

Investigating the Genetic Architecture of Blood Pressure Regulation through PPI Network-based Integration of GWAS and Functional Data

Evridiki-Pandora G. Tsare
Metabolic Engineering & Systems
Biology Laboratory, Institute of
Chemical Engineering Sciences
(ICE-HT), FORTH, Patras, Greece
tsare@upatras.gr

Maria I. Klapa
Metabolic Engineering & Systems
Biology Laboratory, Institute of
Chemical Engineering Sciences
(ICE-HT), FORTH, Patras, Greece
mklapa@iceht.forth.gr

Nicholas K. Moschonas
School of Medicine, Un. of Patras &
Metabolic Engineering & Systems
Biology Laboratory, Institute of
Chemical Engineering Sciences
(ICE-HT), FORTH, Patras, Greece
n_moschonas@med.upatras.gr

ABSTRACT

Introduction: A major goal of medical genetics is to elucidate the genetic architecture of multifactorial diseases. This can be addressed through the integrative analysis of genome-wide association studies (GWAS) and functional data within the human protein–protein interaction (PPI) network, as disease pathophysiology results from disruptions in interacting molecular pathways. This approach requires the development of disease-specific meta-databases that integrate genetic and other biological and functional data, including PPIs, and workflows enabling network-based prioritization of genes, proteins, and pathways. Here, we applied this approach to blood pressure (BP) regulation.

Methods: BP-GWAS data were retrieved and curated from the GWAS Catalog (www.ebi.ac.uk/gwas/) and literature. The BP-associated protein interactome was reconstructed using PICKLE (www.pickle.gr) meta-database [2]. Network analysis and metrics were applied to identify new proteins with high influence on BP. Pathway enrichment and functional analyses were performed using knowledge-based bioinformatics tools.

Results: The development of a systematically curated BP-GWAS meta-database enabled the identification of 1,065 GWAS proteins with PPIs. Reconstruction of the BP-associated PPI network, by extending the GWAS-derived network with the shortest paths, connecting all BP GWAS-proteins into one component, led to the identification of additional 1,443 intermediate protein nodes. Pathway enrichment analysis revealed multiple BP-related biological processes, most emerging after network expansion. A set of 335 BP-proteins was prioritized, by combining an integrated GWAS-based score with two network-based criteria. Functional analysis indicated many BP-proteins as targets of known anti-hypertensive drugs, and associations with other diseases such as metabolic, neurological and cardiovascular.

Conclusion: The proposed workflow could be effectively applied to other multifactorial diseases.

This work summarizes results previously reported in [1].

CCS CONCEPTS

•Applied computing ~ Life and medical sciences ~ Systems biology •Applied computing ~ Life and medical sciences ~Computational biology ~Biological networks •Applied computing ~ Life and medical sciences ~ Bioinformatics

KEYWORDS

Blood pressure regulation, Hypertension, GWAS, Human protein–protein interactions (PPIs), PPI network analysis, Network medicine, Systems medicine, Gene prioritization, Pathway enrichment analysis

1 Implementation of a systematically literature-curated BP-GWAS meta-database

The BP-GWAS meta-database (Fig. 1) systematically stores BP-GWAS data, SNP-transcript associations and eQTL measurements [1]. The SNP–transcript associations link the GWAS data to the genetic information ontology network connecting genes, transcripts and proteins, thereby enabling the integration of any type of biological, and functional data, including PPIs. Our dataset comprises 21,788 SNP–BP trait association p-values (5×10^{-8}) for 6,687 SNPs, of which 1,065 have known interactions in human. To enable gene prioritization prior to network analysis, we defined an integrated GWAS-based score formulated as a weighted summation of key GWAS-derived attributes: the gene-phenotype association p-value, the number of significant SNPs per gene and the number of supporting GWAS publications (Fig. 2) [1]. Based on this scoring scheme, 103 genes were prioritized.

2 Reconstruction of the BP PPI network & extension with new proteins of high probability to be BP-associated

The projection of the BP-GWAS proteins onto the human protein interactome of PICKLE meta-database [2] revealed that $\sim 2/3$ of them form a connected component, referred to as “blue nodes” (BNs), while the remaining proteins are referred to as “green nodes” (GNs) (Fig. 2A).

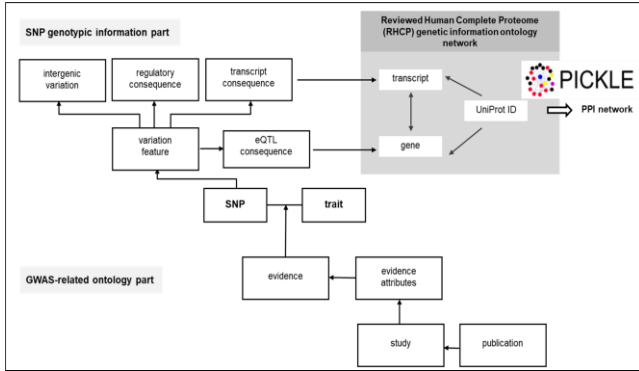


Figure 1: The relational scheme of the BP-GWAS meta-database.

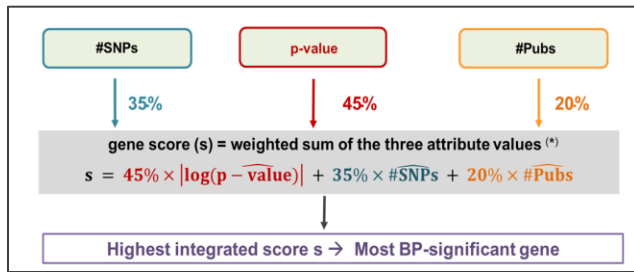


Figure 2: The integrated GWAS prioritization score.

Considering that all BP-GWAS proteins are associated with BP, we assumed that they are functionally connected. We proposed a network expansion approach based on the identification of the shortest paths linking all GWAS-proteins into one component [1]. The shortest-path intermediates (Fig. 2B “yellow nodes”-YNs) have a high probability of being associated with BP due to the “guilt-by-association” principle [3]. The reconstructed protein interaction network was enriched with 1,443 additional proteins.

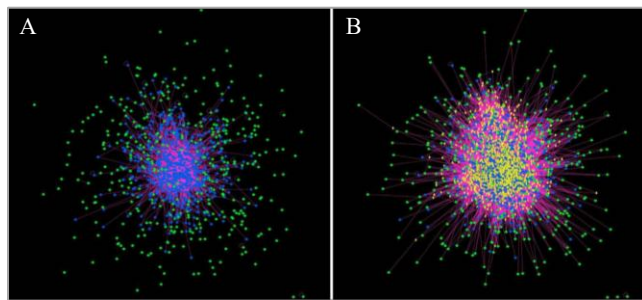


Figure 3: The GWAS-deduced (A), and the reconstructed (B) BP PPI network.

Based on the BP-extended network, we analyzed the role of each protein node using two approaches, which served as the two network-based criteria for protein prioritization [1]. The first network-based criterion uses an integrated metric that considers both the number of interactions of each node and its role in maintaining network connectivity [4]. The second novel network-based prioritization criterion focuses on the *in silico* identified BP-proteins (YNs), that are common neighbors of the GWAS-prioritized (criterion 1) [1]. Using these two network-based

approaches, we identified 106 and 170 prioritized proteins, respectively.

Pathway enrichment analysis revealed that BP-associated proteins are involved in multiple significantly enriched pathways related to cardiovascular function and signaling, including adrenergic signaling, PI3K-Akt, cGMP-PKG, HIF-1, and calcium signaling, as well as pathways associated with hormone regulation and metabolic processes, underscoring the functional relevance of the reconstructed network [1]. Notably, the inclusion of the network-inferred proteins (YNs) enhanced the statistical significance of these associations and enabled the identification of pathways, such as HIF-1 signaling in which 58% of the proteins correspond to YNs (Fig. 4). These pathways would not have been revealed using GWAS-derived proteins alone, highlighting the added value of network-based expansion.

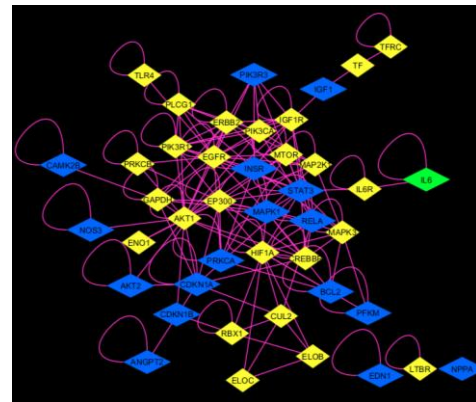


Figure 4: The PPI network of the identified BP-proteins involved in the HIF-1 pathway.

3 Integrated prioritization scheme of the extended BP-protein set

We proposed as the final BP-prioritized protein set the union of the three protein sets identified by the three prioritization criteria [1] (Fig. 5). In total, 335 proteins were prioritized. These proteins were further ranked according to the number of criteria they satisfy. Only one protein, ESR1, satisfies all three prioritization criteria, while 125 proteins satisfy two. The top-ranked proteins following ESR1 include INSR, PTPN11, CDK6, CSK, NOS3, SH2B3, ATP2B1, FES, and FN1. The association of these proteins with BP is supported by functional studies [1].

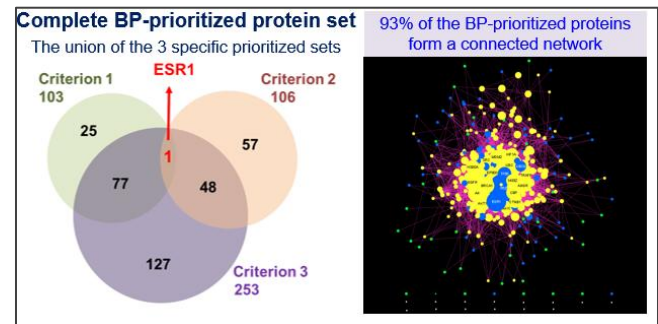


Figure 5: The integrated gene/protein prioritization scheme.

A large proportion of the prioritized proteins participate in pathways identified as significantly enriched in BP-associated proteins, further supporting the biological relevance of the prioritization scheme.

In summary, the integrated set of BP-prioritized proteins constitutes a robust resource for identifying BP-relevant candidates, combining evidence from GWAS with network-based measures of protein influence, thereby increasing confidence in prioritization. In addition, several of the identified BP-proteins correspond to targets of known anti-hypertensive drugs and are associated with multiple comorbid conditions, such as substance use (alcohol and tobacco), metabolic disorders including diabetes and metabolic syndrome, neurological and mood disorders, cardiovascular diseases, cancer, and renal failure. These findings highlight the translational potential of the proposed approach.

The currently available GWAS data are predominantly derived from populations of European ancestry, underscoring the need for increased representation of diverse ancestries to enable more comprehensive and unbiased analyses [5]. In addition, the current analysis does not account for sex-specific genetic architectures due to the limited availability of sex-stratified GWAS data; incorporating such datasets in future work could enable more refined analyses, supported by the structure of the developed meta-database.

The proposed workflow can be readily extended to other multifactorial diseases by [1]: (i) integrating disease-specific curated GWAS findings augmented with gene-variant associations, eQTL, and other functional data, (ii) reconstructing the disease-associated PPI network to enable the identification of novel disease-related genes and pathways, and (iii) prioritizing disease-related gene/proteins based on GWAS and network-related criteria. The developed workflow and meta-database are intended to be made publicly available (e.g., via ELIXIR-GR) to support reproducibility and collaborative research.

ACKNOWLEDGMENTS

This work was supported by a Hellenic Foundation for Research and Innovation (HFRI) PhD scholarship grant to EPGT (#2082), by the BrainPrecision (TAEDR 0535850) project that is funded by the European Union-Next Generation EU, Greece 2.0 National Recovery and Resilience plan, National Flagship Initiative "Health and Pharmaceuticals", by Greek national projects ELIXIR-GR (MIS 5002780), "EATRIS-GR (MIS 5028091), BBMRI-GR (MIS 5028275), "INSPIRED (MIS 5002550) implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", 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), by the EU-funded H2020 projects "DIAGONAL (ID 95152), "JOINTPROMISE: (ID 874837) and "AIDPATH: (ID 101016909).

REFERENCES

- [1] Tsare EPG, Klapa MI, Moschonas NK. Protein-protein interaction network-based integration of GWAS and functional data for blood pressure regulation analysis. *Hum Genomics*. 2024;18(1):15. doi:10.1186/s40246-023-00565-6
- [2] Dimitrakopoulos GN, Klapa MI, Moschonas NK. PICKLE, 2021. 3.0: enriching the human meta-database with the mouse protein interactome extended via mouse-human orthology. Martelli PL, ed. *Bioinformatics*.37(1):145-146. doi:10.1093/bioinformatics/btaa1070
- [3] Oliver S. Guilt-by-association goes global. *Nature*. 2000;403(6770):601-602. doi:10.1038/35001165
- [4] Salavaty A, Ramialison M, Currie PD. Integrated Value of Influence: An Integrative Method for the Identification of the Most Influential Nodes within Networks. *Patterns*. 2020;1(5):100052. doi:10.1016/j.patter.2020.100052
- [5] Tsare EPG, Klapa MI, Moschonas NK. Investigating differences in the GWAS-based protein-protein interaction network of blood pressure regulation due to ancestry or transcript consequence severity. *IFAC-PapersOnLine*. 2024;58(23):133-138. doi:10.1016/j.ifacol.2024.10.023