BIOINFORMATICS APPLICATIONS NOTE Vol. 31 no. 3 2015, pages 440–441
doi:10.1093/bioinformatics/btu623

Systems biology

Advance Access publication October 9, 2014

NOVA: a software to analyze complexome profiling data

Heiko Giese[1], Jörg Ackermann[1], Heinrich Heide[2,3], Lea Bleier[2], Stefan Dröse[2], Ilka Wittig[2,3], Ulrich Brandt[2,4] and Ina Koch[1,*]

[1]Molecular Bioinformatics Group, Institute of Computer Science, Faculty of Computer Science and Mathematics, Cluster of Excellence Frankfurt "Macromolecular Complexes", Goethe-University, Robert-Mayer-Str. 11-15, 60325 Frankfurt am Main, Germany, [2]Molecular Bioenergetics Group, Medical School, Cluster of Excellence Frankfurt "Macromolecular Complexes", Goethe-University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany, [3]Functional Proteomics, SFB815 core unit, Medical School, Goethe-University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany and [4]Nijmegen Centre for Mitochondrial Disorders, Radboud University Medical Centre, 6525 GA, Nijmegen, The Netherlands

Associate Editor: Janet Kelso

ABSTRACT

Summary: We introduce NOVA, a software for the analysis of complexome profiling data. NOVA supports the investigation of the composition of complexes, cluster analysis of the experimental data, visual inspection and comparison of experiments and many other features.

Availability and implementation: NOVA is licensed under the Artistic License 2.0. It is freely available at http://www.bioinformatik.uni-frankfurt.de. NOVA requires at least Java 7 and runs under Linux, Microsoft Windows and Mac OS.

Contact: ina.koch@bioinformatik.uni-frankfurt.de

Received on May 18, 2014; revised on August 6, 2014; accepted on September 12, 2014

1 INTRODUCTION

To facilitate diverse cellular functions, proteins can associate to form complex molecular machineries. Complexome profiling (CP, Heide *et al.* 2012) uses blue-native gel electrophoresis to separate intact proteins and protein complexes up to a molecular weight of 10 MDa (Schägger and Jagow, 1991; Wittig *et al.*, 2006) or even 60 MDa (Strecker *et al.*, 2010) using special large pore gels (LP-BNE). After the separation, the gel strip is cut into 60 equally sized slices. Migration profiles are reconstructed by applying mass spectrometry (LC-MS/MS) to identify the proteins contained in each slice. The relative abundance of each protein is calculated by label-free LC/MS-based quantification (Heide *et al.*, 2012). Through comparison of these profiles by hierarchical clustering, groups of co-migrating proteins are recognized, indicating the composition of quaternary structures and functional complexes (Andersen *et al.*, 2003; Foster *et al.*, 2006; Wessels *et al.*, 2009). The method has been successfully applied to analyze mitochondrial complexes in rats (Heide *et al.*, 2012) and humans (Wessels *et al.*, 2013) as well as to explore complex formation in plants and bacteria (Takabayashi *et al.*, 2013). Typically, experimental datasets contain up to thousands of migration profiles, which cannot be manually processed. Bioinformatic tools are needed to efficiently handle and visualize these data. Though several programs like Cluster 3.0 (de Hoon *et al.*, 2004), Java Treeview (Saldanha, 2004) or the MultiExperiment Viewer (Saeed *et al.*, 2003) have been applied, there is to the best of our knowledge no tool available that supplies all functionalities required for the visualization and evaluation of CP data.

*To whom correspondence should be addressed.

2 FEATURES

We developed NOVA–a flexible interactive tool for the analysis and visual inspection of CP data. Protein abundance profiles obtained by other protein separation techniques, such as density gradient centrifugation or size exclusion chromatography, may also be analyzed with NOVA. Datasets can be imported as XLS, XLSX, CSV or TXT files. NOVA displays the migration profiles as a heat map, see Figure 1A, providing mouse functionality for visual inspection and data management. Links to databases like UniProt (Magrane and UniProt Consortium, 2011) enable the fast access to additional information about the proteins. For the identification of complex formation changes, for example, caused by a knockdown of a specific assembly factor, heat maps of multiple CP data can be compared. NOVA supports the normalization of data by range, maximum values, profile area and unit vector. Migration profiles of selected proteins can be compared, applying the module *profile viewer*, see Figure 1B.

The clustering of migration profiles is the basic step for the prediction of the composition of protein complexes. For this purpose, NOVA implements the distance measures of Euclid, Pearson, Manhattan and hierarchical clustering covering single, complete, average and Wards linkage. A *tree viewer* displays the cluster hierarchy, see Figure 1C. Proteins and/or entire mass ranges can be excluded from the clustering process allowing the investigation of specific data subsets. Clustered data can be exported as XLS, XLSX, CSV or TXT files, figures as JPEG, PNG, TIFF or BMP files.

3 SUMMARY

NOVA is a freely available software tool for the evaluation and visualization of CP datasets. It supports highly flexible and interactive inspection, exploration and analysis of complexome data. The implemented analysis techniques focus on hierarchical clustering methods that are the current standard approach to

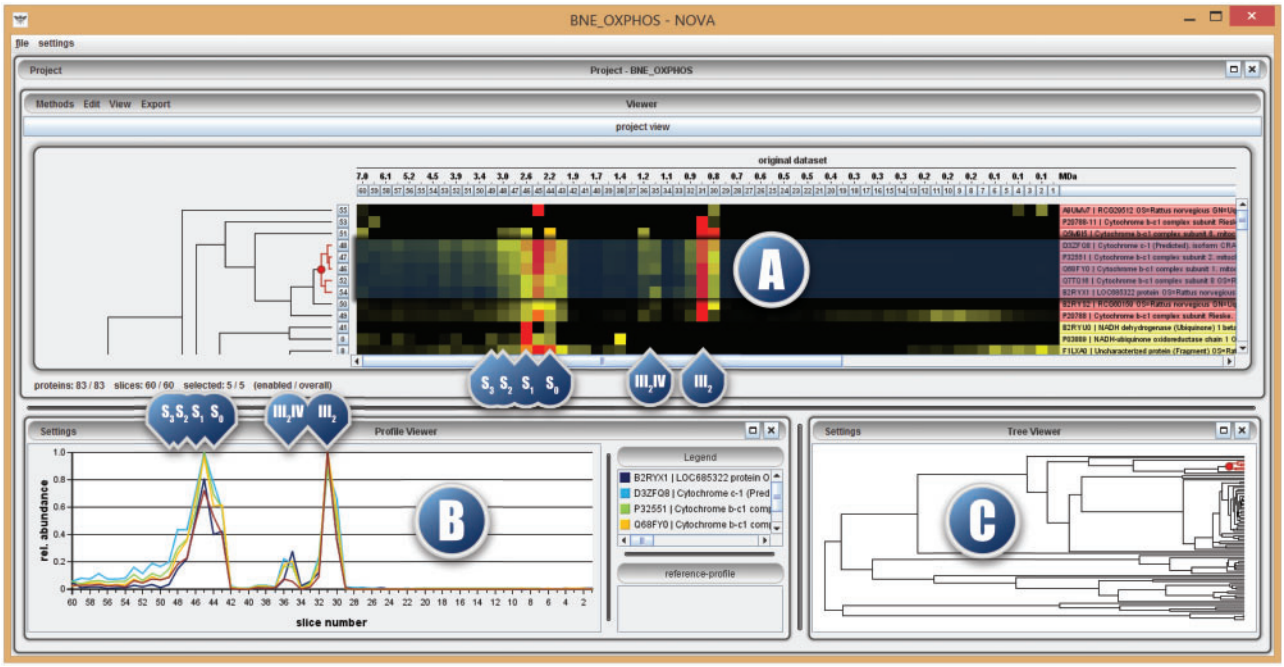

Fig. 1. NOVA's graphical user interface displaying a clustered CP dataset from rat hearth mitochondria (Heide *et al.*, 2012): (**A**) The heat map represents gel migration profiles. Each row displays the profile of an individual protein. A label at the end of each row identifies the protein. The background color of the label indicates a known membership of the protein to a complex (here 10 subunits of respiratory complex C III and 3 subunits of C I). On the top of the heat map, a mass scale shows the expected mass of proteins assembled in the gel slices. The migration profiles are clustered (here Pearson correlation, average linkage). Left, the corresponding cluster subtree is aligned to the heat map. A cluster of proteins in the cluster tree is highlighted, corresponding to the selected rows of the heat map. (**B**) A diagram displays the migration profiles of the selected proteins. Each peak of the consensus profile corresponds to a protein assembly, functional complex or supercomplex. These assemblies can be identified by their distinct mass. For example, the selected proteins are members of the respiratory homodimeric complex III_2, the complex assembly III_2IV and the series of supercomplexes S_0–S_3. (**C**) The complete cluster tree. The *tree viewer* allows to navigate through the heat map and to explore migration profiles of subgroups of proteins

study the protein composition of quaternary structures. NOVA provides functionalities to assist the study of different experimental conditions, knockout experiments and disease-related changes. It is already applied by various research groups working with CP data.

ACKNOWLEDGEMENTS

The authors gratefully thank Valentina Strecker, Kim-Kristin Prior, Jens Einloft and Jörg Kuharev for testing and many valuable suggestions.

Funding: This work was partly supported by the Excellence Initiative of the German Federal and State Governments (EXC 115).

Conflict of interest: none declared.

REFERENCES

Andersen,J.S. *et al.* (2003) Proteomic characterization of the human centrosome by protein correlation profiling. *Nature*, **426**, 570–574.

de Hoon,M. *et al.* (2004) Open source clustering software. *Bioinformatics*, **20**, 1453–1454.

Foster,L.J. *et al.* (2006) A mammalian organelle map by protein correlation profiling. *Cell*, **125**, 187–199.

Heide,H. *et al.* (2012) Complexome profiling identifies TMEM126B as a component of the mitochondrial complex I assembly complex. *Cell Metab.*, **16**, 538–549.

Magrane,M. and UniProt Consortium. (2011) UniProt Knowledgebase: a hub of integrated protein data. *Database*, 2011, 1–13.

Saeed,A.I. *et al.* (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **34**, 374–378.

Saldanha,A.J. (2004) Java Treeview–extensible visualization of microarray data. *Bioinformatics*, **20**, 3246–3248.

Schägger,H. and Jagow,G. (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.*, **199**, 223–231.

Strecker,V. *et al.* (2010) Large pore gels to separate mega protein complexes larger than 10 MDa by blue native electrophoresis: Isolation of putative respiratory strings or patches. *Proteomics*, **10**, 3379–3387.

Takabayashi,A. *et al.* (2013) Protein co-migration database (PCoM -DB) for *Arabidopsis* thylakoids and *Synechocystis* cells. *Springerplus*, **2**, 148.

Wessels,H.J. *et al.* (2009) LC-MS/MS as an alternative for SDS-PAGE in blue native analysis of protein complexes. *Proteomics*, **9**, 4221–4228.

Wessels,H.J. *et al.* (2013) Analysis of 953 human proteins from a mitochondrial HEK293 fraction by complexome profiling. *PLoS One*, **8**, e68340.

Wittig,I. *et al.* (2006) Blue-native PAGE. *Nat. Protoc.*, **1**, 418–428.