Integrated Analysis of Genomic and Genome-Wide Association Studies Identified Candidate Genes for Nutrigenetic Studies in Flavonoids and Vascular Health: Path to Precision Nutrition for (Poly)phenols

Tatjana Ruskovska[1], Filip Postolov[1] and Dragan Milenkovic[2]

[1]Faculty of Medical Sciences, Goce Delcev University, 2000 Stip, North Macedonia

[2]Plants for Human Health Institute, Department of Food Bioprocessing & Nutrition Sciences, NC State University, Kannapolis, NC 28081, USA

UNIVERSITY GOCE DELCEV

NC STATE UNIVERSITY

PURPOSE

Flavonoids exert vasculoprotective effects in humans, but interindividual variability in their action has also been reported.
This study aims to identify genes that are associated with vascular health effects of flavonoids and whose polymorphisms could explain interindividual variability in response to their intake.

METHODS

1. **Systematic literature search** was conducted to identify human intervention studies with flavonoids that analysed global gene expression in peripheral blood cells via microarray analysis and demonstrated **at least one positive effect on vascular function**. Differentially expressed genes (**DEGs**) were extracted from these studies. Upstream Regulators (**URs**) for each set of DEGs were obtained through **Qiagen-IPA tool.**
2. **Genes associated with vascular dysfunction** were independently identified from genome-wide association studies (**GWAS**), focusing on traits such as **hypertension**, **atherosclerosis**, and **arterial stiffness**.
3. **Integrative Analysis** using ***InteractiVenn*** was then performed by comparing DEGs from intervention studies with GWAS-derived genes to identify **overlapping candidate genes relevant to flavonoid-mediated vascular effects**.
4. **Pathway Enrichment Analysis (Gene Trail 3.2.)** and **Functional Analysis** we aimed to select **top priority candidate genes**. For these, we used the National Center for Biotechnology Information (NCBI) ***Variation Viewer*** to identify high-frequency variants and queried ***PharmGKB*** for pharmacogenomic relevance.

RESULTS

By extracting data from the 5 eligible human intervention studies with flavonoids, we obtained **5807 differentially expressed genes (DEGs)**. The number of identified upstream regulators (**URs**) varied across the studies, from **227 to 1407**. The search of the **GWAS Catalog revealed 493 genes associated with vascular dysfunction**. An **integrative analysis** of the transcriptomic data with GWAS genes identified **106 *candidate DEGs* and 42 *candidate URs***, while subsequent **functional analyses** and a search of the literature identified **20 top priority candidate genes**:

ALDH2, **APOE**, **CAPZA1**, **CYP11B2**, **GNA13**, **IL6**, **IRF5**, **LDLR**, **LPL**, **LSP1**, **MKNK1**, **MMP3**, **MTHFR**, **MYO6**, **NCR3**, **PPARG**, **SARM1**, **TCF20**, **TCF7L2**, and **TNF**.

Through the **Variation Viewer database** we identified their **variant types, molecular consequences, most severe clinical significances, and top 10 genetic variants with highest frequencies**, while PharmGKB revealed multiple variants that have already been shown to have specific **pharmacologic relevance**.

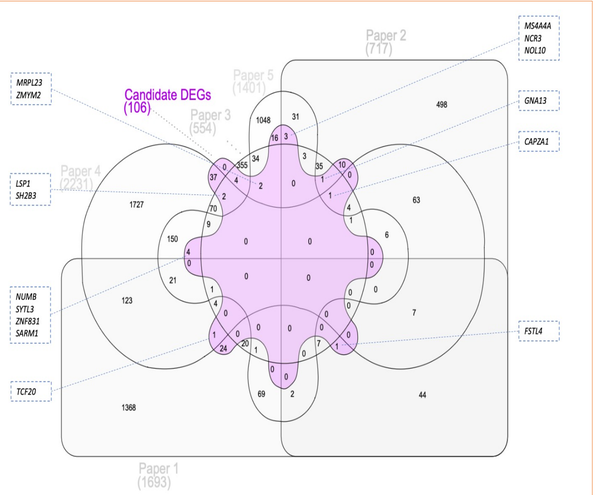
5807 DEGs
227-1407 URs
IDENTIFIED THROUGH SYSTEMATIC LITERATURE SEARCH

493
GWAS GENES ASSOCIATED WITH VASCULAR DYSFUNCTION
HYPERTENSION
ATHEROSCLEROSIS
ARTERIAL STIFFNESS

106
CANDIDATE DEGs
42
CANDIDATE URs

20
TOP PRIORITY CANDIDATE GENES

PAPER	STUDY POPUALTION	BIOACTIVES	OUTCOMES
1	Healthy, middle-aged, moderately overweight men	Hesperidin	↓ DBP
2	Non-obese, healthy male smokers, smoking 10 and more cigarettes per day for at least 5 years	Monomeric and oligomeric flavanols from grape seeds	Improved Vascular Health Index
3	Healthy male volunteers	Wild blueberry anthocyanins	↑ flow-mediated vasodilatation ↓ 24 h SBP
4	Healthy middle-aged men	Cocoa flavanols	↑ Flow-Mediated Vasodilatation ↓ SBP and DBP ↓ Pulse Wave Velocity
5	Healthy, non-smoking women, 3 to 10 years after menopause	Grapefruit juice flavanones	↓ Carotid–Femoral Pulse Wave Velocity



Canonical Pathways	Number of Hits	Genes
Pathways directly involved in vascular dysfunction		
VEGFA-VEGFR2 Signaling Pathway *	6	*CSK, MKNK1, MYO6, PTPRJ, SMARCA2, TNXB*
Regulation of actin cytoskeleton **	4	*BRK1, CSK, GNA13, SOS2*
Adherens junction	3	*PTPRJ, TCF7L2, YES1*
Angiopoietin Like Protein 8 Regulatory Pathway *	3	*LPL, PRKAG2, SOS2*
ECM-receptor interaction	3	*NPNT, TNXB, VTN*
Focal adhesion	3	*SOS2, TNXB, VTN*
Apelin signaling pathway	2	*GNA13, PRKAG2*
Cholesterol metabolism	2	*LDLR, LPL*
Composition of Lipid Particles *	2	*LDLR, LPL*
Fluid shear stress and atherosclerosis	2	*BMPR1B, GSTA4*
Glycerolipid metabolism	2	*ALDH2, LPL*
Metabolic pathway of LDL, HDL and TG, including diseases *	2	*LDLR, LPL*
Platelet activation	2	*GNA13, LYN*
Statin Pathway *	2	*LDLR, LPL*
Pathways involved in inflammation		
Chemokine signaling pathway **	3	*CSK, LYN, SOS2*
Cytokine-cytokine receptor interaction	3	*BMPR1B, GDF10, LTB*
NF-kappa B signaling pathway **	3	*LTB, LYN, TAB2*
Regulation of toll-like receptor signaling pathway *	3	*IRF5, SARM1, TAB2*
B cell receptor signaling pathway	2	*LYN, SOS2*
Interleukin-11 Signaling Pathway *	2	*FES, YES1*
Natural killer cell mediated cytotoxicity	2	*NCR3, SOS2*
Structural Pathway of Interleukin 1 (IL-1) *	2	*MKNK1, TAB2*
TNF signaling pathway	2	*DAB2IP, TAB2*
Toll-like receptor signaling pathway **	2	*IRF5, TAB2*
Cell signaling pathways		
MAPK signaling pathway **	4	*MKNK1, SOS2, STK3, TAB2*
PI3K-Akt signaling pathway **	4	*MCL1, SOS2, TNXB, VTN*
EGF/EGFR Signaling Pathway *	3	*CSK, SOS2, TWIST1*
Insulin signaling pathway	3	*MKNK1, PRKAG2, SOS2*
Sterol Regulatory Element-Binding Proteins (SREBP) signalling *	3	*LDLR, LPL, PRKAG2*
cGMP-PKG signaling pathway	2	*ATP2B1, GNA13*
FoxO signaling pathway	2	*PRKAG2, SOS2*
Jak-STAT signaling pathway	2	*MCL1, SOS2*
Phospholipase D signaling pathway	2	*GNA13, SOS2*
Wnt Signaling Pathway and Pluripotency *	2	*LDLR, TCF7L2*
Antioxidant protection		
NRF2 pathway *	2	*GSTA4, SLC39A8*

Legend: No asterisk—KEGG pathways; * WikiPathways; ** both KEGG and WikiPathways.

Gene	Variant Type (Number)	Molecular Consequences (Number)	Most Severe Clinical Significance (Number)	Number of Associations with Pharmacological Significance	10 Variants with Highest Frequencies				
					Variant ID	Molecular Consequences	Alleles	Alleles with Highest Frequencies	Frequency
ALDH2	single nucleotide variant (172) deletion (3) insertion (2) indel (22)	missense variant (1) intron variant (131) 3 prime UTR variant (27) 500 B downstream variant (1) 2 KB upstream variant (14)	pathogenic (1)	9	rs2296651 rs6489793 rs2106497 rs10774638 rs886205 rs4767939 rs10774637 rs9971942 rs10774639 rs11066028	intron variant 3 prime UTR variant nct specified intron variant 2 KB upstream variant intron variant intron variant nct specified nct specified intron variant	C,A,G T,G T,A,C,G T,A,C A,C,G,T A,G C,G,T C,T G,A,C,T A,C,G,T	C T T T A G T T A A	0.497804 0.497604 0.496406 0.492612 0.491214 0.419728 0.419129 0.410743 0.410743 0.372804
APOE	single nucleotide variant (8)	missense variant (3) synonymous variant (1) intron variant (4) 5 prime UTR variant (1) 500 B downstream variant (1) 2 KB upstream variant (1)	pathogenic (2) drug-response (2)	33	rs405509 rs440446 rs769450 rs429358 rs7412 rs769449 rs1081105 rs877973	2 KB upstream variant missense variant, intron variant, synonymous variant intron variant intron variant missense variant missense variant intron variant 500 B downstream variant 5 prime UTR variant, intron variant	T,G C,G,T G,A T,C C,T G,A A,C,G C,A,T	T C A C T A C A	0.471845 0.373802 0.327276 0.150709 0.0750799 0.0648962 0.0301518 0.0159744
CAPZA1	single nucleotide variant (136) deletion (1) insertion (1) indel (26)	missense variant (1) synonymous variant (2) intron variant (149) nc transcript variant (1) 5 prime UTR variant (3) 3 prime UTR variant (1) 500 B downstream variant (2) 2 KB upstream variant (22)	/	/	rs3013440 rs7524494 rs7413820 rs3110450 rs2925236 rs3013439 rs12246329 rs12246466 rs12246208 rs9429486	intron variant intron variant intron variant intron variant intron variant intron variant intron variant intron variant intron variant intron variant	G,A A,G,T G,A G,A,T G,A,T T,A,G T,A,C,G T,A,C C,T T,A,C,G	G A G G G T T T T T	0.478435 0.478435 0.478435 0.478435 0.478435 0.477935 0.477935 0.477935 0.436502 0.435703
CYP11B2	single nucleotide variant (86) insertion (2) indel (3)	missense variant (9) synonymous variant (6) intron variant (66) 3 prime UTR variant (7) 500 B downstream variant (2) 2 KB upstream variant (5)	pathogenic (1) association (1) benign (17)	3	rs10110732 rs28615142 rs28566703 rs13263682 rs6421 rs79201878 rs80062072 rs74838461 rs28594050 rs6429	intron variant intron variant intron variant intron variant intron variant intron variant intron variant intron variant intron variant intron variant	C,T T,C A,G A,C T,C A,T A,G T,C C,T G,A,C,T	C C G C C T G C T C	0.495208 0.483227 0.482628 0.479433 0.460064 0.45607 0.45607 0.45607 0.445967 0.44389

CONCLUSION

In conclusion, we performed an integrated bioinformatic analysis with large-scale GWAS and transcriptomic data to generate a refined list of candidate causal genes for interindividual variability in response to flavonoid intake. These results should serve as an important resource, facilitating the focusing of nutrigenetic research in the field of plant food bioactives to identify gene variants associated with a better health response to these bioactives, and therefore, build a foundation for precision nutrition research in the field of (poly)phenols.

Abbreviations: DEGs = differentially expressed genes**; URs =** upstream regulators**; GWAS =** genome-wide association studies.

DOI:10.3390/nu16091362

Declaration of Interest: The authors declare no conflicts of interest.