EVALUATING THE EXPANSION OF THE EXPERIMENTALLY DETERMINED HUMAN PROTEIN INTERACTOME BASED ON THE PICKLE META-DATABASE

G.N. Dimitrakopoulos[1,2], A. Gioutlakis[1,2], M.I. Klapa[2] and N.K. Moschonas[1,2,*]
[1]General Biology Laboratory, Department of Medicine, University of Patras, Patras, Greece
[2]Metabolic Engineering & Systems Biology Laboratory, FORTH/ICE-HT, Patras, Greece
(* n_moschonas@med.upatras.gr)

ABSTRACT
Protein InteraCtion KnowLedgebasE (PICKLE) is meta-database containing experimentally detected direct human protein-protein interactions, collected by five primary databases. The usage of reviewed human complete proteome (RHCP) as basis for node set enables us to examine the expansion of the protein interaction network throughout the different PICKLE releases. While the protein interactome is getting enriched with new edges and nodes, we observed that the network structure remains unchanged with regard to its hubs.

INTRODUCTION
The PICKLE meta-database, developed by our groups at University of Patras and FORTH/ICE-HT[1,2], consists of the experimentally-determined direct protein-protein interaction (PPI) network in human and is and is publicly available at www.pickle.gr. PICKLE integrates experimental PPI data from five source databases (BioGRID[3] , DIP[4], HPRD[5], IntAct[6], MINT[6]). This is performed on the basis of the genetic information ontology network of the UniProtKB/Swiss-Prot-defined RHCP, eliminating thus the need for their a priori normalization to a selected level of genetic information (i.e. protein (UniProt), gene or mRNA). This ontological way of PPI data integration from various sources using RHCP as a standardized reference node set, constitutes PICKLE the only PPI meta-database today that enables the evaluation of the expansion of the experimentally determined human protein interactome at both the protein and gene levels. This evaluation, which is the objective of the present study, can be achieved by comparing PICKLE successive releases, which reflect the respective changes in the biological and PPI primary datasets.

MATERIALS AND METHODS
The default ("cross-checked") direct PPI networks at the UniProt level of PICKLE releases 1.0, 2.1, 2.2 and 2.3, integrating primary biological and PPI data from January 2012, February 2015, December 2017 and November 2018, respectively, were compared in this study. These networks exclude experimental PPIs that are of low confidence of being direct based on the available experimental evidence[1]. Network analysis was performed using the Cytoscape software[7].

RESULTS AND DISCUSSION
The current experimentally determined direct PPI network in human, as reconstructed by PICKLE, comprises 178306 PPIs between 15823 UniProt IDs, supported by 39603 publications[8]. The PICKLE network remains double in number of PPIs from any of the largest primary PPI datasets (BioGRID or IntAct), emphasizing the need for integration of source databases for the reliable reconstruction of the currently known human protein interactome (Fig. 1). Moving from PICKLE 2.1 to PICKLE 2.2, we observed an increase of 9.2% in UniProt IDs, 33.2% in PPIs and 7.7% in supporting references, with these changes being 2.5%, 10.7% and 2.8%, respectively, from PICKLE 2.2 to the current interactome. Interestingly, the increase in the experimentally supported PPIs since PICKLE 2.1 is at least three-fold larger than the respective change in the number of supporting references, reflecting the fact that in the last years PPIs are largely determined by high-throughput experiments.

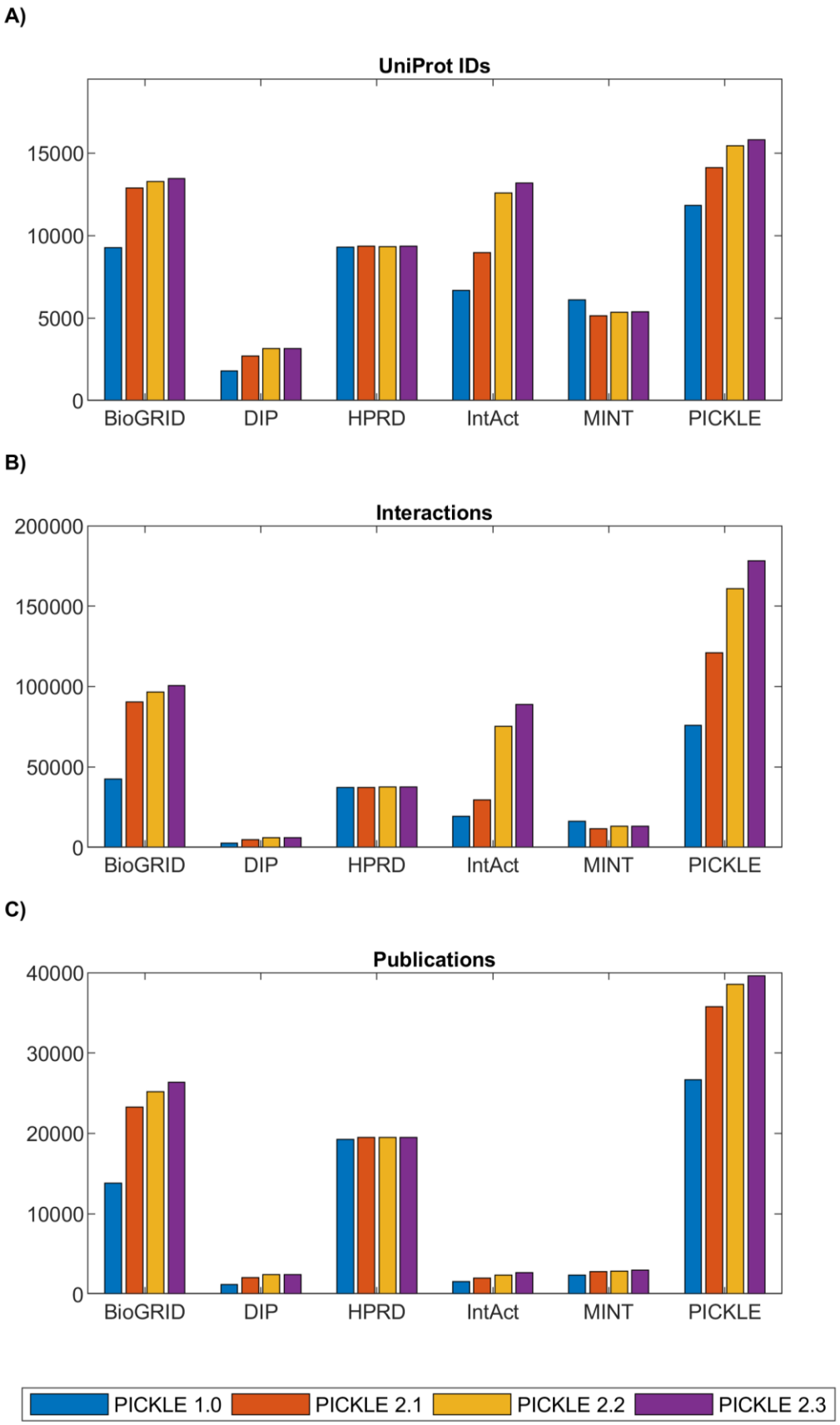

Figure 1. *The expansion of the PICKLE PPI network and the five primary PPI databases it consists of, in terms of A) proteins, B) interactions and C) publications. As basis, the PICKLE cross-checked network at UniProt level was used from versions 1.0, 2.1, 2.2 and 2.3.*

Moreover, we observed a three times smaller increase in the number of newly added protein-nodes compared to the newly determined PPIs, with most of the latter enriching the interactions of already existing proteins in the human interactome. In addition, as we had predicted[2], the vast majority of the newly added proteins have low degree, while none has been identified among the hubs (i.e. with over 300 interactions) (Fig. 2). Specifically, from the 452 new entries in PICKLE 2.3 PPI network, the 349 nodes have 1 – 4 interactions.

In summary, the interactome becomes denser, further supporting the overlapping between biological processes underlying its robustness and flexibility, the disease co-morbidity and the efforts for drug repurposing. However, we note that in the most recent PICKLE version, the "unfiltered" network, which contains the interactions characterized as of lower confidence to be direct, covers 17300 proteins from the set of 20381 proteins in RHCP. Thus, the further expansion of the PPI network over the 15% of RHCP proteins that remain with no known protein interactions seems to require more targeted experiments with respect to the proteins to be investigated and the biological processes to which they participate.

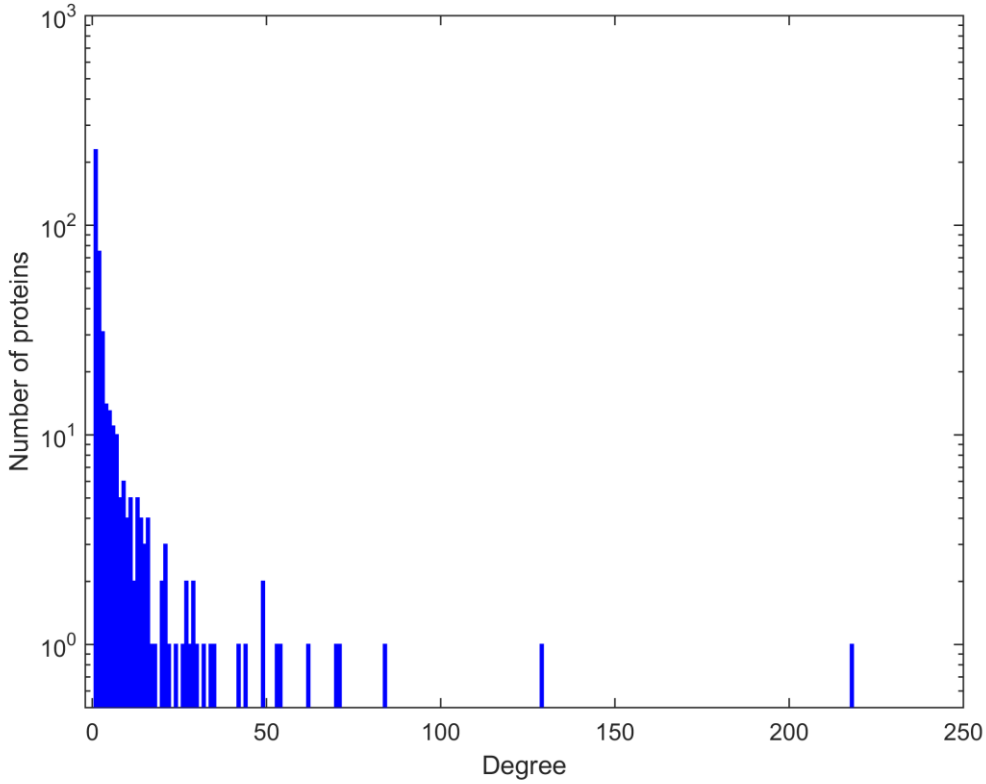

Figure 2. *The degree distribution of the 452 newly inserted proteins in PICKLE 2.3 cross-checked network.*

CONCLUSIONS

Our observations reinforce our previous statement[2] that the structure of the human protein interactome with respect to its hubs, as it can be determined by untargeted high-throughput experiments, has already been largely defined. Identifying its dynamics based on developmental stage-, tissue- and cell-specific omic data in health and disease is the next challenge for network biology and medicine.

ACKNOWLEDGMENTS

This work was supported mainly by the project "ELIXIR-GR: Hellenic Research Infrastructure for the Management and Analysis of Data from the Biological Sciences" (MIS 5002780), and partly by EATRIS-GR (MIS 5028091) and INSPIRED (5002550) projects, which are implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", and partly by "ΒΙΤΑΔ-ΔΕ: Advanced Research Activities in Biomedical and Agroalimentary Technologies" (MIS 5002469) project, implemented under the "Action for the Strategic Development on the Research and Technological Sector", all funded by the Operational Program "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

REFERENCES

[1] A. Gioutlakis, M.I. Klapa, N. K.Moschonas. PLoS ONE, 12(10) (2017) e0186039.
[2] M.I. Klapa, K. Tsafou, E. Theodoridis, A. Tsakalidis, N.K. Moschonas. BMC Syst. Biol. 7:96 (2013).
[3] R. Oughtred, C. Stark, B. J. Breitkreutz, J. Rust, L. Boucher, …, M. Tyers. Nucleic Acids Res., 47(D1) (2019) D529-D541.
[4] L. Salwinski, C.S. Miller, A. J. Smith, F. K. Pettit, J. U. Bowie, D. Eisenberg. Nucleic Acids Res., 32 (2004) D449-51.
[5] T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, …, A. Pandey. Nucleic Acids Res., 37 (2009) D767–72.
[6] S.Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, ..., H. Hermjakob. Nucleic Acids Res. 42 (2014) D358–63.
[7] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, ..., T. Ideker. Genome Research, 13 (2003) 2498-2504.
[8] G. N. Dimitrakopoulos, A. Gioutlakis, M.I. Klapa, N. K.Moschonas. In prep. (2019).