the amount of sample material, shortened the time to collect raw data and substantially decreased the associated costs [4]. Hence, large scale approaches are now accessible by many research laboratories. To harness the true potential of these heterogeneous high-throughput data, software/bioinformatics tools have become indispensable resources for the ensuing analysis [5]. To match the unprecedented growth in data generation, robust software analysis tools are constantly developed by academic and commercial entities [4]. Most of the software tools are developed by specific laboratories or groups or companies wherein they are designed to perform the required analysis for the group. However, the software tools fail to capture "what the community wants in a tool". A "community needs software" may overcome these hurdles and aid in the development of a comprehensive data analysis tool.

Here, we report a novel community needs software initiative in the context of functional or gene set enrichment analysis. To achieve this, we initially reached out to the scientific community through editorial [6], conference participations, social networking sites and communicated with researchers in the OMICS community via e-mail. FunRich, an open access standalone functional enrichment analysis tool [7], was used as a template for this purpose. By various means of communication, we invited the researchers to request for additional features/changes in FunRich software through the online forum (http://www.funrich.org/forum). We had enthusiastic participation from many researchers who requested for additional features/changes in the existing software. By September 2016, we had 54 unique requests/changes from users worldwide (Supplementary Table 1). The features were prioritised based on the number of users per request and were implemented in FunRich tool over the last 18 months. Remarkably, 90% of the requested features have been implemented in the new version of FunRich (Version 3). The updated version of FunRich is now freely available for download (http://www.funrich.org/download) both for academic and commercial users.

To gain biological insights, researchers often rely on functional enrichment analysis of large scale data from high-throughput experiments to identify overrepresented classes. Using FunRich, users can perform functional enrichment analysis with minimal or no support from computational and database experts for more than 13,320 species. The database is integrated from heterogeneous genomic and proteomic resources (>6.8 million annotations). The background database in any analysis tool is critical for the analysis and needs to be constantly updated [8]. However, the currently existing functional enrichment analysis tools do not allow the users to control the databases nor to update them in real time [8]. Using the forum, one of the request from the community pertained to "user controlled databases" and "regular update of background databases". To address this, FunRich now uniquely allows the users to update the background database for 13,320 species from UniProt, Gene Ontology and Reactome in real time (Figure 1). Additionally, the users can build custom databases with tab delimited files and perform the enrichment analysis irrespective of the organism and the type of dataset (e.g. metabolomics). Hence, these database options allows for longer sustainability of FunRich as a tool to perform functional enrichment analysis.

Other popular requests that have been implemented in FunRich include miRNA enrichment analysis (requested by most users – Supplementary Table 1), customisable heat maps, plugin to analyse extracellular vesicle datasets, comparison of oncogenes using COSMIC database and customisable colour for all the publication quality graphs. In miRNA enrichment analysis, users can submit a list of miRNA and identify biological pathways that may be perturbed. Gene set enrichment analysis is normally performed with the number of input genes/proteins and the quantitative data is often ignored. In FunRich, users can upload quantitative data and perform enrichment analysis for gene/protein expression values. For instance, total mRNA/protein abundance of genes involved in Wnt signalling pathway is compared between datasets in addition to number of genes. The quantitative data can also be utilised to generate customisable heat maps. Furthermore, users have complete control on all the graphs where the text and colour can be customised. Based on popular requests, FunRich now allows users to automatically download data from Vesiclepedia, an online compendium that hosts RNA and protein data pertaining to extracellular vesicles including exosomes [9]. The input datasets can be compared with filtered Vesiclepedia data either through enrichment analysis or through Venn diagrams. In addition users can customise the data that can be downloaded from Vesiclepedia by using filters
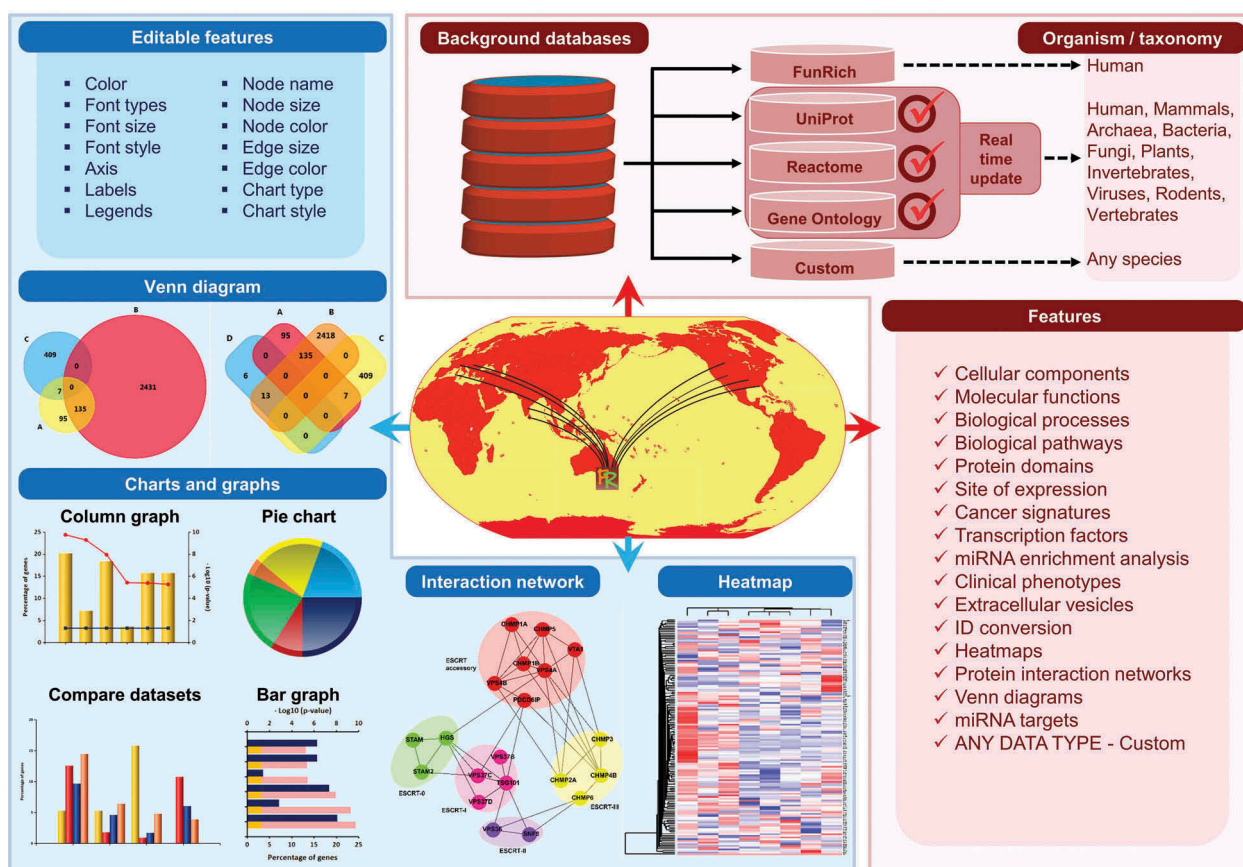
Figure 1. Features available in FunRich. FunRich is a free standalone functional enrichment analysis tool. Users can obtain customisable graphs, charts, interaction networks and heatmaps. All features of these images including color and font is editable and thus allows for quick representation of analysis. In spite of its ease of use, the images produced are of publication quality and can be directly imported into manuscripts. In addition, users have multiple background database options for more than 13,320 species. One of the features requested by the community is the option to update the background databases in real time. FunRich now allows the users to download data from UniProt, Reactome and Gene Ontology databases in real time. Furthermore, users can perform analyses in the context of biological pathways, gene ontology categories, protein domains, site of expression, cancer signatures, transcription factors, clinical phenotypes, extracellular vesicles, miRNA enrichment, protein interaction network and cross database accession conversion. The custom database option allows users to use any data type for any species thereby allowing for flexibility.

based on extracellular vesicles subtype, sample type, isolation method, cargo type and identification method.

Overall, with the involvement of researchers in the early phase of software development, we have developed a comprehensive tool for functional enrichment analysis. As databases are not regularly updated in most of the functional enrichment analysis tools [8], the *community requested features* pertaining to automatic database update and custom database feature will allow for the continuous use of FunRich. Though we have completed 90% of the requests from researchers, we are constantly implementing the newer requests. We envision the development of a web-based version of FunRich and implementation of metabolomic analysis in the near future. With the advent of large volumes of data, it is

critical to build such comprehensive software tools for data analysis. Based on this fruitful experience, we strongly encourage community driven software development for research purposes so as to build comprehensive software tools and to curtail software duplications.

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

ORCID

Mohashin Pathan http://orcid.org/0000-0001-7071-0022
Shivakumar Keerthikumar http://orcid.org/0000-0001-9865-9767
David Chisanga http://orcid.org/0000-0002-0421-3957
Pedro Fonseca http://orcid.org/0000-0001-9685-8092
Pamali Fonseka http://orcid.org/0000-0003-1420-4103
Simona Fontana http://orcid.org/0000-0003-4681-2170
Nunzio Iraci http://orcid.org/0000-0003-2146-9329
Kenneth Kastaniegaard http://orcid.org/0000-0003-3806-5638
Joanna Kowal http://orcid.org/0000-0001-7849-6040
Tommaso Leonardi http://orcid.org/0000-0002-4449-1863
Anders Øverbye http://orcid.org/0000-0002-2564-3729
Héctor Peinado http://orcid.org/0000-0002-4256-3413
Simona Principe http://orcid.org/0000-0003-4681-2170
Cristiana Spinelli http://orcid.org/0000-0002-4655-6173
Joanne L. Welton http://orcid.org/0000-0002-1445-245X

References

[1] Zhang B, Wang J, Wang X, et al. Proteogenomic characterization of human colon and rectal cancer. Nature. 2014;513:382–387.
[2] Gerstein MB, Rozowsky J, Yan -K-K, et al. Comparative analysis of the transcriptome across distant species. Nature. 2014;512:445–448.
[3] Bailey P, Chang DK, Nones K, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531:47–52.
[4] Muir P, Li S, Lou S, et al. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol. 2016;17:53.
[5] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102:15545–15550.
[6] Benito-Martin A, Peinado H. FunRich proteomics software analysis, let the fun begin! Proteomics. 2015;15:2555–2556.
[7] Pathan M, Keerthikumar S, Ang C-S, et al. FunRich: an open access standalone functional enrichment and interaction network analysis tool. Proteomics. 2015;15:2597–2601.
[8] Wadi L, Meyer M, Weiser J, et al. Impact of outdated gene annotations on pathway enrichment analysis. Nat Methods. 2016;13:705–706.
[9] Kalra H, Simpson RJ, Hong J, et al. Vesiclepedia: a compendium for extracellular vesicles with continuous community annotation. PLoS Biol. 2012;10:e1001450.