ORIGINAL ARTICLE

Validating methods for testing natural molecules on molecular pathways of interest in silico and in vitro

KRISTJANA DHULI[1,*], GABRIELE BONETTI[1], KYRYLO ANPILOGOV[2], KAREN L. HERBST[3], STEPHEN THADDEUS CONNELLY[4], FRANCESCO BELLINATO[5], PAOLO GISONDI[5], MATTEO BERTELLI[1,2,6]
[1] MAGI'S LAB, Rovereto (TN), Italy; [2] MAGI EUREGIO, Bolzano, BZ, Italy; [3] Total Lipedema Care, Beverly Hills California and Tucson Arizona, USA; [4] San Francisco Veterans Affairs Health Care System, Department of Oral & Maxillofacial Surgery, University of California, San Francisco, CA, USA7; [5] Section of Dermatology and Venereology, Department of Medicine, University of Verona, Verona, Italy; [6] MAGISNAT, Peachtree Corners (GA), USA

Keywords

Gene expression • Bioinformatics tools • Biochemical pathways • In vitro • Natural molecules

Summary

Differentially expressed genes can serve as drug targets and are used to predict drug response and disease progression. In silico drug analysis based on the expression of these genetic biomarkers allows the detection of putative therapeutic agents, which could be used to reverse a pathological gene expression signature. Indeed, a set of bioinformatics tools can increase the accuracy of drug discovery, helping in biomarker identification. Once a drug target is identified, in vitro cell line models of disease are used to evaluate and validate the therapeutic potential of putative drugs and novel natural molecules. This study describes the development of efficacious PCR primers that can be used to identify gene expression of specific genetic pathways, which can lead to the identification of natural molecules as therapeutic agents in specific molecular pathways. For this study, genes involved in health conditions and processes were considered. In particular, the expression of genes involved in obesity, xenobiotics metabolism, endocannabinoid pathway, leukotriene B4 metabolism and signaling, inflammation, endocytosis, hypoxia, lifespan, and neurotrophins were evaluated. Exploiting the expression of specific genes in different cell lines can be useful in in vitro to evaluate the therapeutic effects of small natural molecules.

Introduction

The post-genomic era is marked by several discoveries in the discipline of molecular medicine that have enabled the recognition of disease-related genes and the subsequent development of targeted therapeutic strategies. Next generation RNA sequencing clearly demonstrates that genes do not function alone, but rather constantly interact with each other. These genetic interactions are crucial for regulating gene expression, and downstream biochemical, and signal transduction pathways [1, 2]. Genes that function in the same biological pathway are regulated by the same transcription factors and tend to show similar expression levels under a particular stimulus [3]. Networking amongst genes can be identified by pathway and network analysis models [4]. The topology of the networks is used to predict gene expression under different circumstances. For instance, under a particular environmental condition the expression of a central or 'hub gene' that acts as a transcription regulatory factor helps determine the expression of the genes that are regulated by it [5-7]. This means that this regulatory gene is acting as a switch to control the expression of its interacting genes. On the other hand, gene network connections might represent a set of genes activated in a particular pathway by a biological process or pathogenic condition, termed gene expression signatures [8]. These signatures indicate a difference in cellular activity and depict the interchange of different biological pathways. The gene expression signatures and contrasting networks can explain how aberrations in gene-gene and gene-environment interactions result in pathological conditions [3]. Consequently, one of the most powerful uses of high throughput genomic, transcriptomic, proteomic, and metabolomics data is the unravelling of the mechanisms underlying diseases by comparing biological pathways in control versus disease states [9]. This makes clear the importance of pathway analyses in deciphering the etiology of a specific disease, in the identification of potential biomarkers, and in targeted drug discovery [9, 10].

Biological pathways from a bioinformatics viewpoint

Biological pathways include are a set of genes or molecules that act in a synergistic fashion to accomplish a biological function. Biological pathways play a vital part in the advancement and survival of an organism and failure in functioning of a pathway results in the onset of disease [11]. Based on the cellular requirements at a particular time, the products of a pathway can manifest differently as structural or functional responses.
Biological pathways can be broadly categorized into metabolic, genetic and cell signalling pathways. These pathways interact with one another, forming a network of interconnected pathways that deal with complex cellular functions and with the regulation of gene expression [12, 13].

Biologic pathway analysis integrates gene ontology and pathway structure information to identify pathways whose activation/inactivation is linked with a specific condition or disease. This makes pathway analysis an important tool in deciphering mechanisms underlying a disease and consequent drug discovery [9, 12]. In fact, it is now clear that complicated diseases are a consequence of dysregulated pathways rather than the dysregulated expression of an individual gene. In fact, a variety of gene pathways may combine to manifest the same condition [14]. In such cases, responses to these disorders are expected to ultimately affect the same cellular system [14]. Pathway-centric models are fundamental in figuring out the mechanisms of complicated diseases and recognition of candidate drug targets. Pathway-centric models represent pathways as graphs of circles or nodes, where larger nodes denote pathways with larger numbers of components, and edges between nodes symbolize interaction between the different pathway nodes [15].

Pathway analysis methods and databases

Differential expression (DE) of genes in experiments comparing two situations – such as two phenotypes, two drugs, two states (control vs disease; treated vs untreated) – and subsequent statistical analysis approaches such as ANOVA [16], t test [17], or Z scores [18] can help identify the genes or set of genes that contribute to the development of a particular phenotype. However, as genes are not expressed alone and are under the control of several regulatory elements, the identification of genes alone cannot elucidate the mechanisms of complex diseases; therefore, knowledge obtained from the DE of genes is studied in the context of information obtained from pathway databases. Pathway analysis coupled with data obtained from DE of genes helps to decipher the mechanisms underlying a particular condition and to identify which pathways are significantly affected. Several studies have reported the use of pathway databases to identify genetic markers, gene signatures, and mechanisms of complex diseases (Tab. I). Important pathway databases used in studying genetic, signalling, and metabolic pathways are presented in Table II. In addition to pathway analysis, network analysis is also carried out to see the interactions between various gene networks which are analyses collected from distinct populations, conditions, or groups [19].

Tab. I. Examples of diseases identified by different methodologies of pathway analysis.

Disease	Software	References
Glioblastoma multiforme	TRED database, eQTL mapping	[15]
Alzheimer's disease	WebGestalt	[20]
Olfactory behaviour	R spider	[21]
Esophageal squamous cell carcinoma	ICSNPathway server with i-GSEA	[22]
White adipocyte insulin resistance	GO analysis; KEGG pathway analysis	[23]
Biliary cirrhosis	LRT and i-GSEA4GWAS	[24]
Bladder cancer	Combined outcomes of GSEA and ARTP	[25]
Bipolar disorder	IPA and GSEA-SNP	[26]
Major depressive disorder	GSEA and statistical analysis	[27]
Schizophrenia	ICSNPathway server with i-GSEA, MAGENTA, ALIGATOR, INRICH and Set Screen	[28, 29]
Coronary heart disease	VSEA in GWAS	[30]

TRED: Trascriptional Regulatory Element Database; eQTL: expression Quantitative Trait Loci; ICSNPathway: Identifiy candidate Causal SNPs and Pathways; i-GSEA: Gene Set Enrichment Analysis; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; LRT: Likelihood Ratio Test; GWAS: Genome-Wide Association Study; ARTP: Adaptive Rank Truncated Product; IPA: Ingenuity Pathways Analysis; SNP: Single Nucleotide Polymorphism; MAGENTA: Meta-Analysis Gene-set Enrichment of variaNT Associations; ALIGATOR: Association LIst Go AnnoTatOR; INRICH: INterval enRICHment analysis; VSEA: Variable Set Enrichment Analysis.

Altered expression of specific genes as biomarkers and their exploitation as therapeutic targets

Biomarkers are biological molecules that act as indicators of normal or pathological processes or pharmacological responses to a directed therapeutic [41]. In addition, biomarkers are used in screening for disease, as diagnostic and prognostic factors, and for selecting patient-specific therapy. Biomarkers are also useful in evaluating the effect of drugs administered to patients or to cell lines for therapeutic and experimental purposes, respectively. Biomarkers must be reliable and reproducible because human health is at stake. Biomarkers can be discovered through gene expression analysis followed by feature selection methods that enable the discovery of a small subset of biomarkers that have the ability to discriminate between molecular subtypes of diseases [42]. Recent studies have evaluated gene expression in peripheral blood mononuclear cells (PBMCs) to identify biomarkers for disease [43], including Crohn's disease [44], Behcet's disease [45] and ulcerative colitis [44]. Relevant to the obesity epidemic, one study identified 9 genes that correlated with obesity indices in humans out of 19 genes differentially expressed in the PBMCs of high fat-fed rats [43]. Another study reported the identification of biomarkers of insulin resistance found in the expression profiles from adipocytes of subjects with insulin resistant obesity using Gene Expression Omnibus. The study identified 10 hub genes (genes with the most interactions with other genes) using various bioinformatic tools, such as GSEA, GO analysis, KEGG pathway analysis [46], and Cytohubba. Moreover, using these biomarkers, potential small molecular compounds that could treat insulin resistance were detected [23].

Bioinformatic tools for identification of drugs-disease-pathway interaction

In the past few decades, the pharmaceutical industry has successfully deployed its one drug, one target model,

Tab. II. Common pathway annotation databases.

Name	Database Description	References
KEGG	Genomic and pathway information in various organisms	[31]
PANTHER v.14	Evolutionary relationships data for protein analysis	[32]
Pathway Ontology	Contains several biological pathways, including altered and disease pathways, and the relationships between them	[33]
BioCarta	Defines gene sets for data analysis	[34]
SPIKE	Focuses on pathways describing cellular responses such as DNA damage response, cell cycle, apoptosis and hearing-related pathways	[35]
GeneOntology	Pioneered use of ontologies in computational biology	[36]
PID	Information about molecular and cellular signalling pathways	[37]
MetaCyc	Metabolic and enzymatic pathways from various organisms	[38]
REACTOME	A platform for annotating and visualising data from several databases	[39]
MSigDB	Collects gene sets by biological functions, GO, KEGG, positions, sequence regulation information	[40]

focusing on druggable genes, genes encoding proteins that can be modulated using experimental small molecule compounds. This model emphasizes only on a small subset of genes affected by the drugs, completely ignoring the mechanisms underlying the action of the drug on these genes and molecular pathways [47]. In addition, this paradigm ignores the function of synergistic molecules from different pathways and their effects on the same subset of genes. This means that although successfully deployed, this paradigm cannot fully explain the drug-target interaction. This is due to the fact that the onset of a disease cannot be reduced to a single change, but rather to a cascade of gene expression alterations under the influence of the physiological environment of the body. Furthermore, the drug itself does not only interact with a single target, but rather with pathways or metabolic patterns of the body [48]. In this scenario, systems bioinformatics holds promise in predicting drug-pathway interactions by elucidating the mechanisms underlying drug activity and its possible side effects. Identifying enrichment pathways or gene sets from drug-induced datasets can lead to the discovery of promising drug targets, with a focus on reducing side effects. In addition, unravelling drug-disease-pathway interactions can provide useful insights of the systemic drug efficacy. Some important pathway databases and networks used for drug-disease-pathway-interactions are presented in Table III.

Cell cultures as a model for studying drug-pathway interactions

Preclinical models, such as cell lines, have been successfully deployed in studying and predicting the response and mechanism of action of drugs on disease-related genes and dysregulated pathways. Cell lines provide a continuous source of biological material for experimentation. The field in which cell line models are most used is cancer research. Indeed, cell lines are used to study the effect of anticancer drugs, as well as to study genetic alterations found in specific types of tumors. In addition, cell line models are particularly useful in cases where it is difficult to obtain clinical samples or where the monetary or human cost of obtaining clinical samples is high [65].

In order to study the effects of anti-tumoral drugs on genetic variants involved in tumour formation, several cell line models are used [66]. Data can also be retrieved from the Cancer Genome Project (CGP) and from the Cancer Cell line Encyclopedia (CCLE) [67], which contain data regarding 36 cancer cell lines [68]. Immortalized cell lines are often used to test drug efficacy and toxicity, or to identify drug-specific biomarkers [69]. A typical example is the Epstein-Barr virus (EBV) transformed Human Lymphoblastoid Cell Lines (LCLs) [69]. The goal of this study is to demonstrate the importance of developing an experimental model to study the effects of natural molecules in cell lines.

Materials and Methods

Bioinformatic study for gene selection

Genes of interest, associated with a specific condition, disease, or process, were chosen by searching GeneCards with specific keywords, specifically, obesity, xenobiotics metabolism, endocannabinoid pathway, leukotriene B4 metabolism and signaling, inflammation, endocytosis, hypoxia, lifespan, and neurotrophins were considered and used as keywords. For each query, we identified a list of genes identified by a score to reflect the association of the records with the query: the genes above a specific score threshold were retained for further study. The resulting list of genes was analyzed in the KEGG database to define common metabolic, gene regulation or signal transduction pathways. Genes already studied in the MAGI laboratory in association with other conditions were also included in the final list. Reference material supporting the genes chosen was obtained from PubMed. Finally, using the free tool STRING the interrelationships between the products of the identified genes was highlighted.

Primer design

For each selected gene, pairs of primers were designed for evaluating gene expression through real time PCR

Tab. III. Databases for drug target discovery.

Database	Database Description	Applications	Reference
Drug-Path	Reports genes that can be upregulated or downregulated by drugs interactions	Retrieval of drug-induced pathway data. Highlights the dysregulated pathways of diseases.	[49]
DGIdb 4.0	Information on drug-gene interactions and druggable genes	Identification of drug targets and studying drug-gene interactions	[50]
PubChem	Information on chemicals and on their toxicity	Identification of chemicals that have potential to be used as drugs	[51]
NCBI dbGaP	Archived genetic data including the relation between phenotype and GWAS	Identification of genes involved in a disease with genotype-phenotype interaction studies	[52]
GWAS Catalog	Metadata of the most significant published results	Identification of disease genes, prioritization of candidate loci, prediction of disease risk and molecular disease mechanisms	[53]
ChEBI	Ontology of chemicals and molecular entities, especially small molecules	Supply of identifiers for unambiguously refer to chemical entities	[54]
DrugBank 5.0	Drug and Drug Target Info. Provides molecular information regarding drugs and their mechanisms of action, interactions with other drugs, and their targets	Study of pharmacological properties of drugs, drug-drug, drug-pathway, drug-food interaction elucidation	[55]
PharmGKB	Aggregated information of genetic variants-drug response interaction	Extraction of interactions between drugs-drugs/genes/pathways/SNP, diseases-pathway/gene-SNP	[56]
STITCH	Chemical interactions using information from molecular pathways, crystal structures and binding experiments	Identification of drug-pathway interactions	[57]
HMDB	Information about small human metabolites	Identification of drug-metabolome interactions	[58]
MetaboLights	Metabolomics experiments used for cross-platform and cross-species studies	Identification of metabolites' structure, biological roles, concentration, and localization in living systems.	[59]
eDGAR	Relationships among genes related to disease-gene associations	Identification of disease-gene association, gene-gene interaction. Detection of functional terms related to groups of genes	[60]
NPASS	Information on activity and sources of natural products	NP(Natural Product)-based drug discovery, mechanism elucidation of NP and in silico algorithms development	[61]
MetaCyc	Metabolic pathways and enzymatic reactions from organisms of all life's domains	Prediction of the metabolic pathways of an organism from its annotated genome	[39]
MassBank Japan	Mass spectral data of biological molecules	Identification of a chemical compound	[62]
HumanCyc	Metabolic pathways and enzymatic reactions	Analysis of omics data for metabolic pathways	[63]
CMap	Gene expression profiles of immortalized human cell lines after chemical treatment	Prediction of the effects and mode of action of drugs; drug repositioning	[64]

experiments. At first, real time PCR primers were retrieved from Harvard Medical School database (https://pga.mgh.harvard.edu/primerbank/index.html). Pairs of primers that produced amplicons with a length of ≤ 200 bp and with the melting temperature of closest to 60°C for both primers were selected.

When primers were not available in the database or did not meet the required criteria, we designed new primers by using the bioinformatics tool Primer3. The criteria for choosing the primer pairs are as follows:

- Primer length: 18-28 nucleotides;
- Resulting amplicon: ≤ 200 bp (optimal 80-120 bp);
- Melting temperature: 60°C (the two primers must not have more than one degree of melting temperature difference from each other);
- GC content: 20-80% (50% was optimal);
- Primers must not contain repeated nucleotide sequences and complementary regions;
- Primers must be designed preferably on different exons or across exon-exon junctions, to limit as much as possible, the amplification of non-specific regions.

The resulting primer pairs were analyzed by PRIMER BLAST to evaluate the specificity of the amplified region.

The ENSEMBL genome browser was used to confirm that the primers mapped at the exon-exon junction or on different exons. However, this was not an exclusion criterion of the primer pair, as for some genes it is not possible to satisfy this characteristic (e.g. monoexonic genes).

RNA extraction, retrotranscription and qPCR

Total RNA was extracted from selected cell lines and blood using the Tempus Spin RNA Isolation Kit, following the manufacturer's protocol. Cell lines were selected referring to GeneCards database "Expression" section, which shows the tissues that express most highly a gene of interest. Between the ones proposed, cell lines already present in MAGI laboratories were used. Blood was collected from patients used as negative control in previous projects [70]. The SuperScript VILO cDNA Synthesis Kit was used to generate first strand cDNA. Quantitative real-time polymerase chain reaction (qPCR) was performed by using the PowerUp SYBR Green Master Mix (Thermo Fisher Scientific, Vilnius, Lithuania) on a QuantStudio 3 Real-Time PCR System, as reported [71].

Polymerase chain reaction (PCR) for the identification of cells expressing genes of interest

The PCR was performed with the aim of verifying that the primers selected for each gene-produced amplicons of the expected length, that there were no non-specific amplifications, and that the gene was expressed in the chosen cell lines.

qPCR primer efficiency evaluation

The evaluation of the efficiency of the primers is a fundamental step in qPCR, especially when studying gene expression, as it allows the correct analysis of data obtained. When the efficiency is calculated with the **ΔΔCt** method, it is assumed that the efficiency of the used primer is comparable to that of the primer for the housekeeping gene. If the efficiencies of the primers were dissimilar, the gene expression analysis could be affected by errors and misleading results would be obtained. The QuantStudio 3 Real-Time PCR System software calculated the efficiency of each primer pair. Primers whose efficiency was in the range between 90% and 110% were selected.

Fig. 1. Target genes evaluated in this work.

OBESITY
CD36, APOE, EP300, ACE, FTO, LIPE, MC4R, PPARGC1A, RETN, UCP1, ADIPOQ, HSD11B1, LEP, LEPR, CYP19A1, ESR1, IL6,GAPDH
XENOBIOTICS METABOLISM
FAAH, FAAH2, PPARA, LTB4R, LTB4R2, IGF1, AR, ADIPOQ, HSD11B1, LEP, LEPR, AHR, AHRR, AKR1C1, AKR1C2, AKR1C3, AKR1C4, ESR2, HPGD, LMNA, NCOA1, PLIN1, PRLR, VDR, CYP19A1, ESR1, ALOX5,GAPDH
ENDOCANNABINOID PATHWAY
CD36, CNR1, GPR18, GPR55, FAAH, FAAH2, PPARA, ALOX5,GAPDH
LEUKOTRIENE B4 METABOLISM AND SIGNALING
IL6, TNF, LTA4H, LTB4R, LTB4R2, ALOX5AP, GAPDH
INFIAMMATION
IL2, NOS1, NOS2, TLR2, TLR4, TLR7, IL6, TNF, APOE,GAPDH,RIPK1
ENDOCYTOSIS
CAV1, CLTB, ERLIN1, ERLIN2, RFTN1, SGMS1,GAPDH
HYPOXYA
ARNT, HIF1A, HIF1AN, HYOU1, VEGFC, VEGFR1(FLT1), VEGFR3(FLT4), EP300, VEGFA,GAPDH
LIFE-SPAN
IL6, TNF, APOE, ATG5, CERS2, COQ7, FCGR2A, SIRT1, SIRT3, SIRT6 , SOD1, TGFB1, IGF1, AR, GAPDH
NEUROTROPHINS
IGF1, BDNF, GDNF, HGF, IGF1R, NGF, NGFR, GAPDH

Results

In this work, a system for studying the expression of specific genes in selected cell lines was validated. 101 sets of primer pairs targeting specific genes (Fig. 1) were tested based on their efficiency values with the following results: 51 validated, 24 non-validated and 26 sets of primers targeting genes that were not expressed in the available cell lines requiring re-testing in other cell lines (Tab. IV).
The levels of expression were evaluated on the basis of the C_t. $C_t \leq 20$ = high expression; $20 < Ct \geq 23$ = high-medium expression; $23 < Ct \geq 26$ = low medium expres-

Tab. IV. List of primers from this work with qPCR efficiency between 90-110%.

Gene	Sequence (5'->3')	Tm	Cell Line/Tissue	Ct	Expression Level
AKR1C1	CCTAAAAGTAAAGCTTTAGAGGCCACC	60	Blood	27	Low
	GAAAATGAATAAGGTAGAGGTCAACATAAT				
AKR1C2	CCTAAAAGTAAAGCTCTAGAGGCCGT	60	Blood	32	Low
	GAAAATGAATAAGATAGAGGTCAACATAG				
AKR1C3	GAGAAGTAAAGCTTTGGAGGTCACA	60	Blood	26	Low-medium
	CAACCTGCTCCTCATTATTGTATAAATGA				
LTA4H	TCTGGGAGGACCGTATGTATG	60	Blood	27	Low
	ATTCCCTGTCCAGCTATGAGAT				
ALOX5	ACTGGCTGAATGACGACTGG	60	Blood	19	High
	CAGGGGAACTCGATGTAGTCC				
LEPR	TCCTCTTCCATCTTATTGCTTGG	60	Blood	24	Low-medium
	TCTTGGGGTTCGGAACATCT				
VEGFC	ATGTGTGTCCGTCTACAGATGT	60	Blood	26	Low-medium
	GGAAGTGTGATTGGCAAAACTGA				

Tab. IV. *Continues.*

Gene	Sequence (5'->3')	Tm	Cell Line/Tissue	Ct	Expression Level
FLT1(VEGFR1)	GAAAACGCATAATCTGGGACAGT	60	Blood	26	Low-medium
	GCGTGGTGTGCTTATTTGGA				
FLT4(VEGFR3)	TGCACGAGGTACATGCCAAC	60	HepG2	27	Low
	GCTGCTCAAAGTCTCTCACGAA				
APOE	GTTGCTGGTCACATTCCTGG	60	Caco2	21	High-medium
	GCAGGTAATCCCAAAAGCGAC				
BDNF	CTACGAGACCAAGTGCAATCC	60	SK-N-SH	27.5	Low
	AATCGCCAGCCAATTCTCTTT				
GAPDH	GGAGTCAACGGATTTGGTCG	60	ALL	17	High
	GACAAGCTTCCCGTTCTCAG				
SOD1	GGTGGGCCAAAGGATGAAGAG	60	Caco2	23	High-medium
	CCACAAGCCAAACGACTTCC				
TNF	GAGGCCAAGCCCTGGTATG	60	Blood	26	Low-medium
	CGGGCCGATTGATCTCAGC				
AHR	CTTAGGCTCAGCGTCAGTTAC	60	Caco2	25	Low-medium
	CGTTTCTTTCAGTAGGGGAGGAT				
ARNT	TGACTCCTGTTTTGAACCAGC	60	HaCaT	28	Low
	CTGCTCACGAAGTTTATCCACAT				
ATG5	AAAGATGTGCTTCGAGATGTGT	60	SK-N-SH	28	Low
	CACTTTGTCAGTTACCAACGTCA				
CERS2	GCTCTTCCTCATCGTTCGATAC	60	Caco2	22	High-medium
	CTTGCCACTGGTCAGGTAGA				
COQ7	GTTGATGGTTACGTTCAGGGT	60	MCF7	27	Low
	TTGTTGTAGTGATGTGCTATGCT				
FAAH	GTGACCTCCTATCTGGCTGAC	60	Caco2	28	Low
	CTCACAGGGACGCCATAGAG				
FAAH2	CATAGGCTTAGTAGGCCGAGC	60	Caco2	27	Low
	CTTTCTCTGTCGGATCAGCTTG				
HIF1A	GAACGTCGAAAAGAAAAGTCTCG	60	Caco2	24.5	Low-medium
	CCTTATCAAGATGCGAACTCACA				
HIF1AN	ACGAGAGGTTCCCTAATTTCCA	60	Caco2	23	High-medium
	ATGCCACCAGTACATTGGGAT				
IGF1R	AGGATATTGGGCTTTACAACCTG	60	Caco2	25	Low-medium
	GAGGTAACAGAGGTCAGCATTTT				
LMNA	AATGATCGCTTGGCGGTCTAC	60	Caco2	25	Low-medium
	CACCTCTTCAGACTCGGTGAT				
NCOA1	AGAGGCAACACGACGAAATAG	60	Caco2	24	Low-medium
	ACACTGCATTACTTCATAACGCT				
NGFR	CCTACGGCTACTACCAGGATG	60	Primary fibroblasts	33.5	Low
	CACACGGTGTTCTGCTTGT				
NOS2	TTCAGTATCACAACCTCAGCAAG	60	Caco2	24.5	Low-medium
	TGGACCTGCAAGTTAAAATCCC				
SIRT3	ACCCAGTGGCATTCCAGAC	60	Caco2	26	Low-medium
	GGCTTGGGGTTGTGAAAGAAG				
SIRT6	CCCACGGAGTCTGGACCAT	60	Caco2	27	Low
	CTCTGCCAGTTTGTCCCTG				
TGFB1	CTAATGGTGGAAACCCACAACG	60	Caco2	25	Low-medium
	TATCGCCAGGAATTGTTGCTG				
TLR2	ATCCTCCAATCAGGCTTCTCT	60	Caco2	28	Low
	GGACAGGTCAAGGCTTTTTACA				
TLR4	AGTTGATCTACCAAGCCTTGAGT	60	Primary fibroblasts	30	Low
	GCTGGTTGTCCCAAAATCACTTT				
VEGFA	AGGGCAGAATCATCACGAAGT	60	Caco2	26	Low-medium
	AGGGTCTCGATTGGATGGCA				
VEGFC	GAGGAGCAGTTACGGTCTGTG	60	Primary fibroblasts	25.5	Low-medium
	TCCTTTCCTTAGCTGACACTTGT				

Tab. IV. *Continues.*

Gene	Sequence (5'->3')	Tm	Cell Line/Tissue	Ct	Expression Level
HYOU1	GAGGAGGCGAGTCTGTTGG	60	Primary fibroblasts	26	Low-medium
	GCACTCCAGGTTTGACAATGG				
IL6	CCTGAACCTTCCAAAGATGGC	60	Primary fibroblasts	29.5	Low
	TTCACCAGGCAAGTCTCCTCA				
FTO	ACTTGGCTCCCTTATCTGACC	60	Caco2	23	High-medium
	TGTGCAGTGTGAGAAAGGCTT				
RETN	CTGTTGGTGTCTAGCAAGACC	60	HL60	27	Low
	CCAATGCTGCTTATTGCCCTAAA				
PPARGC1A	TCTGAGTCTGTATGGAGTGACAT	60	HepG2	28	Low
	CCAAGTCGTTCACATCTAGTTCA				
CYP19A1	TGGAAATGCTGAACCCGATAC	60	HepG2	27	Low
	AATTCCCATGCAGTAGCCAGG				
ESR1	CCCACTCAACAGCGTGTCTC	60	MCF7	26	Low-medium
	CGTCGATTATCTGAATTTGGCCT				
ADIPOR2	CTGGATGGTACACGAAGAGGT	60	Primary fibroblasts	24.5	Low-medium
	TGGGCTTGTAAGAGAGGGGAC				
EP300	AGCCAAGCGGCCTAAACTC	60	Primary fibroblasts	27	Low
	TCACCACCATTGGTTAGTCCC				
RFTN1	ATGGGTTGCGGATTGAACAAG	60	Primary fibroblasts	24	Low-medium
	AGCGGTATTCATAGGACACATCT				
SGMS1	TGTGCCGAGTCTCCTCTGA	60	Primary fibroblasts	24	Low-medium
	CCGTTCTTGTGTGCTTCCAAA				
CLTB	CGAGGAGGCTTTCGTGAAGG	60	Primary fibroblasts	24	Low-medium
	GCAGGCGGGACACATCTTT				
ERLIN1	TGGCTCCTTATGCAGTGTTTG	60	Primary fibroblasts	21	High-medium
	GGGCCATGAGGTTTAAGTCTTTC				
ERLIN2	TCCACCACGAACTGAACCAG	60	Primary fibroblasts	26	Low-medium
	AACAGCTCAATGTAGACCTCTTG				
ACE2	CAAGAGCAAACGGTTGAACAC	60	Caco2	31	Low
	CCAGAGCCTCTCATTGTAGTCT				
RIPK1	GGCATTGAAGAAAAATTTAGGC	60	Blood	22	High-medium
	TCACAACTGCATTTTCGTTTG				

sion; Ct ≥ 27 = low expression. T_m = melting temperature; C_t = cycle threshold; Caco2 = human colorectal adenocarcinoma; HepG2 = human hepatocyte carcinoma; MCF-7 = human breast cancer; SH-SY5Y, SK-N-SH = human neuroblastoma from bone marrow; HaCaT = human keratynocyte; HL60 = human promyelocytic leukemia.

Discussion

Identification of dysregulated gene expression pathways involved in human health and disease has significantly contributed to the testing of new compounds as potential drugs. The study of the genes involved in the conditions considered in this study, such as obesity and inflammation, are essential to learn more about the molecular pathways in these diseases and to potentially find new small molecule compounds that might help prevent or treat these diseases. Many primer pairs resulted in high or medium-high expression levels. Because of the elevated expression levels, these primer pairs can be exploited to evaluate dysregulated gene expression in vitro in various conditions. Following this method, it may also be possible to find and test in vitro new molecules with therapeutic potential that could be included in dietary supplements. Finally, diverse study models can be constructed based on these methods, focusing not only on a particular biochemical pathway-natural molecule interactivity, but also on a wider relationship.

Acknowledgements

This research was funded by the Provincia Autonoma di Bolzano in the framework of LP 15/2020 (dgp 3174/2021).

Conflicts of interest statement

Authors declare no conflict of interest.

Author's contributions

MB: study conception, editing and critical revision of the manuscript; KD, GB, KA, KLH, STC, FB, PG: literature

search, editing and critical revision of the manuscript. All authors have read and approved the final manuscript.

References

[1] Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered 2003;56:73-82. https://doi.org/10.1159/000073735

[2] Greenspan RJ. The flexible genome. Nat Rev Genet 2001;2:383-7. https://doi.org/10.1038/35072018

[3] Grimes T, Potter SS, Datta S. Integrating gene regulatory pathways into differential network analysis of gene expression data. Sci Rep 2019;9:1-12. https://doi.org/10.1038/s41598-019-41918-3

[4] Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet 2004;5:101-13. https://doi.org/10.1038/nrg1272

[5] Mitra K, Carvunis AR, Ramesh SK, Ideker T. Integrative approaches for finding modular structure in biological networks. Nat Rev Genet 2013;14:719-32. https://doi.org/10.1038/nrg3552

[6] Langfelder P, Mischel PS, Horvath S. When is hub gene selection better than standard meta-analysis? PLoS One 2013;8:e61505. https://doi.org/10.1371/journal.pone.0061505

[7] Sikdar S, Datta S. A novel statistical approach for identification of the master regulator transcription factor. BMC Bioinformatics 2017;18. https://doi.org/10.1186/s12859-017-1499-x

[8] Tian S, Wang C, Wang B. Incorporating Pathway Information into Feature Selection towards Better Performed Gene Signatures. Biomed Res Int 2019;2019:2497509. https://doi.org/10.1155/2019/2497509

[9] Iourov IY, Vorsanova SG, Yurov YB. Pathway-based classification of genetic diseases. Mol Cytogenet 2019;12:4. https://doi.org/10.1186/s13039-019-0418-4

[10] Bertelli M, Kiani AK, Paolacci S, Manara E, Dautaj A, Beccari T, Michelini S. Molecular pathways involved in lymphedema: Hydroxytyrosol as a candidate natural compound for treating the effects of lymph accumulation. J Biotechnol 2020;308:82-6. https://doi.org/10.1016/j.jbiotec.2019.11.017

[11] Rastogi SC, Rastogi P, Mendiratta N. Bioinformatics methods and applications: genomics proteomics and drug discovery. 3rd ed. Delhi: PHI Learning Pvt Ltd 2008.

[12] Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ, Lee PW. Metabolic control analysis in drug discovery and disease. Nat Biotechnol 2002;20:243-9. https://doi.org/10.1038/nbt0302-243

[13] Davidov E, Holland J, Marple E, Naylor S. Advancing drug discovery through systems biology. Drug Discov Today 2003;8:175-83. https://doi.org/10.1016/s1359-6446(03)02600-x

[14] Cho DY, Kim YA, Przytycka TM. Chapter 5: Network biology approach to complex diseases. PLoS Comput Biol 2012;8:e1002820. https://doi.org/10.1371/journal.pcbi.1002820

[15] Kim YA, Wuchty S, Przytycka TM. Identifying causal genes and dysregulated pathways in complex diseases. PLoS Comput Biol 2011;7:e1001095. https://doi.org/10.1371/journal.pcbi.1001095

[16] Al-Shahrour F, Díaz-Uriarte R, Dopazo J. Discovering molecular functions significantly related to phenotypes by combining gene expression data and biological information. Bioinformatics 2005;21:2988-93. https://doi.org/10.1093/bioinformatics/bti457

[17] Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS, Park PJ. Discovering statistically significant pathways in expression profiling studies. Proc Natl Acad Sci USA 2005;102:13544-9. https://doi.org/10.1073/pnas.0506577102

[18] Kim SY, Volsky DJ. PAGE: parametric analysis of gene set enrichment. BMC Bioinformatics 2005;6:144. https://doi.org/10.1186/1471-2105-6-144

[19] Grimes T, Potter SS, Datta S. Integrating gene regulatory pathways into differential network analysis of gene expression data. Sci Rep 2019;9:5479. https://doi.org/10.1038/s41598-019-41918-3

[20] Jiang Q, Jin S, Jiang Y, Liao M, Feng R, Zhang L, Liu G, Hao J. Alzheimer's Disease variants with the genome-wide significance are significantly enriched in immune pathways and active in immune cells. Mol Neurobiol 2017;54:594-600. https://doi.org/10.1007/s12035-015-9670-8

[21] Swarup S, Huang W, Mackay TF, Anholt RR. Analysis of natural variation reveals neurogenetic networks for Drosophila olfactory behavior. Proc Natl Acad Sci USA 2013;110:1017-22. https://doi.org/10.1073/pnas.1220168110

[22] Yang X, Zhu H, Qin Q, Yang Y, Yang Y, Cheng H, Sun X. Genetic variants and risk of esophageal squamous cell carcinoma: a GWAS-based pathway analysis. Gene 2015;556:149-52.https://doi.org/10.1016/j.gene.2014.11.049

[23] Liu G, Luo S, Lei Y, Wu J, Huang Z, Wang K, Yang P, Huang X. A nine-hub-gene signature of metabolic syndrome identified using machine learning algorithms and integrated bioinformatics. Bioengineered 2021;12:5727-38. https://doi.org/10.1080/21655979.2021.1968249

[24] Kar SP, Seldin MF, Chen W, Lu E, Hirschfield GM, Invernizzi P, Heathcote J, Cusi D; Italian PBC Genetics Study Group, Gershwin ME, Siminovitch KA, Amos CI. Pathway-based analysis of primary biliary cirrhosis genome-wide association studies. Genes Immun 2013;14:179-86. https://doi.org/10.1038/gene.2013.1

[25] Menashe I, Figueroa JD, Garcia-Closas M, Chatterjee N, Malats N, Picornell A, Maeder D, Yang Q, Prokunina-Olsson L, Wang Z, Real FX, Jacobs KB, Baris D, Thun M, Albanes D, Purdue MP, Kogevinas M, Hutchinson A, Fu YP, Tang W, Burdette L, Tardón A, Serra C, Carrato A, García-Closas R, Lloreta J, Johnson A, Schwenn M, Schned A, Andriole G Jr, Black A, Jacobs EJ, Diver RW, Gapstur SM, Weinstein SJ, Virtamo J, Caporaso NE, Landi MT, Fraumeni JF Jr, Chanock SJ, Silverman DT, Rothman N. Large-scale pathway-based analysis of bladder cancer genome-wide association data from five studies of European background. PLoS One 2012;7:e29396. https://doi.org/10.1371/journal.pone.0029396

[26] Nurnberger JI Jr, Koller DL, Jung J, Edenberg HJ, Foroud T, Guella I, Vawter MP, Kelsoe JR; Psychiatric Genomics Consortium Bipolar Group. Identification of pathways for bipolar disorder: a meta-analysis. JAMA Psychiatry 2014;71:657-64. https://doi.org/10.1001/jamapsychiatry.2014.176

[27] Kao CF, Jia P, Zhao Z, Kuo PH. Enriched pathways for major depressive disorder identified from a genome-wide association study. Int J Neuropsychopharmacol 2012;15:1401-11. https://doi.org/10.1017/S1461145711001891

[28] Lee YH, Kim JH, Song GG. Pathway analysis of a genome-wide association study in schizophrenia. Gene 2013;525:107-15. https://doi.org/10.1016/j.gene.2013.04.014

[29] Duncan LE, Holmans PA, Lee PH, O'Dushlaine CT, Kirby AW, Smoller JW, Öngür D, Cohen BM. Pathway analyses implicate glial cells in schizophrenia. PLoS One 2014;9:e89441. https://doi.org/10.1371/journal.pone.0089441

[30] de las Fuentes L, Yang W, Dávila-Román VG, Gu C. Pathway-based genome-wide association analysis of coronary heart disease identifies biologically important gene sets. Eur J Hum Genet 2012;20:1168-73. https://doi.org/10.1038/ejhg.2012.66

[31] Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017;45:D353-61. https://doi.org/10.1093/nar/gkw1092

[32] Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res 2019;47:D419-26. https://doi.org/10.1093/nar/gky1038

[33] Petri V, Jayaraman P, Tutaj M, Hayman GT, Smith JR, De Pons J, Laulederkind SJ, Lowry TF, Nigam R, Wang SJ, Shimoyama

M, Dwinell MR, Munzenmaier DH, Worthey EA, Jacob HJ. The pathway ontology - updates and applications. J Biomed Semantics 2014;5:7. https://doi.org/10.1186/2041-1480-5-7

[34] Nishimura D. BioCarta. Biotech software & internet report 2001;2:117-20. http://doi.org/10.1089/152791601750294344

[35] Paz A, Brownstein Z, Ber Y, Bialik S, David E, Sagir D, Ulitsky I, Elkon R, Kimchi A, Avraham KB, Shiloh Y, Shamir R. SPIKE: a database of highly curated human signaling pathways. Nucleic Acids Res 2011;39:D793-9. https://doi.org/10.1093/nar/gkq1167

[36] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25-9. https://doi.org/10.1038/75556

[37] Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T, Buetow KH. PID: the Pathway Interaction Database. Nucleic Acids Res 2009;37:D674-9. https://doi.org/10.1093/nar/gkn653

[38] Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. Nucleic Acids Res 2020;48:D445-53. https://doi.org/10.1093/nar/gkz862

[39] Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, Caudy M, Garapati P, Gillespie M, Kamdar MR, Jassal B, Jupe S, Matthews L, May B, Palatnik S, Rothfels K, Shamovsky V, Song H, Williams M, Birney E, Hermjakob H, Stein L, D'Eustachio P. The Reactome pathway knowledgebase. Nucleic Acids Res 2014;42:D472-7. https://doi.org/10.1093/nar/gkt1102

[40] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545-50. https://doi.org/10.1073/pnas.0506580102

[41] Biomarkers Definitions Working Group.. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69:89-95. https://doi.org/10.1067/mcp.2001.113989

[42] Lazar C, Taminau J, Meganck S, Steenhoff D, Coletta A, Molter C, de Schaetzen V, Duque R, Bersini H, Nowé A. A survey on filter techniques for feature selection in gene expression microarray analysis. IEEE/ACM Trans Comput Biol Bioinform 2012;9:1106-19. https://doi.org/10.1109/TCBB.2012.33

[43] Jang K, Tong T, Lee J, Park T, Lee H. Altered Gene Expression Profiles in Peripheral Blood Mononuclear Cells in Obese Subjects. Obes Facts 2020;13:375-85. https://doi.org/10.1159/000507817

[44] Burczynski ME, Peterson RL, Twine NC, Zuberek KA, Brodeur BJ, Casciotti L, Maganti V, Reddy PS, Strahs A, Immermann F, Spinelli W, Schwertschlag U, Slager AM, Cotreau MM, Dorner AJ. Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. J Mol Diagn 2006;8:51-61. https://doi.org/10.2353/jmoldx.2006.050079

[45] Erre GL, Piga M, Carru C, Angius A, Carcangiu L, Piras M, Sotgia S, Zinellu A, Mathieu A, Passiu G, Pescatori M. Global microRNA profiling of peripheral blood mononuclear cells in patients with Behçet's disease. Clin Exp Rheumatol 2015;33:S72-S79.

[46] Bonetti G, Paolacci S, Samaja M, Maltese PE, Michelini S, Michelini S, Michelini S, Ricci M, Cestari M, Dautaj A, Medori MC, Bertelli M. Low Efficacy of Genetic Tests for the Diagnosis of Primary Lymphedema Prompts Novel Insights into the Underlying Molecular Pathways. Int J Mol Sci 2022; 23:7414 https://doi.org/10.3390/ijms23137414

[47] Bolognesi ML, Cavalli A. Multitarget drug discovery and polypharmacology. ChemMedChem 2016;11:1190-2. https://doi.org/10.1002/cmdc.201600161

[48] Schenone M, Dančík V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. Nat Chem Biol 2013;9:232-40. https://doi.org/10.1038/nchembio.1199

[49] [49]Zeng H, Qiu C, Cui Q. Drug-Path: a database for drug-induced pathways. Database (Oxford). 2015;2015:bav061. https://doi.org/10.1093/database/bav061

[50] Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, Griffith M, Griffith OL, Wagner AH. Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. Nucleic Acids Res 2021;49:D1144-51. https://doi.org/10.1093/nar/gkaa1084

[51] Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. PubChem substance and compound databases. Nucleic AcidsRes2016;44:D1202–13. https://doi.org/10.1093/nar/gkv951

[52] Mailman MD, Feolo M, Jin Y, Kimura M, Tryka K, Bagoutdinov R, Hao L, Kiang A, Paschall J, Phan L, Popova N, Pretel S, Ziyabari L, Lee M, Shao Y, Wang ZY, Sirotkin K, Ward M, Kholodov M, Zbicz K, Sherry ST. The NCBI dbGaP database of genotypes and phenotypes. Nat Gen 2007;39:1181-6. https://doi.org/10.1038/ng1007-1181

[53] Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, Lee M, Popova N, Sharopova N, Kimura M, Feolo M. NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res 2014;42:D975-9. https://doi.org/10.1093/nar/gkt1211

[54] Swainston N, Hastings J, Dekker A, Muthukrishnan V, May J, Steinbeck C, Mendes P. libChEBI: an API for accessing the ChEBI database. J Cheminform 2016;8. https://doi.org/10.1186/s13321-016-0123-9

[55] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res 2018;46:D1074-82. https://doi.org/10.1093/nar/gkx1037

[56] Thorn CF, Klein TE, Altman RB. PharmGKB: the Pharmacogenomics Knowledge Base. Methods Mol Biol 2013;1015:311-20. https://doi.org/10.1007/978-1-62703-435-7_20

[57] Kuhn M, von Mering C, Campillos M, Jensen LJ, Bork P. STITCH: interaction networks of chemicals and proteins. Nucleic Acids Res 2008;36:D684-8. https://doi.org/10.1093/nar/gkm795

[58] Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, Cheng D, Jewell K, Arndt D, Sawhney S, Fung C, Nikolai L, Lewis M, Coutouly MA, Forsythe I, Tang P, Shrivastava S, Jeroncic K, Stothard P, Amegbey G, Block D, Hau DD, Wagner J, Miniaci J, Clements M, Gebremedhin M, Guo N, Zhang Y, Duggan GE, Macinnis GD, Weljie AM, Dowlatabadi R, Bamforth F, Clive D, Greiner R, Li L, Marrie T, Sykes BD, Vogel HJ, Querengesser L. HMDB: the Human Metabolome Database. Nucleic Acids Res 2007;35:D521-6. https://doi.org/10.1093/nar/gkl923

[59] Kale NS, Haug K, Conesa P, Jayseelan K, Moreno P, Rocca-Serra P, Nainala VC, Spicer RA, Williams M, Li X, Salek RM, Griffin JL, Steinbeck C. MetaboLights: An Open-Access Database Repository for Metabolomics Data. Curr Protoc Bioinformatics 2016;53:14.13.1-14.13.18. https://doi.org/10.1002/0471250953.bi1413s53

[60] Babbi G, Martelli PL, Profiti G, Bovo S, Savojardo C, Casadio R. eDGAR: a database of Disease-Gene Associations with annotated Relationships among genes. BMC Genomics 2017;18:554. https://doi.org/10.1186/s12864-017-3911-3

[61] Zeng X, Zhang P, He W, Qin C, Chen S, Tao L, Wang Y, Tan Y, Gao D, Wang B, Chen Z, Chen W, Jiang YY, Chen YZ. NPASS: natural product activity and species source database

for natural product research, discovery and tool development. Nucleic Acids Res 2018;46:D1217-22. https://doi.org/10.1093/nar/gkx1026

[62] Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, Ojima Y, Tanaka K, Tanaka S, Aoshima K, Oda Y, Kakazu Y, Kusano M, Tohge T, Matsuda F, Sawada Y, Hirai MY, Nakanishi H, Ikeda K, Akimoto N, Maoka T, Takahashi H, Ara T, Sakurai N, Suzuki H, Shibata D, Neumann S, Iida T, Tanaka K, Funatsu K, Matsuura F, Soga T, Taguchi R, Saito K, Nishioka T. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom 2010;45:703-14. https://doi.org/10.1002/jms.1777

[63] Trupp M, Altman T, Fulcher CA, Caspi R, Krummenacker M, Paley S, Karp PD. Beyond the genome (BTG) is a (PGDB) pathway genome database: HumanCyc. Genome Biol 2010;11(Suppl 1):O12. https://doi.org/10.1186/gb-2010-11-s1-o12.

[64] Lv C, Nagle DG, Zhou Y, Zhang W. Application of Connectivity Map (CMAP) Database to Research on Traditional Chinese Medicines (TCMs). In: Zhang W, ed. Systems Biology and its Application in TCM Formulas Research. London: Academic Press 2018, pp. 113-119.

[65] Voskoglou-Nomikos T, Pater JL, Seymour L. Clinical predictive value of the in vitro cell line, human xenograft, and mouse allograft preclinical cancer models. Clin Cancer Res 2003;9:4227-39.

[66] Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer 2006;6:813-23. https://doi.org/10.1038/nrc1951

[67] Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P Jr, de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palescandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R, Garraway LA. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603-7. https://doi.org/10.1038/nature11003

[68] Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S, Lutz SR, Kogera F, Lawrence K, McLaren-Douglas A, Mitropoulos X, Mironenko T, Thi H, Richardson L, Zhou W, Jewitt F, Zhang T, O'Brien P, Boisvert JL, Price S, Hur W, Yang W, Deng X, Butler A, Choi HG, Chang JW, Baselga J, Stamenkovic I, Engelman JA, Sharma SV, Delattre O, Saez-Rodriguez J, Gray NS, Settleman J, Futreal PA, Haber DA, Stratton MR, Ramaswamy S, McDermott U, Benes CH. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 2012;483:570-5. https://doi.org/10.1038/nature11005

[69] Ziliak D, O'Donnell PH, Im HK, Gamazon ER, Chen P, Delaney S, Shukla S, Das S, Cox NJ, Vokes EE, Cohen EE, Dolan ME, Huang RS. Germline polymorphisms discovered via a cell-based, genome-wide approach predict platinum response in head and neck cancers. Transl Res 2011;157:265-72. https://doi.org/10.1016/j.trsl.2011.01.005

[70] Michelini S, Chiurazzi P, Marino V, Dell'Orco D, Manara E, Baglivo M, Fiorentino A, Maltese PE, Pinelli M, Herbst KL, Dautaj A, Bertelli M. Aldo-Keto Reductase 1C1 (AKR1C1) as the First Mutated Gene in a Family with Nonsyndromic Primary Lipedema. Int J Mol Sci 2020;21:6264. https://doi.org/10.3390/ijms21176264.

[71] Paolacci S, Ergoren MC, De Forni D, Manara E, Poddesu B, Cugia G, Dhuli K, Camilleri G, Tuncel G, Kaya Suer H, Sultanoglu N, Sayan M, Dundar M, Beccari T, Ceccarini MR, Gunsel IS, Dautaj A, Sanlidag T, Connelly ST, Tartaglia GM, Bertelli M. In vitro and clinical studies on the efficacy of α-cyclodextrin and hydroxytyrosol against SARS-CoV-2 infection. Eur Rev Med Pharmacol Sci 2021;25:81-9. https://doi.org/10.26355/eurrev_202112_27337

Correspondence: Kristjana Dhuli, MAGI'S LAB, Rovereto (TN), 38068, Italy. E-mail: kristjana.dhuli@assomagi.org

How to cite this article: Dhuli K, Bonetti G, Anpilogov K, Herbst KL, Connelly ST, Bellinato F, Gisondi P, Bertelli M. Validating methods for testing natural molecules on molecular pathways of interest in silico and in vitro. J Prev Med Hyg 2022;63(suppl.3):E279-E288. https://doi.org/10.15167/2421-4248/jpmh2022.63.2S3.31