INVESTIGATING THE GENETIC ARCHITECTURE OF BLOOD PRESSURE REGULATION BY PPI NETWORK-BASED INTEGRATION OF GWAS AND FUNCTIONAL DATA

Evridiki-Pandora G. Tsare[1,2], Maria I.Klapa[2] & Nicholas K. Moschonas[1,2]

[1] General Biology Laboratory, School of Medicine, University of Patras, Patras, Greece

[2] Metabolic Engineering & Systems Biology Laboratory, Institute of Chemical Engineering Sciences, FORTH, Patras, Greece

ABSTRACT

Introduction: A major objective of medical genomics concerns the determination of the genetic architecture of multifactorial diseases. This effort could be approached by analyzing relevant GWAS and functional data in the context of the human protein interactome, as the specific pathophysiology is the result of interactions of polyprotein pathways. Such an analysis may require the development and curation of disease-specific GWAS meta-databases integrating genetic & eQTL data, linking GWAS to other biological datasets including protein-protein interactions (PPIs), and the identification of workflows for the integrated network analysis of all these data towards gene/protein and pathway prioritization. In this study, we pursued this approach for the analysis of blood pressure (BP) regulation.

Methods: BP-GWAS data were retrieved and curated from GWAS-Catalog (www.ebi.ac.uk/gwas/) and the literature. The BP-associated protein interactome was reconstructed using PICKLE meta-database (www.pickle.gr). Network analysis was carried out and network metrics were employed to determine proteins with high influence to BP. Pathway enrichment and functional analysis was performed by using knowledge-based bioinformatics tools.

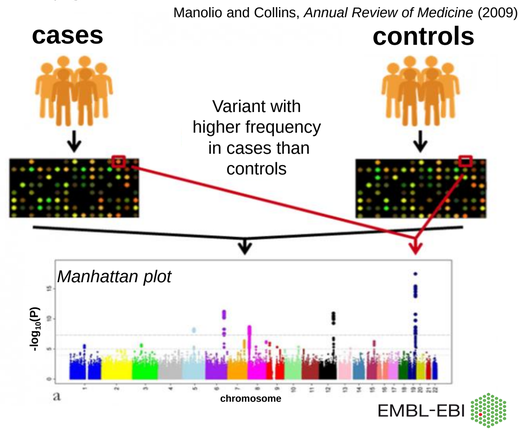
Results: The development of a systematically curated BP-GWAS meta-database by combining GWAS data with their transcript effects and eQTL measurements, led to the projection of these data on the human PPI network. The reconstruction of the BP-protein interactome and the extension of the GWAS-deduced network with the shortest paths, connected all BP-associated GWAS-deduced proteins into one component, i.e.: 1065 GWAS-based and 1443 intermediate protein nodes, and, interestingly, indicated that almost all of BP GWAS proteins are at most, second neighbours. For protein prioritization, we combined a newly proposed integrated GWAS-based scoring criterion with two network-based criteria, one considering the protein role in the reconstructed by shortest path interactome, and one novel, promoting the common neighbors of GWAS-prioritized proteins. A set of 335 protein-nodes were prioritized as BP associated, ranked by the number of satisfied criteria. Pathway enrichment analysis unveiled numerous bioprocesses, which are indeed functionally supported as BP-associated, underlining the importance of exploring GWAS in the context of protein interactome; otherwise, many BP-associated pathways would not have been identified. Moreover, functional analysis indicated many BP-proteins as targets of known anti-hypertensive drugs, and revealed associations with other diseases such as metabolic, neurological and cardiovascular diseases, heart and renal failure, and stroke.

Conclusion: Importantly, the implemented workflow could be effectively applied to GWAS datasets of any multifactorial disease.

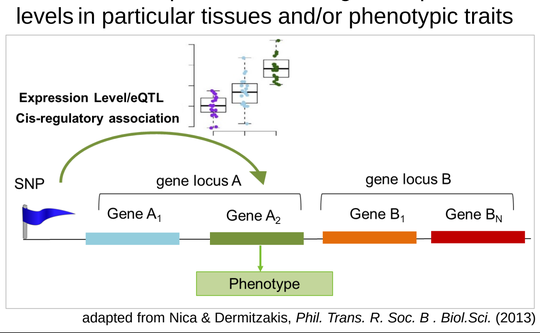
Understanding the Genetics of Complex Diseases: GWAS & eQTLs

GWAS are defined as "any studies of common genetic variation across the entire human genome designed to identify genetic associations with observable traits".

Manolio and Collins, *Annual Review of Medicine* (2009)



eQTL studies concern the discovery of genetic variants that explain variation in gene expression levels in particular tissues and/or phenotypic traits



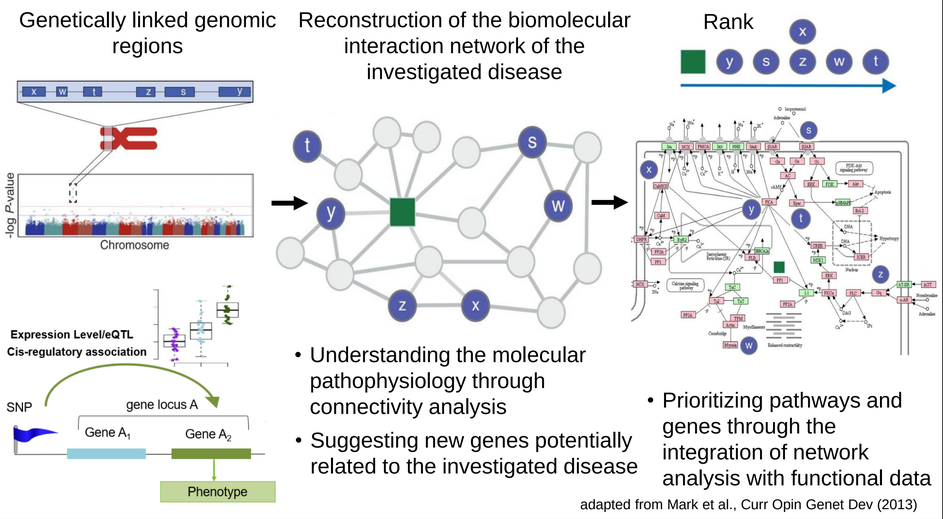
adapted from Nica & Dermitzakis, *Phil. Trans. R. Soc. B . Biol.Sci.* (2013)

GWAS & eQTL data: Do we possess all the information required to determine the molecular mechanisms of multifactorial diseases?

- GWAS and eQTL-identified disease-related genes individually may explain only a small portion of the underlying molecular mechanisms of a specific pathophysiology — Chimusa *et al.*, *Brief Bioinform.* (2019)
- Collecting and analyzing the full GWAS (and eQTL) dataset for a particular complex phenotype is very important, as the specific physiology results from the combined inter-regulation of multiple interacting polyprotein pathways, rather than the isolated effect of certain genes — Yang X., *Trends Mol Med.* (2020); Tsare, Klapa & Moschonas, *Hum Genomics* (2023), in review

It is of value to upgrade the information content of GWAS and eQTL data through their analysis in the context of biomolecular interaction networks

Upgrade the Genetics of Complex Diseases: Networks!



- Understanding the molecular pathophysiology through connectivity analysis
- Suggesting new genes potentially related to the investigated disease
- Prioritizing pathways and genes through the integration of network analysis with functional data

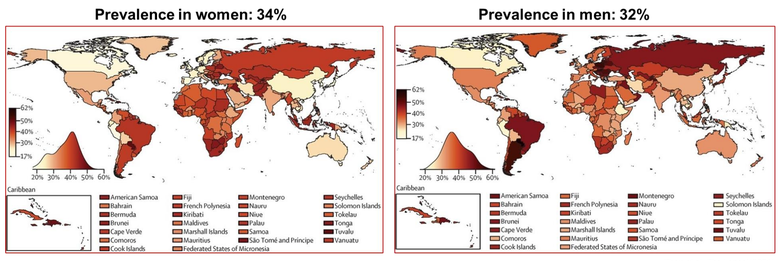
adapted from Mark et al., Curr Opin Genet Dev (2013)

Aim of Study

a. The development of systematic workflows for the PPI network-based integration of GWAS, eQTL and functional data for the molecular analysis of multifactorial disease pathophysiology.
b. Application in blood pressure (BP) regulation analysis

BP a major risk factor worldwide

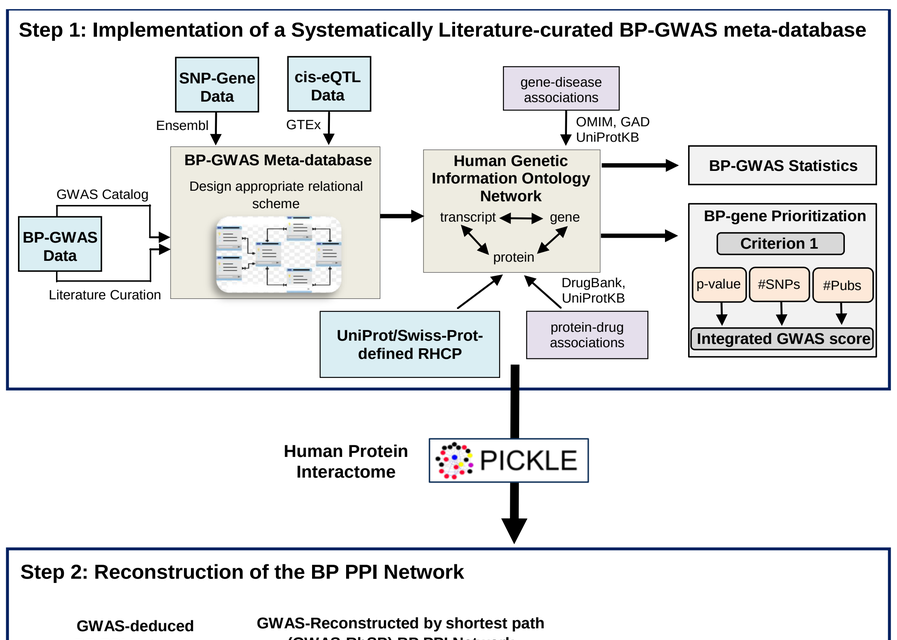
- Affects about ~1 billion people around the world



Zhou et al.,*The Lancet* (2021)
Funded by Word Heath Organization (WHO)

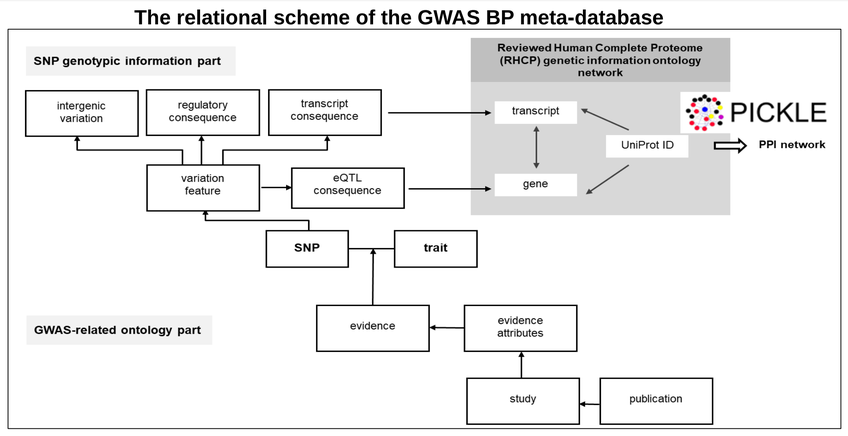
- A large number of genetic loci have been associated with BP through GWAS

Workflow of the integrated BP-GWAS, eQTL, functional data and PPI network analysis



STEP 1: Implementation of a Systematically Literature-curated BP-GWAS meta-database

The relational scheme of the GWAS BP meta-database

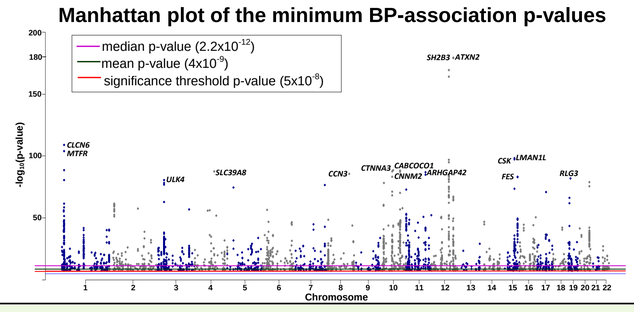


BP meta-database statistics

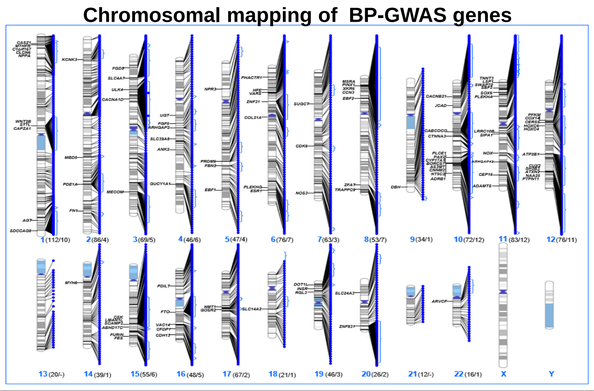
	p-value < 5x10^{-8}
SNP-trait associations	21788
SNPs	6687
Publications	54
Independent studies	151

60% protein-coding → 56% (3738) RHCP-coding → 1167 RHCP-coding genes → GTEx → 57% eGenes
16% non-coding
24% intergenic

Manhattan plot of the minimum BP-association p-values



Chromosomal mapping of BP-GWAS genes



A new gene prioritization criterion based on an integrated GWAS score: **Criterion 1**

#SNPs (35%) × p-value (45%) × #Pubs (20%)

gene score (s) = weighted sum of the three attribute values (*)

$$s = 45\% \times |\log(p - \widehat{value})| + 35\% \times \#\widehat{SNPs} + 20\% \times \#\widehat{Pubs}$$

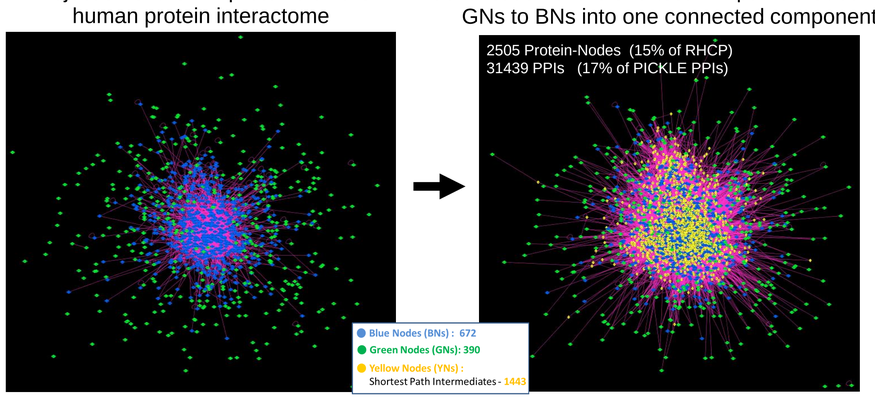
Highest integrated score s → Most BP-significant gene

Criterion 1: 103 genes

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

GWAS-deduced PPI network
Projection of the BP-proteins on the human protein interactome

GWAS-RbSP BP PPI network
Identification of the shortest PPI paths that connect GNs to BNs into one connected component



~2/3 of the BP-GWAS proteins forming a connected component

Almost all BP-GWAS proteins are at most second neighbours in the human protein interactome

Prioritizing BP-proteins based on their role in the GWAS-RbSP BP PPI network: **Criterion 2**

Integrated Value of Influence (IVI)

Hubness Score: Degree, Local H index
Spreading Score: Neighbourhood Connectivity, ClusterRank, Betweenness Centrality, Collective Influence

$$IVI = (DC + LH_{index})((NC + CR)(BC + CI))^{(*)}$$

Salavaty et al. in *Patterns* (2020)

Criterion 2: 106 proteins

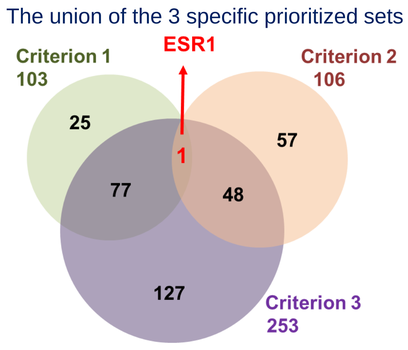
A novel network-based method promoting GWAS-prioritized proteins and their common YN neighbors : **Criterion 3**

- The Criterion 3 prioritized protein-set includes GWAS-prioritized proteins and their common YN neighbors
- The protein-set is ranked based on the node IVI in their subnetwork

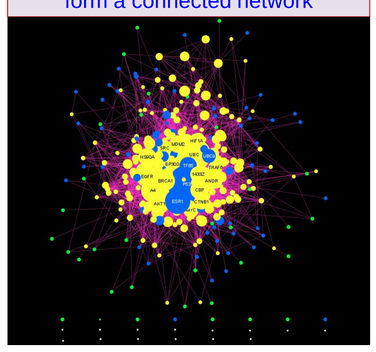
Criterion 3: 253 proteins

STEP 3: Integrated Prioritization and Functional Analysis of the extended BP-protein set

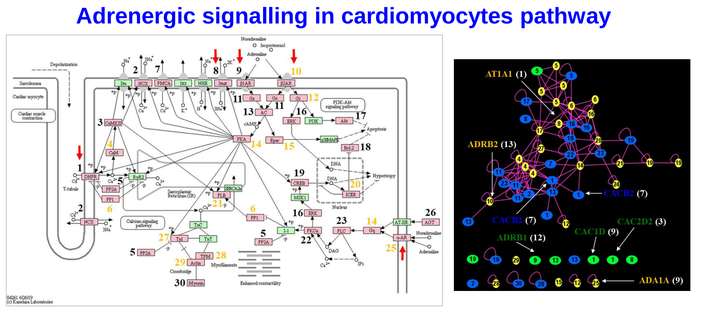
Complete BP-prioritized protein set
The union of the 3 specific prioritized sets



93% of the BP-prioritized proteins form a connected network

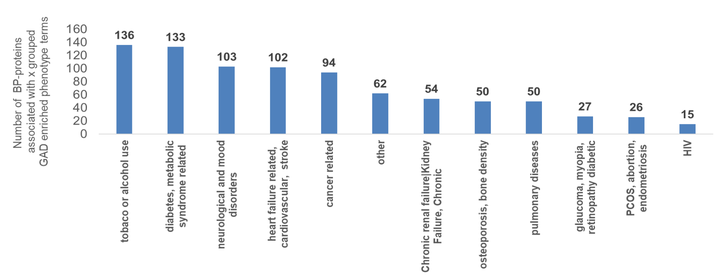


Pathway enrichment analysis
Adrenergic signalling in cardiomyocytes pathway



YNs enhance the statistical significance of the BP-association of the pathways

Comorbidity analysis



BP-Proteins as Antihypertensive Drug Targets

BP-enriched KEGG pathway	Number of Anti-hypertensive BP protein targets	Number of Anti-hypertensive drugs targeting pathway proteins
hsa04924 Renin secretion	9	45
hsa04022 cGMP-PKG signaling pathway	10	39
hsa04020 Calcium signaling pathway	10	32
hsa04024 cAMP signaling pathway	9	30
hsa04261 Adrenergic signaling in cardiomyocytes	8	29
hsa04270 Vascular smooth muscle contraction	6	25
hsa05410 Hypertrophic cardiomyopathy (HCM)	5	24
hsa01100 Metabolic pathways	7	23
hsa05414 Dilated cardiomyopathy (DCM)	5	21

CONCLUSIONS

Tsare, Klapa & Moschonas, *Hum Genomics* (2024)

- We introduced an integrated workflow for upgrading the information content of BP GWAS data and eQTL through PPI network analysis
- We implemented a systematically literature-curated BP-GWAS meta-database, augmented with eQTL data and variant-transcript associations linking to the human genetic information ontology network
- We reconstructed the BP-associated PPI network based on the comprehensive GWAS meta-dataset and a newly proposed extension method connecting all GWAS-proteins into one component through shortest paths
- We developed a gene/protein prioritization scheme based on the combination of an integrated GWAS-based & two network-based criteria
- Functional analysis validated our proposed workflow and extended our knowledge about BP regulation

ACKNOWLEDGEMENTS

- HFRI PhD scholarship grant to E.-P. Tsare
- Brain Precision (TAEDR-0535850)
- ELIXIR–GR (MIS 5002780)
- "BITAD" (MIS 5002469); "ELIXIR-GR" (MIS 5002780); "INSPIRED" (MIS 5002550) and "EATRIS-GR" (MIS 5028091) Greek national projects
- "DIAGONAL" (ID 95152); "JOINTPROMISE" (ID 874837); and "AIDPATH" (ID 101016909) EU-funded H2020 projects