Molecular Omics

RESEARCH ARTICLE

The human GPCR signal transduction network

Panagiota Kontou[1], Athanasia Pavlopoulou[2], Niki Dimou[3], Margarita Theodoropoulou[4], Georgia Braliou[1], Georgios Tsaousis[4], Georgios Pavlopoulos[5], Stavros Hamodrakas[4], Pantelis Bagos[1*]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

The eukaryotic cell surface G protein-coupled receptors (GPCRs) interact with a wide spectrum of ligands. The intracellular transmission of the extracellular signal is mediated by the selective coupling of GPCRs to G proteins, which, in turn, activate downstream effectors. GPCRs are of paramount pharmacological importance, with approximately 40% of all commercial drugs targeting these proteins. Herein, we have made an effort to unravel the molecular mechanisms underlying the GPCR-mediated signaling pathway and the way this pathway is associated with diseases. Network-based approaches were utilized to delineate the GPCR pathway, incorporating data from gene expression profiles across eleven healthy tissues and disease-gene associations from three diverse resources. The associations between the tissue-specific expression profiles of the disease-related genes along with the relative risk of disease development were further investigated. In the GPCR-activated pathway, the signal was found to be amplified at the successive steps of the pathway so that the effector molecules are highly expressed compared to ligands. This amplification effect was more pronounced when the respective genes encoding the particular proteins were associated with diseases. It was also found that co-expressed genes, corresponding to interacting molecules in affected tissues, may constitute powerful predictive markers for disease development. A disease risk prediction model based on tissue-specific expression profiles of the disease-associated genes was also generated. These findings could be applied to clinical settings.

Introduction

G protein-coupled receptors (GPCRs) constitute the largest group of eukaryotic cell surface receptors, spanning the major taxonomic divisions[1], with a common structure of seven transmembrane α-helices[2]. The human genome encodes approximately 800 GPCRs, which, according to a recent classification system, can be grouped into six major classes based both on sequence and function similarity[3]. GPCRs interact with a wide range of native ligands (peptides and proteins, prostaglandins, neurotransmitters, photons, hormones, ions, pheromones)[4, 5]. Upon such interaction, the conformational equilibrium of GPCRs is shifted[6, 7] and, subsequently, the extracellular signal is transmitted within the cell through the coupling with the intracellular heterotrimeric GTP-binding-proteins (G proteins)[8]. These proteins form heterotrimeric complexes composed by the G_α, G_β, and G_γ subunits [9, 10]. GPCRs activate an associated G protein by promoting the exchange of its bound GDP for a GTP. This results to the dissociation of the Gα subunit, along with the bound GTP, from the Gβ/Gγ dimer to further activate downstream effector molecules, such as ion channels and enzymes, and lead to diverse biochemical, physiological and cellular responses [11-13]. GPCR-mediated signal transduction pathways are implicated in a plethora of diseases[14-16]. As a result, GPCRs are of enormous pharmacological importance, with approximately 40% of all marketed drugs targeting these proteins[17, 18]. Moreover, crystallographic studies of GPCRs during the last 15 years has experienced exponential growth, resulting in the determination of more than100 structures of 28 different GPCRs from various classes [19, 20]. This progress on GPCR structural studies shed light on molecular mechanisms of GPCR ligand recognition, activation and allosteric modulation, as well as structural basis of GPCR dimerization[21]. This growth has also triggered the development of specialized protein resources that gather, curate and make available to the public, information regarding GPCR crystal structures, family classification, receptor mutants and information concerning the GPCRs coupling to G-proteins [22-28]. Taken all this into account it is understood that elucidating the underlying molecular mechanisms of the GPCR-mediated signaling pathway and the way this pathway is associated with diseases is of paramount biological and medicinal importance.

The objective of this study is to study the system ligand<GPCR<G protein<effector, henceforth referred to as 'GPCR signal transduction network' or 'GPCR signaling system'. Certain phenotypes or pathological manifestations are attributed to genes that form functional modules, whereas loss of a single gene may disrupt the functional activity of the

[1]*Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia 35 131, Greece*
[2] *Izmir International Biomedicine and Genome Institute (iBG-Izmir), Dokuz Eylül University, 35340, Turkey*
[3]*Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Stavros Niarchos Av. Ioannina 45110, Greece*
[4]*Department of Cell Biology and Biophysics, Faculty of Biology, University of Athens, Panepistimiopolis, Athens 157 01, Greece*
[5]*Berkeley Lab, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA*
**Author to whom correspondence should be addressed*
Department of Computer Science and Biomedical Informatics
University of Thessaly, Papasiopoulou 2-4,
Lamia, 35131, Greece
Tel.: +30-223-106-6914; Fax: +30-223-106-6915.
E-mail: pbagos@compgen.org

complete module[29-31]. To this end, functional networks comprising relationships of the four types of molecules of the GPCR signaling system were utilized. A global all-against-all network was constructed, as well. We further integrated information of tissue-specific gene expression profiles to enhance the biological significance of our study. Tissue-specific networks were generated in order to identify functional modules present in particular tissues. Another objective of this study was to investigate how the tissue-specific expression profiles of the disease-associated genes correlate with the relative risk of developing a disease. Characteristic molecular pathways that are associated with pathophysiological processes were identified, as well. Furthermore, associations between components of the GPCR signaling system and other proteins, as well as known drugs, were examined. Finally, the potential molecular associations among the Gα, Gβ, and Gγ subunits of the G protein heterotrimer were investigated.

Materials and Methods

All human GPCRs, G-proteins (Gα subunit) and effectors, as well as their corresponding pairwise associations, were collected from Human-gpDB [26]. The classification of GPCRs was retrieved from IUPHAR/BPS Guide to PHARMACOLOGY [32] and Human-gpDB [26]. Interactions between peptide ligands and GPCRs were found by an extensive literature search.

Gene expression profiles determined in 11 normal human tissues (hypothalamus, spleen, ovary, lung, liver, kidney, heart, colon, adipose, testes and skeletal muscle) were obtained from the RNA-Seq Atlas [33].

Gene-disease association was investigated by utilizing information from three different publicly available databases: Genetic Association Database (GAD)[34], Genome Wide Association Studies (GWAS)[35] and Online Mendelian Inheritance in Man (OMIM)[36]. Due to disease name heterogeneity and ambiguity, a consistent nomenclature and classification for diseases was needed, and thus the naming conventions described in the International Classification of Diseases (ICD) were used. In order to maintain a uniform nomenclature across all datasets, all gene names were converted to the official HGNC (HUGO Gene Nomenclature Committee)[37] gene symbols. Integrating these data from different sources, a gene-disease network was constructed covering different types of gene-disease associations, ranging from rare monogenic disorders to multifactorial diseases [38].

The diseases were mapped to their corresponding tissues based on information retrieved through database and biomedical literature mining. To this end, PhenoDigm[39, 40], a database that includes curated annotations of tissue-phenotype associations, was searched. In addition, PUBMED was mined using combinations of a given disease term with each one of the 11 tissue terms. The top 20 most relevant articles were read thoroughly to identify any relationships between the given disease and tissue.

For each type of molecule (peptide ligands, GPCRs, Gα, effectors), the number of drugs interacting with them was retrieved from UniProt, which is cross-referenced to DrugBank [41], a knowledgebase that contains information on drugs and drug targets. To this end, all entries in DrugBank referred to each corresponding protein entry in UniProt were counted in order to calculate the total number of interacting drugs.

The well annotated and experimentally verified interactions among the four types of molecules, as well as with other non-GPCR network component proteins, were retrieved from UniProt [42]. In particular, these protein–protein interactions (PPI) were derived automatically from Intact [43], a database of PPI data.

Given that experimentally verified data regarding interactions among the various G_α, G_β, and G_γ subunits that comprise the G-protein heterotrimer are not generally available, the potential interactions among these component proteins were investigated through the construction and analysis of a gene co-expression network [44]. In this way, genes of known similar function can be grouped together based on their co-expression patterns [45]. The expression profiles of the G_α, G_β, and G_γ subunits in the 11 normal human tissues from the RNA-Seq Atlas were used to construct a gene co-expression network. As a gene co-expression similarity measure, the Pearson correlation coefficient was calculated for all pairs of expression profiles ($cor(x_i, x_j)$). This similarity measures the level of concordance among gene expression profiles across tissues. To define a signed co-expression measure among gene expression profiles, a simple modification of the Pearson correlation was used:

$$s_{ij}^{signed} = \frac{1+cor(x_i, x_j)}{2}$$

The resulting similarity matrix was transformed into an adjacency matrix which encodes the connection strength between each pair of nodes. The signum function was used as an adjacency function, which implements 'hard' thresholding (dichotomizing) involving the threshold τ :

$$a_{ij} = signum(s_{ij}, \tau) \equiv \begin{cases} 1 & if \quad s_{ij} \geq \tau \\ 0 & if \quad s_{ij} < \tau \end{cases}$$

Consequently, an unweighted gene co-expression network was constructed, where the different G_α, G_β, and G_γ subunits are represented as nodes and the co-expression relationships as links.

The RNA-Seq Atlas[33] gene expression profiles were visualized graphically with the HeatmapGenerator software package [46]. Cytoscape v.3.2.1 [47] was employed for the statistical analyses, processing and visualization of the network data. The network analysis was focused on the topological properties of the GPCR signal transduction network. The Markov Cluster Algorithm, MCL [48, 49], an unbiased cluster algorithm for graphs implemented in Cytoscape was used to identify clusters in the generated networks.

A logistic regression model was constructed to predict the relative risk of developing a disease for the case of the pairwise interactions of the components of the GPCR signaling system (i.e. ligands-GPCRs, GPCRs-G-proteins and G-proteins-effectors). We calculated robust standard errors accounting for the number of the pairwise interactions. In particular, the model was adjusted for the type of interacting molecules, the tissue, the logarithmic mean of the product of the expression values, the sum of the number of drugs, the sum of the total number of interactions and the sum of the outgoing edges coming from a node and the incoming edges onto a node per tissue of the implicated genes. All statistical analyses were performed with the Stata 13 statistical software package [50].

Results

The components of the GPCR signaling system interact with each other through pairwise associations. Collectively, 964 molecules and 2471 unique pairwise associations among them were retrieved from Humam-gpDB. By comparing the names of the retrieved molecules with the gene terms in RNA-Seq Atlas, 483 unique molecule entries and their corresponding associations were retained. Since in 32 out of the initial 483 corresponding genes the expression value was equal to 0 in all tissues, these entries were removed from the subsequent steps of this study. In addition, 18 molecules that were linked exclusively to these 32 molecules were removed. Therefore, a total of 433 molecules and 1666 pairwise interactions were analyzed in this study. The distribution of the 433 molecules per tissue is shown diagrammatically in Fig. 1A. Of the 433 molecules, over half represent GPCRs (Fig. 1A). On the contrary, G proteins are under-represented, with only 13 members.

Tissue-specificity of gene expression

The tissue-specific expression patterns of the 433 corresponding genes were determined by utilizing the expression data provided for the 11 healthy tissues in RNA-Seq Atlas. Seven genes (i.e. *AGT (angiotensinogen)*, *ANXA1 (annexin A1)*, *APP (amyloid beta precursor protein)*, *C3 (complement component 3)*, *GNAI2 (G protein subunit alpha i2)*, *KNG1 (kininogen 1)* and *SAA1 (serum amyloid A1)*) had extremely high levels of expression. The expression level of a gene in a particular tissue was estimated by calculating the average of the logarithmic values of gene expression per tissue. Paradoxically, G proteins display, on average, the highest level of expression (Fig. 1B). Effectors exhibit slightly lower level of expression compared to G proteins across tissues. On the contrary, GPCRs display the lowest level of expression in all tissues (Fig. 1B). This is probably due to the fact that different ligands display differential binding affinities for GPCRs [51], thereby leading to distinct expression levels. Furthermore, the highest levels of expression for G proteins and effectors were found in the hypothalamus, whereas for ligands and GPCRs were found in the liver and the spleen, respectively (Fig. 1B).

Networks

A joint network including all pairwise molecular interactions across tissues was constructed which is shown in Fig. 2 using four different types of visualization.The joint network is a directed network, where each node has two different degrees (k), that is, number of edges connected to the node. The out-degree (k_{out}) indicates the number of outgoing edges coming from a node, and the in-degree (k_{in}) is the number of incoming edges onto a node. The total degree of the node (k_{tot}) is the sum of its out- and in-degree ($k_{out} + k_{in}$) [52]. The overall topology of the joint network is best described as a highly non-uniform scale-free, in which the probability $P(k)$ that a given node has k edges approximates a power law [53] (Fig. 3A). Therefore, the nodes of the joint network have varying degrees, with few nodes (hubs) having a greater number of edges compared to the rest.

In a similar manner, networks specific for each tissue were generated by taking into account the expression profiles of the genes encoding the four types of molecules in each tissue. The eleven tissue-specific networks displayed the same architecture[53] as the joint network (Fig. 3B, 3C). An unsupervised graph clustering technique [48] was employed for the detection of functional modules in these networks. In each of the tissue-specific networks, a large cluster of associated molecules and several smaller clusters were detected, consistent with their scale-free property [53]. These 11 networks are slightly denser compared to the joint network, since the values of total connectivity/density and centralization (Table S1) are marginally higher compared to the ones of the joint network [52]; this means that more edges are connected to a given node in these networks compared to the same node in the joint network. The above observations lead to the suggestion that specific genes are overexpressed in particular tissues and are associated with each other to form a complex that exerts their functions.

All nodes of the joint network were ranked based on the number of their degree centrality. The top ten ranking hubs per molecule are listed in Table 1. Among them, 40% of ligands and G proteins, as well as 30% of GPCRs and effectors, are associated with diseases. Notably, all ligands belong to the wingless-type MMTV integration site family (WNT)[54, 55], whereas all effectors belong to the tubulin (TUB) family of globular proteins[56] (Table 1). Furthermore, the G proteins GNAI2, GNAI3 (G protein subunit alpha i3), GNAS (GNAS complex locus) had, by far, the highest total degree (k_{tot}) distribution in all tissues (Table 1).

Expression profiles of disease-associated genes

As mentioned above, in a recent study, data for gene-disease associations from three diverse resources were integrated[57], and a list of disease-related genes was created (Table S2). Approximately one-third (28.6%) of the total number of genes encoding the four types of molecules were found to be associated with one or more diseases/classes of diseases (Fig. 1A). Of those, endocrine, nutritional or metabolic diseases and diseases of the circulatory system were the most frequent. Of particular importance, less than half of the total number of genes (i.e. 57 genes) were found to be associated with a tissue-specific disease. This relatively low number is probably due to the fact that the gene expression was investigated in a limited number of 11 tissues that were available in the RNA-Seq Atlas. Notably, a tendency of overexpression of ligands, GPCRs and G proteins in disease states was displayed (Fig. 1B), consistent with the critical role of these molecules in disease [14-16].

Drugs and interactions of the GPCR network components

For each type of molecule (ligands, GPCRs, G proteins and effectors), the number of drugs targeting them and the total number of interactions were retrieved. The highest number of drugs was recorded for GPCRs, from the four types of molecules, with a mean value approximately equal to 8, ranging from 0 to 85, whereas no drug was found to target G proteins. The number of drugs targeting both ligands and effectors varied from 0 to 5. Furthermore, effectors were found to have the highest number of protein-protein interactions (PPIs) with a mean value approximately equal to 28; less than half interactions were calculated for G proteins (mean value equal to 12.5). A mean value of 6 interactions was estimated for ligands, whereas the number of interactions ranged from 0 and 77 for GPCRs.

Gene co-expression patterns

In this study, the corresponding genes of the interacting molecules expressed in the same tissue are termed as co-expressed genes [58]; hereafter referred to as gene pairs. These gene pairs are: ligand-GPCR (20.6%), GPCR-G protein (63.3%) and G protein-effector (16.1%). The gene co-expression patterns were investigated in nondisease, as well as disease states, since pathophenotypes are largely restricted to certain tissues [59]. More than half of the gene pairs (55%) were related to a disease (i.e. either one or both genes of the gene pair were associated with a disease). In one out of ten gene pairs, both genes were related to a disease. Among them, only three gene pairs were associated with the same disease/disease category: AGT - AGTR1 (angiotensin II receptor type 1), associated with disorders resulting from impaired renal tubular function; GRM7 (glutamate receptor, metabotropic 7) - GNAI3, related to depressive episode; LPAR1 (lysophosphatidic acid receptor 1) - GNA12 (G protein subunit alpha 12), associated with height abnormalities. The expression level of the gene pairs was calculated by the product of the expression values of the interacting genes. It was shown that these values fluctuate according to the presence or absence of disease, type of the interacting molecules (e.g. ligand-GPCR, etc.) and tissue; in other words, there is a three-way interaction. In particular, skeletal muscle tissues have the lowest and hypothalamus the highest expression levels. As expected, G proteins-effector pairs present the highest expression levels compared to the other gene pairs. Overall, the expression levels of the gene pairs (especially GPCR-G protein and G protein-effector) associated with diseases appear to be consistently higher compared to those of the corresponding gene pairs that are not involved in diseases across tissues. Notably, in hypothalamus, the expression level of gene pairs is similar for the ones associated with diseases and for those that are not involved in diseases. Furthermore, there is a positive correlation between the gene co-expression patterns and the clinical manifestations of the diseases. For example, the gene pair *AGT - AGTR1* expressed in kidney is associated with a kidney disease.

Identification of associations among the G_α, G_β, and G_γ subunits

The unweighted gene co-expression network among the components of the G protein heterotrimer was constructed by using a threshold value of $\tau = 0.85$. The network consists of 27 G proteins (15 G_α, 5 G_β and 7 G_γ subunits) that are represented by nodes and a total of 88 connections between gene pairs (co-expressed genes) that are denoted by edges. By analyzing the connections between the different G protein subunits, 32 potential heterotrimeric complexes of the G_α, G_β, and G_γ subunits were identified. Co-expression relationships among these complexes were further examined using high confidence curated molecular interactions in String [60]. Of note, using this approach, 17 out of the 32 potential heterotrimeric complexes were also confirmed. Furthermore, 7 out of the 17 complexes were found to consist of subunits all of which are involved in the chemokine signaling pathway (KEGG PATHWAY, map number: hsa04062).

Molecular pathways

Possible molecular pathways, both in disease and nondisease associated genes in certain tissues, in which all four types of serially interacting molecules participate, were identified. The genes encoding the components of the molecular pathway in Fig. 4A are expressed in heart, suggesting that the increased level of expression of these genes is required for the normal activity of the heart. These genes were found to regulate diverse cardiovascular functions. For example, the chemokine CXCL11 is suggested to play a role in heart transplantation models [61], whereas ACKR3 (Atypical Chemokine Receptor 3) is implicated in vasoconstriction[62]. The G protein GNAI1 is involved in the regulation of *in vivo* heart rate dynamics [63]. The effector TUBGCP5 (Tubulin, Gamma Complex Associated Protein 5) is necessary for microtubule nucleation [64]. Microtubules, along with actin, compose the major component of the cardiomyocyte cytoskeleton[65]. Given that TUBGCP5 is required in heart, we propose that it could, also, play a role in the cardiac cytoskeleton.

The molecules participating in the heart-specific pathway shown in Fig. 4B are highly expressed in heart tissue where defects in the respective genes cause heart-specific pathologies, such as cardiovascular diseases. In particular, the gene *EDN1*, which encodes the ligand endothelin-1, is believed to contribute to the pathogenesis of hypertension, heart failure and atherosclerosis [66]. Furthermore, HDL (High-Density Lipoprotein) cholesterol levels are associated with a substitution, K198N, in endothelin-1 [67]. The endothelin receptor type A, encoded by the gene *EDNRA*, mediates most of the vasoconstrictive properties of the ligand endothelin-1. Besides, polymorphisms in *EDNRA* were found to be associated with arterial hypertension[68]. The *ADRA2C* gene encodes the alpha-adrenoreceptor 2C, which is required for the regulated release of neurotransmitters from noradrenergic neurons in the heart. Polymorphisms in the *ADRA2C* gene are linked to systemic hypertension[69]. Besides, the protein encoded by the gene *GNAI2* is activated by both receptors, EDNRA and ADRA2C. It is proposed that GNAI2 is implicated in the pathogenesis of hypertension, since individuals with single nucleotide polymorphisms in the gene *GNAI2* were found to have an increased risk of developing essential hypertension [70]. Polymorphisms in the gene encoding the effector CSK (C-Src tyrosine kinase) are associated with a significantly increased risk of hypertension [71].

These data are particularly important, since these four genes likely contribute to other phenotypes, as well, like the phenotypes of the metabolic syndrome (LDL cholesterol, hyperlipidemia and type II diabetes) [72]. These genes could also provide the foundation for targeted studies to detect gene-gene interactions in genetic association studies [73]. The major obstacle in these studies is that, on a genomic scale, simultaneous analysis of all pairs of genes is computationally intensive and statistically unreliable, as the type I error would be too large due to multiple comparisons.

Disease risk prediction model

A logistic regression model was constructed to predict the relative risk of developing a disease in the case of co-expressed gene-pairs. We chose the pairs as the units of analysis and not the single genes, in order to capture the flow and direction of information from one level to the other (i.e. from ligands to GPCRs, from GPCRs to G proteins and so on). Towards this end, we calculated robust standard errors, in order to account for the dependencies that arise from the fact that the expression data are provided for the 11 healthy tissues within the same gene-pair. In particular, this model was adjusted for the type of interacting molecules, the tissue and the logarithmic mean of

the product of the expression values of the interacting genes, including highly significant three-way interactions between these terms. The sum of the number of drugs, the sum of the total number of PPIs and the sum of the outgoing edges coming from a node and the incoming edges onto a node per tissue of the interacting molecules were also evaluated as covariates. Overall, the interactions between the molecules GPCRs-G proteins and G proteins-effectors demonstrate a greater probability of developing a disease with increasing expression levels (Fig. S1). The estimated disease risk probability goes above 0.9 when the expression of the GPCR-G protein pair reaches its peak value, in ovary and skeletal muscle (Fig. S1). This trend, however, is not observed in hypothalamus with prediction probabilities almost uniformly equally to 0.63 in GPCRs-G proteins with even decreasing estimates in colon, hypothalamus and testes in G proteins-effectors. Of particular interest, there is a negative correlation between the probability of ligands-GPCRs interacting molecules that are implicated in a disease and the expression values in all tissues examined here (Fig. S1). In addition, we estimated an 23% increase in the likelihood of developing a disease for every 10 additional PPIs for a given pair and a 15% decrease for every 10 additional outgoing and incoming edges of the gene pairs. The drugs targeting GPCR network components apparently do not alter the probabilities of presenting a disease. Of particular note, our model indicated a probability of developing a disease greater than 0.8 in 34 interactions between gene pairs that have not been associated with diseases thus far. GNAQ (Guanine Nucleotide Binding Protein (G Protein), Q Polypeptide) participates in 11 of these cases suggesting that the role of this protein in diseases should be further investigated.

Discussion

This is the first, to our knowledge, *in silico* study of the GPCR signaling network. First, a single global network comprising functional relationships among the components of the GPCR signaling system was generated. The global network analyses, however, ignore the biological processes that contribute to certain phenotypes taking place within specific tissues[74]. Therefore, information of tissue-specific expression profiles was utilized. To this end, 11 tissue-specific functional networks were created in order to investigate the dynamic potential of the different tissues. The joint and the tissue-specific networks have the same topological properties, leading to the suggestion that the mode of transmitting the extracellular signals through the ligand<GPCR <G-protein<effector system is the same. These networks are scale-free, suggesting that they are dominated by few highly connected nodes that hold the individual networks together. Other prominent biological scale-free networks include the protein-protein interactions network [75], and the metabolic and the biochemical networks [76].

Genetic diseases exhibit pathological manifestations that are often restricted to specific tissues and, therefore, defects in the genes or pathways responsible for these diseases may have varying impact on different tissues [59, 74]. In this study, the gene expression in tissues affected by a disease was elevated as compared to genes in non- affected tissues, albeit in an inconsistent manner, as in the case of effectors. This corroborates previous studies which showed that the genes with higher expression levels under normal conditions are those which are most likely to have polymorphisms predisposing to disease [77] and, also, the proteins with more PPIs are more likely to be implicated in diseases [78]. To provide further support to the latter hypothesis, the disease-associated G-protein GNAI2 was among the ones with the highest number of interactions.

It is postulated that protein complexes play an important role in disease formation [59]. Motivated by this hypothesis, the expression profiles of gene pairs, were investigated, in this study. Uniform patterns of expression of gene-pairs across tissues were observed. For example, the expression of the G-protein-effector pair was found to be higher compared to the GPCR-G-protein pair across tissues, consistent with the amplification of the transmitted extracellular signal along the GPCR signaling cascade. Of particular importance, it was shown that 916 disease-associated gene-pairs, representing 55% of the total gene-gene associations, span a wide variety of diseases or disease categories. The expression of genes associated with diseases appears to be consistently higher compared to that of the corresponding gene pairs that are not involved in diseases across tissues. This is in agreement with the hypothesis that inactivation of genes the functions of which are essential to a certain tissue can have detrimental effects on this tissue that could be reflected in tissue-specific pathophenotypes [79]. Of importance, gene pairs were found to constitute robust predictive factors for the risk of developing a disease. However, the hypothalamus-specific gene pairs were found to be poor predictive factors for the disease risk. Of importance, the hypothalamus secretes different hormones which stimulate or inhibit the production of other hormones within tissues throughout the body [80].

Both potential associations among the Gα, Gβ, and Gγ subunits of the G protein heterotrimer, as well as associations of Gα with GPCRs and effectors, were identified in disease-affected tissues. Given that no commercial drugs have been developed for the G proteins investigated in this study, it would be intriguing to speculate that the aforementioned G proteins could represent attractive targets for therapeutic interventions. Likewise, the interactions among the protein molecules encoded by gene pairs that are potential predictive markers of disease risk could be taken into consideration for rational drug design [81].

Taken together, the above findings highlight the association between disease-related genes, co-expressed gene-pairs and tissue specificity. Acquiring functional information for specific diseased tissues would be very important in the identification of biomarkers for disease diagnosis, prognosis and monitoring. Tissue-specific networks revealed molecular pathways that participate in the same disease or disease category that would not be detected in the global network. Only a few disease-associated pathways restricted to a specific tissue were identified. This is probably due to the fact that signal transduction pathways traverse multiple tissues [74]. A theoretical limitation of this study was the low availability of tissue-specific gene expression data. Besides, the genes encoding the molecules under study were acquired from Human-gpDB which contains peptide and protein sequences. Thus, there are other types of molecules (such as neurotransmitters, ions, photons, etc.) information of which was not included in this study. Besides, no unconventional relationships, such as GPCR-effector [82], were identified. Nevertheless, in this study, the first network-based approach for investigating the relationships among the components of the GPCR signaling network, is presented.

Conclusions

Concluding, this is the first, in silico study of the GPCR signaling network. A global network comprising functional relationships among the components of the GPCR signaling system was generated. Therefore, information of tissue-specific expression profiles was utilized. To this end, 11 tissue-specific functional networks were created in order to investigate the dynamic potential of the different tissues. The functional relationships among the components of the GPCR pathway and their association with diseases were further elucidated, in this study. A disease risk prediction model based on tissue-specific expression profiles of the disease-associated genes was also generated. These findings could be very important in the identification of biomarkers for disease diagnosis, prognosis and monitoring.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was funded by the SYNERGASIA 2009 PROGRAMME. This Programme is co-funded by the European Regional Development Fund and National resources (Project Code 09SYN-13-999), General Secretariat for Research and Technology of the Greek Ministry of Education and Religious Affairs, Culture and Sports.

References

1. N. King, C. T. Hittinger and S. B. Carroll, *Science*, 2003, **301**, 361-363.
2. K. L. Pierce, R. T. Premont and R. J. Lefkowitz, *Nature reviews. Molecular cell biology*, 2002, **3**, 639-650.
3. , !!! INVALID CITATION !!! (Attwood and Findlay 1994; Foord, et al. 2005; Kolakowski 1994).
4. U. Gether, *Endocrine reviews*, 2000, **21**, 90-113.
5. Y. Okuno, A. Tamon, H. Yabuuchi, S. Niijima, Y. Minowa, K. Tonomura, R. Kunimoto and C. Feng, *Nucleic acids research*, 2008, **36**, D907-912.
6. K. Kristiansen, *Pharmacology & therapeutics*, 2004, **103**, 21-80.
7. S. Bockenhauer, A. Furstenberg, X. J. Yao, B. K. Kobilka and W. E. Moerner, *J Phys Chem B*, 2011, **115**, 13328-13338.
8. C. R. McCudden, M. D. Hains, R. J. Kimple, D. P. Siderovski and F. S. Willard, *Cellular and molecular life sciences : CMLS*, 2005, **62**, 551-577.
9. B. R. Conklin and H. R. Bourne, *Cell*, 1993, **73**, 631-641.
10. S. Rens-Domiano and H. E. Hamm, *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 1995, **9**, 1059-1066.
11. B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski and S. Filipek, *Current medicinal chemistry*, 2012, **19**, 1090-1109.
12. T. M. Cabrera-Vera, J. Vanhauwe, T. O. Thomas, M. Medkova, A. Preininger, M. R. Mazzoni and H. E. Hamm, *Endocrine reviews*, 2003, **24**, 765-781.
13. N. Wettschureck and S. Offermanns, *Physiol Rev*, 2005, **85**, 1159-1204.
14. R. T. Dorsam and J. S. Gutkind, *Nature reviews. Cancer*, 2007, **7**, 79-94.
15. M. D. Thompson, G. N. Hendy, M. E. Percy, D. G. Bichet and D. E. Cole, *Methods in molecular biology*, 2014, **1175**, 153-187.
16. H. F. Vischer, M. Siderius, R. Leurs and M. J. Smit, *Nature reviews. Drug discovery*, 2014, **13**, 123-139.
17. J. P. Overington, B. Al-Lazikani and A. L. Hopkins, *Nature reviews. Drug discovery*, 2006, **5**, 993-996.
18. A. L. Hopkins and C. R. Groom, *Nature reviews. Drug discovery*, 2002, **1**, 727-730.
19. D. Zhang, Q. Zhao and B. Wu, *Molecules and cells*, 2015, **38**, 836.
20. V. Katritch, V. Cherezov and R. C. Stevens, *Annual review of pharmacology and toxicology*, 2013, **53**, 531.
21. B. K. Kobilka, *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 2007, **1768**, 794-807.
22. V. Isberg, B. Vroling, R. van der Kant, K. Li, G. Vriend and D. Gloriam, *Nucleic Acids Res.*, 2013, gkt1255.
23. V. Isberg, S. Mordalski, C. Munk, K. Rataj, K. Harpsøe, A. S. Hauser, B. Vroling, A. J. Bojarski, G. Vriend and D. E. Gloriam, *Nucleic Acids Res.*, 2016, **44**, D356-D364.
24. C. Munk, V. Isberg, S. Mordalski, K. Harpsøe, K. Rataj, A. Hauser, P. Kolb, A. Bojarski, G. Vriend and D. Gloriam, *British journal of pharmacology*, 2016.
25. M. C. Theodoropoulou, P. G. Bagos, I. C. Spyropoulos and S. J. Hamodrakas, *Bioinformatics*, 2008, **24**, 1471-1472.
26. V. P. Satagopam, M. C. Theodoropoulou, C. K. Stampolakis, G. A. Pavlopoulos, N. C. Papandreou, P. G. Bagos, R. Schneider and S. J. Hamodrakas, *Database : the journal of biological databases and curation*, 2010, **2010**, baq019.
27. J. Kazius, K. Wurdinger, M. van Iterson, J. Kok, T. Bäck and A. P. IJzerman, *Human mutation*, 2008, **29**, 39-44.
28. A. J. Harmar, R. A. Hills, E. M. Rosser, M. Jones, O. P. Buneman, D. R. Dunbar, S. D. Greenhill, V. A. Hale, J. L. Sharman and T. I. Bonner, *Nucleic Acids Res.*, 2009, **37**, D680-D685.
29. K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal and A. L. Barabasi, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 8685-8690.
30. M. Oti and H. G. Brunner, *Clinical genetics*, 2007, **71**, 1-11.
31. L. H. Hartwell, J. J. Hopfield, S. Leibler and A. W. Murray, *Nature*, 1999, **402**, C47-52.
32. C. Southan, J. L. Sharman, H. E. Benson, E. Faccenda, A. J. Pawson, S. P. Alexander, O. P. Buneman, A. P. Davenport, J. C. McGrath, J. A. Peters, M. Spedding, W. A. Catterall, D. Fabbro, J. A. Davies and I. Nc, *Nucleic acids research*, 2016, **44**, D1054-1068.
33. M. Krupp, J. U. Marquardt, U. Sahin, P. R. Galle, J. Castle and A. Teufel, *Bioinformatics*, 2012, **28**, 1184-1185.
34. K. G. Becker, K. C. Barnes, T. J. Bright and S. A. Wang, *Nature genetics*, 2004, **36**, 431-432.
35. D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, A. Klemm, P. Flicek, T. Manolio, L. Hindorff and H. Parkinson, *Nucleic acids research*, 2014, **42**, D1001-1006.

36. J. S. Amberger, C. A. Bocchini, F. Schiettecatte, A. F. Scott and A. Hamosh, *Nucleic acids research*, 2015, **43**, D789-798.
37. K. A. Gray, B. Yates, R. L. Seal, M. W. Wright and E. A. Bruford, *Nucleic acids research*, 2015, **43**, D1079-1085.
38. P. I. Kontou, A. Pavlopoulou, N. L. Dimou, G. A. Pavlopoulos and P. G. Bagos, *Gene*, 2016, **In press**.
39. A. Oellrich, P. Sanger Mouse Genetics and D. Smedley, *Database : the journal of biological databases and curation*, 2014, **2014**, bau017.
40. D. Smedley, A. Oellrich, S. Kohler, B. Ruef, P. Sanger Mouse Genetics, M. Westerfield, P. Robinson, S. Lewis and C. Mungall, *Database : the journal of biological databases and curation*, 2013, **2013**, bat025.
41. V. Law, C. Knox, Y. Djoumbou, T. Jewison, A. C. Guo, Y. Liu, A. Maciejewski, D. Arndt, M. Wilson, V. Neveu, A. Tang, G. Gabriel, C. Ly, S. Adamjee, Z. T. Dame, B. Han, Y. Zhou and D. S. Wishart, *Nucleic acids research*, 2014, **42**, D1091-1097.
42. C. UniProt, *Nucleic acids research*, 2015, **43**, D204-212.
43. S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R. C. Lovering, B. Meldal, A. N. Melidoni, M. Milagros, D. Peluso, L. Perfetto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni and H. Hermjakob, *Nucleic acids research*, 2014, **42**, D358-363.
44. A. J. Butte and I. S. Kohane, *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 2000, 418-429.
45. M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein, *Proceedings of the National Academy of Sciences of the United States of America*, 1998, **95**, 14863-14868.
46. B. B. Khomtchouk, D. J. Van Booven and C. Wahlestedt, *Source code for biology and medicine*, 2014, **9**, 30.
47. P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome research*, 2003, **13**, 2498-2504.
48. A. J. Enright, S. Van Dongen and C. A. Ouzounis, *Nucleic acids research*, 2002, **30**, 1575-1584.
49. J. H. Morris, L. Apeltsin, A. M. Newman, J. Baumbach, T. Wittkop, G. Su, G. D. Bader and T. E. Ferrin, *BMC bioinformatics*, 2011, **12**, 436.
50. StataCorp.2013, *Journal*.
51. X. Zhang and U. S. Eggert, *Molecular bioSystems*, 2013, **9**, 586-595.
52. A. L. Barabasi and Z. N. Oltvai, *Nature reviews. Genetics*, 2004, **5**, 101-113.
53. A. L. Barabasi, *Science*, 2009, **325**, 412-413.
54. R. Nusse and H. Varmus, *The EMBO journal*, 2012, **31**, 2670-2684.
55. R. Nusse and H. E. Varmus, *Cell*, 1992, **69**, 1073-1087.
56. E. Nogales, K. H. Downing, L. A. Amos and J. Lowe, *Nature structural biology*, 1998, **5**, 451-458.
57. P. I. Kontou, A. Pavlopoulou, N. L. Dimou, G. Pavlopoulos and P. G. Bagos, *Data in Brief*, 2016, **8**, 1036-1039.
58. L. J. Heyer, S. Kruglyak and S. Yooseph, *Genome research*, 1999, **9**, 1106-1115.
59. K. Lage, E. O. Karlberg, Z. M. Storling, P. I. Olason, A. G. Pedersen, O. Rigina, A. M. Hinsby, Z. Tumer, F. Pociot, N. Tommerup, Y. Moreau and S. Brunak, *Nature biotechnology*, 2007, **25**, 309-316.
60. D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K. P. Tsafou, M. Kuhn, P. Bork, L. J. Jensen and C. von Mering, *Nucleic acids research*, 2015, **43**, D447-452.
61. N. Mitsuhashi, G. D. Wu, H. Zhu, M. Kearns-Jonker, D. V. Cramer, V. A. Starnes and M. L. Barr, *Molecular and cellular biochemistry*, 2007, **296**, 1-9.
62. H. H. t. Bach, Y. M. Wong, A. Tripathi, A. M. Nevins, R. L. Gamelli, B. F. Volkman, K. L. Byron and M. Majetschak, *Molecular medicine*, 2014, **20**, 435-447.
63. Z. Zuberi, L. Birnbaumer and A. Tinker, *American journal of physiology. Regulatory, integrative and comparative physiology*, 2008, **295**, R1822-1830.
64. B. R. Oakley, *Trends in cell biology*, 1992, **2**, 1-5.
65. V. Sequeira, L. L. Nijenkamp, J. A. Regan and J. van der Velden, *Biochimica et biophysica acta*, 2014, **1838**, 700-722.
66. R. P. Lifton, A. G. Gharavi and D. S. Geller, *Cell*, 2001, **104**, 545-556.
67. G. Pare, D. Serre, D. Brisson, S. S. Anand, A. Montpetit, G. Tremblay, J. C. Engert, T. J. Hudson and D. Gaudet, *American journal of human genetics*, 2007, **80**, 673-682.
68. A. V. Benjafield, K. Katyk and B. J. Morris, *Clinical genetics*, 2003, **64**, 433-438.
69. J. Savva, K. Alfakih, S. L. Galloway, A. S. Hall, R. M. West, S. G. Ball, A. J. Balmforth and A. Maqbool, *Blood pressure*, 2012, **21**, 116-121.
70. C. Menzaghi, G. Paroni, C. De Bonis, T. Soccio, A. Marucci, S. Bacci and V. Trischitta, *Journal of the American Society of Nephrology : JASN*, 2006, **17**, S115-119.
71. B. Xi, Y. Shen, K. H. Reilly, X. Wang and J. Mi, *Metabolism: clinical and experimental*, 2013, **62**, 196-203.
72. K. G. Alberti, P. Zimmet, J. Shaw and I. D. F. E. T. F. C. Group, *Lancet*, 2005, **366**, 1059-1062.
73. A. K. Manning, M. LaValley, C. T. Liu, K. Rice, P. An, Y. Liu, I. Miljkovic, L. Rasmussen-Torvik, T. B. Harris, M. A. Province, I. B. Borecki, J. C. Florez, J. B. Meigs, L. A. Cupples and J. Dupuis, *Genetic epidemiology*, 2011, **35**, 11-18.
74. Y. Guan, D. Gorenshteyn, M. Burmeister, A. K. Wong, J. C. Schimenti, M. A. Handel, C. J. Bult, M. A. Hibbs and O. G. Troyanskaya, *PLoS computational biology*, 2012, **8**, e1002694.
75. J. C. Nacher, M. Hayashida and T. Akutsu, *Bio Systems*, 2009, **95**, 155-159.
76. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabasi, *Nature*, 2000, **407**, 651-654.
77. I. P. Gorlov, G. E. Gallick, O. Y. Gorlova, C. Amos and C. J. Logothetis, *PloS one*, 2009, **4**, e6511.
78. H. Jeong, S. P. Mason, A. L. Barabasi and Z. N. Oltvai, *Nature*, 2001, **411**, 41-42.
79. L. Duret and D. Mouchiroud, *Molecular biology and evolution*, 2000, **17**, 68-74.
80. S. Melmed, K. S. Polonsky, P. R. Larsen and H. M. Kronenberg, *Williams Textbook of Endocrinology*, W.B. Saunders Company, 12th edn., 2011.
81. M. Skwarczynska and C. Ottmann, *Future medicinal chemistry*, 2015, **7**, 2195-2219.
82. M. J. Marinissen and J. S. Gutkind, *Trends in pharmacological sciences*, 2001, **22**, 368-376.

83. G. A. Pavlopoulos, S. I. O'Donoghue, V. P. Satagopam, T. G. Soldatos, E. Pafilis and R. Schneider, *BMC systems biology*, 2008, **2**, 104.

Figures

Fig. 1: Gene expression data. A) Distribution of the 433 retrieved molecules based on the type of molecule and the disease state per tissue. B) Logarithmic mean of the expression value of the four types of molecules per tissue in normal and diseased states.

Fig. 2: Joint human GPCR signal transduction network with four different types of visualization. (A) 3D network visualized by Arena3D [83]. 126 Ligands (orange), 223 GPCRs (purple), 13 G proteins (pink) and 71 Effectors (green) are connected with 1666 connections across layers (B) 3D network visualized by Arena3D [83]. White color indicates the 124 molecules that are implicated in a disease and their interactions. (C) 2D network visualized by Cytoscape. (D) 2D network visualized by Cytoscape. The order of the four types of interacting molecules corresponds to their order in the signaling network. Ligands (green); GPCRs (orange); G proteins (pink); effectors (blue). The nodes represent molecules and the links (edges) associations between molecules.

Fig. 3: Networks' degree distributions. (A) Degree distribution *P(k)* of the out- and in-degree nodes in the joint GPCR signaling network. The degree distribution results to a straight line on the double-logarithmic plot, indicative of scale-free property. (B) Degree distribution *P(k)* of the in-degree and (C) Degree distribution *P(k)* of the out-degree nodes in the tissue-specific GPCR signaling networks.

Fig. 4: Molecular pathways. (A) Example of pathways in which all four molecules are implicated in the same class of disease. (B) Example of pathways where none of the molecules are implicated in diseases. Ligand (green); GPCRs (orange); G protein (pink); effector (blue). The second pathway lacks a ligand.

Tables

Table 1: The top ten hubs of the joint network. The molecules associated with diseases are indicated by an asterisk.

Ligands	GPCRS	G proteins	Effectors
WNT11*	FZD1	GNAI1	TUBA1A
WNT2B	FZD8	GNAI2*	TUBA1B
WNT4*	CCR10	GNAI3*	TUBA1C
WNT5A	FZD4*	GNAO1	TUBA3C
WNT5B	FZD3	GNAQ	TUBA3D
WNT16*	FZD5	GNAZ	TUBA3E
WNT3*	FZD6	GNA11	TUBA4A
WNT7B	FZD7	GNAS*	TUBB3*
WNT9A	TSHR*	GNA14	TUBA8*
WNT9B	F2R*	GNA15*	TUBB1*