FORTH
INSTITUTE OF CHEMICAL ENGINEERING SCIENCES

Investigating differences in the GWAS-based protein-protein interaction network of blood pressure regulation due to ancestry or transcript consequence severity

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

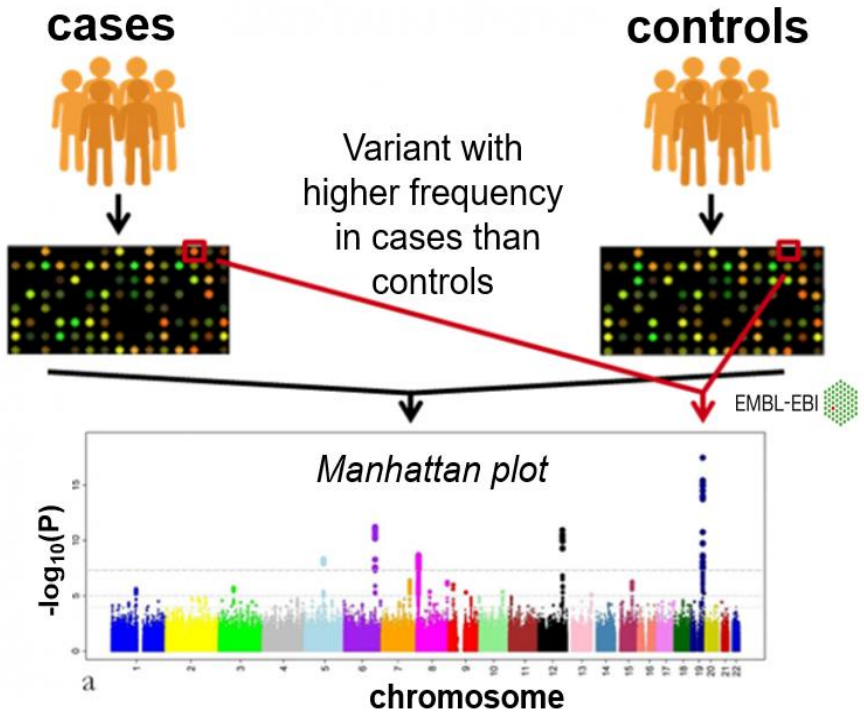
[1] Metabolic Engineering & Systems Biology Laboratory, FORTH/ICE-HT, Patras, Greece
[2] General Biology Laboratory, School of Medicine, University of Patras, Greece

IFAC INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL
FOSBE 2024
Corfu Island, Greece • September 8 - 11

Understanding the Genetics of Complex Diseases: GWAS

Genome Wide Association Studies (GWAS) have been defined as “any studies of common genetic variation across the entire human genome designed to identify genetic associations with observable traits”.

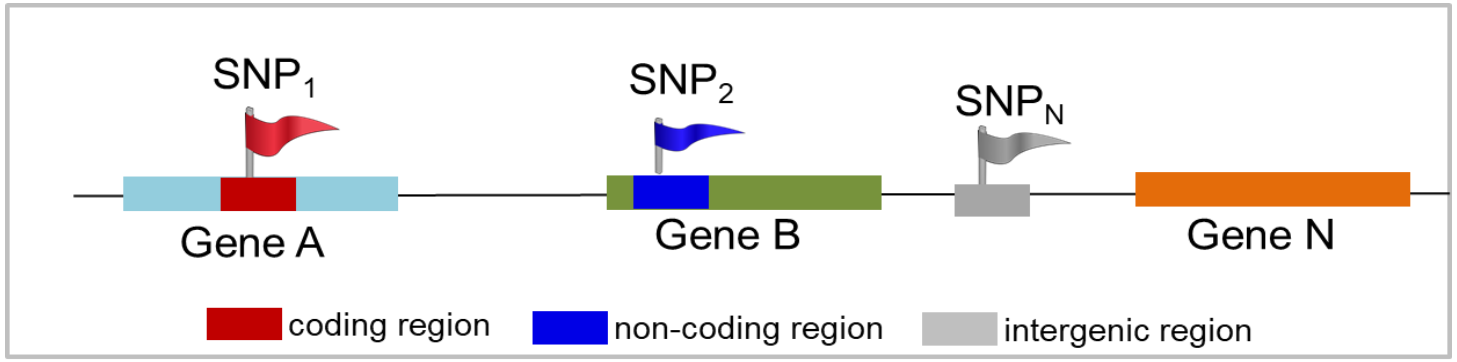
Manolio and Collins (2009). *Annual Review of Medicine*. 60: 443-456



- Single-nucleotide polymorphisms (SNPs) are the most common type of genetic variation among people. A SNP is a germline substitution of a single nucleotide at a specific position in the genome
- Statistical analysis indicates the probability of a SNP to be associated with a trait/phenotype variant (SNP-trait association p-value)

Variant-gene associations and variant consequences on genes and/or proteins

- SNPs may occur within coding sequences of genes, non-coding regions of genes, or in the intergenic regions (regions between genes).



- There are genetic approaches that can identify the consequences of a variant that may have on gene(s)..

Examples of variant consequences on genes in order of severity

more severe to less severe

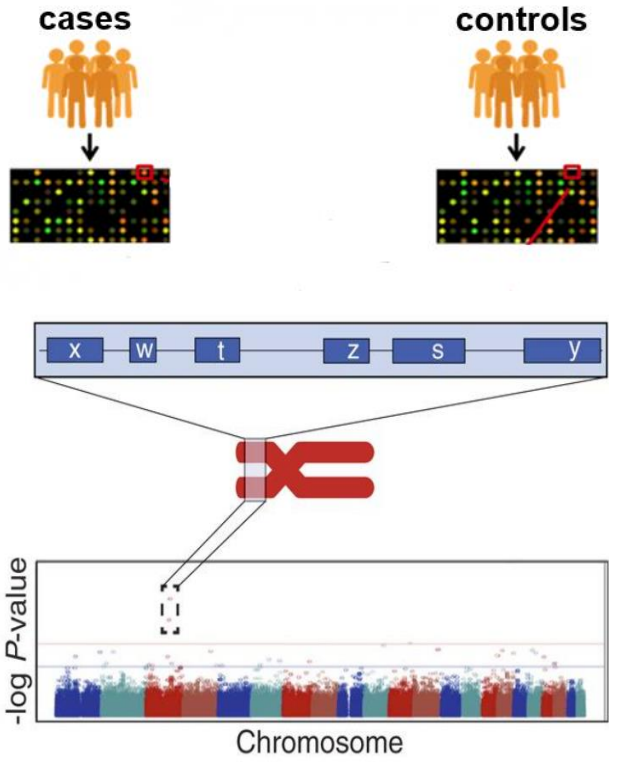
Consequence Type	Description
transcript_ablation	A feature ablation whereby the deleted region includes a transcript feature
splice_acceptor_variant	A splice variant that changes the 2 base region at the 3' end of an intron
splice_donor_variant	A splice variant that changes the 2 base region at the 5' end of an intron
stop_gained	A sequence variant whereby at least one base of a codon is changed, resulting in a premature stop codon, leading to a shortened transcript
⋮	⋮
inframe_insertion	An inframe non synonymous variant that inserts bases into in the coding sequence
inframe_deletion	An inframe non synonymous variant that deletes bases from the coding sequence
missense_variant	A sequence variant, that changes one or more bases, resulting in a different amino acid sequence but where the length is preserved
⋮	⋮
intergenic_variant	A sequence variant located in the intergenic region, between genes

(https://www.ensembl.org/)

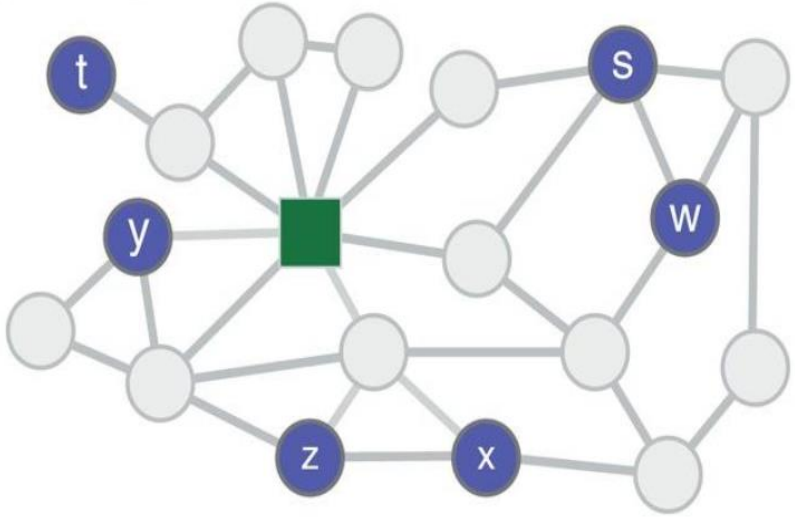
Upgrade the Genetics of Complex Diseases: Networks!

Network analysis of GWAS

Genetically linked genomic regions

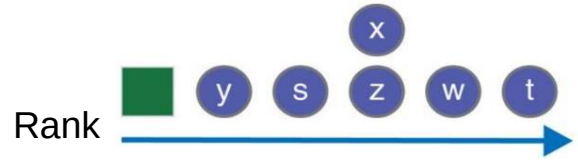


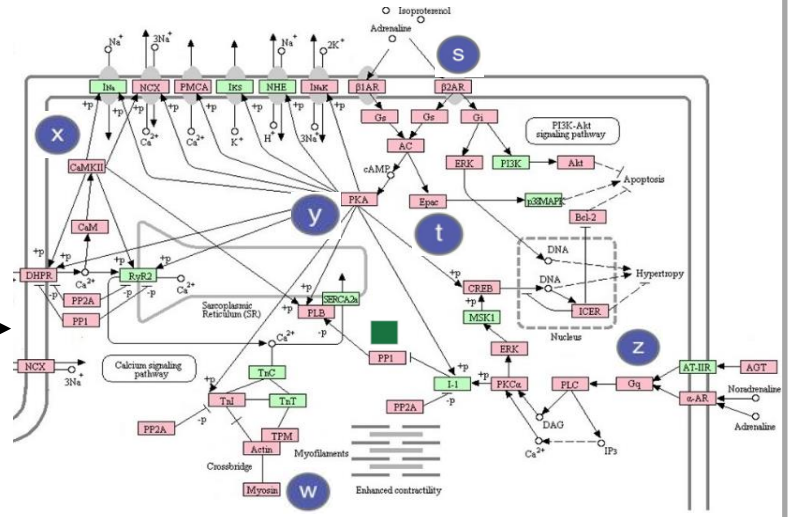
Reconstruction of the biomolecular interaction network of the investigated disease



- Understanding the molecular pathophysiology through connectivity analysis
- Suggesting new genes/proteins potentially related to the investigated disease

Rank





- Prioritizing pathways and proteins through the integration of network analysis with functional data

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

Protein-Protein Interaction (PPI) Network describes the cell physiology at protein level. Proteins are the main regulators of the majority of biological processes.

Tsare *et al. Human Genomics* (2024) 18:15
https://doi.org/10.1186/s40246-023-00565-6

Human Genomics

RESEARCH

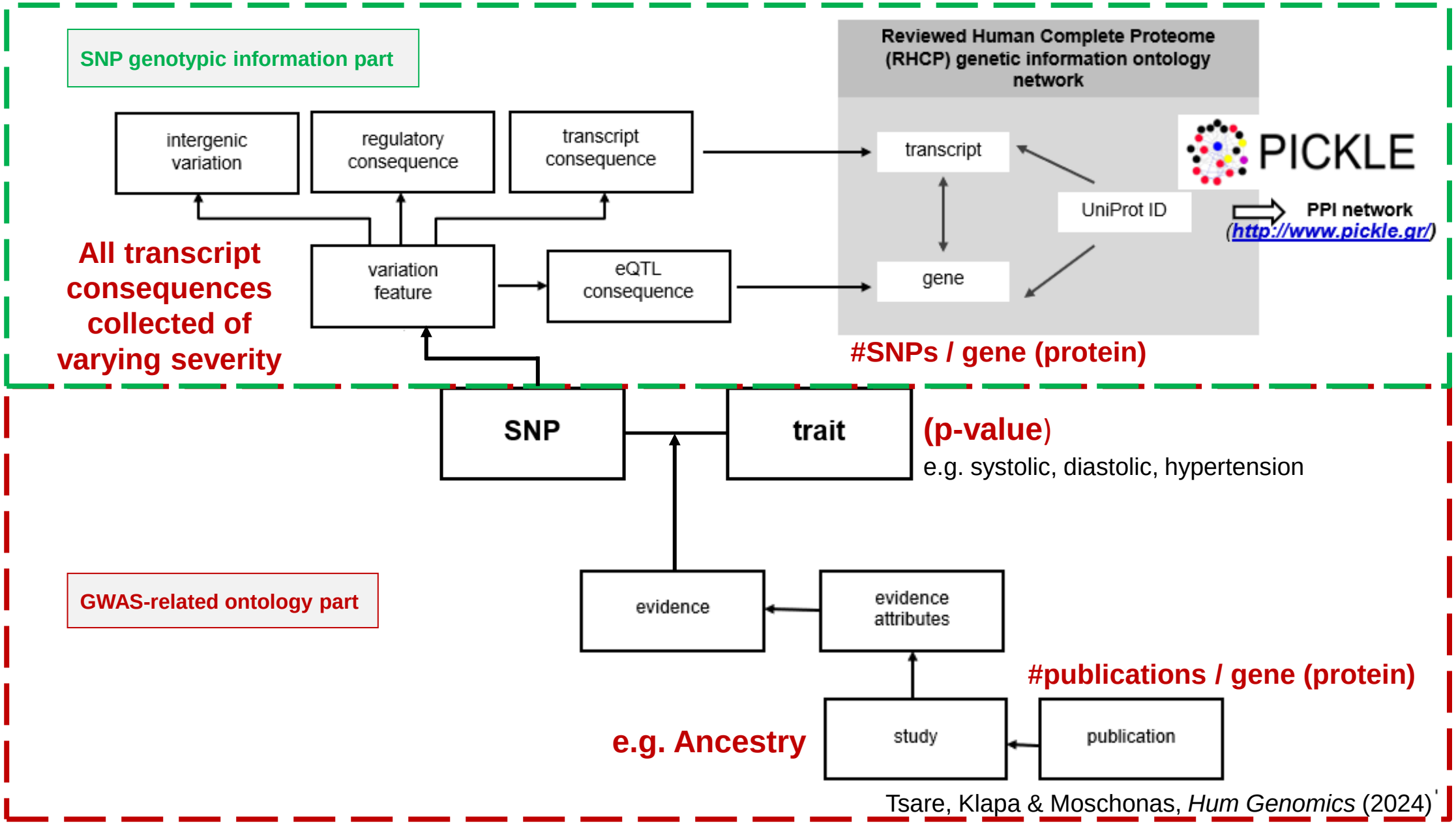
Open Access

Protein–protein interaction network-based integration of GWAS and functional data for blood pressure regulation analysis

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

- We developed a standardized BP regulation GWAS meta-database which collects all the associated BP-SNPs along with other biological data connecting SNPs with genes and proteins
- We reconstructed the extended PPI network of BP regulation revealing proteins that have not been reported as BP-related based on GWAS data
- We developed a gene/protein prioritization scheme based on the combination of an integrated GWAS-based & two network-based criteria

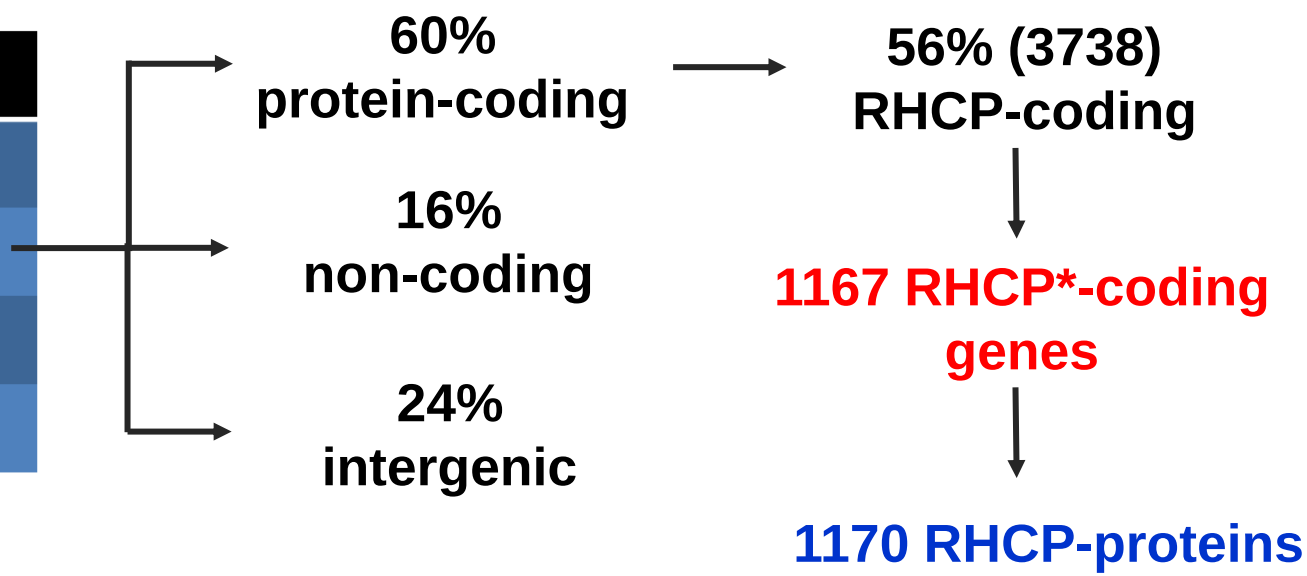
A standardized GWAS meta-database for blood pressure regulation (BP)



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

BP meta-database statistics for current GWAS significance threshold (p-value < 5x10-8)

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



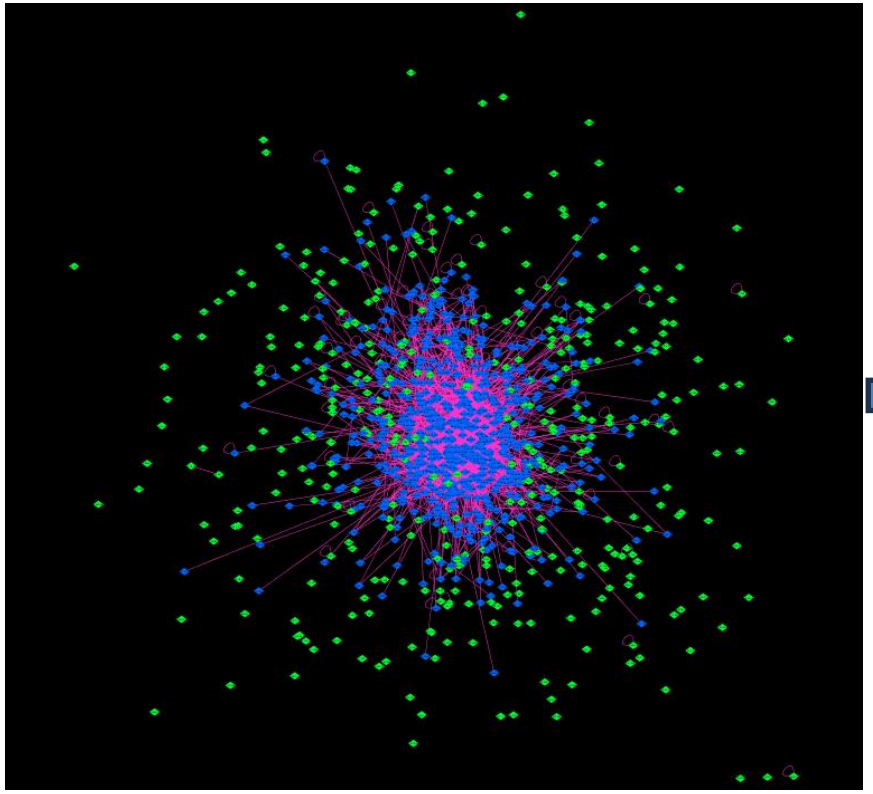
98% of the significant SNP-trait associations are based on our systematic and extended manual literature curation

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

*RHCP: Reviewed Human Complete Proteome

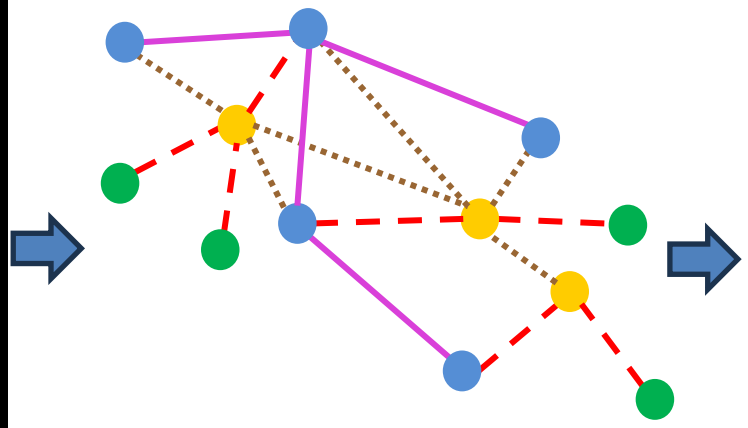
The reconstructed BP PPI Network by shortest path approach

The GWAS-deduced BP PPI network
(**1065 proteins – 1700 PPIs**)



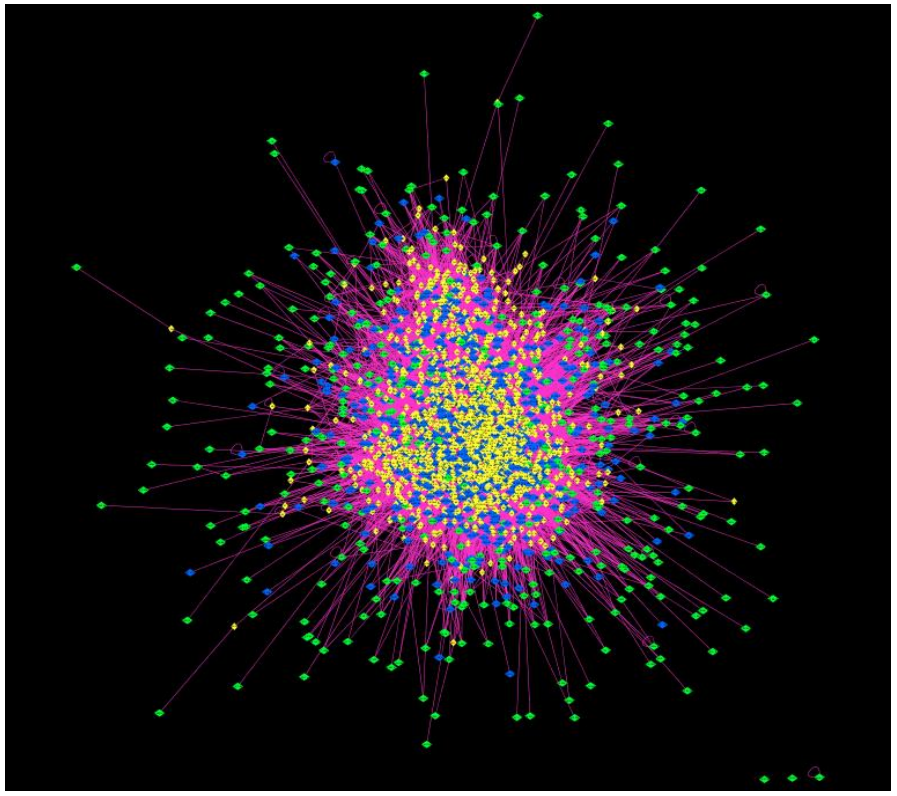
Connected GWAS-Proteins (672)
Non Connected GWAS-Proteins (393)

"Guilt by association" principle



Yellow Nodes (YNs) :
Shortest Path Intermediates
SPs between GNs & BNs

The reconstructed BP PPI network
(**2505 proteins – 31439 PPIs**)



Shortest Path Intermediates - **1443**

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

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

Integrated gene/protein prioritization method based on combined GWAS and network analysis criteria

GWAS-based

Criterion 1

p-value | #SNPs | #Pubs

Gene prioritization using a combined gene score

Network-based

Criterion 2

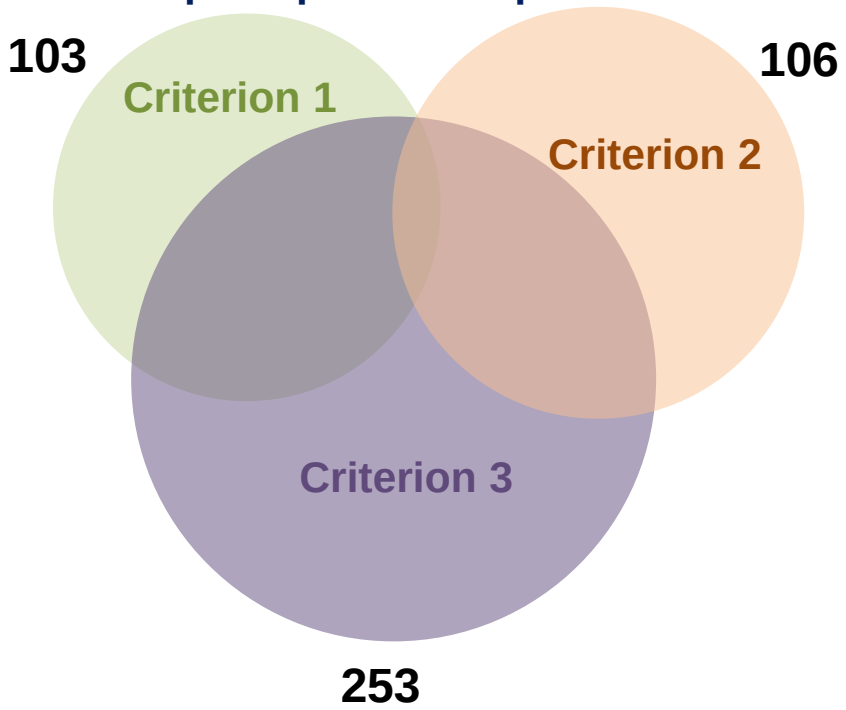
Prioritizing BP-proteins based on role in the extended PPI network

Criterion 3

Promoting GWAS-prioritized proteins and their common YN neighbors

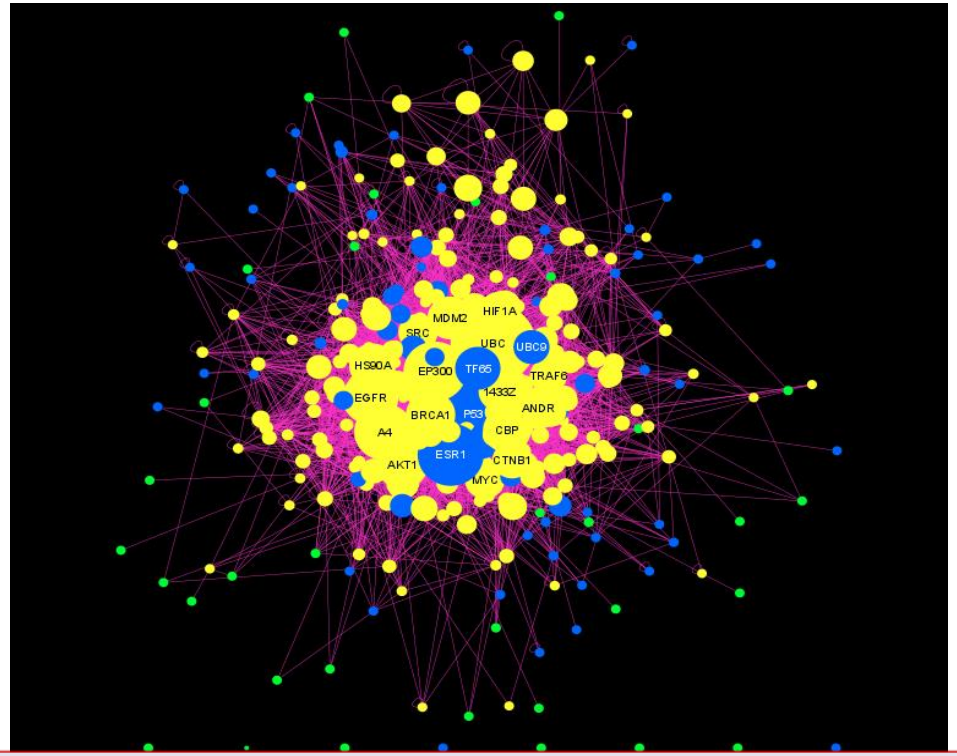
Protein prioritization using PPI network analysis metrics

Complete prioritized protein set



Union

335 proteins



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

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

Proposed Ranking of the prioritized proteins

The prioritized proteins are ranked based on the **number of satisfied prioritization criteria**

The top-10 BP prioritized proteins

Protein Entry Name	Gene Symbol	Criterion 1	Criterion 2	Criterion 3	Overall Ranking
ESR1	*ESR1*	✓	✓	✓	1
INSR	*INSR*	✓		✓	2
PTN11	*PTPN11*	✓		✓	3
CDK6	*CDK6*	✓		✓	4
CSK	*CSK*	✓		✓	5
NOS3	*NOS3*	✓		✓	6
SH2B3	*SH2B3*	✓		✓	7
ATP2B1	*ATP2B1*	✓		✓	8
FES	*FES*	✓		✓	9
FINC	*FN1*	✓		✓	10

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

- Estrogen receptor 1 (ESR1) was the only protein satisfying all three criteria
- The top-10 BP prioritized proteins are functionally supported to be associated with BP regulation mechanisms

Objectives

We aimed at investigating the BP PPI network, resulting from the GWAS data with respect to:

- **Ancestry-specific differences**, focusing on the most two abundant GWAS sub-sets (i.e. European & Asian)

- The **variant consequence severity**, excluding SNPs that are involved only in “modifier” variant categories (i.e. variants identified mainly as non-coding or associated with non-coding genes and of difficult to predict impact or unknown impact)

BP-GWAS data analysis based on ancestry

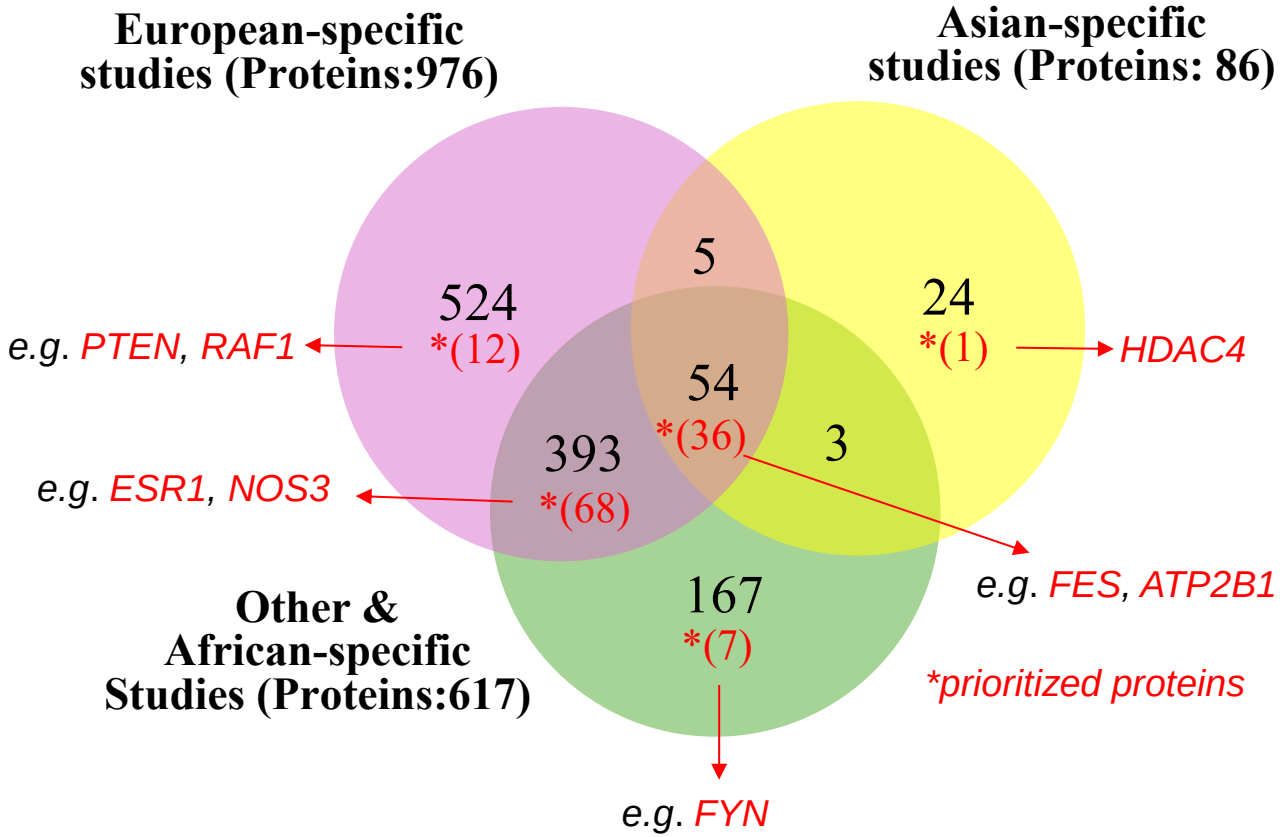
Ancestry-based BP-GWAS data statistics

(Total SNPs:6687, RHCP-SNPs: 3738, Proteins: 1170)

	Total $p<5x10^{-8}$	RHCP-protein coding $p<5x10^{-8}$	
Ancestry	#SNPs (specific)	#SNPs (specific)	#Proteins (specific)
African	86 or 1% (37)	25 (12)	23 or 2% (8)
Asian	727 or 11% (145)	373 (70)	86 or 7% (24)
European	5495 or 82% (1943)	3110 (1054)	976 or 83% (524)
Other*	4544 (951)	2592 (526)	617 (167)

*Other involves mixed-ancestry studies

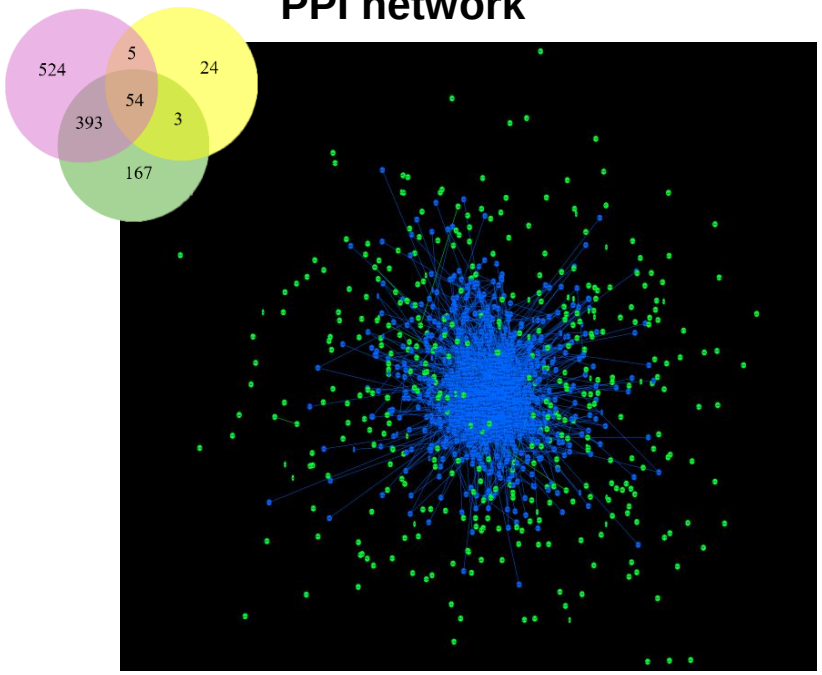
Ancestry-based Venn diagram of the BP-GWAS proteins



More GWAS on non-European ancestries are needed to validate any ancestry-specific variants or proteins, but current results highlight genetic differences in BP across ancestries.

BP GWAS-deduced PPI networks of European-specific and Asian-specific studies

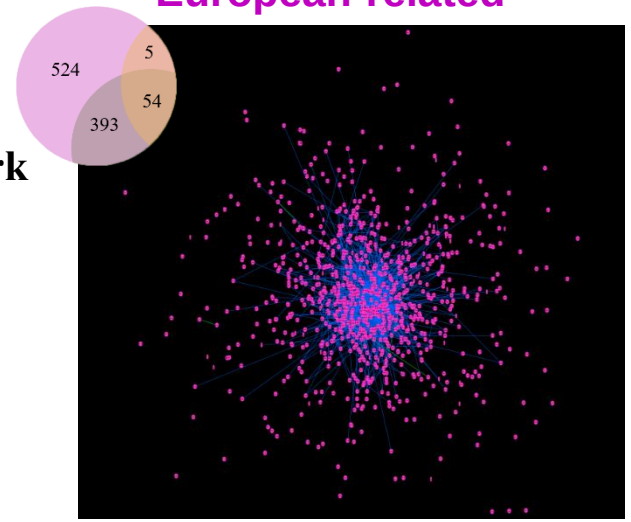
BP GWAS-deduced PPI network



BNs: 672 GNs: 393
edges: 1707

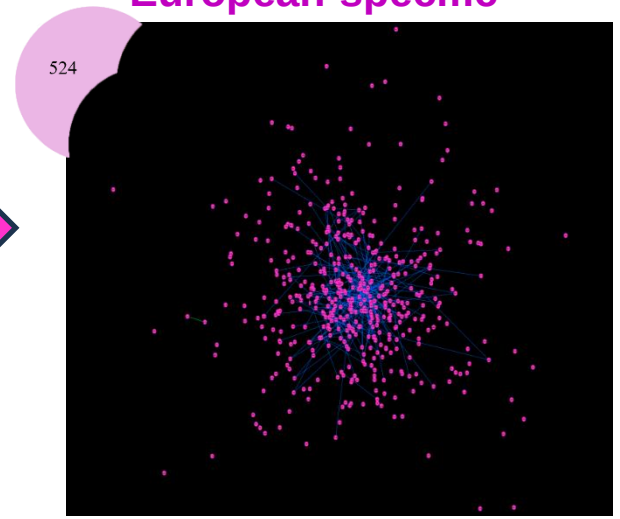
European PPI network

European-related



BNs: 561 GNs: 333
edges: 1138

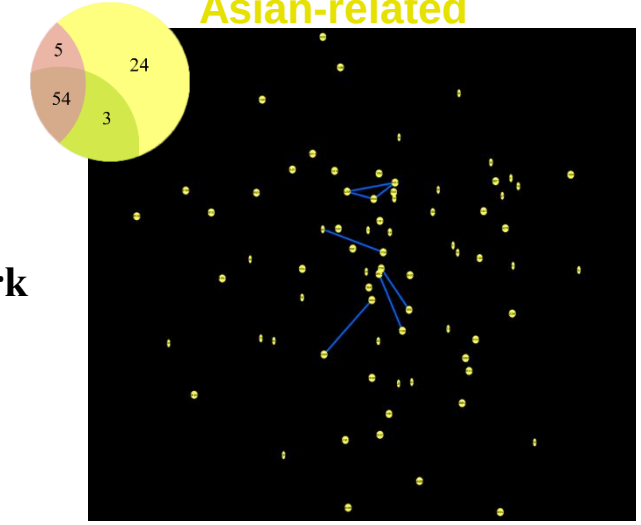
European-specific



BNs: 301 GNs: 173
edges: 293

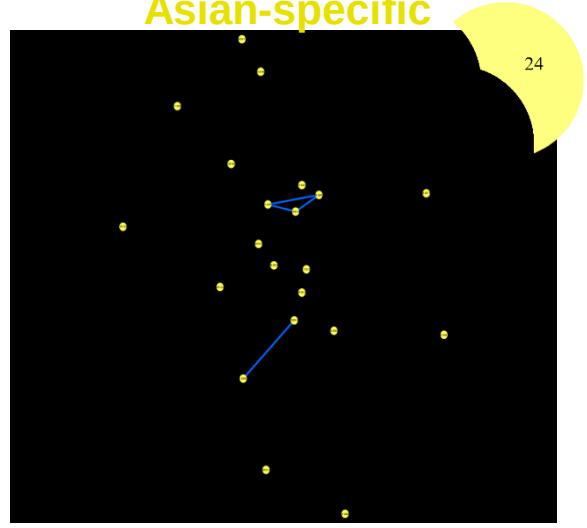
Asian PPI network

Asian-related



BNs: 44 GNs: 32
edges: 7

Asian-specific



BNs: 15 GNs: 6
edges: 4

Pathway enrichment analysis of the European-specific, Asian-specific and common datasets

- Thirty KEGG-defined pathways were significantly enriched (q<0.05) in European-specific BP GWAS-proteins
  - ➢ Renin & insulin secretion, aldosterone & cortisol synthesis and secretion, vascular smooth muscle contraction, cGMP-PKG signaling, cAMP signaling, PI3k-Akt signaling, cell-cell junctions

- No significantly enriched pathways are identified in the Asian-specific and common protein sets based on q<0.05

- By using p-value<0.05 as significance threshold
  - ➢ Eleven pathways are significantly enriched in Asian-specific BP GWAS-proteins e.g., cGMP-PKG signaling, PI3k-Akt signaling, Gap junctions
  - ➢ Five pathways are significantly enriched in common proteins, e.g., calcium signaling, aldosterone & cortisol synthesis and secretion

These results further support the need to analyze the GWAS data in the context of pathways and networks unravelling connections and related mechanisms even in relatively sparse datasets

BP-GWAS data analysis based on stricter variant consequence severity threshold

RHCP-associated BP GWAS data statistics based on severity level (Total SNPs: 3738, Total Proteins: 1170)

more severe to less severe

Variant consequence (impact)	total #SNPs up to the particular level	total #Proteins up to the particular level
stop gained (high)	5	5
inframe deletion (moderate)	6	6
missense (moderate)	137	118
splice region (low)	151	130
synonymous (low)	183	157
5 prime UTR (modifier)	224	184
3 prime UTR (modifier)	352	258
intron (modifier)	3738	1170
NMD transcript (modifier)	3738	1170

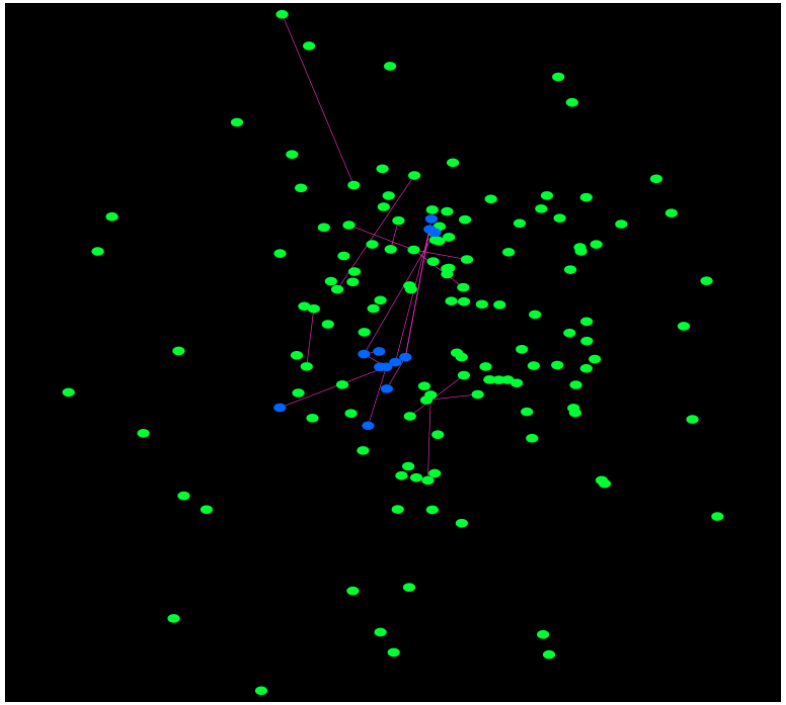
"modifier" variant categories*

* *non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact that mapped on intergenic or intronic regions*

- The vast majority of BP SNPs with RHCP-coding transcript consequences (~85%, 3168/3738) are only intron variants
- 912/1170 BP GWAS-proteins are associated with intron variants and 856/912 (73%) are considered as BP-related based on this type of SNPs only.

Reconstruction of the BP PPI network based on stricter variant consequence severity threshold

The GWAS-deduced BP PPI network
of the proteins associated with SNPs in
non-modifier consequence categories
(142 nodes, 21 edges)

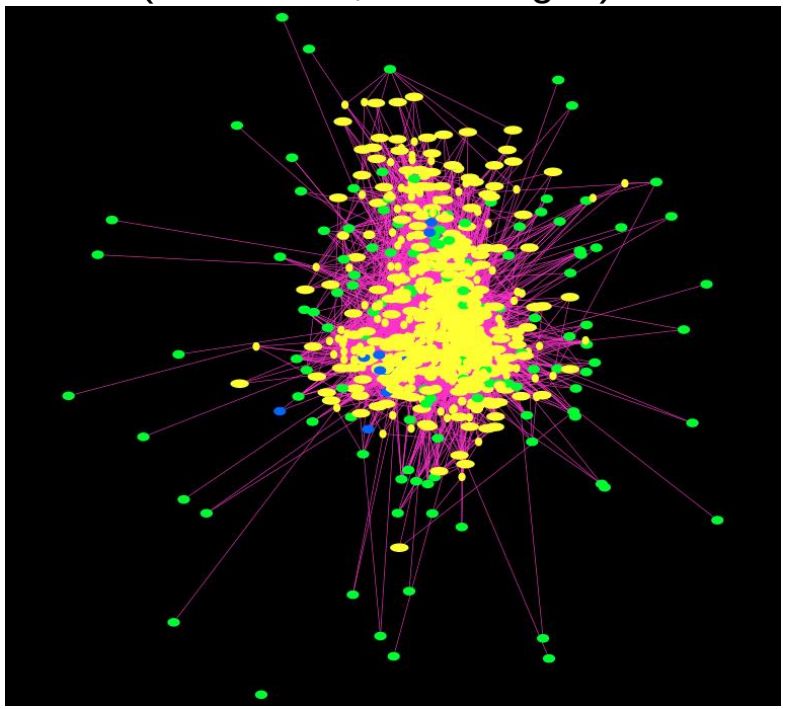


Connected GWAS-Proteins (12)
Non Connected GWAS-Proteins (130)

"Guilt by association" principle
shortest path approach

Tsare, Klapa & Moschonas,
Hum Genomics (2024)

The reconstructed BP PPI network
of the proteins associated with SNPs in
non-modifier consequence categories
(797 nodes, 9826 edges)



Shortest Path Intermediates - **655**

Pathway enrichment analysis indicated enrichment in most of the same pathways, that have been strongly associated with BP, as the full BP PPI network, presenting the same perspective despite their differences in starting proteins and size.

This result supports the need to upgrade the information content of GWAS data through network analysis

Conclusions

Based on our BP GWAS meta-database which enables the selection of GWAS-data of different ancestries and/or different variant consequences severities and the analysis of the GWAs data in the context of PPI networks:

- We identify the European- and the Asian-specific BP PPI networks, supporting the fact that most available BP-GWAS data are of European-ancestry.
- Despite its small size, the Asian-specific BP GWAS dataset pointed to some pathways, suggesting pathway-level BP (de)regulation is less dependent on dataset size.
- As more BP GWAS data from diverse ancestral background become available our meta-database can contribute to more specific studies that may lead to valuable ancestry-specific insights for BP
- We reconstructed the BP PPI network of the most impactful SNPs, which despite smaller than the full, revealed the same BP-significant pathways
- This study supports the significance of integrating genetic with functional knowledge in the context of biomolecular networks as this combined approach can diminish the impact of false positives in the involved datasets

Thank you for your attention!

Supported by:
- Hellenic Foundation for Research and Innovation (HFRI) PhD scholarship grant
- Brain Precision (TAEDR 0535850)
- ELIXIR-GR (MIS 5002780)
- EATRIS-GR (MIS 5028091)
- BBMRI-GR (5028275)
- INSPIRED (MIS 5002550)
- H2020 DIAGONAL (# 95152)
- H2020 JOINTPromise (# 874837)
- H2020 AIDPATH (#101016909)

GSRI GENERAL SECRETARIAT FOR RESEARCH AND INNOVATION

H.F.R.I. Hellenic Foundation for Research & Innovation

Brain Precision TAEDR 0535850	BBMRI MIS 5028275	ELIXIR-GR MIS 5002780	INSPIRED MIS 5002550	EATRIS-GR MIS 5028091	DIAGONAL #95152	JOINTPromise #874837	AIDPATH #101016909

Greece 2.0 NATIONAL RECOVERY AND RESILIENCE PLAN

Funded by the European Union NextGenerationEU

European Union European Regional Development Fund

HELLENIC REPUBLIC MINISTRY OF DEVELOPMENT AND INVESTMENTS SPECIAL SECRETARIAT FOR ERDF & CF PROGRAMMES MANAGING AUTHORITY OF EPAnEK

EPAnEK 2014-2020 OPERATIONAL PROGRAMME COMPETITIVENESS ENTREPRENEURSHIP INNOVATION

ΕΣΠΑ 2014-2020 ανάπτυξη - εργασία - αλληλεγγύη Partnership Agreement 2014 - 2020

Co-financed by Greece and the European Union

DIAGONAL

Joint Promise

AIDPATH

European Commission

Horizon 2020 European Union funding for Research & Innovation