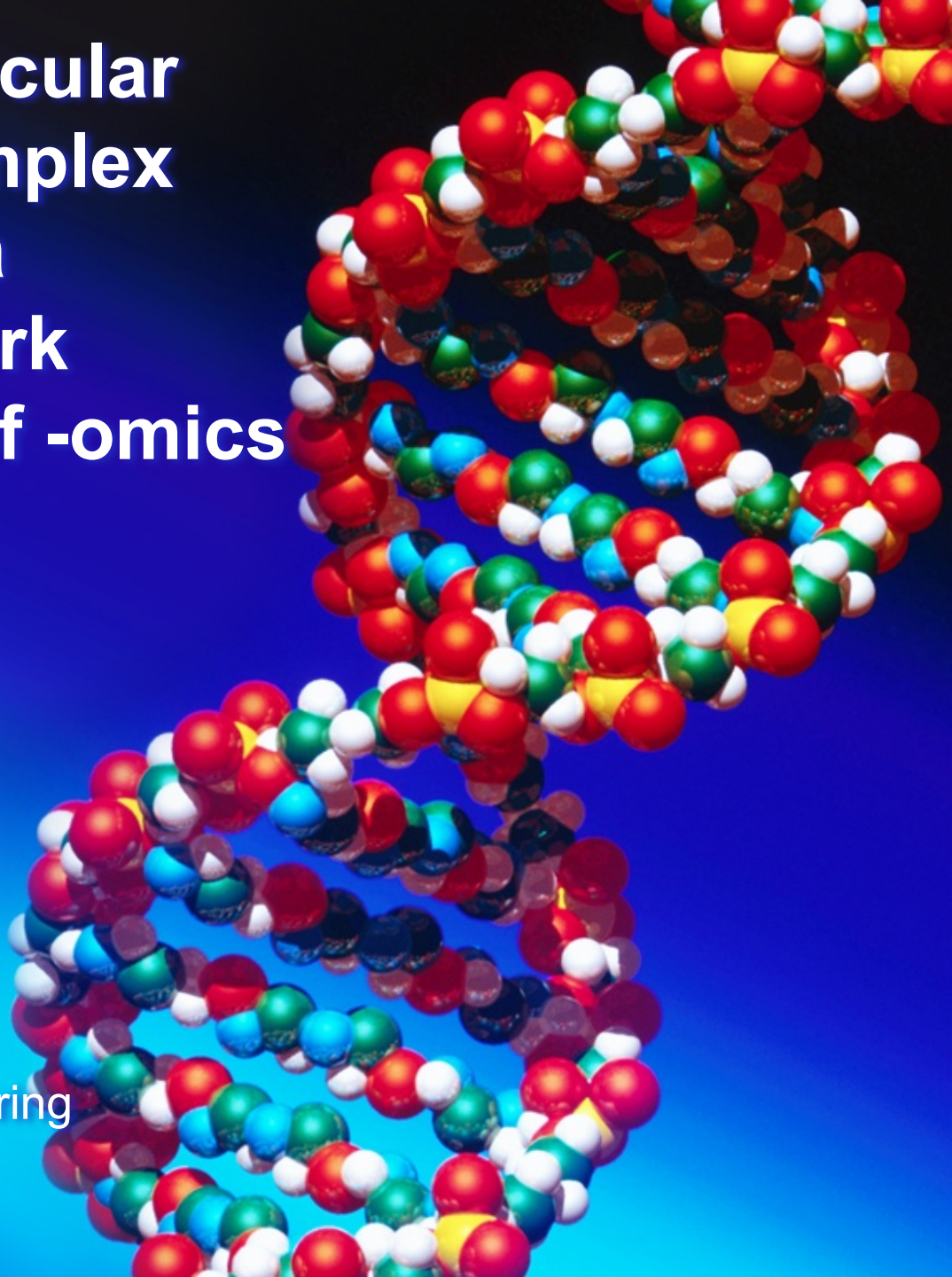
Dissecting the Molecular Mechanisms of Complex Diseases Through a Pathway and Network Oriented Analysis of -omics Data



Burcu Bakır-Güngör, Ph.D.

Assistant Professor
Department of Computer Engineering
Abdullah Gül University
Email: burcu.gungor@agu.edu.tr

Burcu Bakır-Güngör, Ph.D.

AGÜ

Asisst. Prof.	Abdullah Gül University	Dept. of Computer Engineering
Senior Researcher	TÜBİTAK, BİLGEM, UEKAE	Advanced Genomics and Bioinformatics Research Center
Asisst. Prof.	Bahçeşehir University	Dept. of Computer Engineering, Dept. of Genetics & Bioinformatics
Senior Researcher	Medical College of Wisconsin, Human and Molecular Genetics Center	Bioinformatics Research Center, Rat Genome Database (P.I. Howard Jacob)
PhD	Georgia Institute of Technology, Sabancı University	Bioinformatics (P.I. Mark Borodovsky, Ugur Sezerman)
Masters degree	Georgia Institute of Technology	Bioinformatics (P.I. Mark Borodovsky)
Research Intern	The State University of NewYork, University at Buffalo	The Buffalo Center of Excellence in Bioinformatics (P.I. Jeff Skolnick)
Undergraduate	Sabancı University	Biological Sciences and Bioengineering

Georgia Tech

MEDICAL COLLEGE OF WISCONSIN
knowledge changing life

RGD

University at Buffalo

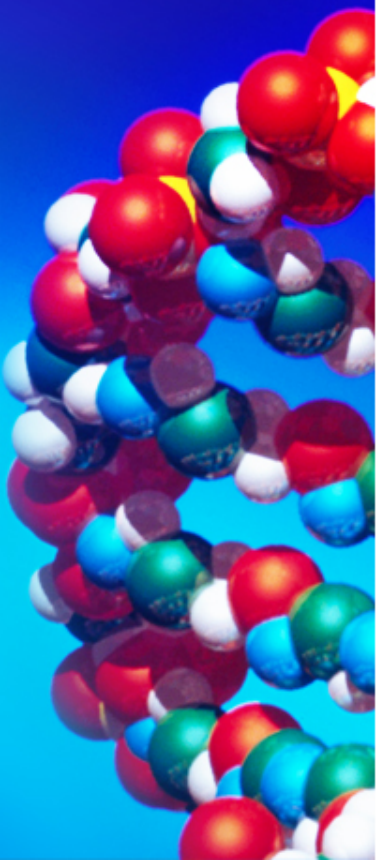
HMGC
Human and Molecular Genetics Center
Leading Personalized Healthcare

BAU
Bahçeşehir University

IGBAM
Advanced Genomics and Bioinformatics Research Center

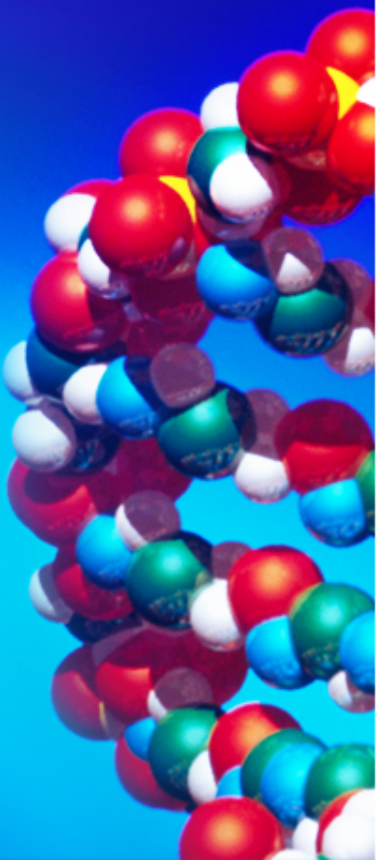
TÜBİTAK BİLGEM

Sabancı Üniversitesi



Outline

- Background
  - Genome-wide association studies (GWAS)
  - Difficulties in mining GWAS data
  - Motivation for pathway and network oriented analysis of GWAS
- Pathway and Network Oriented GWAS Analysis (PANOGA)
- Results and Discussions on:
  - Rheumatoid Arthritis, Epilepsy, Intracranial Aneurysm, Behcet's disease dataset
- Integrative analysis of transcriptomics and epigenomics data using PANOGA
- Integrative -omics Data Analysis
- Ongoing Research



Background

- The main goal of human genetics is to understand the inherited basis of human variation in phenotypes, elucidating human physiology and disease.
- Extensive studies are currently being performed to associate disease susceptibility with genetic variations, such as single nucleotide polymorphisms (SNPs).
- Genome-wide association studies (GWAS) emerged as a major tool to identify disease susceptibility loci and have been successful in detecting the association of a number of SNPs with complex diseases.

Genome-wide association studies (GWAS)

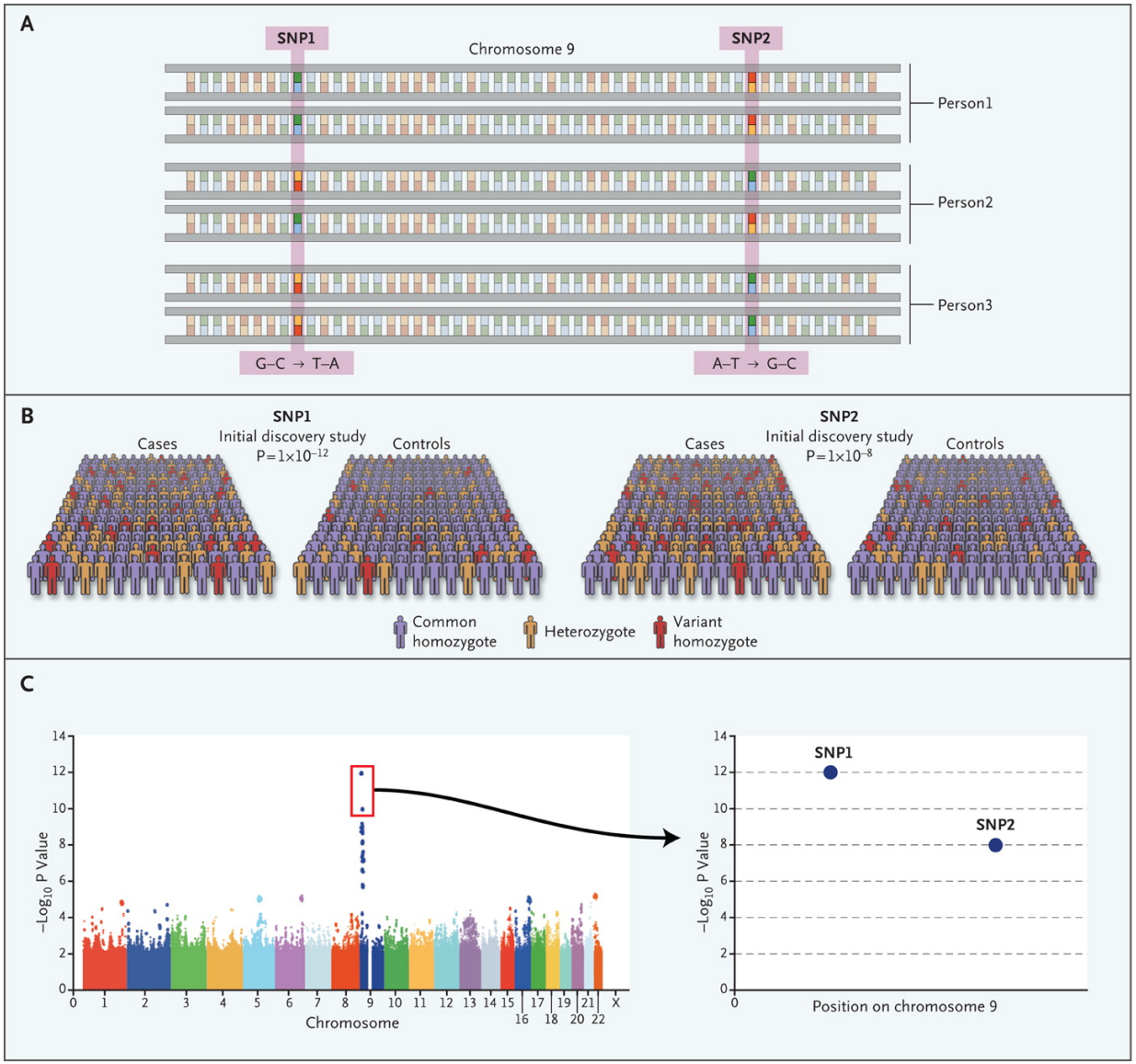
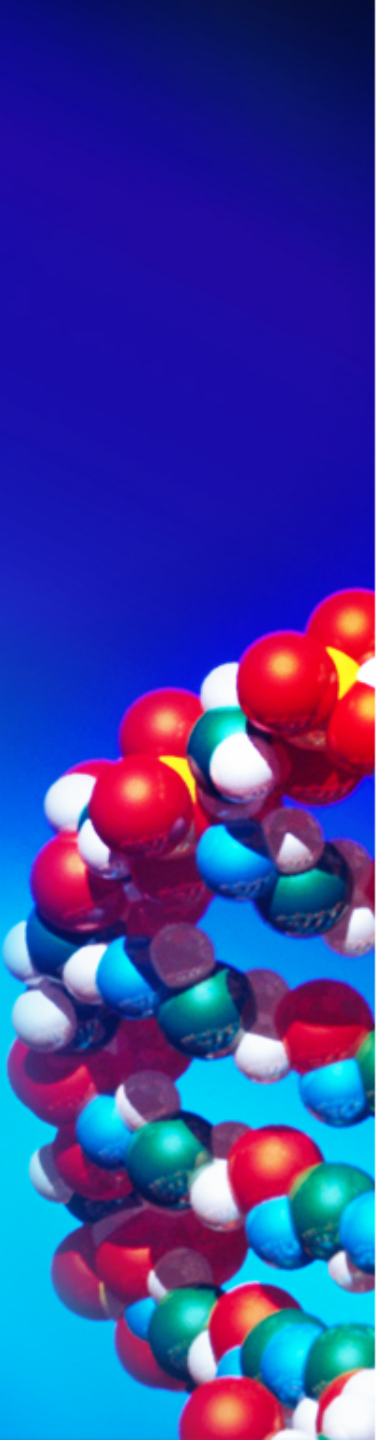


Figure 1. Genome-wide association studies (GWAS)
Excerpted from Genomewide Association Studies and Assessment of the Risk of Disease, Manolio TA. N Engl J Med 2010;363:166-176.

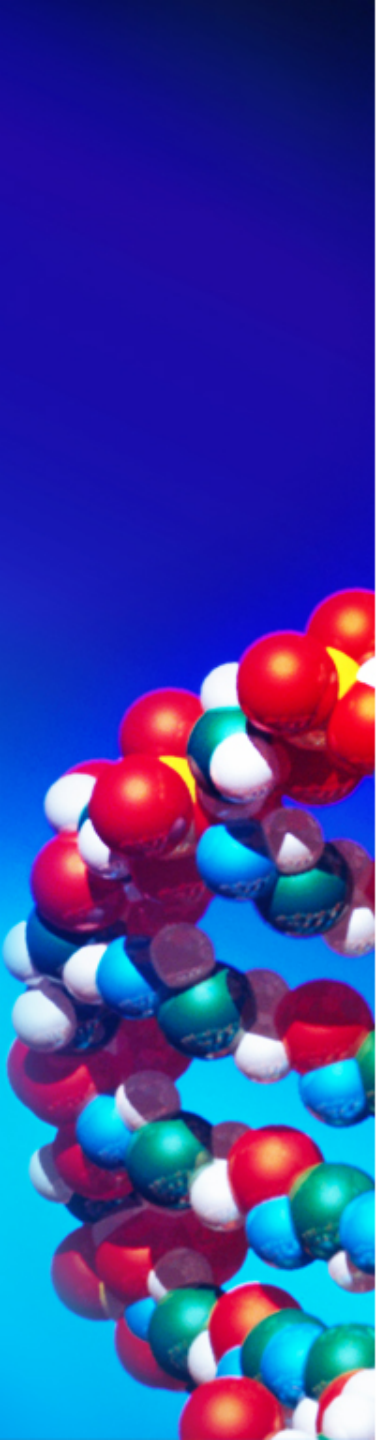


Motivation for pathway and network oriented analysis of GWAS

- Testing only for association of a single SNP is insufficient to dissect the complex genetic structure of common diseases.

- Extracting biological insight from GWAS and understanding the principles underlying the complex phenomena that take place on various biological pathways remain a major challenge.

- Recent studies have shown that the full potential of GWAS can only be achieved by integrating pathway and network based analysis.



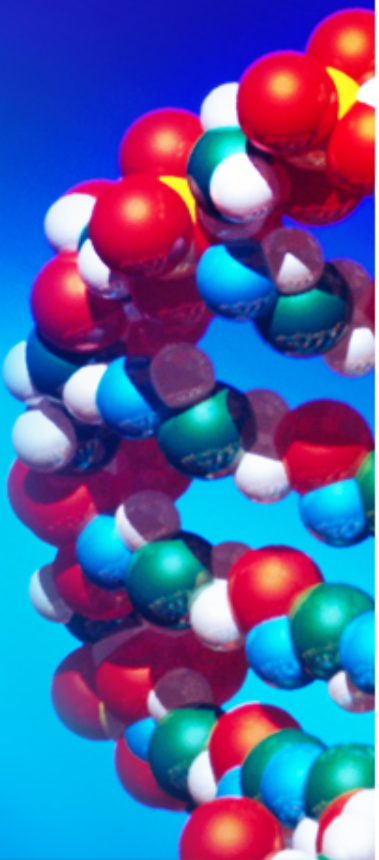
Pathway and Network Oriented GWAS Analysis (PANOGA)

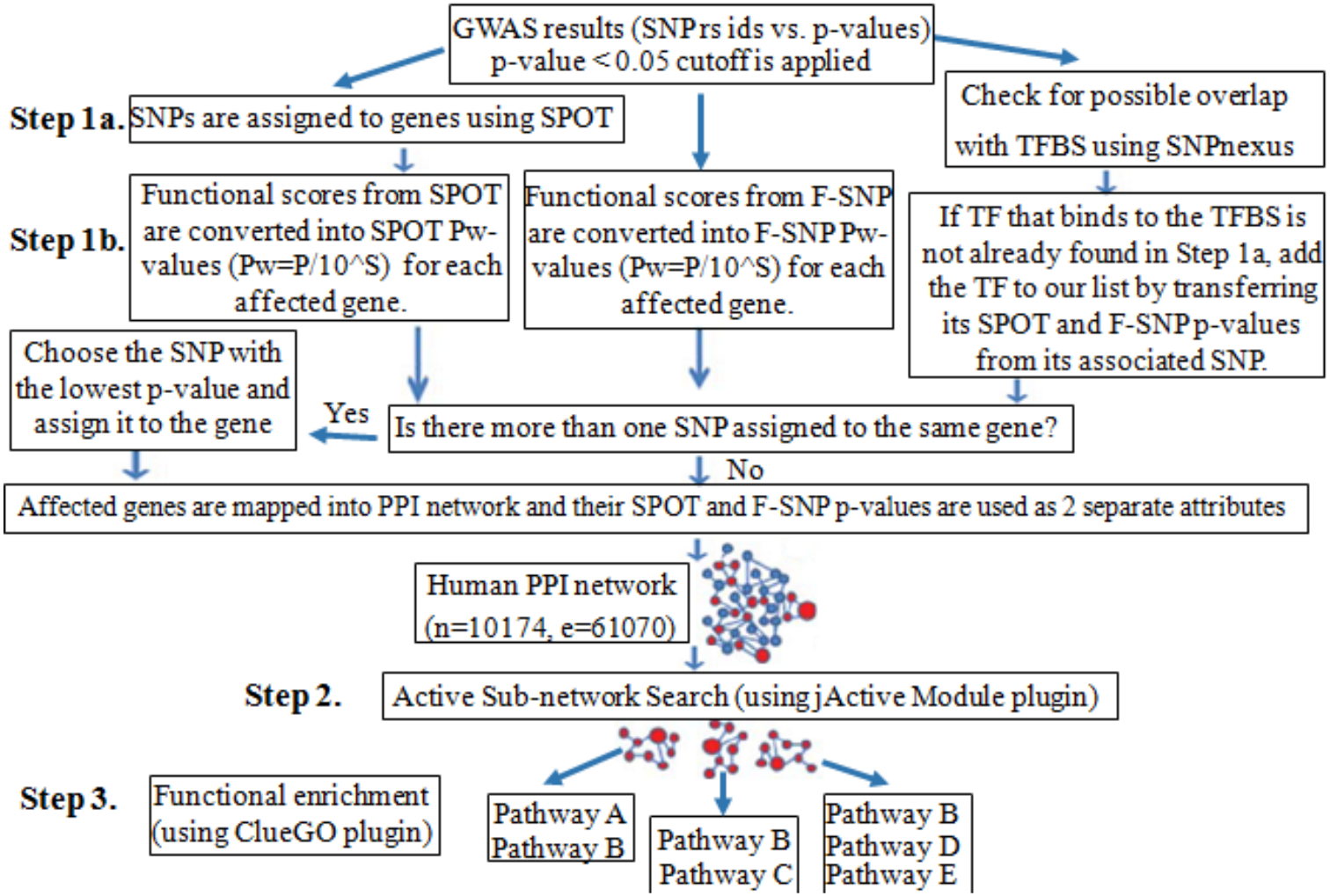
- Developed a novel methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context.

- In this methodology, we incorporated SNP functional properties, protein-protein interaction networks, pathway classification tools into GWAS.

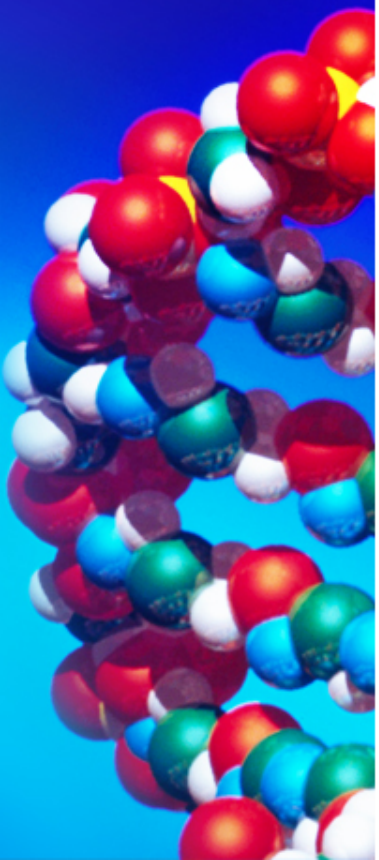
- Hence, leading molecular pathways, which cannot be picked up using traditional analyses were identified.

Our Methodology (PANOGA)





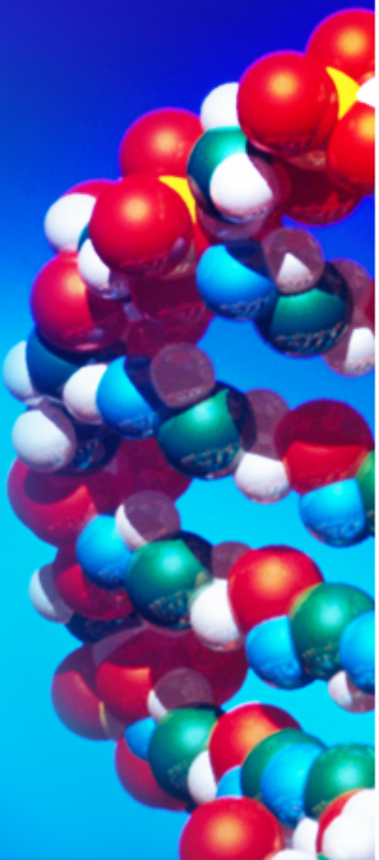
B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.



SNP Functionalization

Functional Category	Tool	Description	Meta-tool
Protein Coding	LS-SNP, SNPs3D, SIFT, SNPeffect	SNP annotation tool, Impact of nsSNPs on protein function, Prediction of amino acid substitution effects, SNP annotation with human disease	F-SNP
Protein Coding	PolyPhen	Prediction of amino acid substitution effects	SPOT, F-SNP
Protein Coding, Splicing Regulation, Transcriptional Regulation	Ensembl	Extensive genomic database including SNPs and gene transcripts	F-SNP
Splicing Regulation	ESEfinder, ESRSearch, PESX, RescueESE	Exonic splice sites, Exonic-splicing regulatory (ESR) sequences, Exon splicing enhancers/silencers, Exonic splice sites	F-SNP
Transcriptional Regulation	Consite TFSearch	Conserved transcription factor binding sites, Transcription factor binding sites	F-SNP
Transcriptional Regulation	SNPnexus	Conserved transcription factor binding sites	SNPnexus
Transcriptional Regulation, Conserved Region	GoldenPath	MicroRNA, cpgIslands, evolutionary conserved regions	F-SNP
Conserved Region	ECRBase	Evolutionary conserved regions	SPOT
Post-translation	KinasePhos, OGPET, Sulfinator	Phosphorylation sites, Prediction of O-glycosylation sites in proteins, Tyrosine sulfination sites	F-SNP
Genomic Coordinates	dbSNP	General SNP/gene transcript properties	SPOT
Genomic Coordinates	UCSC	Extensive genomic database including SNPs and gene transcripts	F-SNP
LD estimation	HapMap, Haploview	Dense genotyping on multiple populations, useful for LD estimates Estimation of r^2 LD coefficients for each population	SPOT

B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.



Our Methodology (ctd.)

- SNP-wise weighted p-value calculation:
  - Combines functional, genomic information of a SNP with genotypic p-values of association for each tested SNP ($P_w = P/10^{FS}$)
- Assigning SNPs to genes:
  - Considering all known SNP/gene transcript associations, the gene with the highest priority is chosen.
- Active sub-network searches
  - By using the gene-wise weighted p-values, active sub-networks are found in the human PPI network.
- Functional Enrichment of Sub-networks
  - Evaluate whether the identified sub-networks are biologically meaningful.

Rheumatoid Arthritis (RA)

- A chronic, systemic inflammatory disorder that usually affects joints.
- About 1% of the world's population is affected by RA.
- Known variants explain ~20% of the genetic burden of RA.
- Additional variations remain to be discovered.

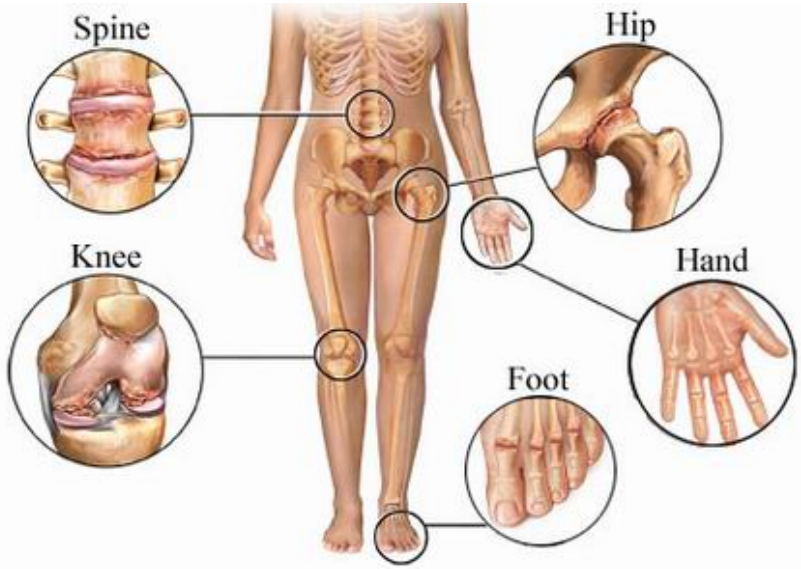


Figure 3. Rheumatoid Arthritis

http://rheumatoidarthritismedications.com/Rheumatoid-Arthritis.jpg

Rheumatoid Arthritis (RA) dataset

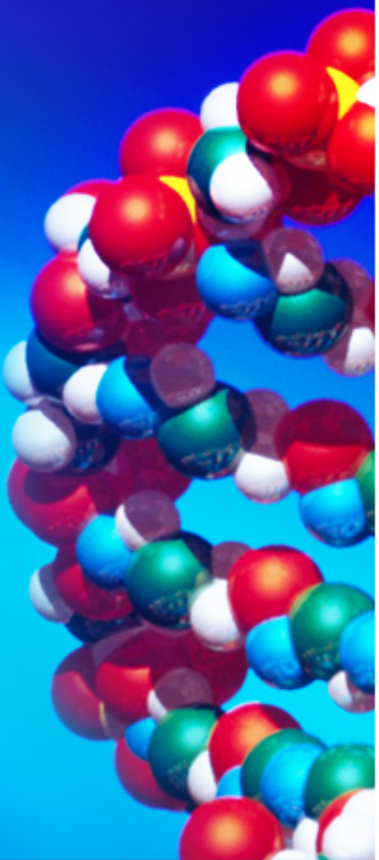
- Wellcome Trust Case Control Consortium (WTCCC) dataset

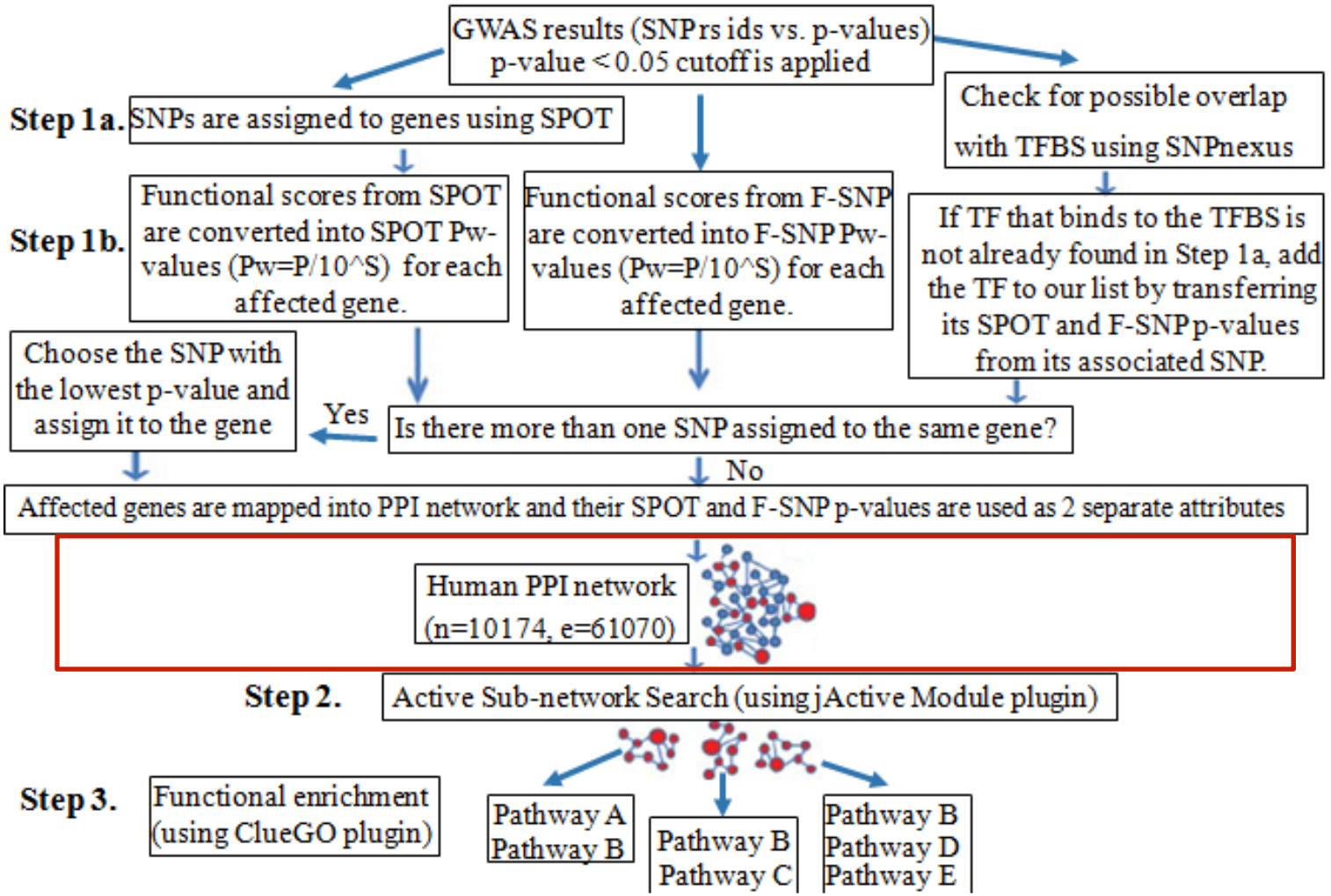
# of Cases	# of Controls	# of genotyped SNPs	Platform
1,999	3,004	500,475	Affymetrix GeneChip Human Mapping 500 K Array Set

Table 1. Summary of Rheumatoid Arthritis (RA)dataset.

- 25,027 SNPs were included with $P<0.05$.

Our Methodology (PANOGA)





B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.

Mapping to Human PPI network

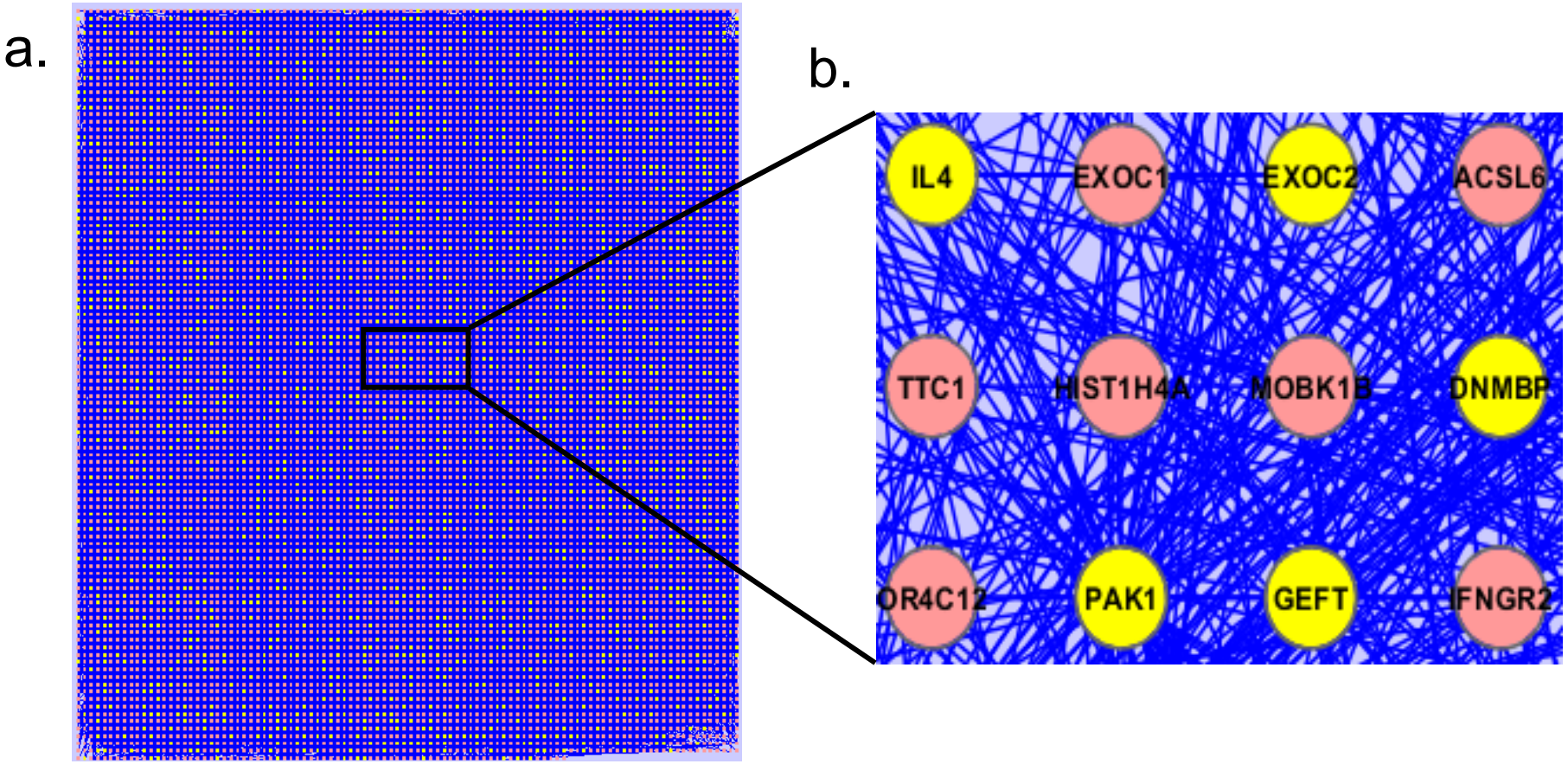
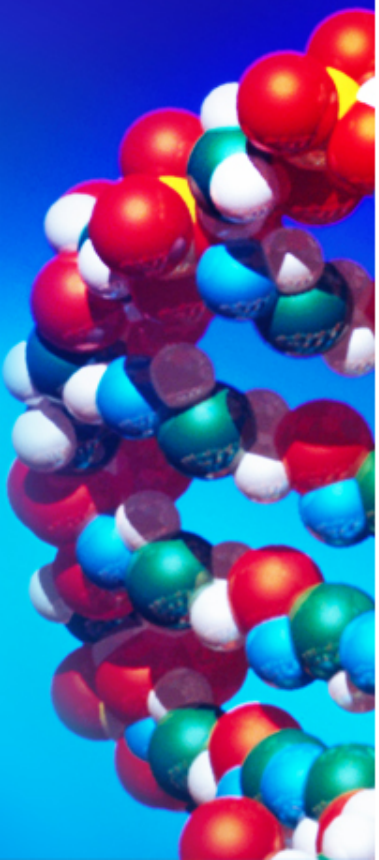
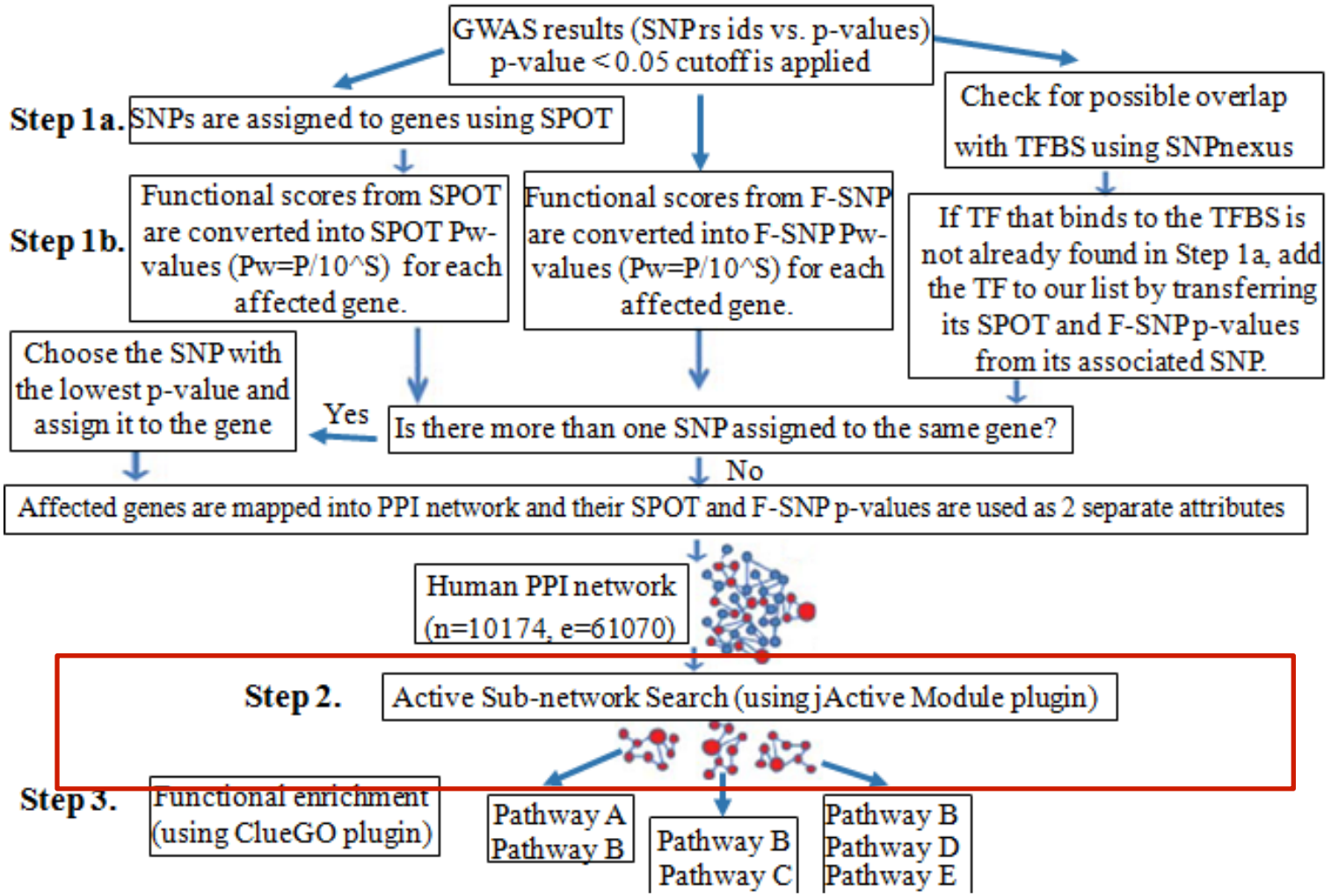


Figure 2. a. Human PPI network including 10,174 nodes and 61,070 edges. **b.** Zoomed in view of the human PPI network. 25,027 SNPs from RA GWAS dataset are assigned to 4,094 genes (shown in yellow).

B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.

Our Methodology (PANOGA)





B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.

Identified Subnetworks

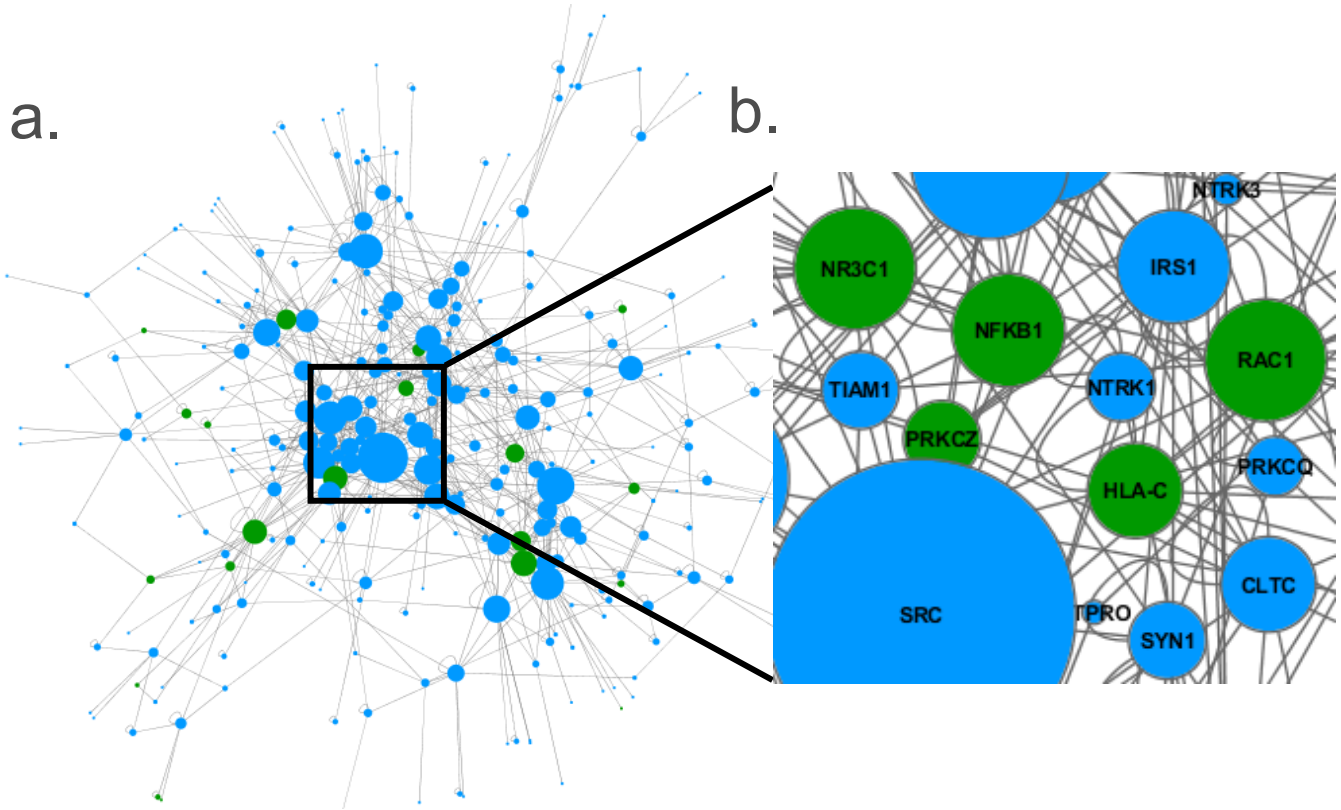
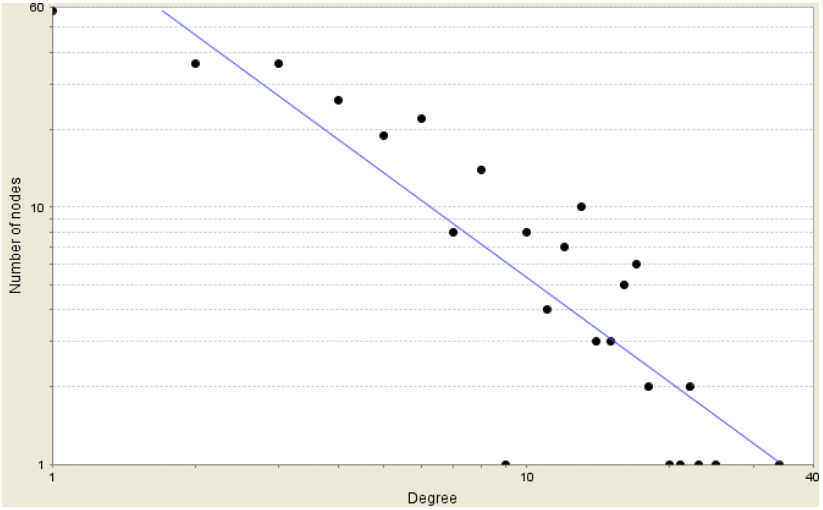


Figure 3. **a.** The highest scoring sub-network is composed of 275 nodes and 778 edges (as found in active sub-network search step). Node size is shown as proportional to the degree of a node. **b.** 20 genes known in literature as associated with RA are shown in green.

B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.

Identified vs. Random Subnetworks

a.



b.

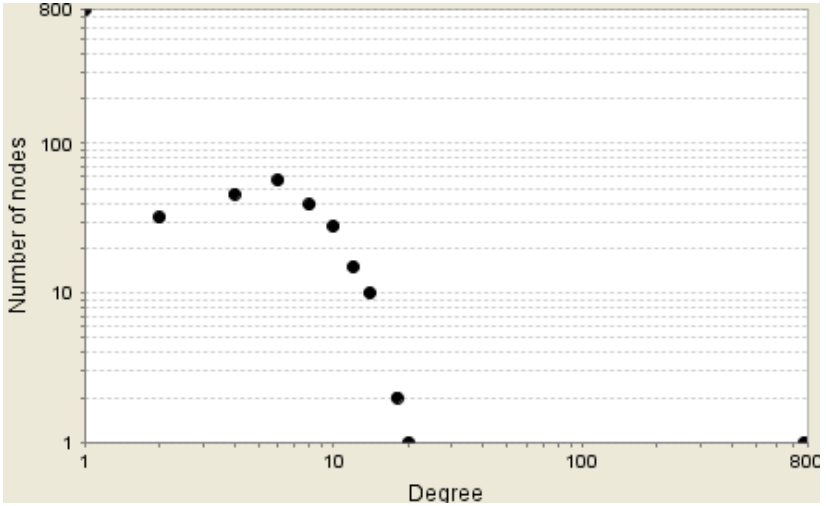
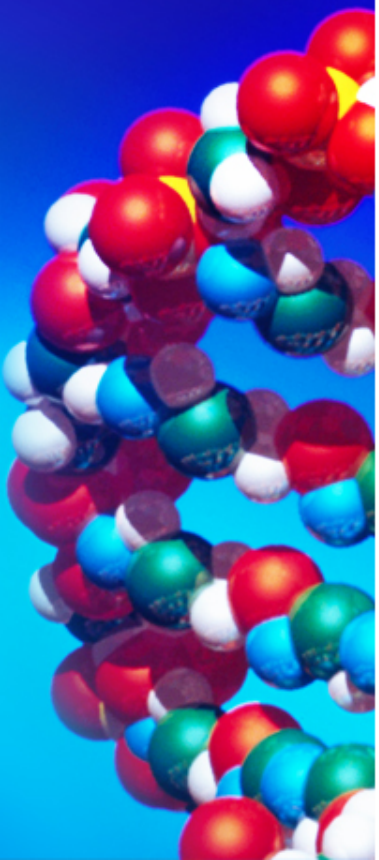
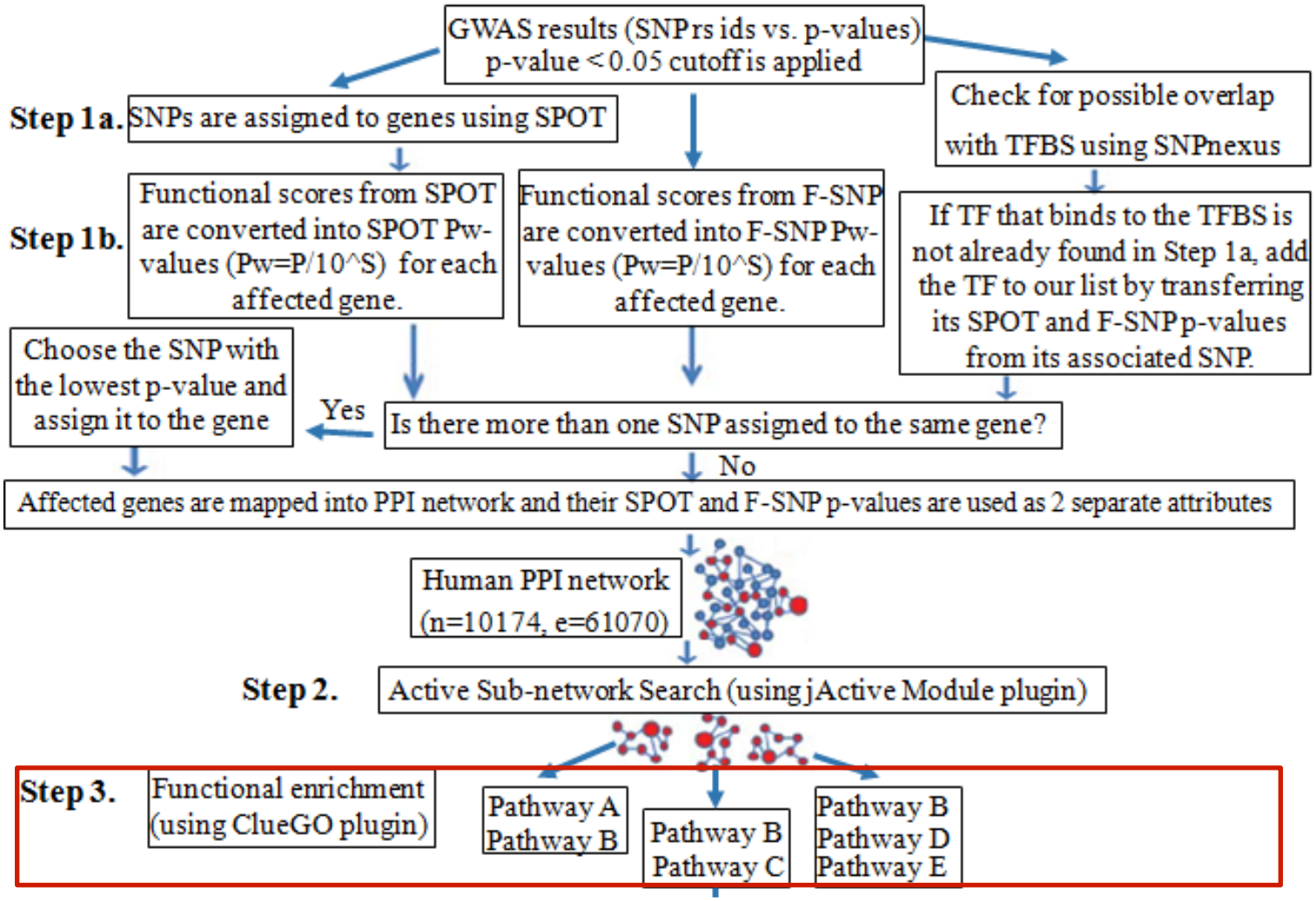


Figure 4. **a.** Node degree distribution of the highest scoring sub-network follows a power-law ($P(k)=ax^{-\gamma}$, a= 120.03, γ=1.353, R^2=0.773, Correlation= 0.891 in log log scale), showing that our network displays scale-free properties, as expected from a biological network. **b.** Node degree distribution of a same sized random network, obtained using Erdos-Renyi algorithm.

B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.

Our Methodology (PANOGA)





B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.

Comparative Pathway Enrichment Results of RA

KEGG Term	Number of Genes Found						Term Pvalue Corrected Bonferroni
	Baran-zini et.al.	Martin et.al.	Wu et.al	Zhang et.al.	PANOGA (only GWAS pvalues)	PANOGA (w/ regional scores)	
Focal adhesion	0	0	36	32	22	30	9,33E-11
ErbB signaling pathway	0	0	23	0	18	20	2,13E-10
Tight junction	0	0	0	5	20	22	1,80E-08
Chemokine signaling pathway	0	0	0	0	24	26	2,31E-08
Adherens junction	0	0	0	18	16	17	1,16E-07
Bacterial invasion of epithelial cells	0	0	0	0	15	16	1,57E-007
Neurotrophin signaling pathway	0	0	0	0	20	20	2,36E-07
Long-term potentiation	0	22	0	7	14	15	3,67E-07
Pathways in cancer	0	0	0	0	29	32	1,12E-06
Chronic myeloid leukemia	4	0	21	18	10	14	1,44E-06
Cell adhesion molecules (CAMs)	8	26	0	10	12	18	1,42E-05
Leukocyte transendothelial migration	0	24	14	0	17	17	1,72E-05
T cell receptor signaling pathway	4	21	16	16	13	16	2,70E-05
Toll-like receptor signaling pathway	0	0	22	6	7	13	1,97E-03
Antigen processing and presentation	6	0	0	3	11	11	2,08E-03
Allograft rejection	0	0	0	0	8	8	2,16E-03
MAPK signaling pathway	0	0	43	34	16	20	6,13E-03
Type I diabetes mellitus	5	0	0	1	8	8	6,24E-03
Apoptosis	0	18	12	11	6	11	6,84E-03
Jak-STAT signaling pathway	0	25	0	16	13	15	7,41E-03
Prostate cancer	0	0	22	0	10	11	5,04E-02
Calcium signaling pathway	0	35	0	4	15	16	1,63E-01
VEGF signaling pathway	3	0	15	13	8	9	2,71E-01

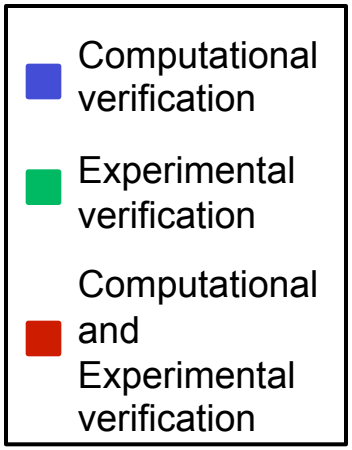


Table 2. Comparison of found KEGG pathways with previous studies in terms of number of genes associated within each KEGG term. Blue denotes computationally found pathways, green denotes experimentally verified RA associated pathways, and red denotes both experimental and computational verification.

Functionally Grouped Annotation Network of RA

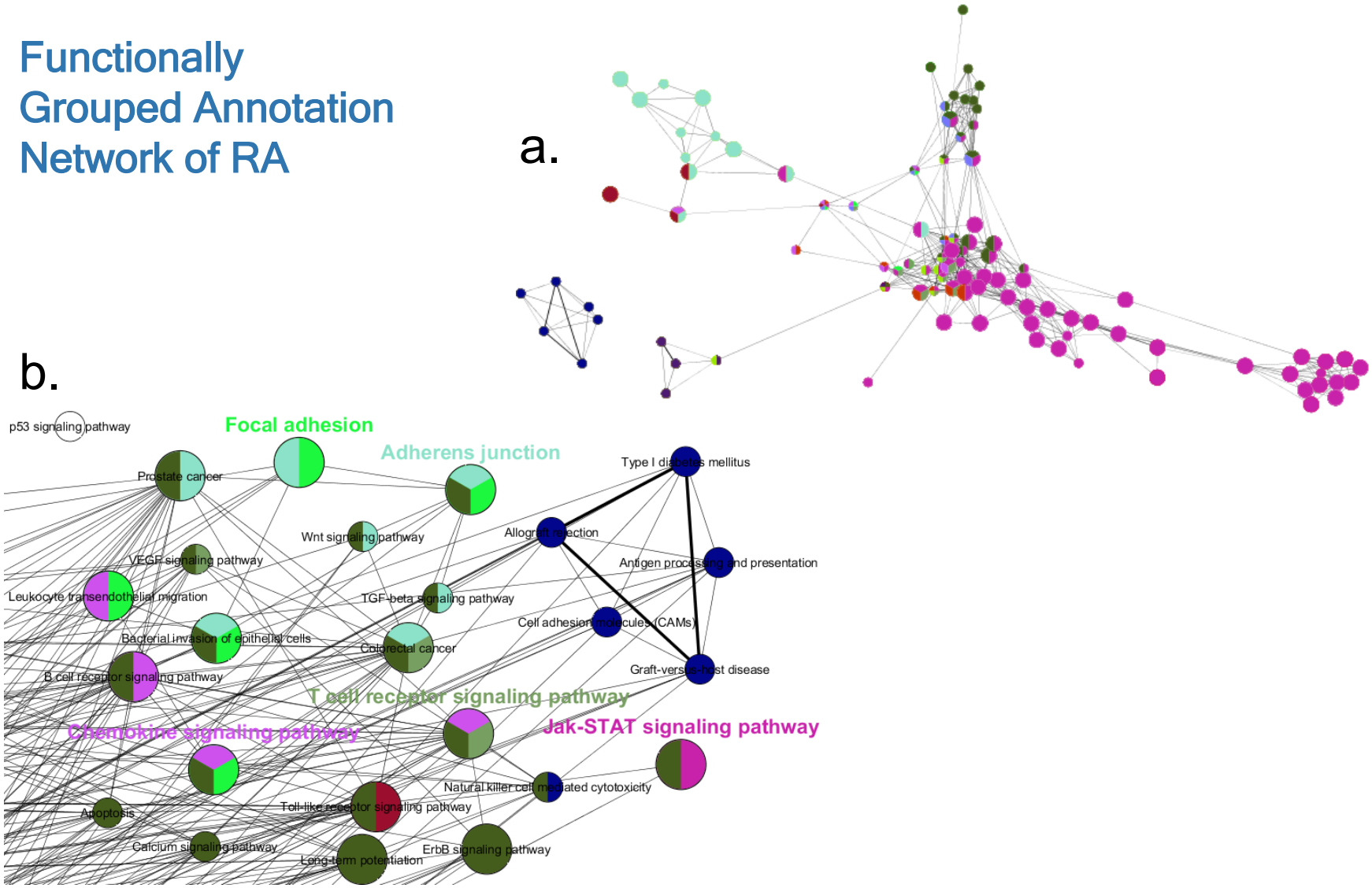


Figure 5. **a.** Functionally grouped annotation network of our highest scoring sub-network. The relationships between the KEGG terms (nodes) were based on the similarity of their associated genes. The size of the nodes reflected the statistical significance of the terms (term p-values corrected with Bonferroni). Edges represent the existence of shared genes. The thickness of the edges is proportional to the number of genes shared and calculated using kappa statistics. The grouped terms (according to their kappa scores) were shown in same color. **b.** Zoomed in view of the functional annotation network. The most significant pathway term of the group with the lowest term p-value (the group leading term) was shown in bold using the group specific color.

Comparative Evaluation of RA Associated Pathways

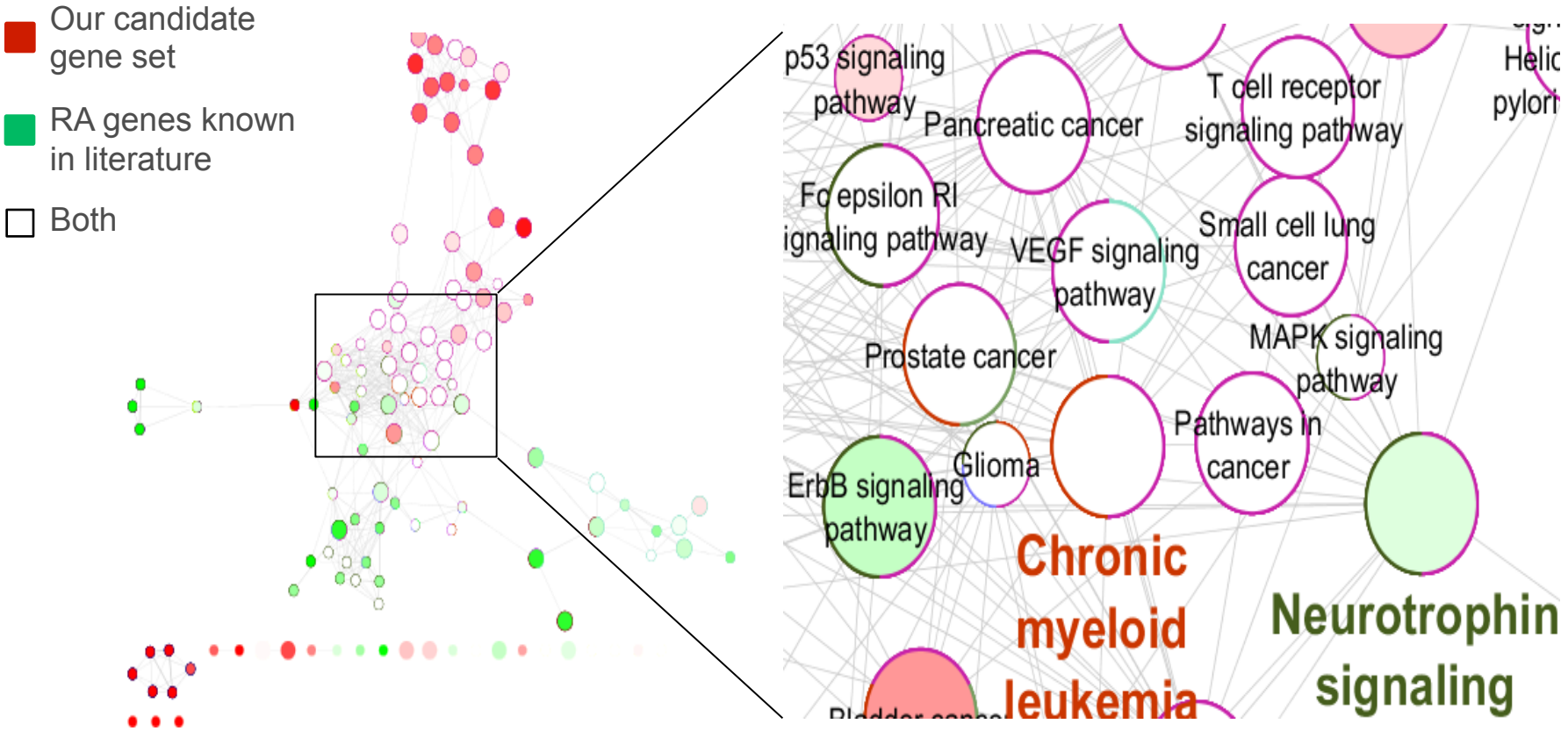
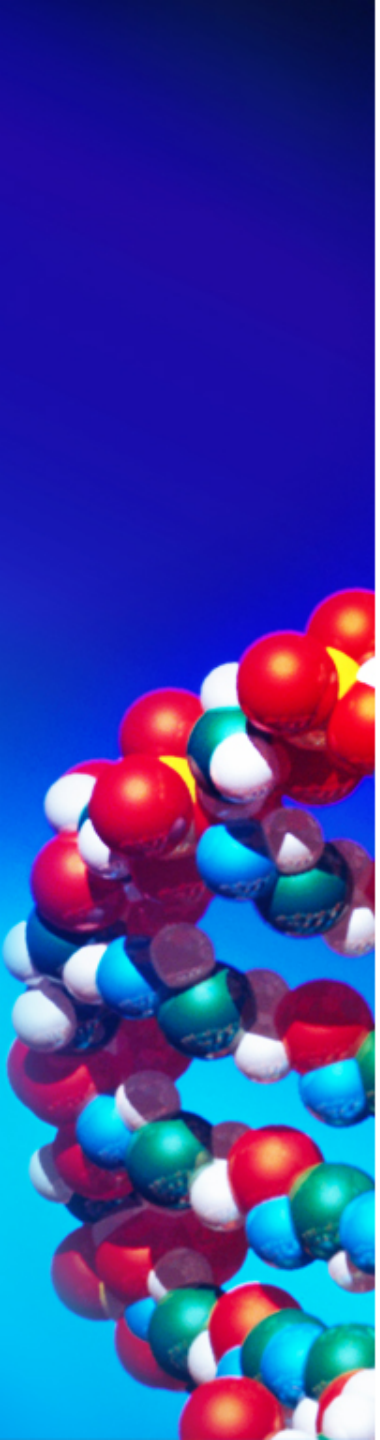


Figure 6. Comparison of KEGG pathway terms with literature verified RA genes/our gene set were shown in green/red, respectively. Nodes represent the identified pathway terms from any one of the two sets. The color gradient showed the gene proportion of each set associated with the term. White color represented equal proportions from the two comparison sets. The size of the nodes reflected the statistical significance of the terms (term p-values corrected with Bonferroni). Edges represented the existence of the shared genes between the pathway terms and node border colors mapped to the group colors.

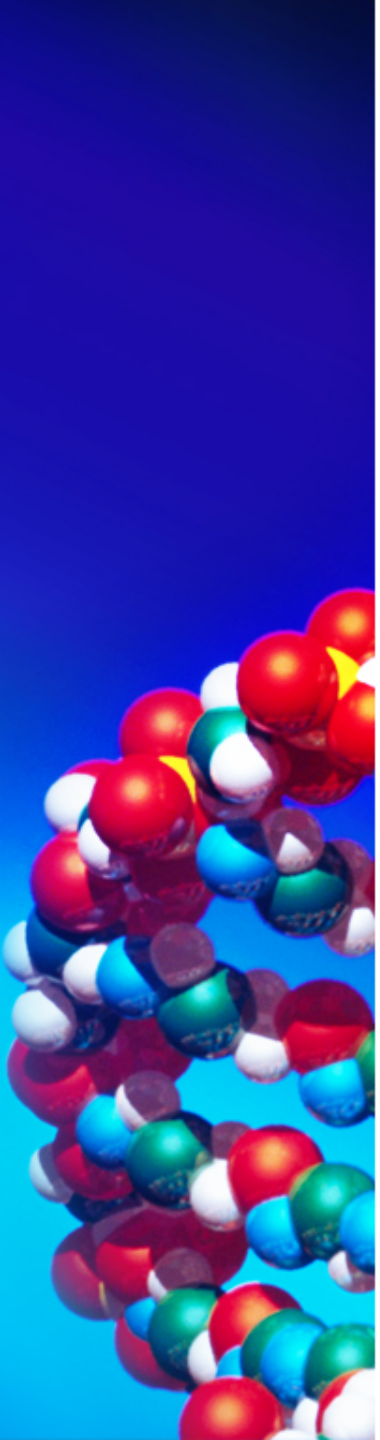


Insights

- We present PANOGA, pathway and network oriented GWAS analysis, that challenges to identify disease associated KEGG pathways by combining nominally significant evidence of genetic association with current knowledge of biochemical pathways, protein-protein interaction networks, and functional information of selected SNPs.

- We identified both previously known and additional KEGG pathways as associated with RA.

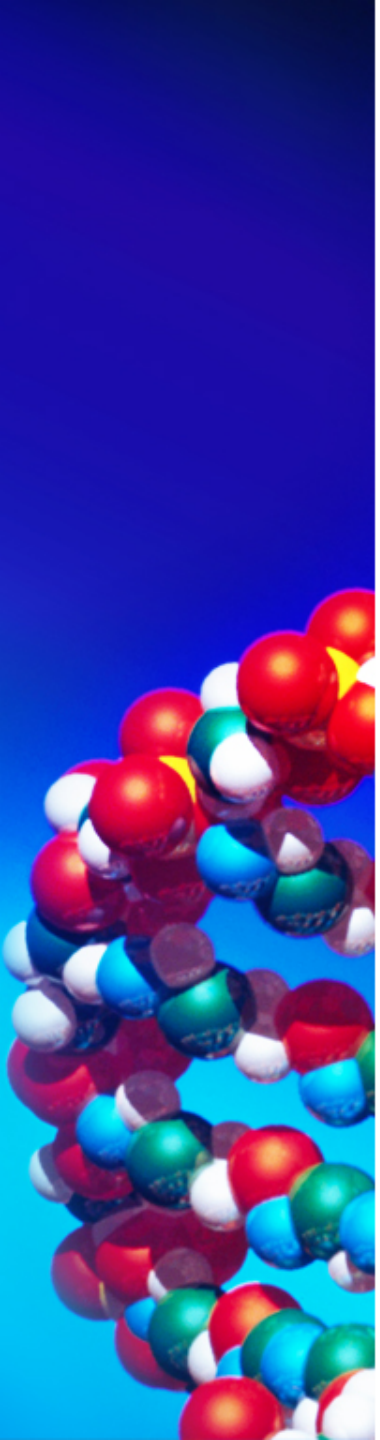
B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.



Insights (ctd.)

- The KEGG functional enrichment of the RA specific drug target genes included these additionally found pathway terms.
- Among the previously known pathways, we identified additional genes as associated with RA.
- Using our highest scoring sub-network, we generated functionally grouped pathway network of RA.

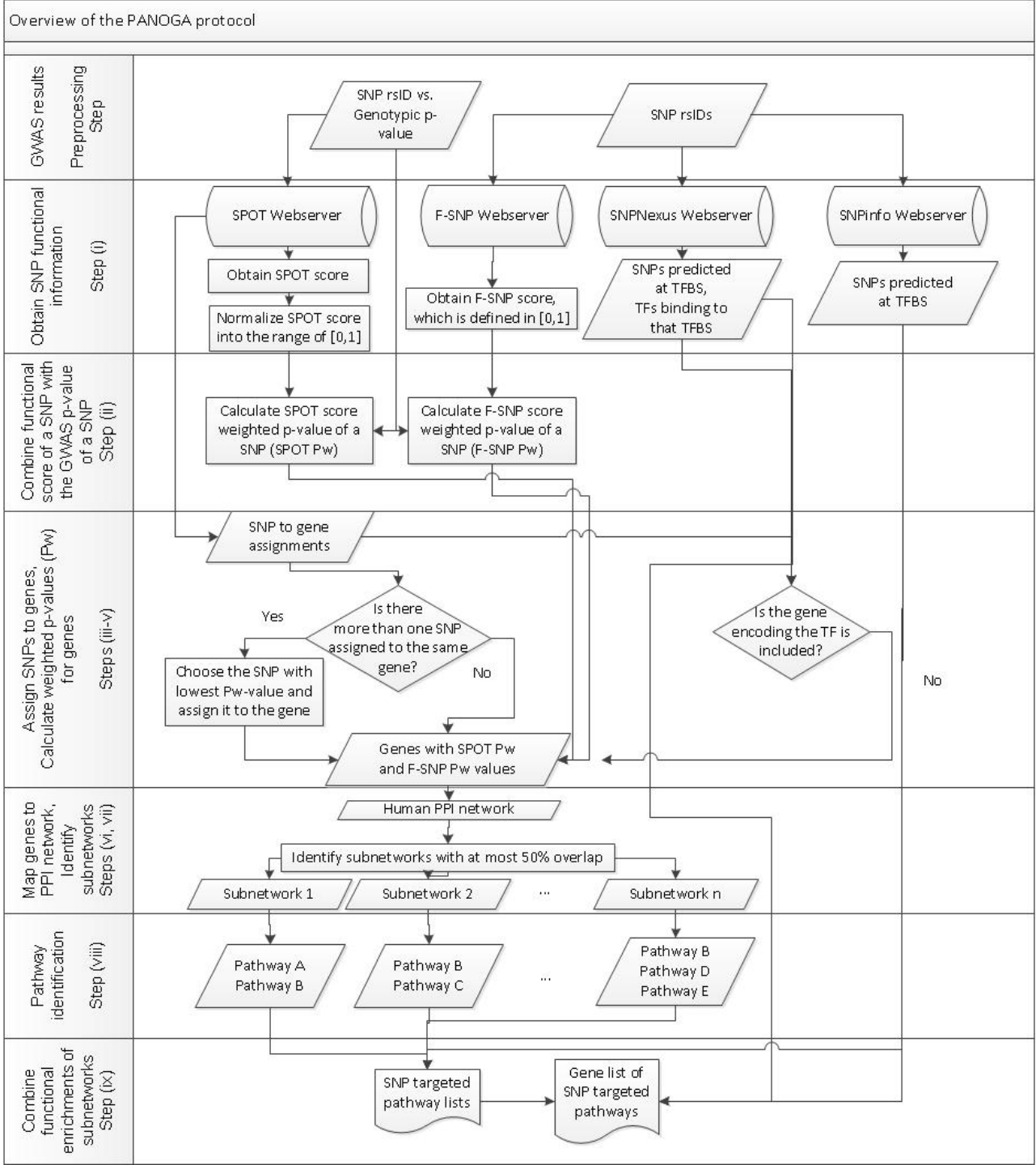
B. Bakir-Gungor, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277.



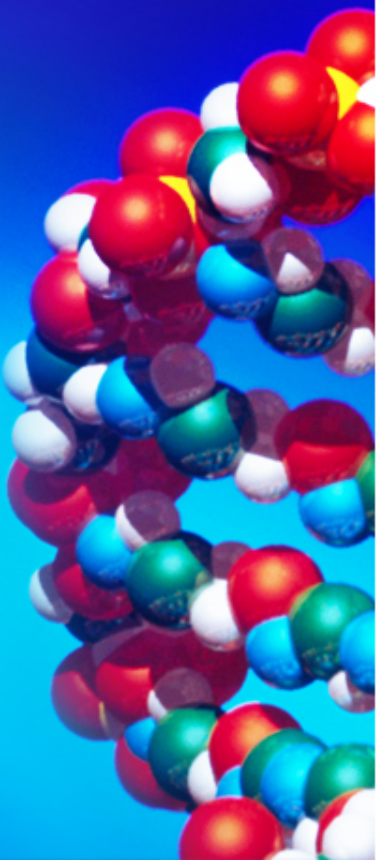
Improvements in the methodology

- Instead of focusing only on the highest scoring sub-network, the functional enrichments of the generated sub-networks are combined.
- The SNPs in the affected genes and pathways are identified.
- Pathway and gene based presentation options.
- The effect of the overlap between sub-networks is evaluated.

Overview of the PANOGA Protocol



B. Bakir-Gungor, O.U. Sezerman, "*Identification of SNP Targeted Pathways From Genome-wide Association Study (GWAS) Data*", 2012, **Nature Protocol Exchange**.

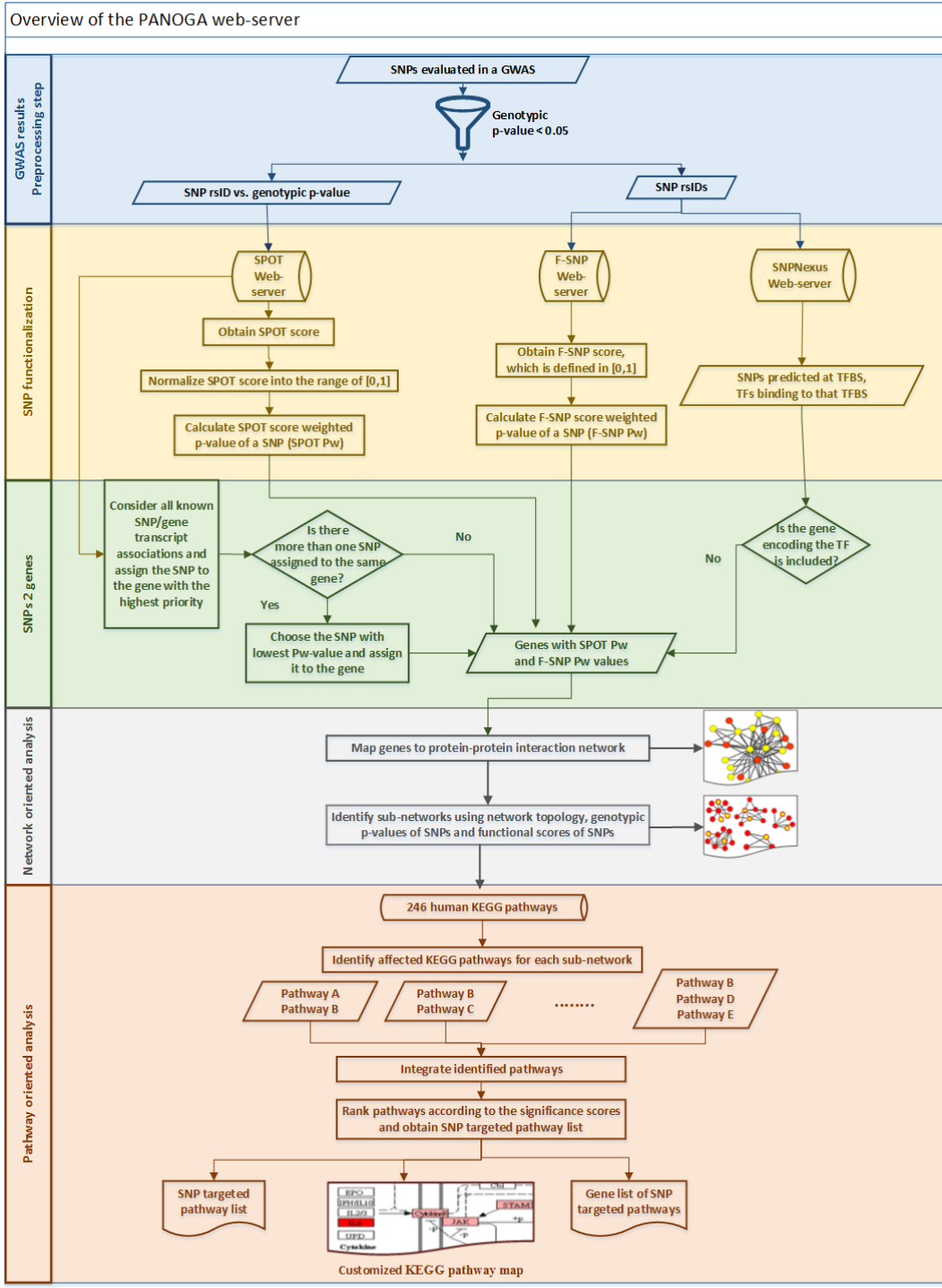


Insights

- PANOGA protocol represents a feasible solution for the identification of pathway markers to bridge the gap between GWAS and biological mechanisms of complex diseases (Bakir-Gungor and Sezerman, 2012).
- Since our method can be easily applied to GWAS datasets of other diseases, it will facilitate the identification of disease specific pathway combinations.
- Due to its modular design pattern, PANOGA protocol gives flexibility to the user.

Overview of the PANOGA Web-server

To present the geneticists a fully automated option, we implemented PANOGA as a web-server.



B. Bakir-Gungor, E. Egemen, O.U. Sezerman, *"PANOGA: a web-server for identification of SNP targeted pathways from genome-wide association study data"*, 2014, **Bioinformatics**, 30(9): 1287-1289.

PANOGA: a web server for identification of SNP-targeted pathways from genome-wide association study data

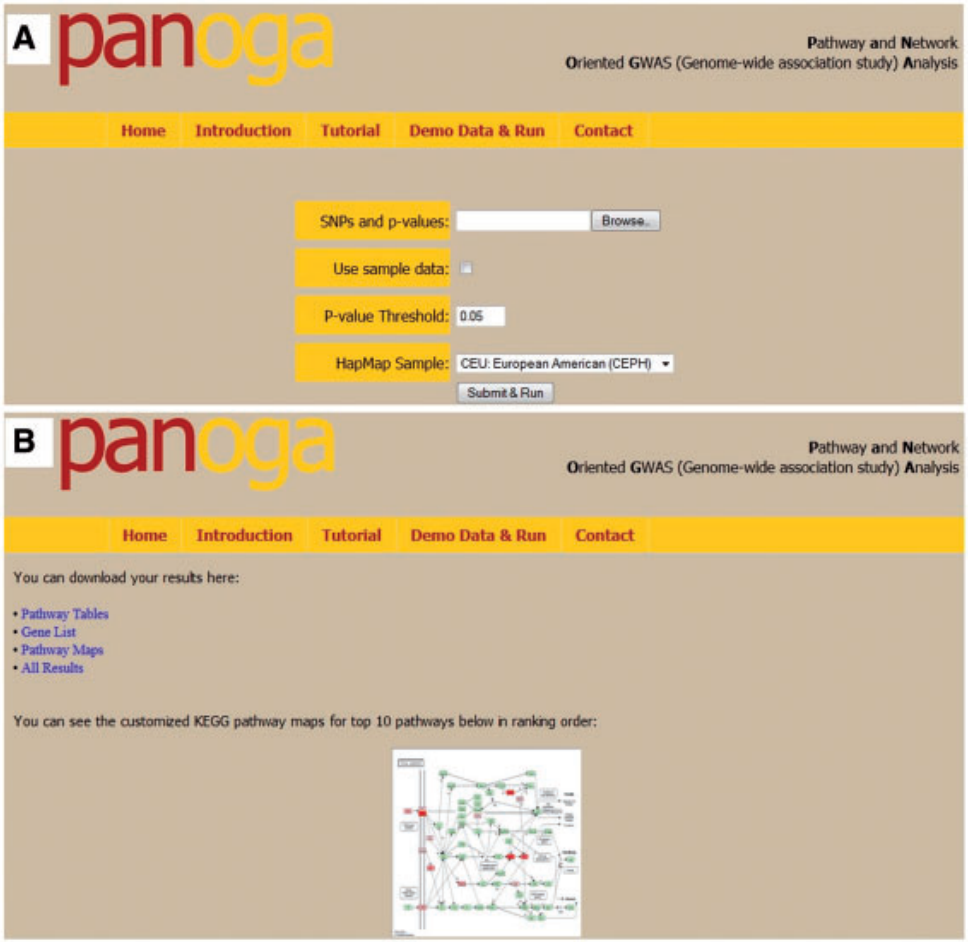


Fig. 7. A snapshot of the web server input (A) and results page (B). A link from the results page for customized KEGG pathway maps opens the zoomed-in version for visual display (C). The shade of red color in genes indicates the number of targeted SNPs (typed in the GWAS of disease) per base pair of the gene. Red refers to the highest targeted gene, whereas white refers to a gene product not targeted by the SNPs.

B. Bakir-Gungor, E. Egemen, O.U. Sezerman, *"PANOGA: a web-server for identification of SNP targeted pathways from genome-wide association study data"*, 2014, **Bioinformatics**, 30(9): 1287-1289.

Epilepsy

- Epilepsy is a abnormal discharge. Characterized by recurrent and spontaneous seizures.
- Common neurologic disorder that affects around 1% of the world population, including one in 200 children.
- In partial epilepsy (PE), seizure affects only one part of the brain.
- It can run in families.

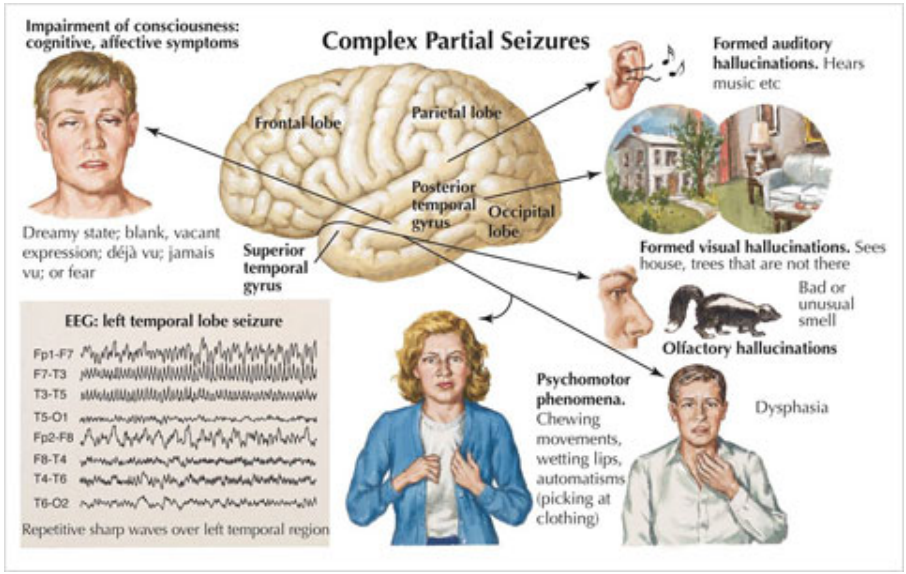


Figure 13: Partial Epilepsy

http://trialx.com/curetalk/wp-content/blogs.dir/7/files/2011/05/diseases/Partial_Epilepsy-3.jpg

Partial Epilepsy Dataset

# of Cases	# of Controls	# of genotyped SNPs	Platform
3,445	6,935	528,745 SNPs	Illumina, Human610-Quadv1 genotyping chips

Table 3. Summary of Partial Epilepsy (PE) dataset.

- Cochran–Mantel–Haenszel test results were used as the genotypic p-values of the identified SNPs.
- Using $P<0.05$ cutoff:
  - 28,450 SNPs were included.

KEGG Term	Term Pvalue Corr Bonf	Wang et al. Study	OMIM	GWAS on PE	CNV Study on Epilepsy	Candidate Gene List	EpiGAD	Rogic et al. Study
Complement and coagulation cascades	2,16E-025	-	Y	-	-	-	-	Y
Cell cycle	1,03E-024	-	Y	-	-	-	-	Y
Focal adhesion	7,10E-023	Y	Y	Y	-	-	-	Y
ECM-receptor interaction	1,62E-022	Y	Y	-	-	-	-	Y
Jak-STAT signaling pathway	1,16E-021	Y	Y	-	-	-	-	Y
MAPK signaling pathway	2,32E-019	Y	Y	Y	-	Y	Y	Y
Proteasome	1,15E-018	-	-	-	-	-	-	-
Ribosome	1,57E-018	-	-	-	-	-	-	Y
Calcium signaling pathway	5,73E-018	Y	Y	Y	Y	Y	Y	Y
Regulation of actin cytoskeleton	9,23E-018	Y	Y	-	Y	-	-	Y
Adherens junction	1,01E-017	-	-	Y	-	-	-	Y
Pathways in cancer	3,94E-017	Y	Y	Y	-	-	-	Y
Gap junction	6,32E-017	Y	Y	Y	-	-	-	Y
Apoptosis	3,72E-016	Y	Y	-	-	-	-	Y
Long-term depression	2,90E-015	Y	Y	Y	Y	Y	Y	Y
Axon guidance	4,01E-015	-	-	-	-	-	-	Y
Fc gamma R-mediated phagocytosis	2,22E-014	Y	Y	Y	Y	-	-	Y
Tight junction	2,82E-014	Y	Y	Y	-	-	-	Y
ErbB signaling pathway	4,04E-014	Y	Y	Y	-	-	-	Y
Wnt signaling pathway	6,28E-014	Y	Y	Y	-	Y	-	Y
Chemokine signaling pathway	9,60E-014	Y	-	Y	Y	-	-	Y
GnRH signaling pathway	1,22E-013	Y	Y	Y	-	-	-	Y
Pentose phosphate pathway	1,29E-013	-	-	-	-	-	-	-
Long-term potentiation	2,28E-013	Y	Y	Y	-	Y	-	Y
Neurotrophin signaling pathway	3,24E-013	Y	Y	-	-	-	-	Y
Glycolysis / Gluconeogenesis	4,29E-013	Y	Y	-	-	-	-	Y
Notch signaling pathway	9,33E-013	-	-	-	-	-	-	-
Dilated cardiomyopathy	1,40E-012	-	Y	Y	-	Y	-	Y
TGF-beta signaling pathway	2,32E-012	-	-	-	-	-	-	Y

Table 4. Comparison of the top 30 SNP-targeted pathways with the pathways of the known genes, as associated to PE. Red color indicates pathway is found in at least 3 studies.

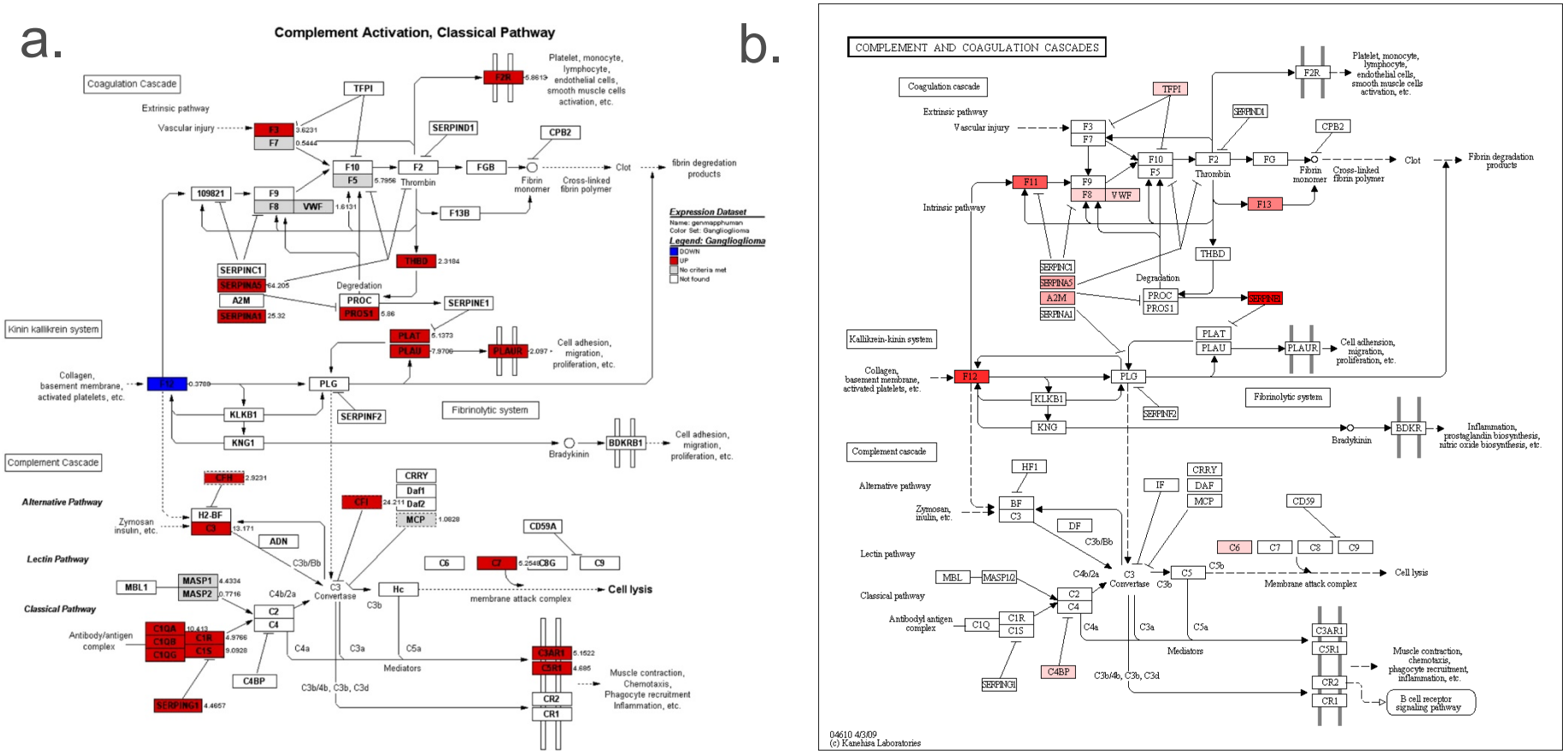


Figure 8. The complement and coagulation cascade (a) Up and down-regulated genes are shown in red and in blue, respectively, as a result of microarray analysis for epilepsy-associated gangliogliomas. (b) The shade of red color in genes indicates the number of GWAS targeted SNPs per base pair of the gene. Red refers to the highest targeted gene, whereas white refers to a gene product, not targeted by the SNPs.

***B. Bakir-Gungor**, et. al.*, "The Identification of SNP Targeted Pathways in Partial Epilepsies Using Genome-wide Association Data", 2013, **Epilepsy Research**, 105(1-2):92-102.

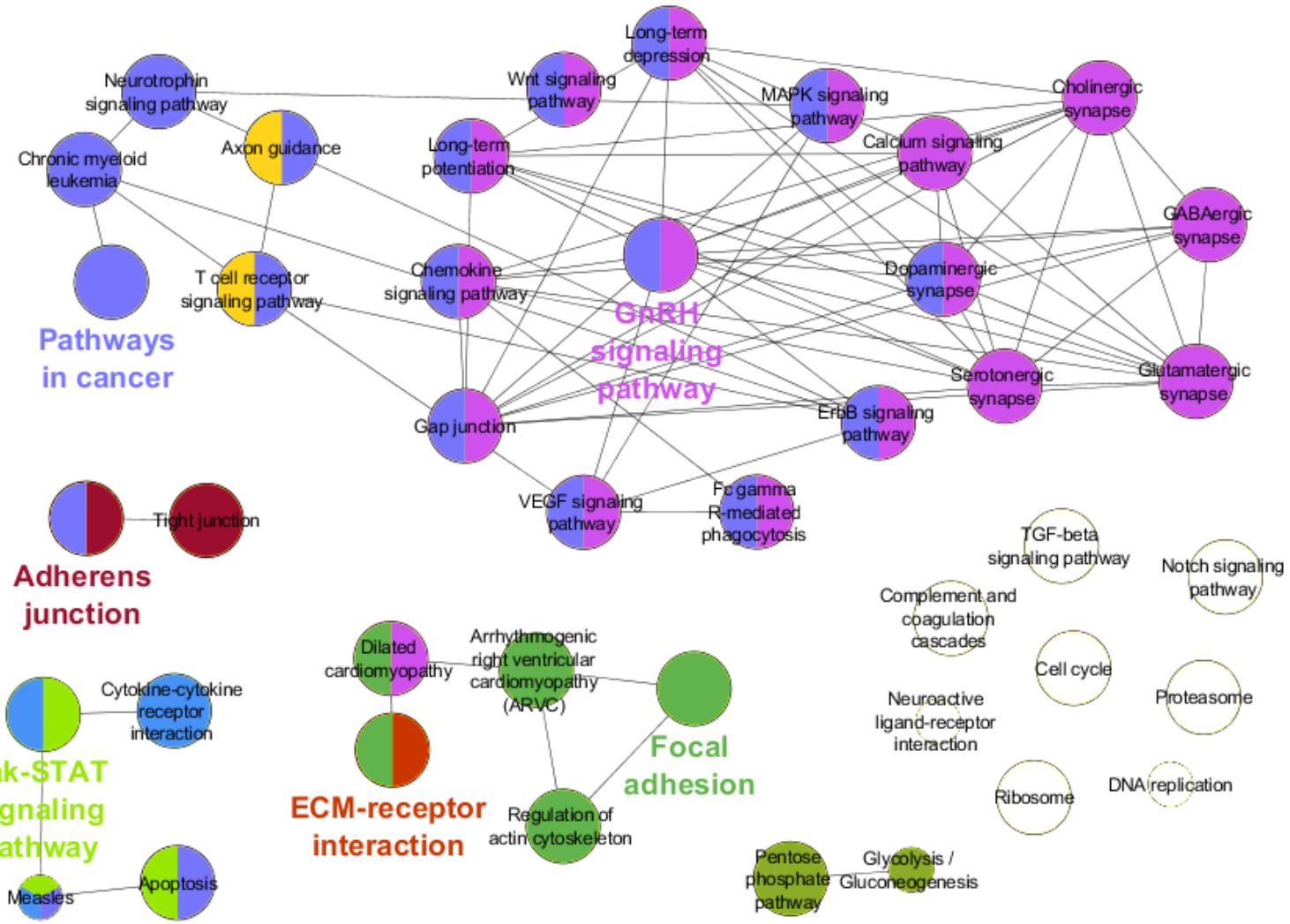
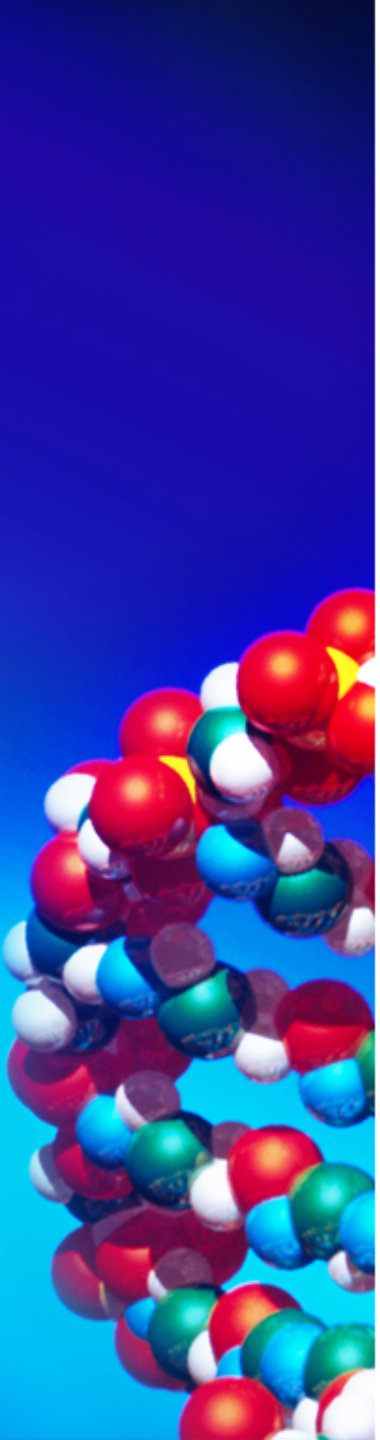


Figure 9. Functionally grouped annotation network of the identified pathways for epilepsy dataset. The pathways are grouped based on the similarity of their SNP targeted genes.

***B. Bakir-Gungor**, et. al.*, "The Identification of SNP Targeted Pathways in Partial Epilepsies Using Genome-wide Association Data", 2013, **Epilepsy Research**, 105(1-2):92-102.



Insights

- We showed that PANOGA was able to identify significant pathways, explaining the pathogenesis of the epilepsy.
- The relation between these pathways and partial epilepsies was supported by previous studies in literature.
- 20 out of the top 30 affected pathways were found to be common with at least three different studies, among the seven studies compared.
- Hence, we emphasize the importance of pathway-oriented analysis to enlighten disease development mechanisms (Bakir-Gungor, et al., 2012).

A Two Stage Genetic Algorithm Approach to Active Subnetwork Search

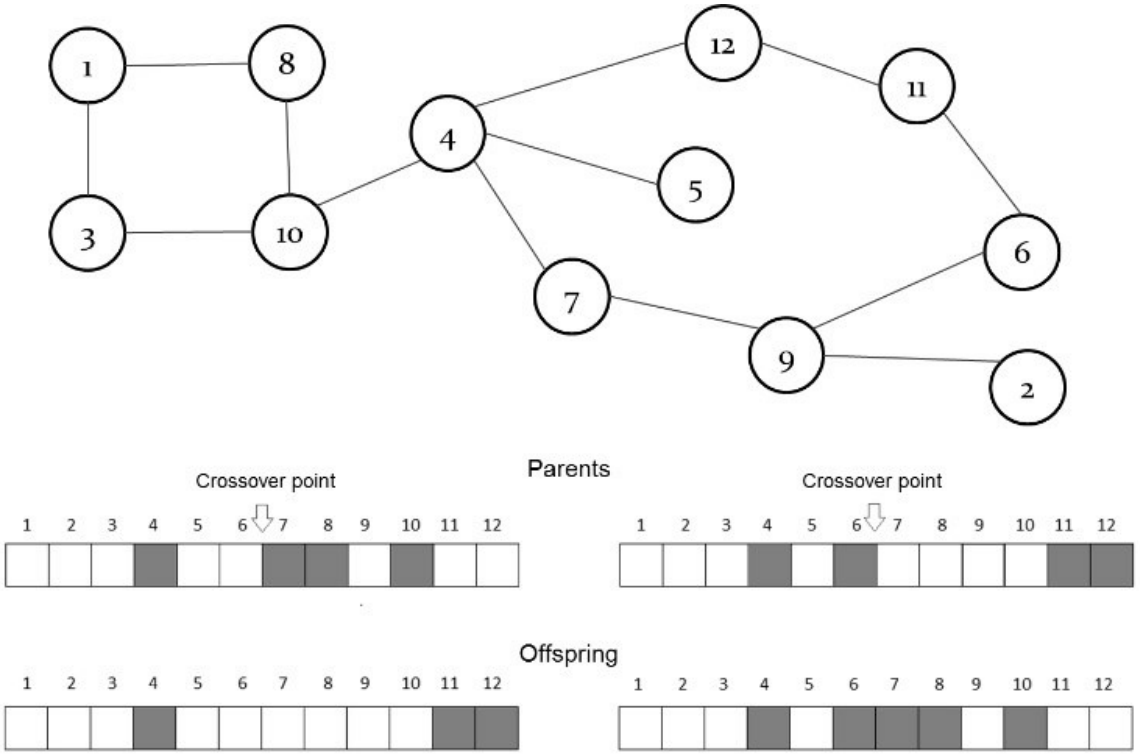


Figure 10. Binary vector chromosome representation and crossover operation in ActiveSubnetworkGA. The graph represents the PPI network. Two subnetworks and their offspring are given underneath in binary vector representation form. Dark cells represent the nodes that are included in the subnetwork. Vector swapping is applied as the crossover operator.

O. Ozisik, **B. Bakir-Gungor,** B. Diri, O.U. Sezerman, “Active Subnetwork GA: A Two Stage Genetic Algorithm Approach to Active Subnetwork Search”, 2017, ***Current Bioinformatics***, 12(4): 320-328(9).

Scoring sub-networks

Significance value p_i of each gene is converted to z-score using equation (1). Here Φ^{-1} is the inverse normal cumulative distribution function.

$$z_i = \Phi^{-1}(1 - p_i) \tag{1}$$

Collective z-score z_A of a subnetwork is calculated by equation (2)

$$z_A = \frac{1}{\sqrt{k}} \sum_{i \in A} z_i \tag{2}$$

It is important to see whether the score z_A is higher than expected relative to a random set of genes drawn from the same experimental data. For this purpose, gene sets consisting of k genes are randomly selected, their z_A scores are calculated, and mean (μ_k) and standard deviation (σ_k) are found. The corrected subnet score is calculated by equation (3).

$$s_A = \frac{z_A - \mu_k}{\sigma_k} \tag{3}$$

O. Ozisik, **B. Bakir-Gungor,** B. Diri, O.U. Sezerman, “Active Subnetwork GA: A Two Stage Genetic Algorithm Approach to Active Subnetwork Search”, 2017, ***Current Bioinformatics***, 12(4): 320-328(9).

Active Sub-network Search Algorithms

2.4. Active Subnetwork Search Algorithms

Currently, there are three algorithms implemented in the pathfindR package for active subnetwork search: greedy algorithm, simulated annealing algorithm and genetic algorithm.

2.4.1. Greedy Algorithm

Greedy algorithm is the problem-solving/optimization concept that chooses locally the best option in each stage with the hope of reaching the global optimum. In active subnetwork search, this is generally applied by starting with a significant seed node and considering addition of a neighbor in each step to maximize the subnetwork score. In pathfindR, we used the approach in Chuang et al.[13]: This algorithm considers addition of a node within a specified distance d to the current subnetwork. In our method maximum depth from the seed can also be set. With the default parameters, our greedy method considers addition of direct neighbors (d=1) and forms a subnetwork with a maximum depth of 1 for each seed. Because the expansion process runs for each significant seed node, several overlapping subnetworks emerge. In pathfindR, overlapping subnetworks are handled by discarding a subnetwork that overlaps with a higher scoring subnetwork more than a threshold, which is set to 0.5 by default.

O. Ozisik, **B. Bakir-Gungor,** B. Diri, O.U. Sezerman, “Active Subnetwork GA: A Two Stage Genetic Algorithm Approach to Active Subnetwork Search”, 2017, ***Current Bioinformatics***, 12(4): 320-328(9).

Active Sub-network Search Algorithms

2.4.2. Simulated Annealing Algorithm

Simulated annealing improves the greedy search by accepting non-optimal actions to increase exploration in the search space. The probability of accepting a non-optimal action decreases in each iteration. In active subnetwork search context, the search begins with a set of randomly chosen genes (that will be referred to as genes in "on" state), connected components in this candidate solution are found and the scores are calculated. In each iteration the state of a random node is changed from on to off, vice versa, connected components are found in the new solution and their scores are calculated. If the score improves, the change is accepted, if the score decreases, the change is accepted with a probability proportional to the temperature parameter that decreases in each step.

O. Ozisik, **B. Bakir-Gungor,** B. Diri, O.U. Sezerman, "Active Subnetwork GA: A Two Stage Genetic Algorithm Approach to Active Subnetwork Search", 2017, ***Current Bioinformatics***, 12(4): 320-328(9).

Intracranial Aneurysm (IA)

- A cerebrovascular disease that affects around 1 per 50 people.
- Major public health concern, since rupture of an IA leads to stroke and death.
- To identify IA related genetic factors, DNA linkage, candidate gene, genetic association and GWAS have been used.
- Four recent GWAS identified some variants associated with IA, which collectively explain only 10% of the familial risk of IA.

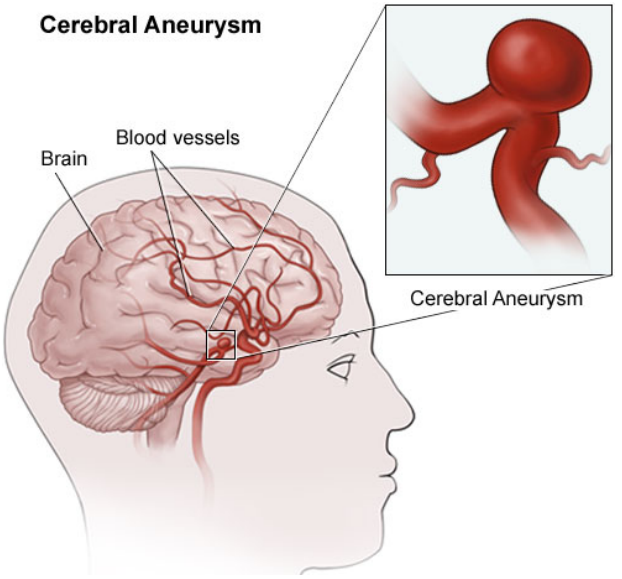
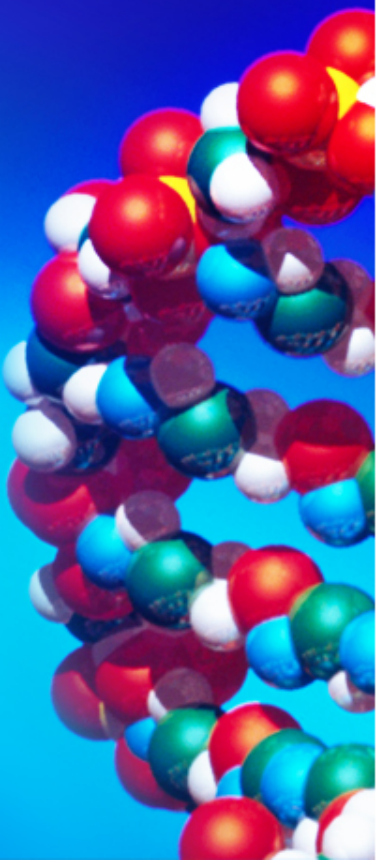


Figure 16: Intracranial (Cerebral) Aneurysm http://www.yalemedicalgroup.org/stw/images/161464.jpg



Intracranial Aneurysm Datasets from Two Different Populations

Population	# of Cases	# of Controls	# of genotyped SNPs	Platform
European	2,780	12,515	832,000	Illumina
Japanese	1,069	904	312,712	Illumina

Table 5. Summary of Intracranial Aneurysm (IA) dataset.

- In both datasets, each SNP's genotypic p-value of association is calculated via Cochran-Armitage trend test.
- Using $P<0.05$ cutoff:
  - 44,351 SNPs were included for EU population,
  - 14,034 SNPs were included for JP population.

	P-values		Rank		# of Associated SNPs in GWAS		# of Common SNPs in GWAS	# of SNP Targeted Genes (STGs)		# of Com-mon STGs	% Common Genes in Both Populations		Common SNPs in GWAS
KEGG Term	EU	JP	EU	JP	EU	JP		EU	JP		EU	JP	
MAPK signaling pathway *	3.53E-27	2.70E-18	1	8	133	43	1	14	18	2	14.29	11.11	rs791062
Cell cycle	2.35E-25	2.81E-19	2	4	76	18	1	11	10	2	18.18	20	rs744910
TGF-beta signaling pathway *	6.26E-24	2.41E-17	3	9	126	20	3	15	9	5	33.33	55.56	rs2053423. rs1440375. rs744910
ErbB signaling pathway	9.52E-22	2.47E-15	4	16	50	15	0	6	4	0	0	0	
Focal adhesion *	9.55E-22	5.60E-21	5	2	117	45	1	21	14	5	23.81	35.71	rs4678167
Proteasome	2.36E-21	4.55E-11	6	35	32	1	0	6	1	0	0	0	
Adherens junction*	4.91E-19	2.58E-21	7	1	85	34	1	13	11	2	15.38	18.18	rs1561798
Notch signaling pathway	2.14E-18	4.74E-12	8	31	26	13	0	8	4	1	12.5	25	
Regulation of actin cytoskeleton *	2.28E-18	4.05E-17	9	10	102	36	1	18	14	1	5.556	7.143	rs4678167
Neurotrophin signaling pathway	2.49E-18	1.93E-18	10	7	68	14	0	7	7	1	14.29	14.29	

Table 6. The top 10 KEGG pathways identified for both populations in IA. 7 out of the top 10 pathways, identified in both populations are shown in red.

B. Bakir-Gungor, O.U. Sezerman, "*The Identification of Pathway Markers in Intracranial Aneurysm Using Genome-wide Association Data from Two Different Populations*", 2013, **PLoS ONE,** 8(3): e57022.

	P-values		Rank		# of Associated SNPs in GWAS		# of Common SNPs	# of SNP Targeted Genes	# of Com-	% Common Genes in Both Populations	
KEGG Term	EU	JP	EU	JP	EU	JP					
MAPK signaling pathway *	3.53E-27	2.70E-18	1	8	133	43					
Cell cycle	2.35E-25	2.81E-19	2	4	76	18					
TGF-beta signaling pathway *	6.26E-24	2.41E-17	3	9	126	20					
ErbB signaling pathway	9.52E-22	2.47E-15	4	16	50	15					
Focal adhesion *	9.55E-22	5.60E-21	5	2	117	45					
Proteasome	2.36E-21	4.55E-11	6	35	32	1					
Adherens junction*	4.91E-19	2.58E-21	7	1	85	34					
Notch signaling pathway	2.14E-18	4.74E-12	8	31	26	13					
Regulation of actin cytoskeleton *	2.28E-18	4.05E-17	9	10	102	36					
Neurotrophin signaling pathway	2.49E-18	1.93E-18	10	7	68	14					

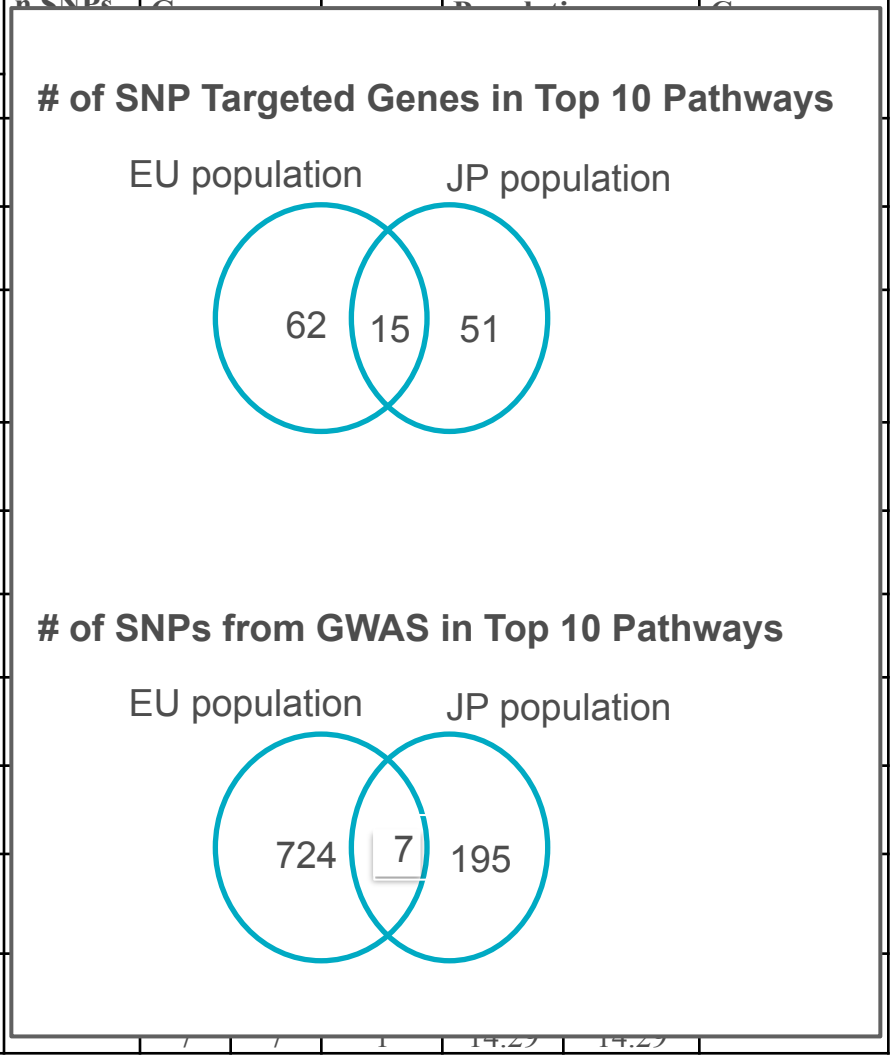


Table 6. The top 10 KEGG pathways identified for both populations in IA. 7 out of the top 10 pathways, identified in both populations are shown in red.

B. Bakir-Gungor, O.U. Sezerman, "*The Identification of Pathway Markers in Intracranial Aneurysm Using Genome-wide Association Data from Two Different Populations*", 2013, **PLoS ONE,** 8(3): e57022.

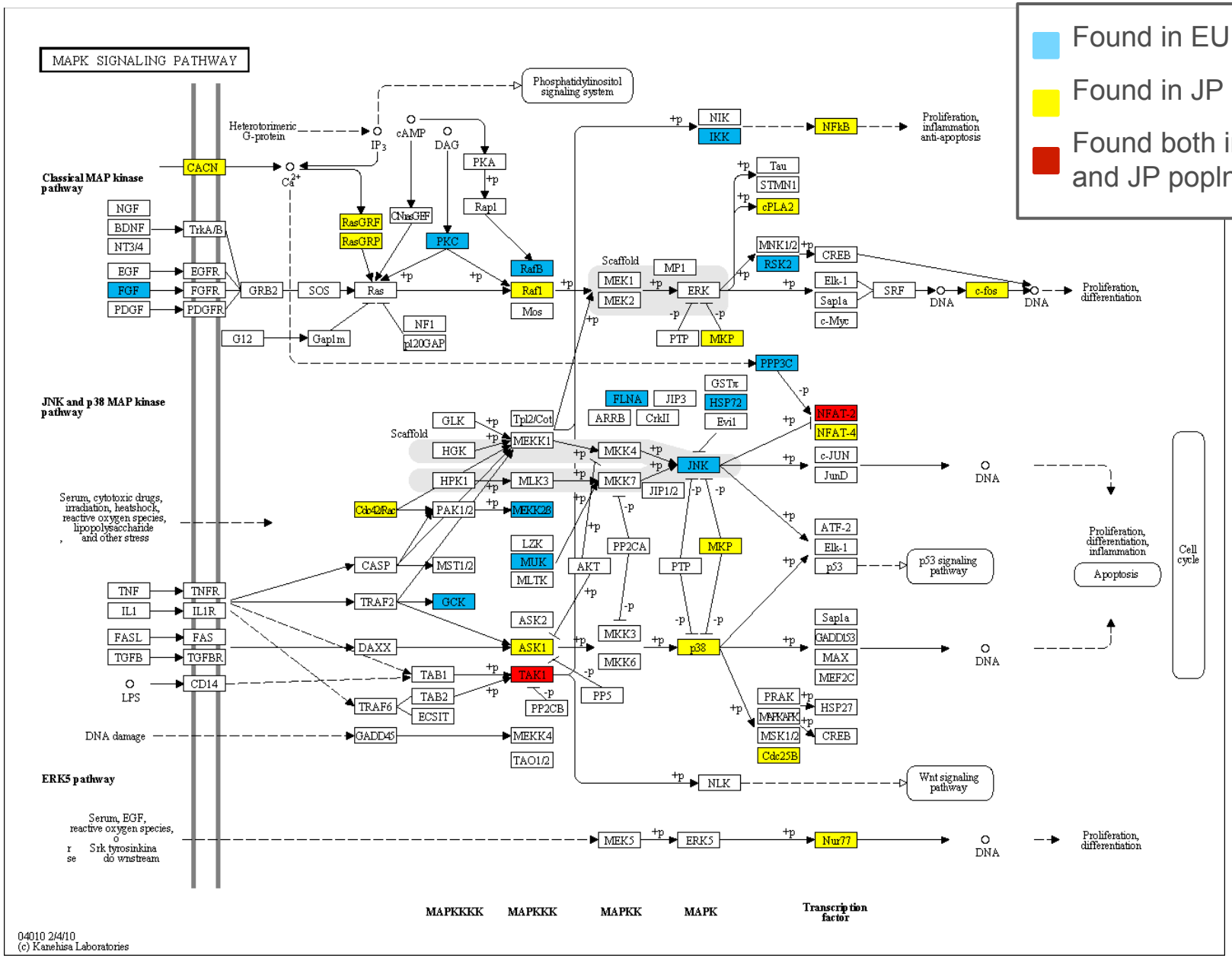
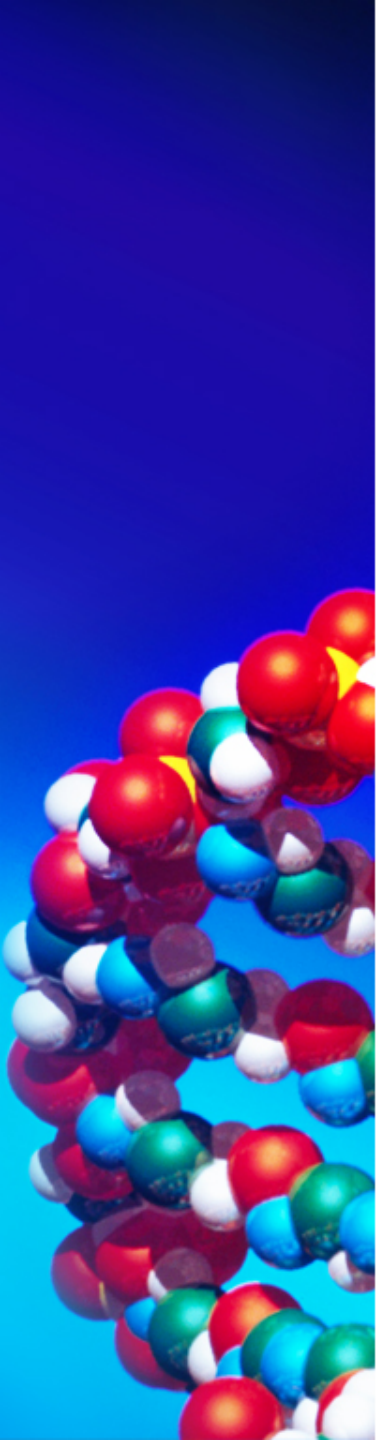


Figure 11. KEGG pathway map for MAPK signaling pathway. The set of genes shown in blue includes genes that are found for EU dataset; yellow includes genes that are found for JP dataset; red includes genes that are found both by EU and JP GWAS of IA.



Insights

- Via applying PANOGA on two aneurysm GWASs, conducted on European and Japanese populations, we have shown that 7 of the top 10 affected pathways are common between these two populations.

- The probability of getting 7 common pathways out of randomly selected 10 pathways from existing 246 human KEGG pathways is $2.24E^{-39}$.

- The relation between these pathways and the IA is supported by previous studies in literature.

- Although different SNP targeted genes are affected on each population, these genes map to the same pathways among different populations (Bakir-Gungor and Sezerman, 2012).

Analysis of IA transcriptomics data using PANOGA

	KEGG Term P-values Corrected with Bonferroni			Rankings		
KEGG Term	Gene Expression	GWAS EU	GWAS JP	Gene Expression	GWAS EU	GWAS JP
Ribosome	7.91E-23	1.40E-08	5.93E-19	1	73	5
Spliceosome	7.40E-17	2.05E-13	4.72E-13	2	33	27
RNA transport	3.97E-14	6.26E-09	-	3	69	-
Complement and coagulation cascades	6.05E-13	7.00E-14	1.06E-09	4	31	48
T cell receptor signaling pathway	7.86E-12	1.62E-16	1.97E-15	5	17	15
ErbB signaling pathway	5.70E-09	9.52E-22	2.47E-15	6	4	16
Chronic myeloid leukemia	6.70E-09	2.62E-18	8.13E-11	7	11	36
Natural killer cell mediated cytotoxicity	9.96E-09	2.56E-07	1.29E-09	8	81	50
RNA degradation	1.44E-08	3.44E-11	1.66E-07	9	44	67
Osteoclast differentiation	1.45E-08	8.12E-15	4.97E-10	10	26	43
Neurotrophin signaling pathway	6.68E-08	2.49E-18	1.92E-18	11	10	7
Adherens junction *	1.74E-07	4.91E-19	2.58E-21	12	7	1
mRNA surveillance pathway	3.59E-07	-	-	13	-	-
Pyruvate metabolism	1.87E-06	-	5.82E-05	14	-	92
Toll-like receptor signaling pathway	3.26E-06	9.18E-13	1.50E-10	15	35	38
Small cell lung cancer	3.55E-06	-	1.01E-08	16	-	55
Proteasome	4.19E-06	2.35E-21	4.54E-11	17	6	35
Focal adhesion *	8.57E-06	9.55E-22	5.60E-21	18	5	2
Fc gamma R-mediated phagocytosis	1.47E-05	4.00E-09	1.32E-13	19	66	22
Toxoplasmosis	2.68E-05	1.06E-08	-	20	72	-

Table 7. The top 20 over-represented KEGG pathways identified for gene expression data of IA. Pathways shown in red are identified in top 20 lists of at least two studies.
* Pathway found to be associated with aneurysm related diseases in KEGG Disease Pathways Database.

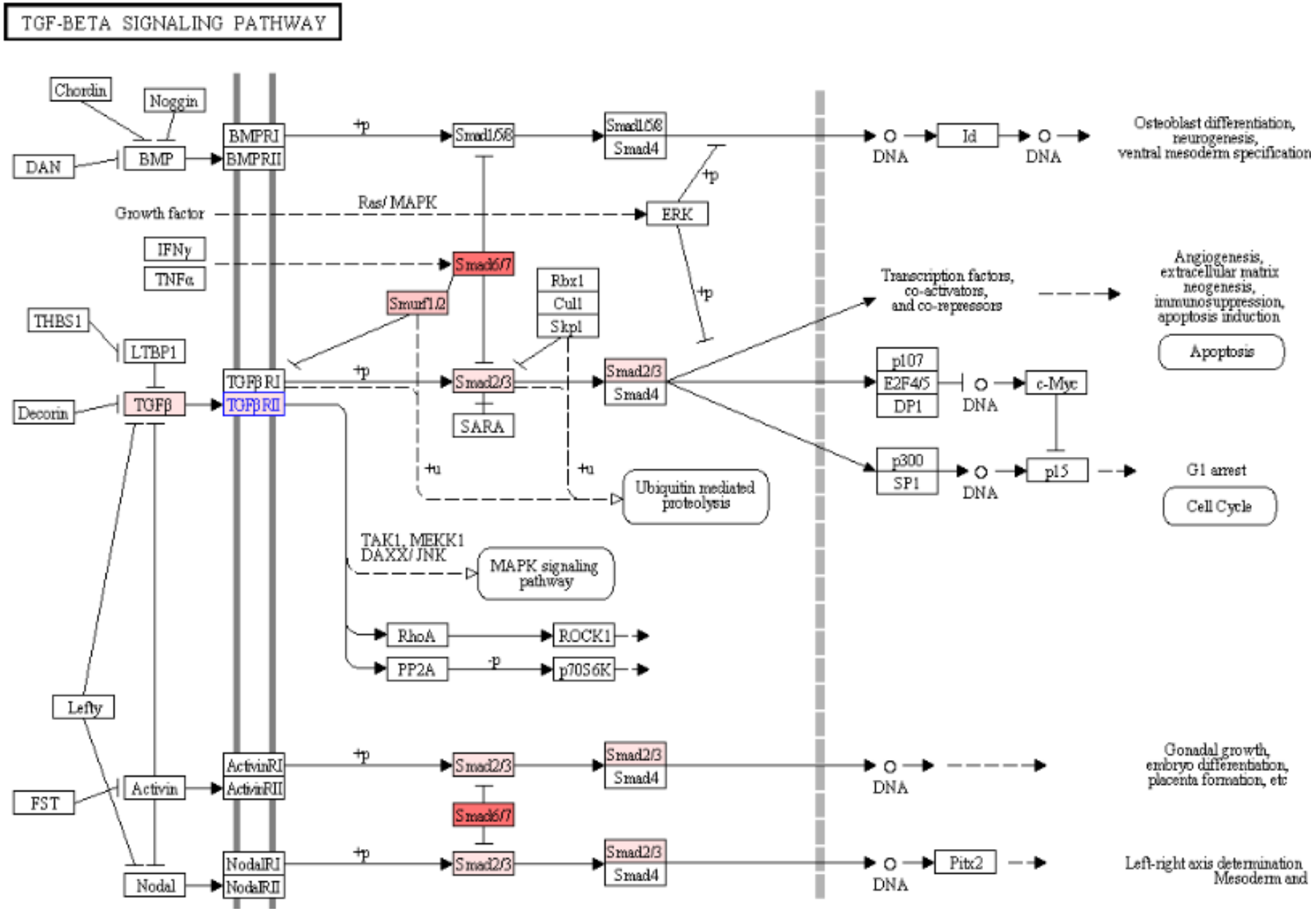


Figure 12. KEGG pathway map for TGF-beta signaling pathway. The shade of red color in genes map to the number of targeted SNPs per base pair of the gene. Blue border indicates that the gene is found to be differentially expressed.

Behçet's disease

- A chronic systemic disease, characterized by recurrent inflammatory attacks affecting multiple organs.
- Widespread in countries along the ancient silk route from Japan to the Middle East and the Mediterranean.
- Known variants account for less than 20% of the genetic risk.

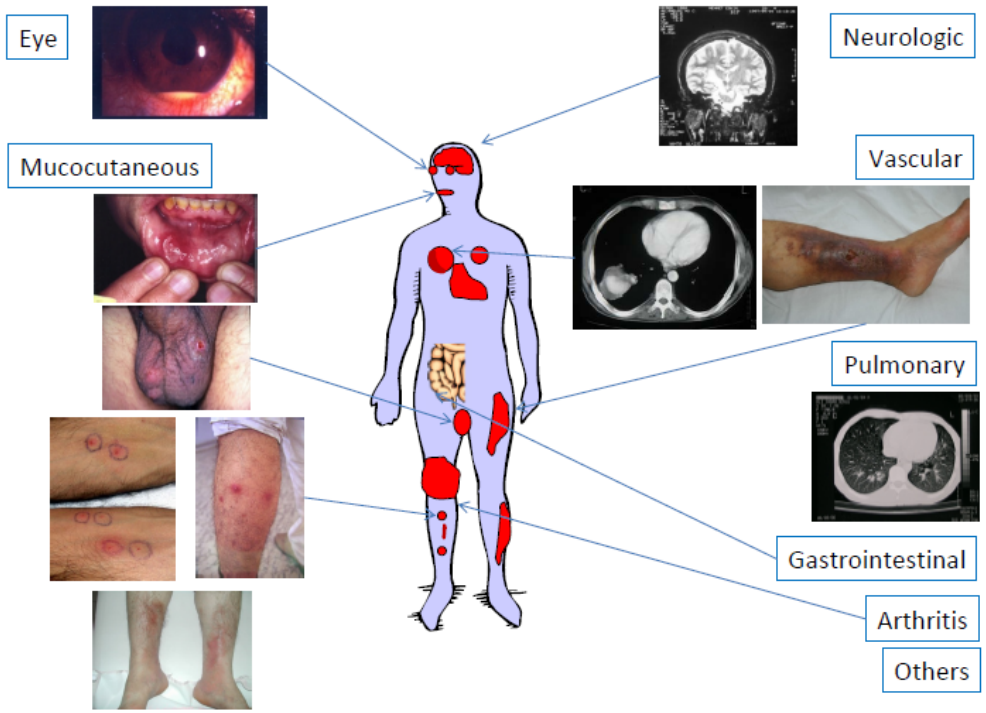
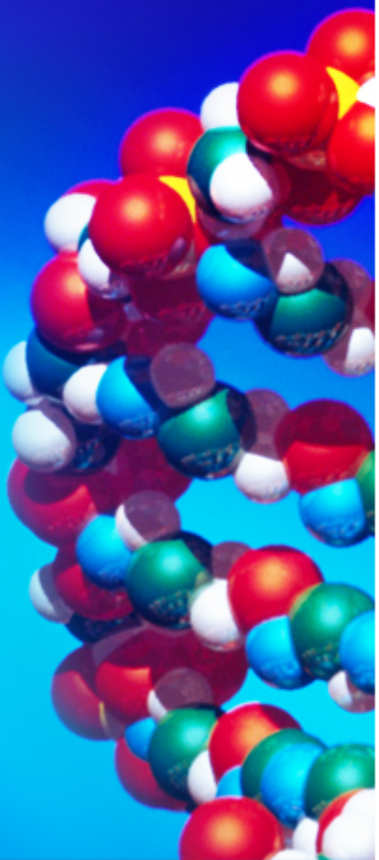


Figure 13. Behcet's Disease

http://excellence-in-rheumatology.org/sites/default/files/presentations/GUL.pdf

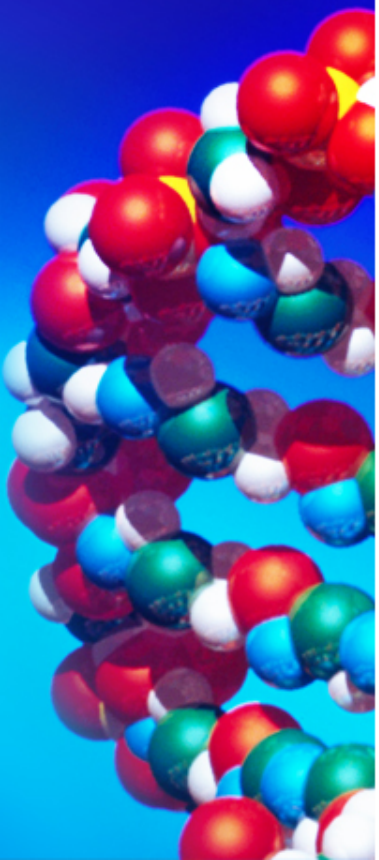


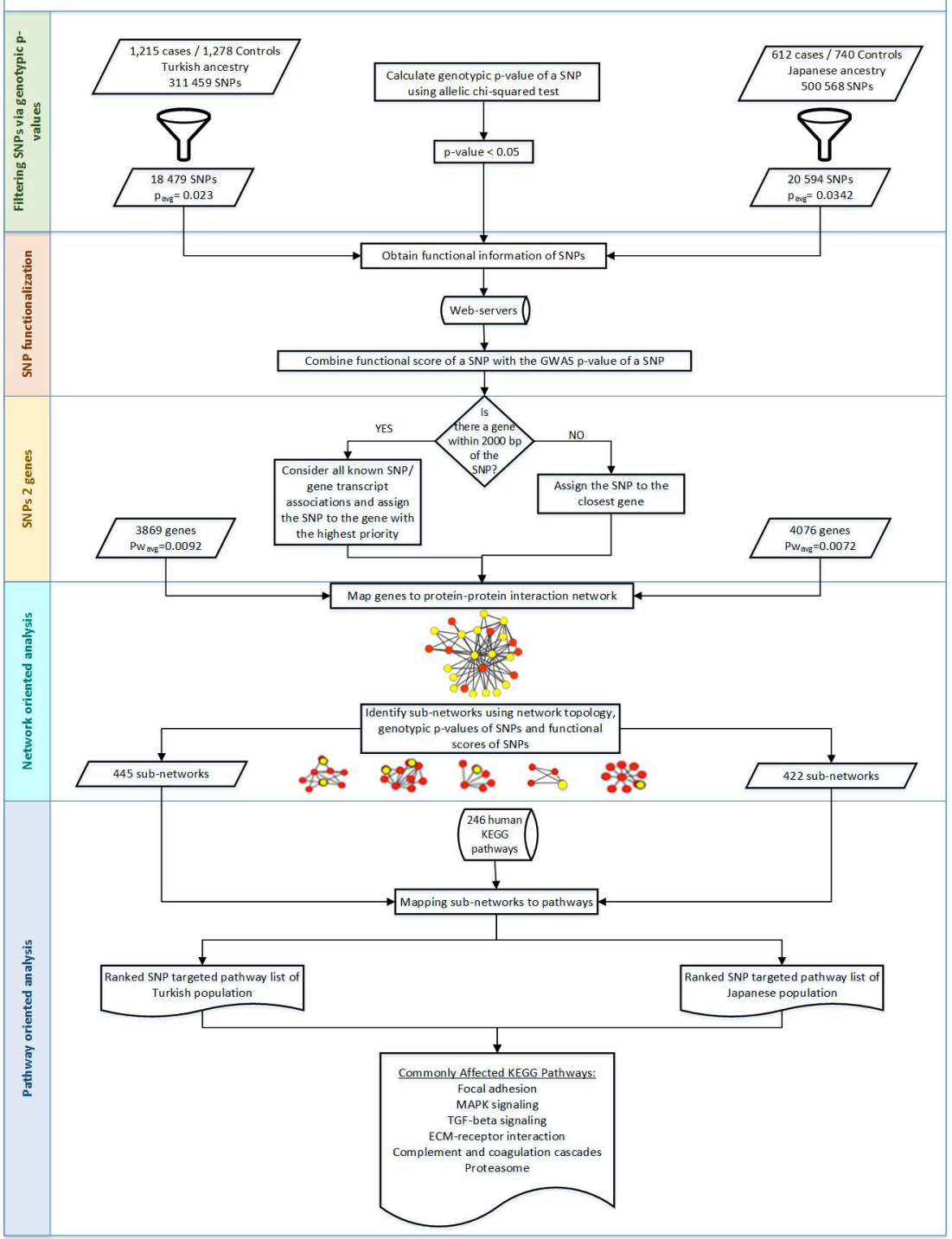
Behcet's disease dataset from two different populations

Population	# of Cases	# of Controls	# of genotyped SNPs	Platform
Turkish	1,215	1,278	311,459	Illumina, Infinium assay
Japanese	612	740	500,568	Affymetrix Gene Chip Human Mapping 500K

Table 8. Summary of Behcet's disease dataset.

- In both datasets, each SNP's genotypic p-value of association is calculated via calculated via allelic chi-squared test.
- Using $P<0.05$ cutoff:
  - 18,479 SNPs were included for TR population,
  - 20,594 SNPs were included for JP population.




Filtering SNPs via genotypic p-values
1,215 cases / 1,278 Controls
Turkish ancestry
311 459 SNPs
Calculate genotypic p-value of a SNP using allelic chi-squared test
612 cases / 740 Controls
Japanese ancestry
500 568 SNPs
p-value < 0.05
18 479 SNPs
p_{avg}= 0.023
20 594 SNPs
p_{avg}= 0.0342
SNP functionalization
Obtain functional information of SNPs
Web-servers
Combine functional score of a SNP with the GWAS p-value of a SNP
SNPs 2 genes
Is there a gene within 2000 bp of the SNP?
YES
NO
Consider all known SNP/ gene transcript associations and assign the SNP to the gene with the highest priority
Assign the SNP to the closest gene
3869 genes
Pw_{avg}=0.0092
4076 genes
Pw_{avg}=0.0072
Network oriented analysis
Map genes to protein-protein interaction network
Identify sub-networks using network topology, genotypic p-values of SNPs and functional scores of SNPs
445 sub-networks
422 sub-networks
Pathway oriented analysis
246 human KEGG pathways
Mapping sub-networks to pathways
Ranked SNP targeted pathway list of Turkish population
Ranked SNP targeted pathway list of Japanese population
Commonly Affected KEGG Pathways:
Focal adhesion
MAPK signaling
TGF-beta signaling
ECM-receptor interaction
Complement and coagulation cascades
Proteasome

	P-values		Rank		# of Associated SNPs in GWAS		# of SNP Targeted Genes (STGs)		# of Com-mon STGs	% Common Genes in Both Populations		Is Common Genes more than 50% in any population?
KEGG Term	TR	JP	TR	JP	TR	JP	TR	JP		TR	JP	
Focal adhesion	9,92E-27	9.47E-23	1	2	102	131	29	24	8	27.58	33.33	N
MAPK signaling pathway	2,05E-23	2.14E-17	2	6	72	121	20	27	4	19.99	14.81	N
Jak-STAT signaling pathway	3,68E-21	6.36E-14	3	17	49	68	20	14	6	29.99	42.85	N
TGF-beta signaling pathway	4,05E-21	1.87E-21	4	3	43	71	15	16	12	79.99	74.99	Y
ECM-receptor interaction	1,43E-20	1.26E-18	5	5	56	49	18	15	9	49.99	59.99	Y
Axon guidance	7,68E-19	5.02E-7	6	74	49	99	11	15	2	18.18	13.33	N
Complement and coagulation cascades	1,00E-18	2.35E-16	7	10	22	29	10	8	3	29.99	37.49	N
Antigen processing and presentation	1,79E-18	1.37E-9	8	43	161	53	14	10	7	49.99	69.99	Y
Proteasome	1,97E-18	1.34E-24	9	1	17	9	4	6	1	24.99	16.66	N
Autoimmune thyroid disease	5,75E-18	7.15E-7	10	76	162	44	15	8	6	39.99	74.99	Y

Table 9. The top 10 KEGG pathways identified for both populations in Behcet's disease. 6 out of the top 10 pathways, identified in both populations are shown in red.

B. Bakir-Gungor, *et. al.*, "*Identification of Possible Pathogenic Pathways in Behçet's Disease Using Genome-wide Association Study Data from Two Different Populations*", 2015, **Eur. Journal of Human Genetics,** 23(5):678-87.

	P-values		Rank		# of Associated SNPs in GWAS	# of SNP	# of	% Common	Is Common
KEGG Term	TR	JP	TR	JP	TR	J			
Focal adhesion	9,92E-27	9.47E-23	1	2	102				
MAPK signaling pathway	2,05E-23	2.14E-17	2	6	72				
Jak-STAT signaling pathway	3,68E-21	6.36E-14	3	17	49				
TGF-beta signaling pathway	4,05E-21	1.87E-21	4	3	43				
ECM-receptor interaction	1,43E-20	1.26E-18	5	5	56				
Axon guidance	7,68E-19	5.02E-7	6	74	49				
Complement and coagulation cascades	1,00E-18	2.35E-16	7	10	22				
Antigen processing and presentation	1,79E-18	1.37E-9	8	43	161				
Proteasome	1,97E-18	1.34E-24	9	1	17				
Autoimmune thyroid disease	5,75E-18	7.15E-7	10	76	162				

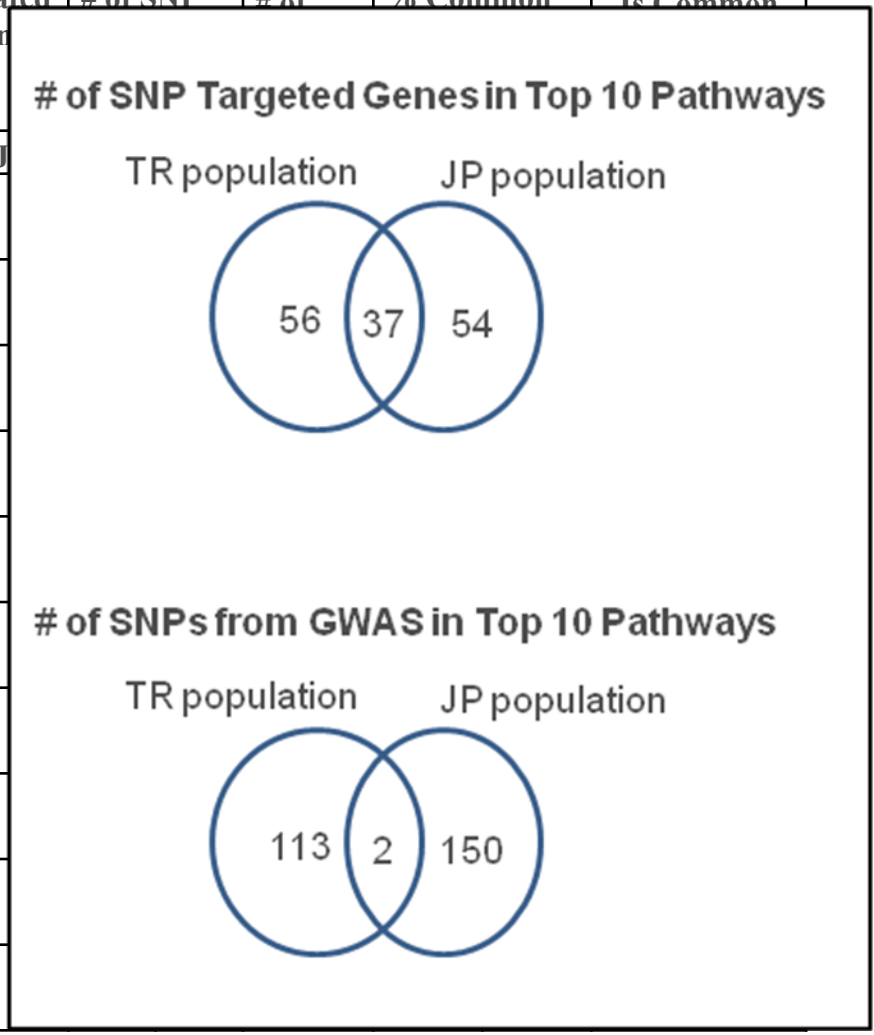


Table 9. The top 10 KEGG pathways identified for both populations in Behcet's disease. 6 out of the top 10 pathways, identified in both populations are shown in red.

B. Bakir-Gungor, *et. al.*, "*Identification of Possible Pathogenic Pathways in Behçet's Disease Using Genome-wide Association Study Data from Two Different Populations*", 2015, **Eur. Journal of Human Genetics,** 23(5):678-87.

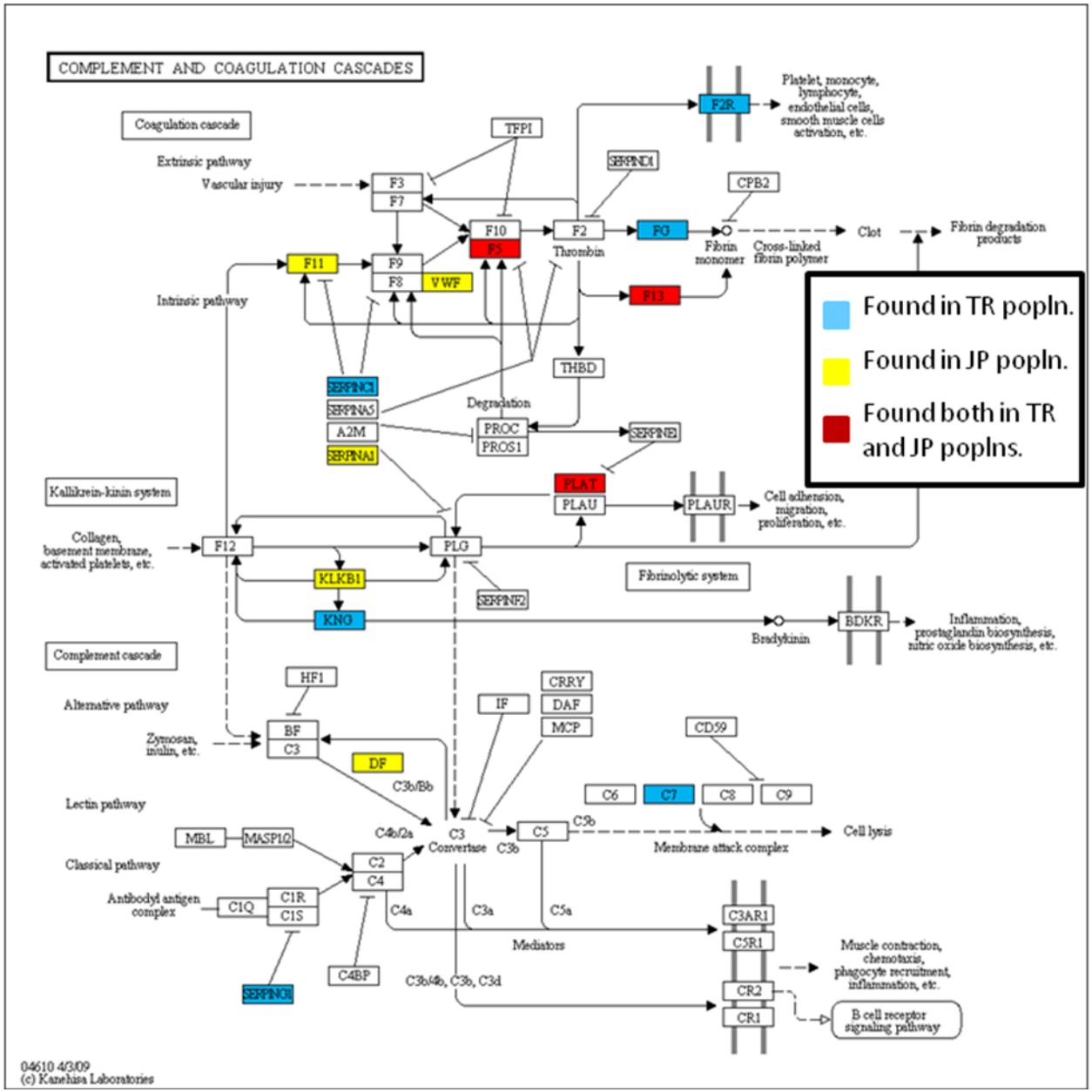


Figure 14. KEGG pathway map for complement and coagulation pathway. The set of genes shown in blue includes genes that are found for TR dataset; yellow includes genes that are found for JP dataset; red includes genes that are found both by TR and JP GWAS of BD.

European Journal of Human Genetics (2015) 23, 1279–1280

www.nature.com/ejhg

LETTERS

'Epistatic interactions between autoimmunity and genetic thrombophilia'

European Journal of Human Genetics (2015) **23,** 1279; doi:10.1038/ejhg.2014.287; published online 28 January 2015

In the recent article by Bakir-Gungor *et al*,[1] a novel method of analysis is proposed to elucidate the genetic pathways that are considered essential in the phenotypic expression of complex diseases, such as Behçet's disease (BD). The combined analysis of the data of two genome-wide association studies (GWAS) that were conducted in the Turkish and Japan populations with BD[2,3] reveals a shared pathway between the complement and the coagulation cascade.

On the basis of the epidemiological data and the diagnostic assessment of three patients with major vessel thrombosis who were hospitalized in our department, we have previously formulated the medical hypothesis that the occurrence of genetic thrombophilia and certain features of the complex spectrum of BD in selected patients with thrombosis may not represent a coincidental coexistence, but rather the core features of a genetically based distinct nosological entity.[4] The role of synergistic epistasis is considered the key in this

The theoretical background of this hypothesis seems to be supported in a preliminary stage through the scientific work of Bakir-Gungor *et al*,[1] and although its data cannot result in safe and comprehensive conclusions, we are strong advocates of similar future studies. Beside, there is now a body of evidence that imply the epistatic interaction between inherited thrombophilia and autoimmunity. In a recent experimental study by Katzav *et al*,[5] it has been demonstrated that when heterozygous and homozygous factor V-Leiden transgenic mice were immunized with antiphospholipid antibodies there have been various autoimmune responses resulting in neurodegenerative manifestations. Further research in the field of GWAS with the methodology presented by Bakir-Gungor *et al*.[1] seems to be the future direction to elucidate the pathways in complex diseases and therefore individualize the treatments or even re-evaluate the classification of certain diseases.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Dimitrios Stoimenis[1], Nikolaos Petridis[1] and Nikos Papaioannou*,[1]
[1]*Department of Internal Medicine, General Hospital "G. Papanikolaou", Thessaloniki, Greece*
E-mail: papaioannou.nikos@gmail.com

1 Bakir-Gungor B, Remmers EF, Meguro A *et al*: Identification of possible pathogenic pathways in Behçet's disease using genome-wide association study data from two different populations. *Eur J Hum Genet* 2015; **23**: 678–687.

Reply to Stoimenis *et al*

European Journal of Human Genetics (2015) **23,** 1279–1280; doi:10.1038/ejhg.2014.288; published online 14 January 2015

We appreciate the comments made by Stoimenis *et al*[1] on our recently published article,[2] describing the application of our novel analysis method to Behçet's disease (BD) genome-wide association study data obtained from the Japanese and Turkish populations. In this study, we analyzed the data in a pathway-related context to identify the disease-related pathways targeted by the single-nucleotide polymorphisms (SNPs).[2] Among the identified pathways, Stoimenis *et al*[1] focus on the complement and coagulation pathway since they identified three BD patients with major vessel thrombosis.

We fully agree with Stoimenis *et al* that in specific ethnic populations there exists a strong prevalence of vascular thrombosis in BD and that there is a positive association between the inherited procoagulant factors and thrombosis in BD.[3] According to our analysis, complement and coagulation pathway ranks as the seventh affected pathway in the Turkish population, whereas it ranks tenth in the Japanese population. Commonly targeted genes in this pathway consist of *PLAT*, *F5* and *F13A1*. All these genes have been previously identified to be associated with BD and thrombosis.[3] Especially the mutations in *F5* gene in Turkish population have been identified to increase the risk of venous thrombosis.[4] Coagulation factor XIII protein is a crucial protein complex in the final step of blood coagulation process. It is made up of two domains produced by two separate genes, *F13A* and *F13B*. *F13A* gene is targeted by the SNPs in both populations, whereas *F13B* gene is targeted only in the Turkish population, creating a higher risk of venous thrombi.[5] All the population-specific SNPs targeting this pathway have different functional impacts yielding to different rankings of this pathway in both the populations.

In conclusion, we agree with Stoimenis *et al* that our method can elucidate the commonly targeted pathways as well as the population-specific pathways that we strongly believe will be the future direction of analysis to elucidate the marker pathways in complex diseases.

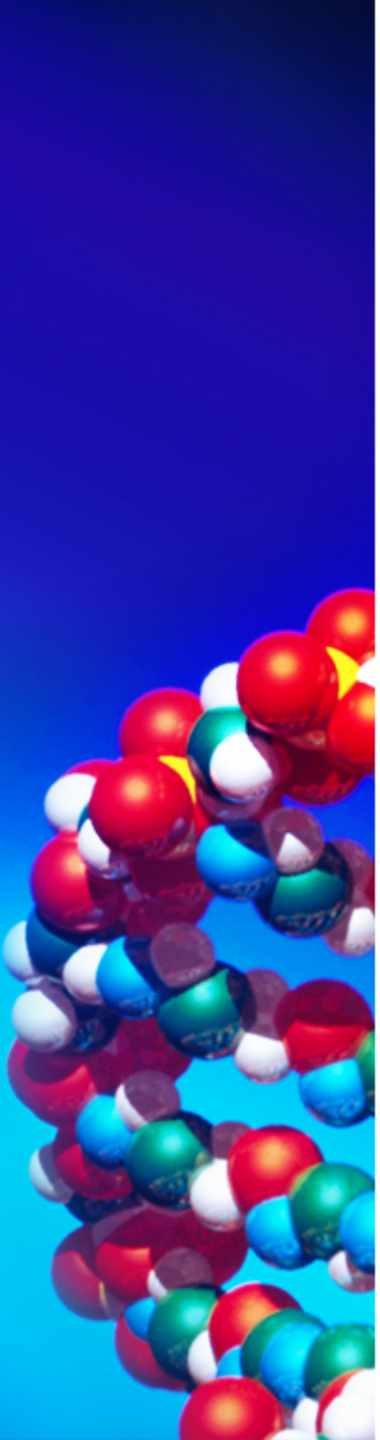
CONFLICT OF INTEREST

The authors declare no conflict of interest.

Burcu Bakir-Gungor*[,1], Elaine F Remmers[2], Akira Meguro[3], Nobuhisa Mizuki[3], Daniel L Kastner[2], Ahmet Gul[4] and Osman Ugur Sezerman[5]

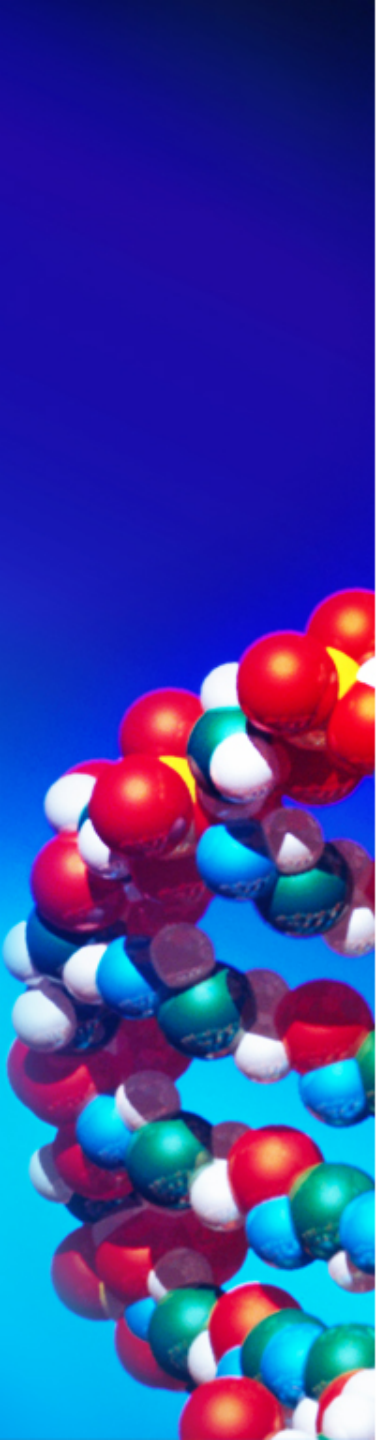
[1]*Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Abdullah Gul University, Kayseri, Turkey;*
[2]*Inflammatory Disease Section, Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA;*
[3]*Department of Ophthalmology, Yokohama City University School of Medicine, Yokohama, Kanagawa, Japan;*
[4]*Division of Rheumatology, Department of Internal Medicine, Istanbul University, Istanbul, Turkey;*
[5]*Department of Biological Sciences and Bioengineering, Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey*
E-mail: burcu.gungor@agu.edu.tr

1 Stoimenis D, Petridis N, Papaioannou N: Epistatic interactions between autoimmunity and genetic thrombophilia. *Eur J Hum Genet* 2015; **23**: 1279.
2 Bakir-Gungor B, Remmers EF, Meguro A *et al*: Identification of possible pathogenic pathways in Behçet's disease using genome-wide association study data from two different populations. *Eur J Hum Genet* 2015; **23**: 678–687.
3 Leiba M, Sidi Y, Gur H, Leiba A, Ehrenfeld M: Behçet's disease and thrombophilia. *Ann Rheum Dis* 2001; **60**: 1081–1085.
4 Gul A, Ozbek U, Ozturk C *et al*: Coagulation factor V gene mutation increases the risk of venous thrombosis in Behcet's Disease. *Br J Rheumatol* 1996; **35**: 1178–1180.
5 Aleman MM, Byrnes JR, Wang JG: Factor XIII activity mediates red blood cell retention in venous thrombi. *J Clin Invest* 2014; **124**: 3590–3600.



Insights

- On Behçet's disease datasets, the identified pathways between two populations show more commonality than individual genes or SNPs. (the probability of getting 6 out of top 10 pathways from existing 246 human KEGG pathways is $2.44E^{-36}$).
- The pathways are critical to elucidate the mechanisms underlying diseases and show higher conservation within and across populations.
- Each individual has a unique combination of factors involved in disease development mechanism.
- But, most of the targeted pathways that need to be altered by these factors are expected to be the same.



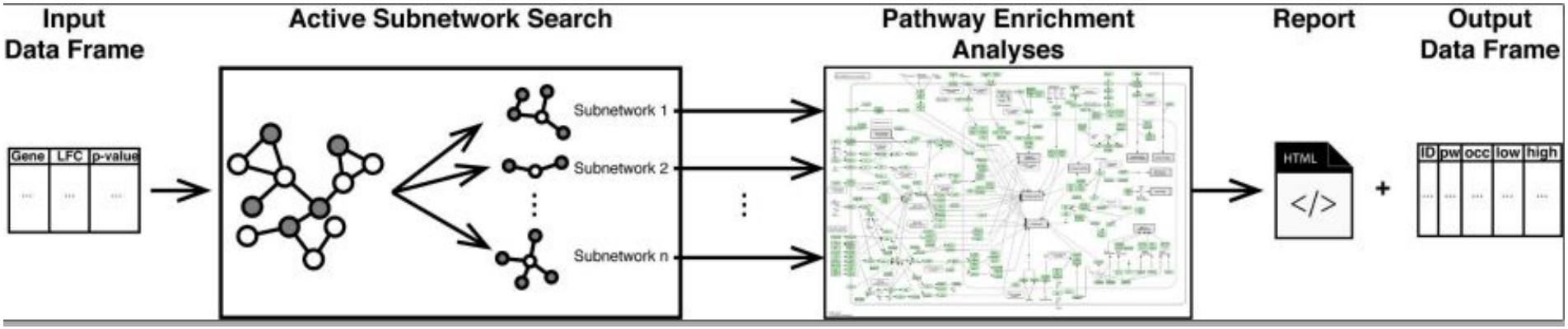
Contributions

- "Fortunately, a portion of the unaccounted 85 to 90% disease variation lies hidden in GWAS datasets but can be revealed using NEW strategies." (Schadt *et al*, Science Translational Medicine).

- For GWAS analysis of complex diseases, novel disease-susceptibility genes and mechanisms can only be identified by looking beyond the tip of the iceberg (the most significant SNPs/genes).

- Our results show that incorporating SNP functional properties, protein-protein interaction networks into GWAS can dissect leading molecular pathways, which cannot be picked up using traditional analyses.

pathfindR- Pathway Enrichment Analysis Utilizing Active Subnetworks



Ulgen E, Ozisik O, Sezerman O.U, pathfindR- An R Package for Pathway Enrichment Analysis Utilizing Active Subnetworks. BioRxiv., 2018.

Integrative analysis of transcriptomics and epigenomics data using PANOGA

Table 7 Pathways that are shared by at least three types of cancers

Pathway name	Breast	Thyroid	Prostate	Colon
ErbB signalling pathway	7.75E-12	1.03E-09	9.74E-09	4.42E-07
Complement and coagulation cascades	2.70E-12	1.90E-13	–	7.84E-13
ECM-receptor interaction	7.38E-10	–	4.69E-11	1.20E-13
Focal adhesion	9.32E-09	5.30E-12	–	2.64E-07
Chronic myeloid leukaemia	8.94E-06	1.41E-09	2.10E-04	–
Neurotrophin signalling pathway	1.37E-05	1.84E-08	–	1.72E-06
T cell receptor signalling pathway	1.11E-06	1.77E-05	1.48E-07	–
Glioma	5.49E-08	1.47E-04	7.90E-04	–
ıcer	2.69E-05	2.31E-05	3.04E-06	–
vasion of epithelial cells	1.22E-05	2.31E-06	3.51E-04	–
cell lung cancer	4.71E-04	3.49E-05	6.57E-04	–
genic right ventricular cardiomyopathy (ARVC)	2.63E-04	–	5.82E-05	1.73E-04

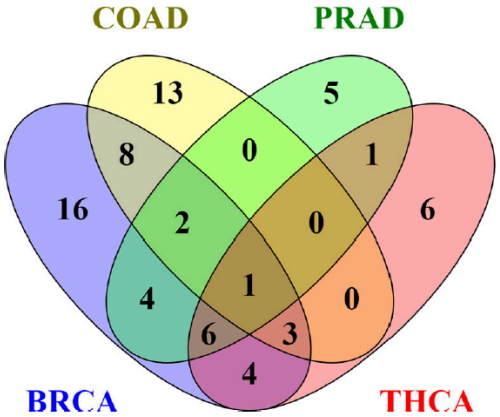
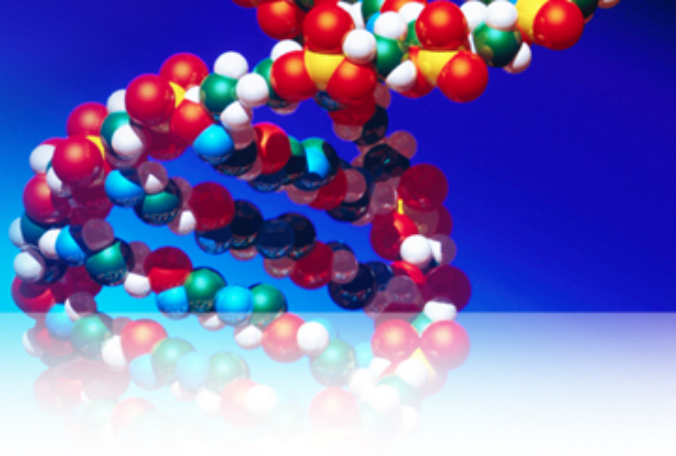


Fig. 6 Venn diagram showing significantly altered pathways for all cancer types. The Venn diagram of significantly altered pathways

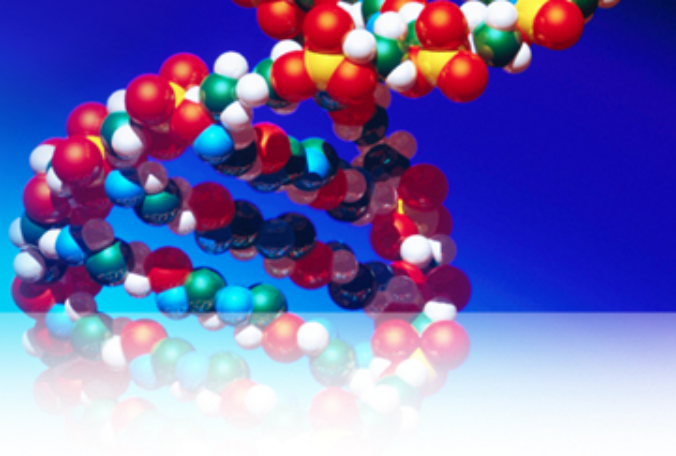
ficantly altered (Bonferroni score < 0.01) pathways that are shared by at least three types of cancers
onferroni scores associated with each dataset. Most interestingly, the ErbB signalling pathway was
nificantly altered for all cancer types in our analysis; hence, the ErbB pathway may be the key factor
ous types of cancers

Summary



- "Therapeutics of the future likely will be designed via keeping cellular networks and pathways in mind." (Collins *et al*, Science Translational Medicine).
- In complex diseases, while individual SNPs/genes are not shared by most of the patients, pathways show more commonality, especially across populations.
- We introduced pathway marker concept to the literature, which explains universal disease development mechanism.
- As a potential application, each population may search for disease causing factors targeting the genes within these marker pathways.

Summary

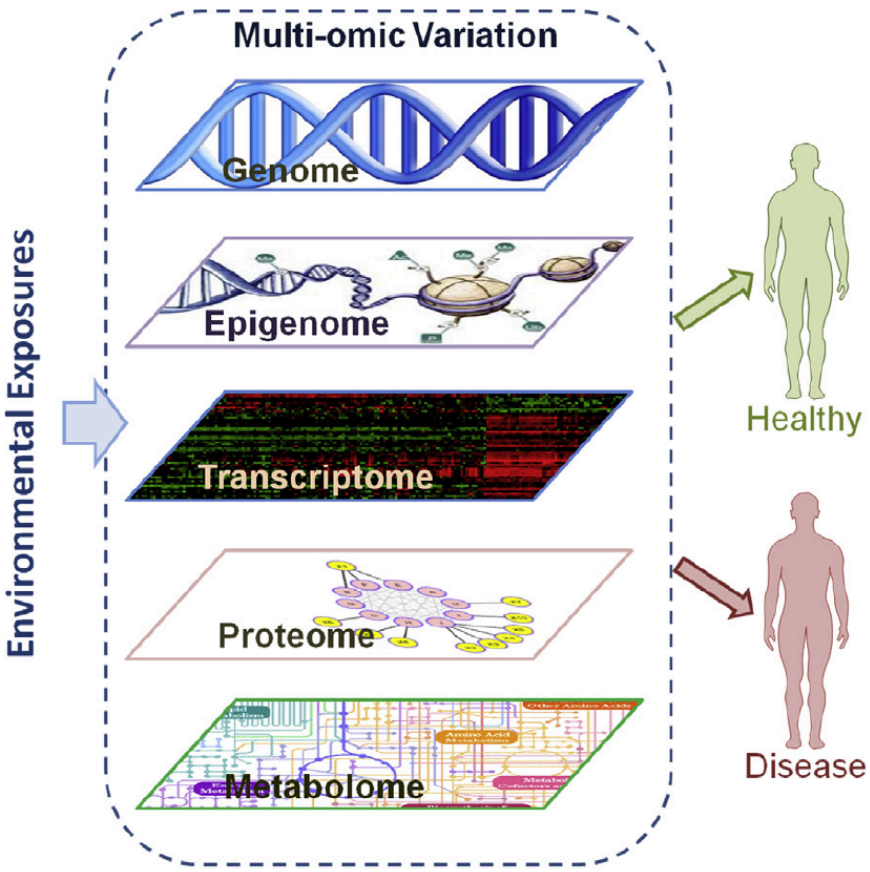


- Pathway markers can also be extended to individual level to identify modifications occurring on the genes within these pathways.

- To understand individual disease development mechanisms, marker pathways can be scanned for an individual for alterations in the functions of the genes contained within.

- Thus, determining the disease-causing factors will provide a valuable insight for personalized therapy targets that would rectify the impact of these function altering factors.

Current Research Interest: Integrative –omics Data Analysis



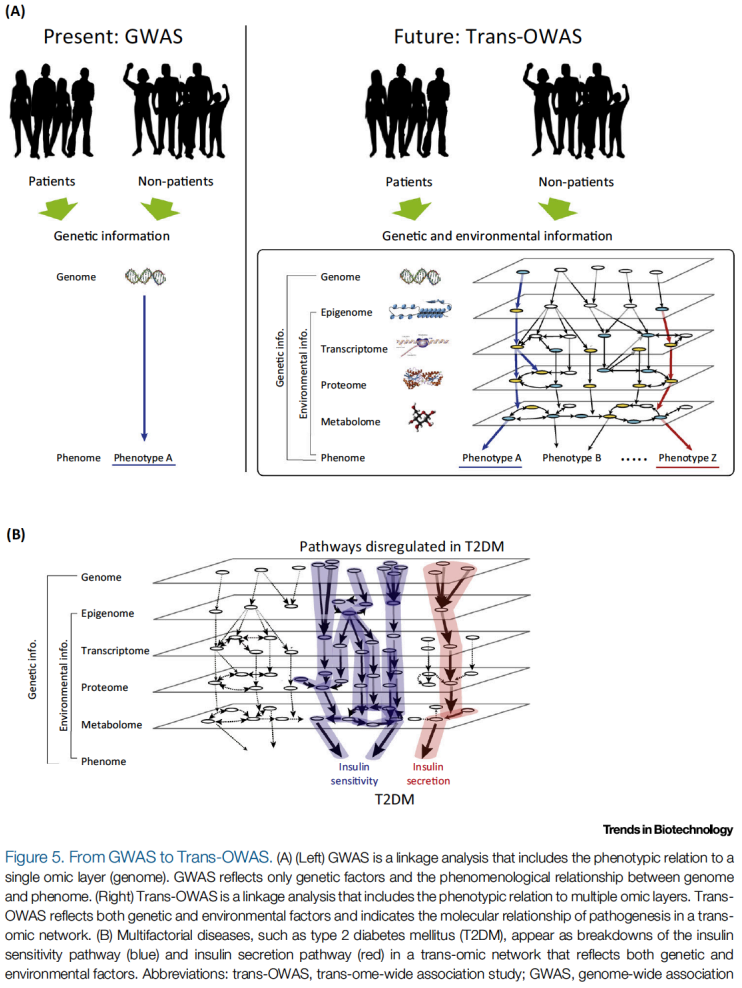


Figure 5. From GWAS to Trans-OWAS. (A) (Left) GWAS is a linkage analysis that includes the phenotypic relation to a single omic layer (genome). GWAS reflects only genetic factors and the phenomenological relationship between genome and phenome. (Right) Trans-OWAS is a linkage analysis that includes the phenotypic relation to multiple omic layers. Trans-OWAS reflects both genetic and environmental factors and indicates the molecular relationship of pathogenesis in a trans-omic network. (B) Multifactorial diseases, such as type 2 diabetes mellitus (T2DM), appear as breakdowns of the insulin sensitivity pathway (blue) and insulin secretion pathway (red) in a trans-omic network that reflects both genetic and environmental factors. Abbreviations: trans-OWAS, trans-ome-wide association study; GWAS, genome-wide association

Yugi K., et. al, "Trans-Omics: How To Reconstruct Biochemical Networks Across Multiple 'Omic' Layers.", Trends Biotechnol. 2016 Apr;34(4):276-290.

Multidimensional -omics Data Integration Methods

TABLE 1 | Comparison of multidimensional data integration methodologies discussed in the manuscript.

Method category	Brief description	Advantages	Limitations	Representative tools
Clustering/ dimensionality reduction-based approaches	Transform data into common space through graph or kernel-based methods	Easy to implement using common statistical techniques; retain within-data properties; robust to different units of measurements and different data sets from the public domain	Cross-data interaction may be altered; application limited to visual overview of data and detection of subpopulations	Clustering-based: iCluster (21); ICM (22); TMD (23); SNF (24) Dimensionality reduction: Biofilter (25); CIA/MCIA (26); FALDA (27); GMDR (28)
Predictive modeling approaches	Machine learning based methodologies to predict prognosis or diagnosis and discover biomarkers	High predictive power; versatile methodologies; data-driven approach (does not require preexisting knowledge of omics interaction)	Overfitting issue; can require high computational power; does not integrate biological knowledge; higher accuracy requires larger data sets	Camelot (29); Kernel fusion (30); sMBPLS (31); MDI (32); PARADIGM (33); DIVIAN (34)
Pairwise omics data integration	Centered on interaction information between pairs of omics data	Easy to implement; reflects inter-omics interaction; causal implication	Available data dominated by expression quantitative trait loci (eQTLs); low robustness of *trans*-association signal	MERLIN (35); RAREMETAL (36); EMMA (37); GEMMA (38); PLINK (39); Matrix eQTL (40); SMR (41)
Network-based approaches	Reduce data complexity by converging multi-omics information onto networks	Networks can accommodate multiple layers of data; intuitive depiction and visualization of regulatory circuits	Computationally expensive; difficult to model feedback loops in multidimensional space	Weighted gene coexpression network analysis (42); MEGENA (43); Bayesian networks (44); TIGRESS (45); ARACNE (46); TIE* (47); GENIE3 (48); mixOmics (49)
Composite approaches	Flexible integration of multiple integration models	Flexibility and adaptability to diverse research needs	Few well-acknowledged frameworks available	Analysis Tool for Heritable and Environmental Network Associations (50, 51); Mergeomics (3, 52)

Arneson D, et. al, “Multidimensional Integrative Genomics Approaches to Dissecting Cardiovascular Disease”, 2017, Front Cardiovasc Med. 2017 Feb 27;4:8.

Multidimensional -omics Data Integration Methods

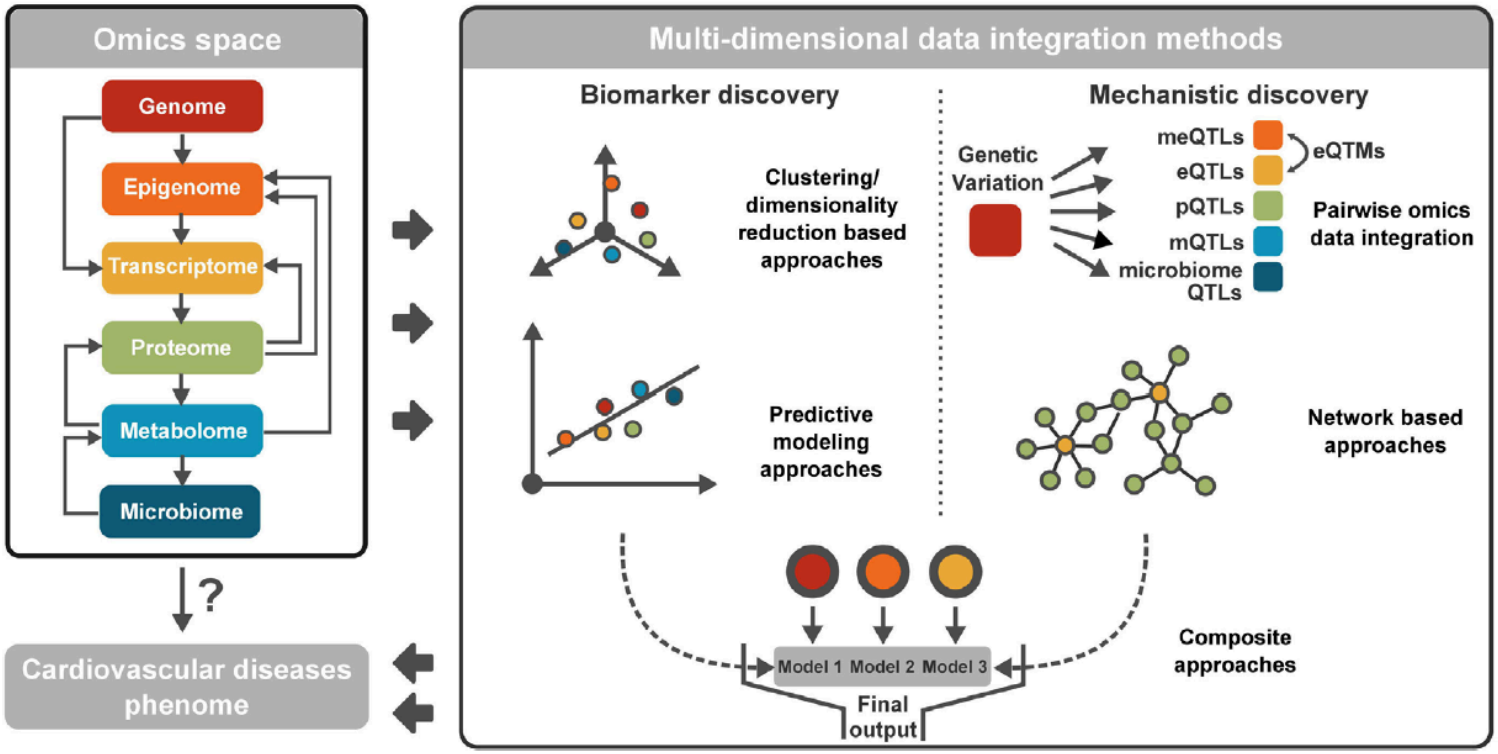
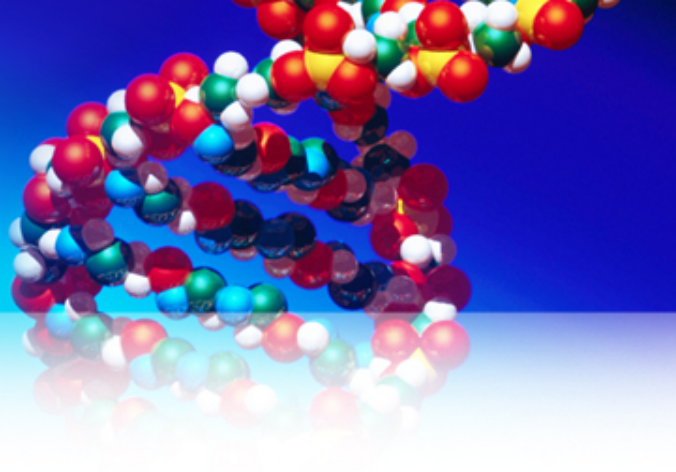


FIGURE 1 | Summary of different omics data types and multidimensional data integration methods. Cardiovascular disease (CVD) involves various omics spaces and complex inter-omics interactions. To discover accurate biomarkers and disentangle disease mechanisms of CVD, multidimensional data integration methods are available, broadly categorized into clustering/dimensionality reduction-based approaches, predictive modeling approaches, pairwise omics data integration, network-based approaches, and composite approaches integrating multiple modeling approaches.

Arneson D, et. al, "Multidimensional Integrative Genomics Approaches to Dissecting Cardiovascular Disease", 2017, Front Cardiovasc Med. 2017 Feb 27;4:8.

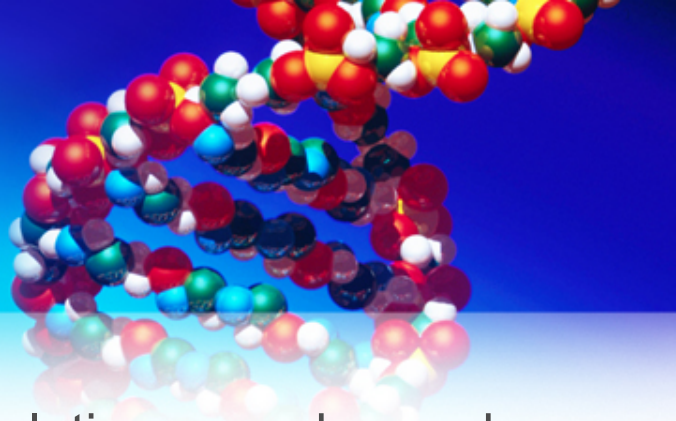
Conclusions



- PANOGA (Pathway and Network-Oriented GWAS Analysis) combines nominally significant evidence of genetic association with current knowledge of biochemical pathways, protein–protein interaction networks, and functional information of selected single nucleotide polymorphisms (SNP).

- With its multifactorial basis, we have shown on four complex diseases that PANOGA has a good potential to decipher the combination of biological processes underlying the disease.

Conclusions

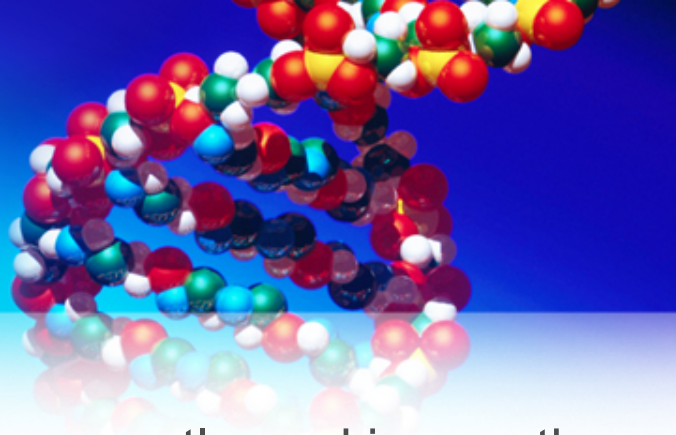


- Via comparing GWASs of two different populations, we have shown that the few SNPs that are identified in GWAS and their associated genes are mostly targeting the same pathway combinations, and these biological pathways show higher conservation across populations.

- If the combination of these pathways does not function properly, a specific disease may develop.

- Although PANOGA is originally developed to identify disease-associated pathways via further analyzing GWAS data, later it is shown to work well on different -omics datasets.

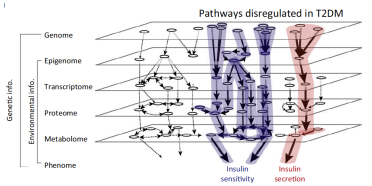
Conclusions

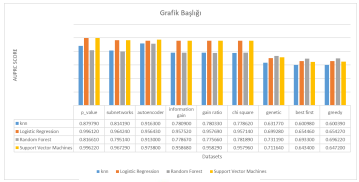


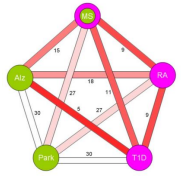
- Using different –omics datasets, our group is currently working on the development of methodologies to extend this approach to individual level to identify specific modifications occurring on the genes within these identified pathways.
- Dissecting the individual disease development mechanisms will provide a valuable insight for discovering individualized therapy targets and will pave the way towards personalized medicine applications.
- This approach would enable biomedical researchers to identify affected pathways and function-altering factors within these pathways.
- For diagnostic purposes, the identification of the disease-related pathways is also instrumental in the determination of biomarkers at different levels (e.g., SNPs, gene expression levels, protein levels in serum, miRNA levels, metabolite concentration).

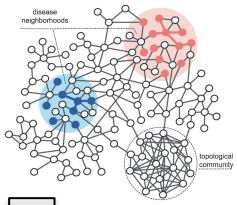
Ongoing Research

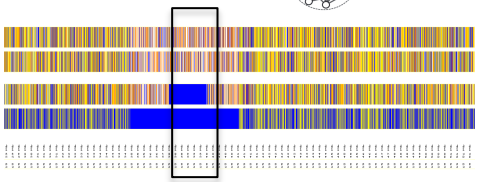
- Merge–omics using Pathway and Network Oriented Integration
- The Identification of Discriminative Single Nucleotide Polymorphisms for the Classification of Behçet's Disease
- Identification of Commonly Affected Pathways in Psychiatric Diseases
- Comparative Analysis of Disease Specific Sub-Network Identification Algorithms
- Homozygous Stretch Identification from Next Generation Sequencing data (HomSI)
- Machine Learning Analysis of Inflammatory Bowel Disease-Associated Metagenomics Dataset

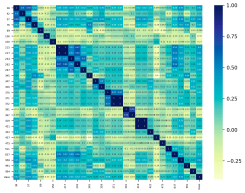


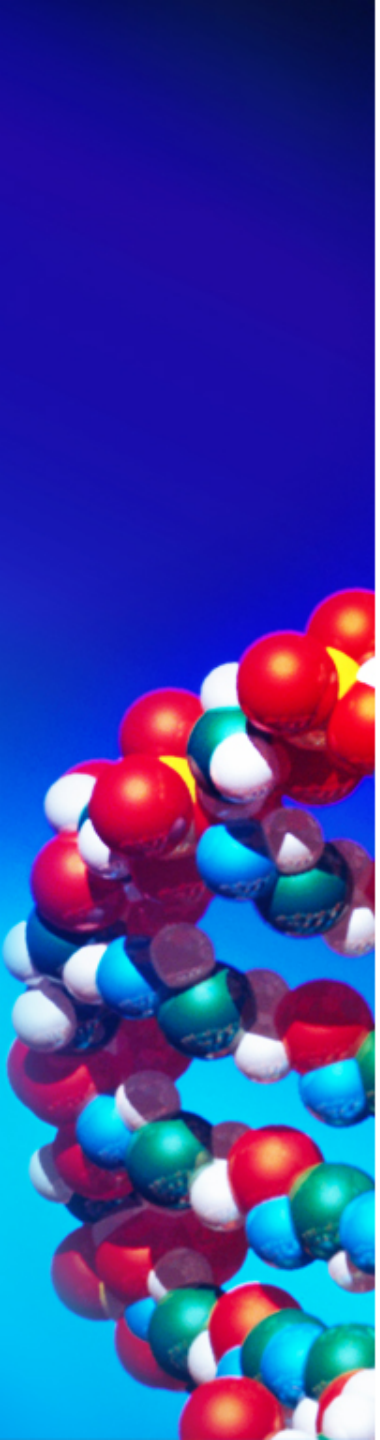












Acknowledgements

- Uğur Sezerman
- Scott Saccone
- Phil Hyoun Lee
- Claude Chelala
- Gabriela Bindea
- Christine Nardini
- Sergio Baranzini

PPI network dataset:

- Albert-László Barabási
- Michael Cusick

Microarray dataset:

- Boris Krischek
- Hirofumi Nakaoka

GWAS datasets:

- John Bowes
- Simon Potter
- Dalia Kasperaviciute
- Murat Günel
- Katsuhito Yasuno
- Ituro Inoue
- Ahmet Gül
- Akira Meguro

NGS datasets:

- Fowzan S. Alkuraya
- Boris Dominik Seelow
- Ian M. Carr
- Fatih Bayraklı

Acknowledgements

References

- ***B. Bakir-Gungor,*** E.F. Remmers, A. Meguro, N. Mizuki, D.L. Kastner, A. Gul, O.U. Sezerman, "*Identification of Possible Pathogenic Pathways in Behçet's Disease Using Genome-wide Association Study Data from Two Different Populations"*, 2015, **European Journal of Human Genetics,** 23(5):678-87.
- ***B. Bakir-Gungor,*** E.F. Remmers, A. Meguro, N. Mizuki, D.L. Kastner, A. Gul, O.U. Sezerman, "*Reply to Stoimenis et. al."*, 2015, **European Journal of Human Genetics,** 23(10):1280**.**
- ***B. Bakir-Gungor,*** E. Egemen, O.U. Sezerman, *"PANOGA: a web-server for identification of SNP targeted pathways from genome-wide association study data*", 2014, **Bioinformatics**, 30(9): 1287-1289.
- ***B. Bakir-Gungor***, O.U. Sezerman, "*The Identification of Pathway Markers in Intracranial Aneurysm Using Genome-wide Association Data from Two Different Populations*", 2013, **PLoS ONE,** 8(3): e57022.
- ***B. Bakir-Gungor***, B. Baykan, S. Ugur İseri, F.N. Tuncer, O.U. Sezerman, "The Identification of SNP Targeted Pathways in Partial Epilepsies Using Genome-wide Association Data", 2013, **Epilepsy Research**, 105(1-2): 92-102.
- ***B. Bakir-Gungor***, O.U. Sezerman, "*A New Methodology to Associate SNPs with Human Diseases According to Their Pathway Related Context*", 2011, **PLoS ONE**, 6(10): e26277**.**
- ***B. Bakir-Gungor,*** O.U. Sezerman, "*Identification of SNP Targeted Pathways From Genome-wide Association Study (GWAS) Data"*, 2012, **Nature Protocol Exchange**. doi: 10.1038/protex.2012.019.
- Gormez, Z., Bakir-Gungor, B., Sagiroglu, M.S. (2014) HomSI: a homozygous stretch identifier from next-generation sequencing data, *Bioinformatics*, **30**(3), 445-7.
- O. Ozisik, **B. Bakir-Gungor,** B. Diri, O.U. Sezerman, "Active Subnetwork GA: A Two Stage Genetic Algorithm Approach to Active Subnetwork Search", 2017, ***Current Bioinformatics***, 12(4): 320-328(9).
- Ulgen E, Ozisik O, Sezerman O.U, pathfindR- An R Package for Pathway Enrichment Analysis Utilizing Active Subnetworks. BioRxiv., 2018.

Thanks

Burcu Bakir-Gungor, Ph.D.

Assistant Professor
Department of Computer Engineering
Abdullah Gul University

burcu.gungor@agu.edu.tr

Visit us at Cappadocia