BIOMEDICAL INFORMATICS

GENE LIST AUTOMATICALLY DERIVED FOR YOU (GLAD4U): DERIVING AND PRIORITIZING GENE LISTS FROM PUBMED LITERATURE

JEROME JOURQUIN

Thesis under the direction of Professor Bing Zhang

Answering questions such as "Which genes are related to breast cancer?" usually requires retrieving relevant publications through the PubMed search engine, reading these publications, and manually creating gene lists. This process is both time-consuming and prone to errors.

We report GLAD4U (Gene List Automatically Derived For You), a novel, free web-based gene retrieval and prioritization tool. The quality of gene lists created by GLAD4U for three Gene Ontology terms and three disease terms was assessed using "gold standard" lists curated in public databases. We also compared the performance of GLAD4U with that of another gene prioritization software, EBIMed.

GLAD4U has a high overall recall. Although precision is generally low, its prioritization methods successfully rank truly relevant genes at the top of generated lists to facilitate efficient browsing. GLAD4U is simple to use, and its interface can be found at: http://bioinfo.vanderbilt.edu/glad4u.

Approved _____ Date _____

GENE LIST AUTOMATICALLY DERIVED FOR YOU (GLAD4U): DERIVING AND PRIORITIZING GENE LISTS FROM PUBMED LITERATURE

By

Jérôme Jourquin

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Informatics

May, 2010

Nashville, Tennessee

Approved:

Professor Bing Zhang

Professor Hua Xu

Professor Daniel R. Masys

ACKNOWLEDGEMENTS

I would like to express profound gratitude to my advisor, Dr. Bing Zhang, for his invaluable support, supervision and suggestions throughout this research work. I would like to deeply thank my committee, Drs. Hua Xu and Daniel R. Masys, for their guidance and support of my master's research.

I would like to thank Dr. Vito Quaranta for his constant support. Dr. Lourdes Estrada was instrumental in the success of my part-time study plan, and this work greatly benefited from the assistance of Ms. Toni Shepard, Brandy Weidow, Jillanne Shell and Jing Hao. Overall, I am also deeply in debt with the entire Quaranta laboratory for their understanding of my morphing into a graduate student. Thank you Walter and Shawn for your patience in answering all the math questions I threw your way. Thank you Manisha, Jill and Brandy for listening and enduring my ups and downs for the last two and a half years.

To the members of the Zhang laboratory—Jing Li, Naresh Prodduturi, Mingguang Shi and Dexter Duncan—thank you so much for your help and feedbacks on my master’s research.

I would also like to thank the Department of Biomedical Informatics. I appreciate the support of the students, faculty, and staff. In particular, I would like to thank Dr. Cynthia Gadd who welcomed me into the program and kept up with my progress, and Ms. Rischelle Jenkins, Claudia McCarn and Nelda Fowles for their assistance throughout this project. Similarly, I appreciate all the help and training that I received from Drs.

David Tabb, Dan Liebler and Robbert Slebos as well as Scott Sobecki, Joseph Burden and William Schultz.

Finally, I am grateful to my family and friends for all their encouragements and support. You have all helped me succeed. My parents have been faithful supporters and expect to be partly paid back with the Commencement celebrations. I am most grateful to Steve for being my number one fan. His strong support, sound advices and continuous encouragement were critical for my research and the realization of this manuscript. Thank you to all my friends, especially to Bart and Sophie, who took my defense as an excuse for re-visiting the South. I have a special 'thank you' addressed to the Matrisian laboratory for "adopting" me; your support has meant a lot!

Thank you to Mr. Steven Goodman and Ms. Brandy Weidow for proofreading the manuscript. This work was supported by the National Institutes of Health (NIH)/ National Institute of General Medical Sciences (NIGMS) through grant R01GM088822, the NIH/National Cancer Institute (NCI) through grant U54CA113007, and the NIH/ National Institute of Mental Health (NIMH) through grant P50MH078028.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACCRE ..... Advanced Computing Center for Research & Education

API ..... Application Programming Interface

CSV ..... Comma-Separated Values

DAG ..... Directed Acyclic Graph

GAD ..... Genetic Association Database

GLAD4U ..... Gene List Automatically Derived for You

GLAD4U Counts ..... GLAD4U prioritization algorithm using counts

GLAD4U Hypergeometric ..... GLAD4U prioritization algorithm using the hypergeometric test

GO ..... Gene Ontology

GOTM ..... Gene Ontology Tree Machine

KEGG ..... Kyoto Encyclopedia of Genes and Genomes

IT ..... Information Technology

MAP ..... Mean Average Precision

MeSH ..... Medical Subject Headings

MIM ..... Mendelian Inheritance in Man

NLP ..... Natural Language Programming

NCBI ..... National Center for Biotechnology Information

OMIM ..... Online Mendelian Inheritance in Man

PMIDs ..... publication identification IDs

SRM ..... Selected Reaction Monitoring

URL ..... Uniform Resource Locator

CHAPTER I

INTRODUCTION

The physical development and phenotype of organisms can be thought of as a product of genes interacting with each other and with the environment. Therefore, it is common for a scientist to ask questions like “Which genes are related to breast cancer?”, “Which genes are involved in embryonic development?”, and “Which genes are functionally related to TP53?”.

The current answers to these questions are primarily contained in the articles indexed in the MEDLINE database. Traditionally, answering these questions requires individuals to retrieve relevant publications through the National Center for Biotechnology Information (NCBI)’s PubMed search engine and then to create gene lists by manually extracting gene-centered information from retrieved literature. This process is not only time-consuming, but also prone to errors. First, it is difficult to ascertain whether all relevant literature is processed. Second, it is unlikely that all relationships in a publication will be detected. Third, individual researchers tend to extrapolate based on domain knowledge.

Over the past decade, bioinformatics approaches have been developed to address these issues. One of the most successful projects in this area is the Gene Ontology (GO) project [1]. GO produces a structured, precisely defined, and controlled vocabulary (i.e., GO terms) for describing the roles of genes and gene products in different species. Genes are associated with GO terms through manual curation as well as computational

inference. A researcher can now access to the GO website [2] to obtain a list of genes related to a GO term of interest. However, because the GO vocabulary only describes gene products in terms of their associated biological processes, cellular components and molecular functions, users are limited by questions linked to this limited vocabulary. Moreover, processes, functions or components that are unique to diseases, such as oncogenesis, are not included in GO because "causing" cancer is not the normal function of any gene.

A useful resource specifically designed for disease studies is the Online Mendelian Inheritance in Man (OMIM [3]) project, "a comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes." It contains information on all known Mendelian disorders. However, information on complex diseases, such as cancer and diabetes, is lacking in OMIM.

In addition to manual curation, text-mining tools have been developed to assist gene list creation [4]. As an example, EBIMed [5, 6] combines text-mining with co-occurrence-based analysis to generate a prioritized list of genes for a user-provided query. Specifically, EBIMed collects MEDLINE records and available full text documents for a user-provided query, identifies protein names, drugs, species, or GO terms in the documents, and prioritizes genes/proteins based on the number of co-occurrences of the different pairs (protein/protein, protein/drug, protein/species, protein/GO term) in the sentences of the documents in which they appear. EBIMed and similar tools, such as FACTA [7] and SciMiner [8], provide more flexible ways to create gene lists that are not limited to certain aspects of biology. Nevertheless, they usually

require heavy computation, and the relevance of the resultant gene lists to the input queries has not been systematically evaluated.

This Master's thesis work reports GLAD4U (Gene List Automatically Derived For You), a new web-based gene retrieval and prioritization tool. GLAD4U takes advantage of existing resources at the National Center for Biotechnology Information (NCBI) to ensure computational efficiency. It provides a simple Google-like interface for easy usage and interpretation. The quality of gene lists created by GLAD4U was assessed using corresponding "gold standard" lists curated in GO, GAD (Genetic Association Database [9]), and OMIM. The performance of GLAD4U was also compared with EBIMed.

CHAPTER II

BACKGROUND

Literature review

This section will highlight web applications, methods, and resources that are relevant to building and evaluating GLAD4U. It is not aimed at being an exhaustive review of the literature on methods available to process information to create gene lists, but will instead provide a snapshot of relevant resources to gene prioritization and evaluation, organized into four categories: gene prioritization with co-occurrence networks, ontologies, integration-aimed software and miscellaneous.

All the applications, methods, and resources described in this section are reported in Table 1, with links and supporting publications.

Gene prioritization using co-occurrences

PubMatrix [10] is an application that mines PubMed using two lists of keywords (search terms and modifier terms) to create a frequency matrix of co-occurrences, which is directly used to prioritize the keywords. Ranking concepts by their co-occurrence in the literature is also at the core of another method, **PubGene** [11], which builds gene-centric literature networks. In these networks, genes and proteins are connected depending on: 1) their co-occurrences in the literature and 2) their similarity of sequences.

Table 1: Applications, methods and resources cited in the Literature review section.

Name	*Source(s) used*
Gene prioritization using co-occurrences	
ALIBABA [12, 13]	Pubmed
EBIMed [6, 14]	Pubmed, UniProt, Gene Ontology
FACTA [7, 15]	Pubmed
PubGene [11, 16]	Pubmed
Pubmatrix [10, 17]	Pubmed
SciMiner [8, 18]	Pubmed, Gene Ontology, HUGO, MeSH
Terminologies/ ontologies	
eVOC [19, 20]	eVOC
Gene Ontology [1, 2]	
Online Mendelian Inheritance in Man [3, 21]	
Textpresso [22, 23]	Pubmed
XplorMed [24, 25]	Pubmed
Resources integration	
Caesar [26]	PageRank, HITS and K-Step Markov model
Endeavour [27, 28]	Gene Ontology, EST expression, InterPro, KEGG, literature, microarray expression data, cis-regulatory motif predictions
GeneSeeker [29-31]	GDB, MIMMAP, MGD, GXD, SWISSPROT, TrEMBL, MLC, OMIM, TBASE, Medline
PolySearch [32, 33]	Pubmed, DrugBank, SwissPort, HGMD, Entrez SNP, OMIM, GAD, HapMap
ToppGene [34, 35]	Gene Ontology, Mouse Genome Informatics [ontology, mouse gene phenotype annotation, orthologous genes from human], KEGG, BioCarta, BioCyc, Reactome, GenMAPP, MSigDb, UniProt, InterPro, Pfam, SMART, PROSITE, Gene3D, ProDom, Pubmed, HPRD, BIND, BioGRID
Miscellaneous	
Anni 2.1 [36, 37]	UMLS thesaurus, Medline, Gene Ontology
GoPubMed [38, 39]	
iHOP [40, 41]	Pubmed
MiSearch [42, 43]	Pubmed
POCUS [44]	Ensembl
PROSPECTR [45, 46]	Ensembl
SUSPECTS [47, 48]	Ensembl

AliBaba [12] is a Java application, which also uses PubMed along with pattern-matching and co-occurrence filtering text-mining approaches to build a concept-centered network of related publications. AliBaba especially focuses on the representation of the results, emphasizing that it is first of all a visualization tool.

FACTA [7] (Finding Associated Concepts with Text Analysis) uses co-occurrence statistics to mine the biomedical literature and present rank tables of query-associated concepts: one table per category and six categories total (gene/protein, disease, symptom, drug, enzyme, compound). FACTA was created with execution speed as one of its priorities (FACTA runs on local copies of resources used for analysis). Unfortunately, FACTA results comprise proteins identified with UniProt IDs as well as proteins identified with "HumanDB" IDs. The latter is not accessible for retrieving protein information, which limits the use of the tool, and it is impossible to integrate FACTA results with other information.

EBIMed [14] is a web-based application that text-mines PubMed to retrieve pairs between any of the proteins, genes, drugs and species that it finds. These pairs are ranked using the Lucene score [49], and are presented in tables, from the most relevant concepts (higher scores) to the least (lower scores). EBIMed defaults to retrieving 500 PubMed abstracts per query, with a maximum retrieval of 10,000 abstracts, a limit imposed because of computation costs. This limitation possibly infers the relevance of the results, as the 10,000 most recent indexed abstracts are not necessarily the most relevant abstracts for a query. Thus, the gene set returned by EBIMed might be incomplete.

SciMiner [8] is a web- and stand-alone application that can be used to mine the biomedical literature for queried genes, and enrich the list with relevant targets (other genes, proteins, pathways). One important feature of this tool is its visualization—powered by Cytoscape [50]—of the molecular interaction networks of the targets. Like EBIMed, this tool is limited to 500 new publications processed per query. Of note, all searches that we attempted exceeded this limit, for which the tool did not run. With no

way to circumvent this limitation—such as running with the first 500 new publications retrieved—it is difficult to design a query to use with this method, beyond the suggested query from the website.

Terminologies/ontologies

The **Gene Ontology** Consortium created and maintains the GO, a group of three ontologies: molecular function, biological process and cellular component. A controlled vocabulary ("terms") is used to annotate genes and gene products according to their functions [1]. GO data is, in part, manually curated and contains 29,959 terms, distributed as follow: 18,581 terms in the biological process ontology, 2,690 in the cellular component ontology and 8,688 in the molecular function ontology (as of 03/05/2010). It can be queried through the web (AMIGO, [51]) or locally implemented after download [52]. Represented as a Directed Acyclic Graph (DAG), terms can appear at multiple levels, in multiple ontologies, and they are linked by relationships such as "is_a" and "part_of."

Started at Johns Hopkins University by Dr. McKusick [21], the **OMIM** [3] was designed as a repository of all known Mendelian disorders—disorders defined by the generational transmission of inherited characters—associated with the genes involved in these disorders. Each Mendelian disorder-associated phenotype or phenotype locus receives a unique identification (MIM ID), which is then preceded by a symbol: an asterisk (*) if the entry is associated with genes of known sequence, a pound (#) if the entry is a phenotype and does not represent a unique locus, a plus (+) if the entry is a phenotype and the associated genes are of known sequence, a percent sign (%) if the

entry describes a confirmed Mendelian phenotype or phenotypic locus with no known underlying molecular basis, and a caret symbol (^) if the entry is deprecated. Finally, if no symbol is added, it means that a Mendelian basis is suspected but not established. The database contains 19,911 MIM IDs: 13,059*, 343#, 2,722+, 1,788% (as of 03/05/2010).

eVOC [19] is a controlled vocabulary aimed at unifying terms describing human anatomy, systems, cell types, diseases and developmental stages. Comparable in design to the more complex GO [53], it was used by Tiffin *et al.* to unify the literature findings into common terms to create disease gene candidates successfully [54]. The prioritization process is mainly dependent on the annotation frequency of both the eVOC vocabulary terms and the genes associated with eVOC terms, with each node score being defined as the sum of all the scores of the genes at this node and of all the scores of the children nodes.

XplorMed [24] is an application that functions as a layer over PubMed. Upon query, XplorMed fetches publications from PubMed and organizes them according to the Medical Subject Headings (MeSH) categories. Unfortunately, the user interface is less than satisfactory because the results present PubMed publication IDs (PMIDs), information that is not readily decipherable by experimentalists.

Textpresso [22] is an ontology that defines a worm-centered vocabulary. It uses the literature, and breaks down the text into "bins" (sentences, words) to create gene-gene interactions. Since its publication, Textpresso has been implemented with different organisms: from its original *Caenorhabditis elegans* to other invertebrates and rat.

Integration

GeneSeeker [29, 30] is a modular web application that derives genetic information from multiple online databases. Thus, it needs neither data warehousing nor content update. As a text-mining application, GeneSeeker provides a filtering layer on top of all information contained in the databases that it queries. Through integration, GeneSeeker is able to present a curated list of candidate genes, smaller than a simple union of the candidate genes retrieved from the different sources, yet containing the genes of importance.

Endeavour [27] is a three-step application prioritizing genes: 1) training to recognize patterns, 2) ranking genes by individual source, according to the training set, and finally, 3) integrating multi-source data ranking. One advantage of querying multiple sources is that the application never relies on any one particular source, while at the same time, it allows the outcome to integrate the most up-to-date data. A similar approach was later adopted by different algorithms: **Caesar** (CAndidatE Search And Rank), a Perl script that uses text- and data-mining to rank genes [26], and a kernel-based algorithm utilizing HITS, K-Step Markov model, and Google's PageRank—the algorithm developed by Google to power its result-ranking algorithm behind its search engine—to rank genes of interest [55].

ToppGene [34] (Transcriptome Ontology Pathway PubMed-based prioritization of GENEs) uses multiple resources, including the NCBI's gene-to-publication mapping, to annotate all genes with their related publications. The unique feature of ToppGene is that it uses mouse phenotype data in addition to other human data sources to enrich its gene prioritization results.

PolySearch [32] is a modular application that, given a concept (or seed), will retrieve human diseases, genes, mutations, drugs or metabolites associations. The user interface is simple ("series of text boxes and pull-down menus"), yet PolySearch has a probability approach ("Given X, find all Y's") that can overwhelm the average experimentalist trying to design queries.

Miscellaneous

POCUS [44] (Prioritization Of Candidate genes Using Statistics) uses probability to compute over-representation of functional annotation between loci for the same disease. Thus, from one gene, the application will present a functionally enriched outcome with other disease-linked genes. There is no web implementation of POCUS, which limits its usability. Similar applications include **PROSPECTR** [45], a machine-learning classifier to discover disease candidate genes through their sequence, and **SUSPECTS** [47], which prioritizes disease candidate genes in large chromosomal regions of interest. These three applications identify potential unknown gene annotations from information already known for neighboring genes or similar sequences.

iHOP [40] (information Hyperlinked Over Proteins), is a network built upon sentences and genes. Interestingly, the authors of this tool found that any two biological concepts can be linked by an average of four steps in their network. It is thus hypothesized that iHOP will help researchers to navigate the literature more easily, making it a good example of information extraction and organization in sentence-gene relationships.

MiSearch [42] is a web application that queries PubMed to retrieve publications. Its unique feature is the ability to adapt its ranking algorithm to the previous queries built by a user. Although it requires a user to open an account to access this feature (a free process), MiSearch is supposed to learn which areas the users are interested in, and rank publications accordingly. MiSearch is limited to the first 5,000 publications returned by PubMed.

Anni 2.1 [36] is an application that executes as a layer to Medline to retrieve documents and to create associations between the concepts mined. Coded with Java, Anni 2.1 is a web- and platform-independent software. The authors of this tool showed that Anni 2.1 was able to extract new associations from the literature simply through its visualization tools (e.g., matrix, projection). A similar categorization of the results is produced by **GoPubMed** [38], where relevant publications are classified within the GO concepts tree.

Preliminary results: Targeted users interviews

When we planned the implementation of GLAD4U, we sought out our targeted users' input. As a preliminary study to assess the needs for and features of GLAD4U, we conducted a small number of free-story interviews of experimentalists—the GLAD4U targeted users. It is important to note that the experimentalists who we interviewed neither tried GLAD4U nor saw the implementation plans. Instead, they were told about the application and its various potential features, whether they were already implemented or not (see Appendix 1 for a survey template adapted from these interviews). Nonetheless, interviewing these potential users was important because they were in the

practice of manually extracting gene lists from the scientific literature in their every-day workflow.

Table 2: Clusters table resulting from free-story interviews analysis.

Time	*Human Factors/ Look/ Output*	*Functionalities/ Options*	*Links/ Outsources*
• Currently, time committed in gathering information • Time of algorithm execution • Time to display first results vs. time to compute all results	• Usability • Clearly stated prioritization rules, as well as any bias • Gene-centered vs. high-throughput publications • Integrated information vs. links opening in new windows • Review vs. research papers	• Exhaustive list of genes • Filters (species, techniques, disease, pathway, etc.) • Number of genes per page, number of supporting publications or genes • Post-output customization • Journal impact factors	• Check list, offer verification tools (locally and remotely) • Gene descriptions • Scale-up (proteins, pathways, etc.) • Consortiums (breast cancer…)

From the clusters table resulting from the analysis of the interviews (Table 2), it was easy to extract three main axes that needed to be part of GLAD4U:

- The speed of the algorithm,
- The clarity of the method used to retrieve and prioritize genes and publications,
- The usability of the application:
    - Parameterization of the algorithm,
    - Clear presentation of results,
    - Ability to outsource the results to third-party tools for further analyses.

Following these axes that were uncovered during our interviews, we drew the following working hypothesis and designed the three Specific Aims defining GLAD4U, the application at the center of this Master's thesis:

Hypothesis: We can build Gene List Automatically Derived For You (GLAD4U) that solely uses PubMed to automatically generate prioritized list of genes upon request.

Specific Aims:

Specific Aim 1. Define and evaluate GLAD4U publication and gene retrieval algorithm,

Specific Aim 2. Increase GLAD4U performance by implementing and evaluating gene prioritization algorithms,

Specific Aim 3. Implement GLAD4U methods as a web user-interface.

CHAPTER III

MATERIALS AND METHODS

Gold standard generation

Gene Ontology terms

To generate the GO-based gold standards, we used two resources: GO and the NCBI (Figure 1). From GO, we downloaded the full GO tree ("gene_ontology.1-2.obo" [52]) and parsed the file to associate each GO ID with its children GO IDs. From the NCBI, we downloaded the GO ID to Entrez-Gene IDs mapping file ("gene2go.gz" [56]). We wrote a Perl script to combine the two resources, including the information about parent-child relationships among the GO terms, i.e., genes with granular annotations were associated with their parent terms. The resulting link table included 9,055 GO IDs (as of 12/20/2009). From this table, we picked three terms with the following two qualities: 1) the term name is short—composed of one to two words, and 2) the term is associated with at least a hundred genes. Following these conditions, our GO-derived gold standards were: apoptosis (GO:0006915, 1,039 genes), cell adhesion (GO:0007155, 785 genes), and DNA repair (GO:0006281, 282 genes) (see Appendix 2 for a list of all genes in each gold standard).

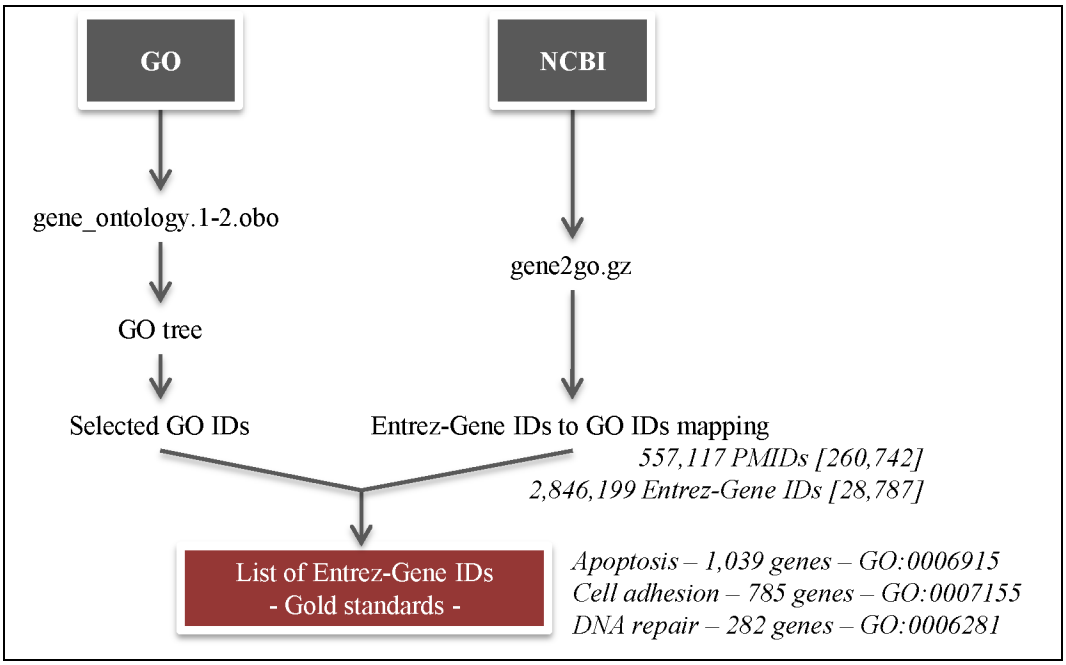

Figure 1: Generating gold standards for GO terms.

Numbers in square brackets indicate the number of publications (PMIDS) and genes (Entrez-Gene IDs) limited to Homo sapiens. *The three GO-derived gold standards are presented with names, numbers of genes in the standard, and their GO IDs. Numbers obtained on 12/20/2009.*

Disease terms

Four resources were used to generate the disease-centered gold standards: GAD [9]), UniProt ([57]), OMIM [3], and the NCBI (Figure 2). With the same rationale used to pick the three GO-derived gold standards, we picked three diseases with a short name and associated with a high number of genes in the "mim2gene" [56]: hypertension, obesity, and schizophrenia. Each of these diseases was then associated with all MIM IDs prefixed with "%" and "#" (phenotypes) and containing the disease name in their title. Corresponding gene IDs were mapped by parsing the file "mim2gene" [56] (as of 12/22/2009). Using GAD [58], we identified all genes associated to a disease. Because GAD uses UniProt IDs to identify disease-associated genes, we used UniProt to map

these IDs to Entrez-Gene IDs. For each disease term, the lists obtained with OMIM and GAD were merged to serve as a gold standard: hypertension (20 MIM IDs, 87 genes), obesity (17 MIM IDs, 111 genes), and schizophrenia (14 MIM IDs, 94 genes) (see Appendix 2 for a list of all genes in each gold standard).

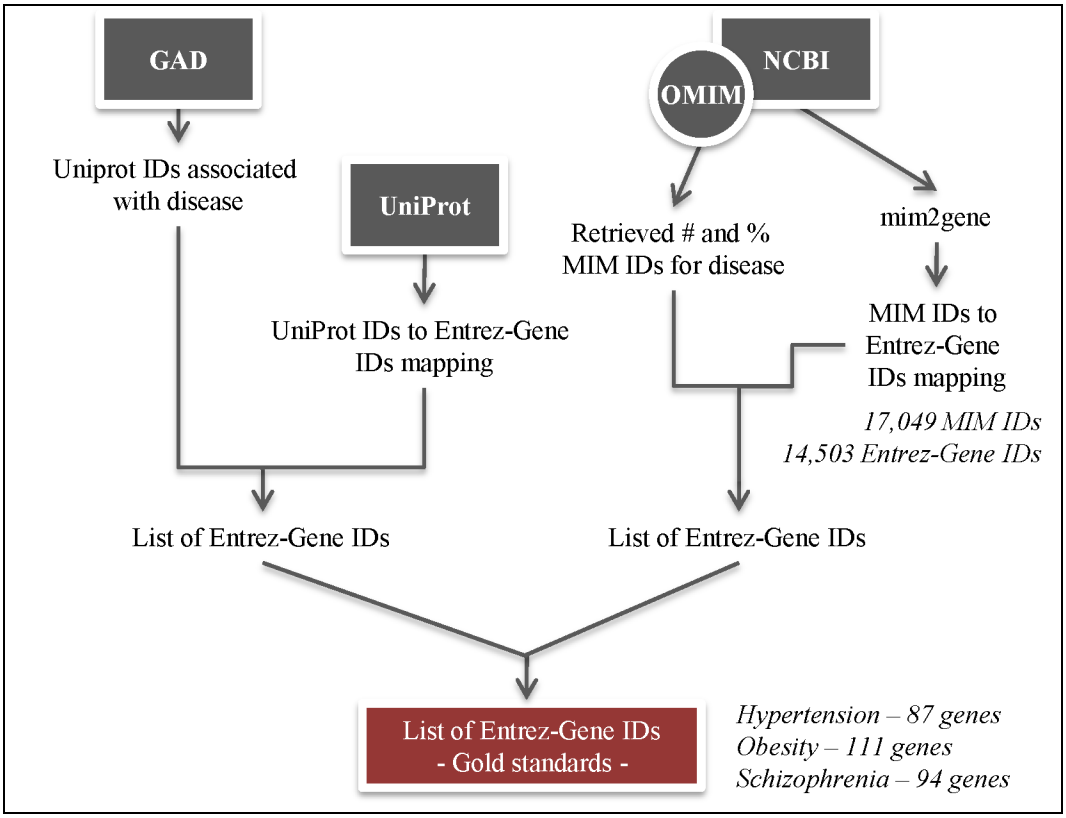

Figure 2: Generating gold standards for disease terms.

The three disease-derived gold standards are presented with names and numbers of genes in the standard. Numbers obtained on 12/22/2009.

Extracting EBIMed-prioritized gene lists

Each gold standard term (apoptosis, cell adhesion, DNA repair, hypertension, obesity and schizophrenia) was used as an input to EBIMed. HTML-formatted tables of

results were saved and parsed with a Perl script to retain gene IDs and their corresponding ranking scores (Figure 3). All IDs not mapped to human genes or those mapped to "obsolete" or "discontinued" UniProt IDs were removed. EBIMed uses UniProt IDs as identifiers, and we used UniProt to map these IDs to Entrez-Gene IDs. As of 12/20/2009, EBIMed-derived lists were: apoptosis (1,469 genes), cell adhesion (1,725 genes), DNA repair (1,100 genes), hypertension (135 genes), obesity (350 genes), and schizophrenia (382 genes) (see Appendix 3 for the complete list of genes returned by EBIMed for each query).

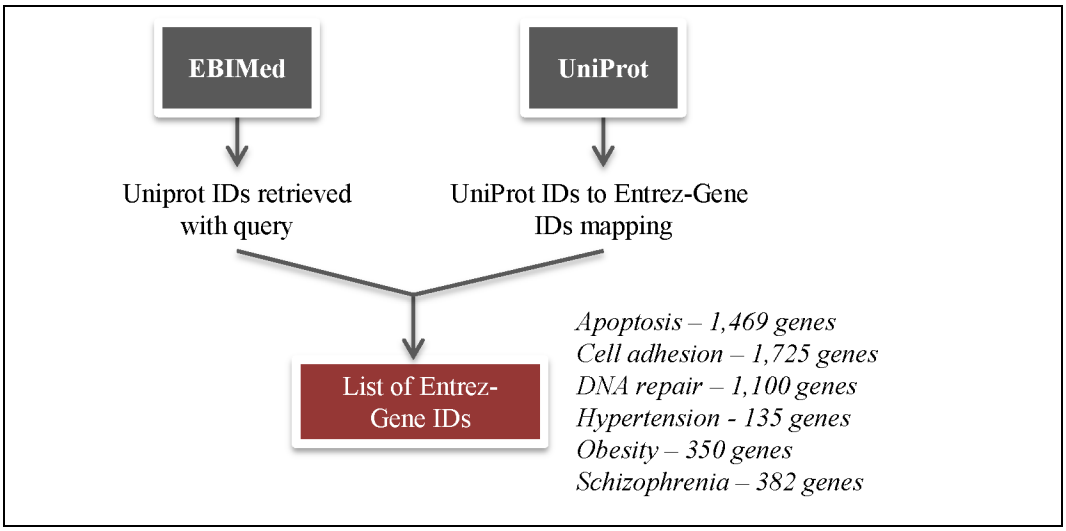

Figure 3: Extracting EBIMed-prioritized gene list.

The six EBIMed-derived gene lists are presented with names and numbers of genes in the lists. Numbers obtained on 12/20/2009.

Publication and gene retrieval

GLAD4U relies on the eSearch application programming interface (API) developed by the NCBI for retrieving publications from the MEDLINE database [59]

(Figure 4). For a user query, eSearch returns an XML file containing the number of publications retrieved by the query and all publication identification IDs (PMIDs). The XML file is parsed to get the list of PMIDs associated with a user query.

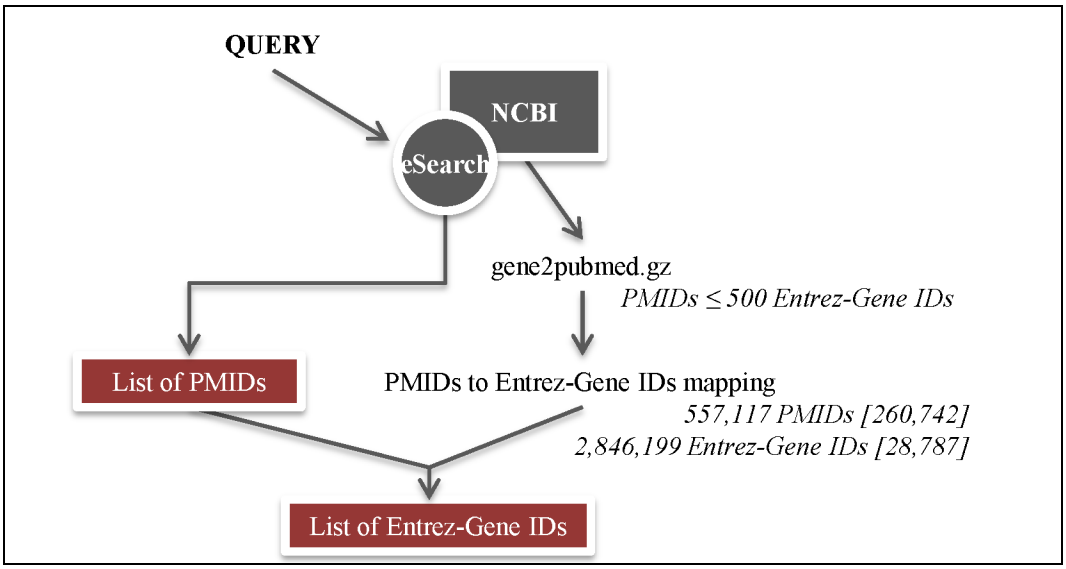

Figure 4: GLAD4U publication and gene retrieval method.

Only the publications related to less or equal to 500 Entrez-Gene IDs were kept in the gene-to-publication mapping table. Numbers in square brackets indicate the number of publications (PMIDS) and genes (Entrez-Gene IDs) limited to Homo sapiens. *Numbers obtained on 12/20/2009.*

Genes associated with PMIDs are retrieved based on the gene-to-publication link table provided by Entrez-Gene [56] (Figure 4). Links between Entrez-Gene IDs and PMIDs are created based on both manual curation within the NCBI and integration of information from other public databases. Publications linked to more than 500 genes are removed from the link table because they lack specificity. After this process, the link table included 2,846,199 genes and 557,117 publications for all organisms, among which 28,787 genes and 260,742 publications were related to human (as of 01/13/2010).

Gene prioritization

We studied two methods to prioritize the retrieved genes based on 1) publication counts, and 2) the hypergeometric test.

GLAD4U Counts

To prioritize using counts ("GLAD4U Counts"), each gene receives a score equal to the number of publications describing it in the link table. For a given query *Q* and a gene *G*, if *j* is the number of publications in the gene-to-publication link table involving the gene *G* (gene-relevant publications), then the scores are calculated as follows:

$S_G = j$

GLAD4U Hypergeometric

Another method ("GLAD4U Hypergeometric") uses the hypergeometric test to prioritize all retrieved genes. Specifically, for a given query *Q* and a gene *G*, let *n* be the number of publications retrieved for the query and present in the gene-to-publication link table (query-relevant publications) and *k* be the number of query-relevant publications that involves the gene *G*. Let us further assume that there are *m* publications in the gene-to-publication link table, *j* of which involve the gene *G* (gene-relevant publications). This method calculates the probability of observing *k* or more query-relevant publications for the gene by chance, based on the hypergeometric test, and scores the gene using the following formula:

$S_G = -\log_{10}(f(m,n,j,k))$, where:

$$f_{(m,n,j,k)} = \sum_{i=k}^{\min(n,j)} \frac{\binom{m-j}{n-i}\binom{j}{i}}{\binom{m}{n}}.$$

Performance evaluation

We used GO and disease terms as queries to evaluate the performance of the GLAD4U algorithms. Retrieval performance was evaluated using performance, recall and F-measures. We used the precision-recall curve, mean average precision (MAP) and precision at k=50 and k=100 to evaluate the performance of our gene prioritization methods [60], and compared it to the performance of the ranked lists generated by EBIMed [6]. All performance values are expressed in the text as mean ± standard deviation.

GLAD4U web user-interface implementation

The GLAD4U user interface was developed in HTML and PHP languages. The scripts to deploy the algorithm on web servers, as well as the generation of hypergeometric test scores, were written in Perl. JQuery was used to implement user-features, such as the ability to hide/show options and functions.

CHAPTER IV

RESULTS

Definition and evaluation of GLAD4U publication and gene retrieval algorithm

Rationale

Searching the scientific literature, there is a lack of applications that extract gene-centered information from the literature, and build a concept-dependent candidate gene list. Using existing resources (i.e., the manually curated NCBI databases), it is possible to build customized gene lists based on user requests, solely based on the scientific literature (PubMed).

Results and Discussion

GLAD4U relies on the NCBI eSearch API to find publications related to a user query and on the gene-to-publication link table to identify genes from the retrieved publications. We used three GO biological process terms (apoptosis, cell adhesion, and DNA repair) and three disease terms (hypertension, obesity, and schizophrenia) as queries to evaluate the overall quality of the retrieved gene lists. For each query, using a corresponding gene list curated by GO or GAD/OMIM as a gold standard, we calculated the precision, recall, and F-measure of the retrieved gene list. As shown in Table 3, gene lists retrieved for all queries showed very high recall (0.90±0.03 for GO terms and 0.96±0.05 for disease terms). In contrast to the high recall, the precision was generally low (0.16±0.04 for GO terms and 0.06±0.02 for disease terms), leading to low F-

measures (0.27±0.05 for GO terms and 0.12±0.03 for disease terms). EBIMed's recall is consistently lower than GLAD4U (0.47±0.15 for GO terms and 0.44±0.11 for disease terms). However, its precision is higher than GLAD4U (0.20±0.05 for GO terms and 0.16±0.04 for disease terms), resulting in better F-measures (0.27±0.03 for GO terms and 0.23±0.04 for disease terms).

Table 3: Overall quality of the retrieved gene lists.

Query	*GO/ MIM gene count*	*GLAD4U gene count*	*EBIMed gene count*		*GLAD4U*	*EBIMed*
Apoptosis	1039	5282 (949) [176854]	1469 (387) [10000]	Precision	0.1797	0.2634
				Recall	0.9134	0.3725
				F-measure	0.3003	0.3086
Cell adhesion	785	3783 (681) [117692]	1725 (305) [10000]	Precision	0.1800	0.1769
				Recall	0.8675	0.3885
				F-measure	0.2982	0.2431
DNA repair	282	2203 (262) [57139]	1100 (180) [10000]	Precision	0.1189	0.1636
				Recall	0.9291	0.6383
				F-measure	0.2109	0.2605
Hypertension	87	1779 (78) [308962]	135 (27) [10000]	Precision	0.0438	0.2000
				Recall	0.8965	0.3103
				F-measure	0.0836	0.2432
Obesity	111	1393 (110) [128830]	350 (59) [10000]	Precision	0.0790	0.1686
				Recall	0.9910	0.5315
				F-measure	0.1463	0.2560
Schizophrenia	94	1471 (92) [86619]	382 (44) [10000]	Precision	0.0625	0.1152
				Recall	0.9787	0.4681
				F-measure	0.1176	0.1849

Numbers in parentheses indicate the number of genes overlapping between the GLAD4U or EBIMed lists and the corresponding gold standard, numbers in square brackets indicate the number of publications retrieved by the query (as of 12/22/2009).

The low precision of GLAD4U may be partially attributed to the incompleteness of the annotation in GO and GAD/OMIM. However, it is likely that the original gene lists include many irrelevant genes. In this case, a prioritization step that ranks truly relevant genes at the top of a list would certainly facilitate more efficient browsing.

Limitations

The speed of the application is a major issue of interest that we identified based on our three interviewees (see Preliminary results section and Table 2). An alternative to improving the speed of our application would be to implement a local version of MEDLINE. Although this approach would require frequent updates to keep the scientific literature database current, it would also reduce the time to retrieve publications by approximately ten-fold. This modification in GLAD4U would also solve another potential limitation: the dependence on the NCBI's API to PubMed. Another drastic improvement of the application speed would occur if it were redeveloped in C++, a computer language known for its memory and speed efficiency.

Increase of GLAD4U performance by the implementation and evaluation of gene prioritization algorithms

Rationale

GLAD4U consistently retrieves most of the genes present in the gold standards (high recall), but it also retrieves many other genes not present (low precision). Thus, GLAD4U results will greatly benefit from adding a prioritization method: bringing the relevant genes (i.e., those present in the gold standard) to the top of the list.

Results and Discussion

We studied the performance of two methods to prioritize gene lists. The first, "GLAD4U Counts," is based solely on the number of supporting publications as commonly implemented in other software [10, 61]. The second, "GLAD4U

Hypergeometric," is proposed in this study, which is based on the Hypergeometric test (see the Materials and Methods section for details). We used the above-mentioned three GO terms and three disease terms as queries to evaluate the performance of our prioritization methods. We also included the prioritized gene lists returned by EBIMed for comparison.

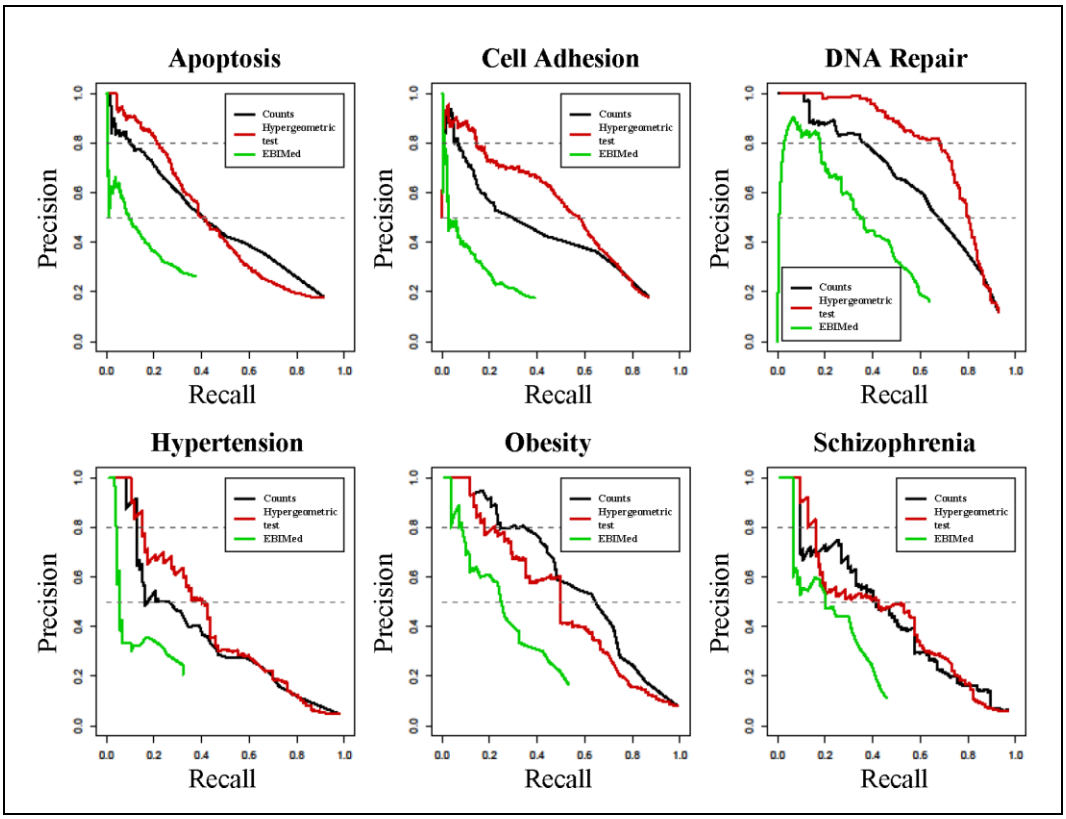

Figure 5: Precision-recall curves for different prioritization methods.

Precision-recall curves for GLAD4U Counts, GLAD4U Hypergeometric and EBIMed are colored in black, red, and green, respectively. Dashed lines correspond to the precision levels of 0.8 and 0.5.

Figure 5 depicts the precision-recall curves from this comparative evaluation. For all queries, based on manual inspection of the curves, both GLAD4U Counts and GLAD4U Hypergeometric methods outperformed EBIMed, especially at the high

precision range. Between the two GLAD4U methods, the Hypergeometric method performed better than the Counts method for GO term queries, while their performances were comparable for disease term queries. The superior overall performance of the two GLAD4U methods over EBIMed was further evaluated by computing MAP, a quantitative measure of quality across all recall levels (Table 4). In this analysis, GLAD4U Counts and Hypergeometric methods scored better than EBIMed (0.48±0.10, 0.52±0.12, and 0.21±0.09, respectively), with GLAD4U Hypergeometric performing the best.

The precision-recall curves and the MAP scores factor in precision at all recall levels. For ranked gene lists, particularly in web-based applications, this variable may not be of interest to users. In most scenarios, what matters may be the number of relevant genes on the first page or the first several pages. "Precision at k" is usually used to measure precision at a fixed low level of retrieved results, i.e., the top k results [60]. To this end, we calculated precisions for the top 50 (k=50) and top 100 (k=100) genes for all three methods, for each query (Table 4). GLAD4U Counts and GLAD4U Hypergeometric methods maintained higher precisions for the top 50 genes compared to EBIMed (0.74±0.15, 0.77±0.20, and 0.54±0.18, respectively), as well as for the top 100 genes (0.64±0.20, 0.69±0.25, and 0.42±0.20, respectively). Although the MAP-based comparison may be biased against EBIMed owing to its low overall recall, precision at 50 and 100 only focus on the top ranking genes and are not affected by the overall recall. These results suggest that GLAD4U can produce lists where relevant genes are ranked at the top.

Table 4: Comparison of different prioritization methods.

	Apoptosis	*Cell Adhesion*	*DNA Repair*	*Hypertension*	*Obesity*	*Schizophrenia*
GLAD4U Counts						
MAP	0.4616	0.4067	0.6205	0.3729	**0.5911**	**0.4479**
Precision at k=50	0.8400	0.8000	0.8800	0.5000	**0.8000**	**0.6200**
Precision at k=100	0.8400	0.7500	0.8400	0.3600	**0.5900**	0.4500
GLAD4U Hypergeometric						
MAP	**0.4650**	**0.5054**	**0.7561**	**0.4234**	0.5056	0.4429
Precision at k=50	**0.9400**	**0.9000**	**1.0000**	**0.5800**	0.6800	0.5400
Precision at k=100	**0.9000**	**0.8700**	**0.9800**	**0.3800**	0.5288	**0.4900**
EBIMed						
MAP	0.1567	0.1256	0.3517	0.1336	0.2673	0.2318
Precision at k=50	0.6200	0.4800	0.8400	0.3137	0.5652	0.4423
Precision at k=100	0.5980	0.4848	0.6700	0.2821	0.1586	0.3200

Numbers in bold indicate the maximum value among the three methods.

Although precision was less than perfect even for the top ranking genes, we noticed that many false-positives could be explained by the incompleteness of the gold standards. Table 5 lists the first 10 genes—along with their first 10 supporting publications—returned by the GLAD4U Hypergeometric method that were not in the corresponding gold standards for the terms "apoptosis" and "hypertension." Taking the first and last genes in the list as examples, for each term (i.e., MDM2 and TXN for apoptosis, and REN and ACE2 for hypertension), we found strong evidence in the most recent supporting publications for linking these non-gold standard genes to the query. MDM2 has antiapoptotic effects, and its direct interaction and regulation of p53 define it as an oncogene [62]. It translocates to the nucleus to interact with p53 and p300 and promotes cell growth by initiating p53 degradation [63, 64]. Its expression is directly linked to prostate cancer patient outcome, potentially predicting the therapeutic response [65]. TXN also has antiapoptotic effects—comparable to BCL2 [66]—and protects cells from Fas-mediated oxidative stress [67] in cancers [68] and cerebral ischemia [69]. It

likely inhibits apoptosis in different ways, as it has been found to down-regulate ASK1 activity and the tumor suppressor PTEN expression [70].

Table 5: First 10 genes retrieved by GLAD4U and not listed in the gold standard lists.

Rank	*Entrez-Gene ID (Gene symbol)*	*Score*	*PMIDs*
		Apoptosis	
44	4193 (MDM2)	40.6211	19759023, 19657064, 19648117, 19639206, 19573080, 19541936, 19541625, 19524506, 19521721, 19470936
46	4609 (MYC)	36.5254	19815507, 19573080, 19407242, 19336552, 19332536, 19330811, 19180571, 19170058, 19143767, 19134217
47	1432 (MAPK14)	36.1962	19748889, 19723092, 19698994, 19628771, 19543489, 19497411, 19468799, 19467570, 19433314, 19343039
56	6774 (STAT3)	30.8516	19724924, 19684620, 19626047, 19623660, 19595668, 19578748, 19503092, 19484147, 19457567, 19429240
76	5580 (PRKCD)	20.5587	19667069, 19581935, 19563780, 19477272, 19146951, 19037087, 19002183, 18952226, 18637130, 18434324
78	29126 (CD274)	20.0903	19826049, 19794071, 19759858, 19739236, 19684086, 19116915, 19081139, 19056397, 18981087, 18941206
84	142 (PARP1)	18.0948	19723035, 19655414, 19557639, 19529948, 19513550, 19506301, 19372636, 19368128, 19281796, 19144573
90	3621 (ING1)	15.9815	19085961, 18836436, 18801192, 18691180, 18655775, 18533182, 18388957, 17585055, 17379210, 16607280
94	406991 (MIR21)	15.2447	19826040, 19682430, 19641183, 19633292, 19597153, 19578724, 19559015, 19546886, 19509158, 19509156
95	7295 (TXN)	15.2083	19566940, 19328186, 19120277, 18983687, 18848838, 18497292, 17896802, 17823364, 17724081, 17652454
		Hypertension	
10	5972 (REN)	54.7687	19891555, 19673942, 19536175, 19509012, 19369955, 19243623, 19126660, 19082699, 18856058, 18722896
12	3291 (HSD11B2)	44.4121	19150652, 18837962, 18573267, 18178212, 17551100, 16872738, 16778331, 16109323, 16061836, 15673310
16	4878 (NPPA)	27.9005	19635983, 19479237, 19430483, 19346663, 19330901, 19219041, 19146799, 19131662, 18633189, 18212314
17	155 (ADRB3)	27.5873	19842096, 19779464, 19479237, 19131662, 18724972, 18510051, 18088254, 17785925, 17626108, 17439327
19	4879 (NPPB)	26.2767	19919978, 19662018, 19635983, 19430483, 19413180, 19391062, 19387960, 19387249, 19327608, 19262210
20	4524 (MTHFR)	25.1380	19853876, 19824427, 19810824, 19776634, 19776610, 19742390, 19717029, 19479237, 19430483, 19394322
22	1401 (CRP)	23.9612	19615354, 19410251, 19375128, 19282863, 19262552, 19244088, 19193941, 19139603, 19075099, 19072030
23	1584 (CYP11B1)	23.7840	19820005, 19567537, 19082699, 18663314, 18294861, 17980006, 17296872, 17121536, 17075029, 16984984
27	6296 (ACSM3)	17.9180	19262474, 18519841, 18192838, 17278971, 17070428, 15361761, 14567496, 12484505, 12046348, 11772874
31	59272 (ACE2)	15.5964	19684612, 19289653, 19286756, 19077694, 18926157, 18258853, 18022600, 17504232, 17473847, 17303661

Regarding hypertension, REN is part of the renin-angiotensin system (RAS) and is regulated in part by ACE2. Both proteins are important regulators of blood pressure and are involved in the onset of hypertension [71, 72]. Thus, inhibiting the regulators of the RAS—such as ACE and its counterpart ACE2 [73, 74]—is a common treatment for hypertension [71, 75]. Interestingly, both REN and ACE2 present polymorphisms, which seem linked to therapeutic response to hypertension [73, 76-81].

From these publications, we believe that MDM2 and TXN should be part of the gold standard for apoptosis, and that REN and ACE2 should be part of the standard for hypertension. These results accentuate the incompleteness of the gold standard and suggest that GLAD4U can facilitate the completion of gold standard lists, and in the task of automatic gene-centered annotations.

We also looked at genes not retrieved by GLAD4U, but present in the gold standards. For examples, we used the two gold standards previously investigated: apoptosis and hypertension for which GLAD4U "missed" 90 genes (representing 8.66% of the gold standard) and nine genes (representing 10.34% of the gold standard), respectively. For each of these two gold standards, we looked at the publications retrieved from PubMed using several of the genes (NMNAT3, ATP7A, ARNT2 and KRT8, LPCAT3, KRT18 for apoptosis and hypertension, respectively) as queries (Table 6). We then compared the set of publications with the set retrieved by querying PubMed with apoptosis and hypertension, accordingly—as a note, the number in this table vary from Table 3 due to the different date at which the queries were done. The common publications between the two sets (gene and query) were analyzed (Table 6).

Table 6: Genes not retrieved by GLAD4U, but listed in the gold standard lists.

Entrez-Gene ID (Gene symbol)	*Publication count*	*Overlapping PMIDs*
Apoptosis [176698]		
349565 (NMNAT3)	14 (1)	20117162
538 (ATP7A)	384 (3)	19578756, 18779302, 19470734
9915 (ARNT2)	66 (3)	19401220, 12527906, 10215907
Hypertension [311186]		
3856 (KRT8)	118 (0)	
10162 (LPCAT3)	4 (0)	
3875 (KRT18)	70 (0)	

Numbers in parentheses indicate the number of publications overlapping between the publications retrieved by the gene and by the query, using PubMed. Numbers in square brackets indicate the number of publications retrieved by the query (as of 03/12/2010).

In the context of apoptosis, the publications retrieved for the genes in the gold standard, but not retrieved by GLAD4U, do not clearly associate them with apoptosis. For example, NMNAT3 (nicotinamide mononucleotide adenylytransferase 3) promotes axonal protection against exogenous oxidants, and the effects are likely to be driven by the tight connection between NMNAT3, mitochondria, and neuronal cell death [82]. ATP7A is required for transport of copper [83] and—along with ATP7B—has been positively associated with the degree of cisplatin resistance in tumor cell lines and clinical specimens [83-85]. Although ATP7A overexpression in ovarian carcinoma cells renders them resistant to cisplatin [84], another study found that ATP7A gene silencing had no significant effect on sensitivity to cisplatin *in vitro*, a role that was instead attributed to ATP7B [85]. Finally, ARNT2 is a transcription factor that, in response to low oxygen concentrations, binds to hypoxia and EWS/ATF-1 responsive elements of genes such as vascular endothelial growth factor (VEGF), glucose transporters, and glycolytic enzymes [86, 87]. Further, ARNT2 seems involved in tumor angiogenesis and

growth [86], but the mechanism of its action remains unknown [88]. Nonetheless, it is regulated during the cell cycle and regulates other genes—possibly related to apoptosis—involved in the cell cycle [88].

Interestingly, none of the publications related to the genes in the hypertension gold standard and not retrieved by GLAD4U were in the set of publications related to hypertension. This difference explains why GLAD4U did not retrieve these genes: No publications clearly linked these genes to the query.

Limitations

One of the main limitations is our ability to evaluate GLAD4U prioritization algorithms; even if the MAP measure is a good alternative when gold standards are unranked, the evaluation of GLAD4U would be better if we had corresponding prioritized gold standards. Ironically, the gold standards themselves are another major limitation. Simply put, the performance evaluation of GLAD4U is only as good as our gold standards. And, as exemplified by the results, the gold standards are far from perfect, especially in the case of disease terms.

Implementing gene prioritization methods helped increasing the relevance of the gene lists, especially in the higher positions of the lists. But the precisions at k=50 and k=100 are still not maximal: because of the high value of the recall, GLAD4U retrieved virtually all of the relevant genes, but did not rank all of them at the top of the lists. Thus, exploring different prioritization methods would help take full advantage of the gene set retrieved by GLAD4U.

Implementation of GLAD4U methods as a web user-interface

Rationale

GLAD4U methods build relevant literature-derived gene lists. Its implementation as an application will be an important addition to existing tools. We chose to embed the methods as a web user-interface because 1) GLAD4U is dependent on the NCBI web service and thus, Internet connection is one of the requirements for GLAD4U, and 2) providing web access to our methods ensures that most of our targeted users will be able to use GLAD4U.

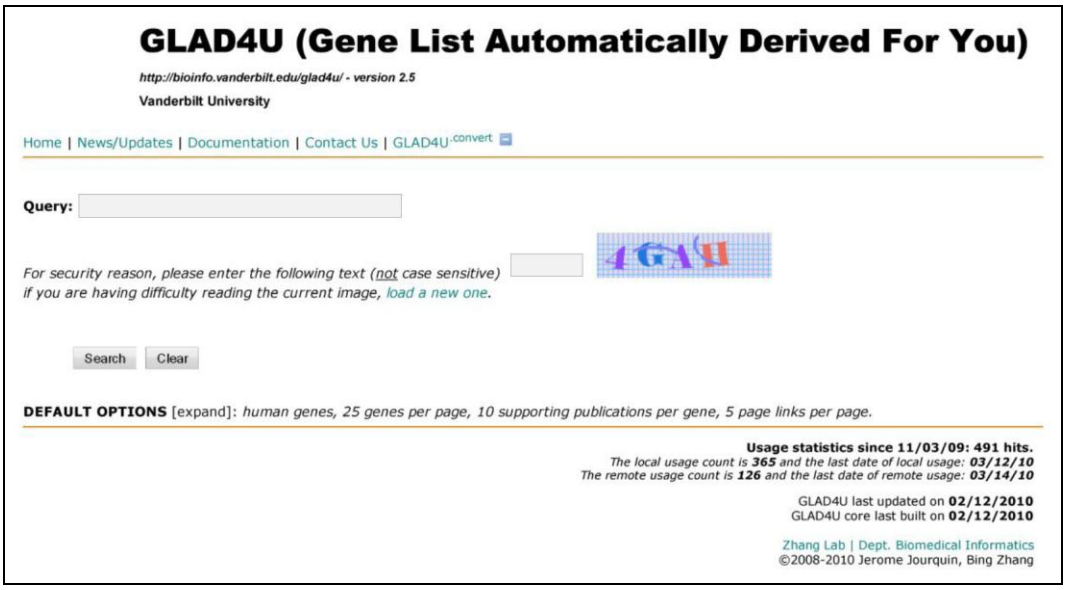

Figure 6: GLAD4U user interface – submitting query.

To submit a query to GLAD4U, the user simply enters it in the Query field, matches the security code, and submits the form. The default options are displayed, but can be changed by expanding the section. Various usage statistics are also displayed.

Results and Discussion

GLAD4U uses a simple query interface for users to submit their queries (Figure 6). Any queries that are valid in a PubMed search can be used in GLAD4U. In the query interface, users can also modify the default parameters of the application, including search space (all species or restricted to human genes), the number of genes to present per result page, the maximum number of publications supporting each gene returned in the result page, and the number of pages to build for each of the algorithm runs.

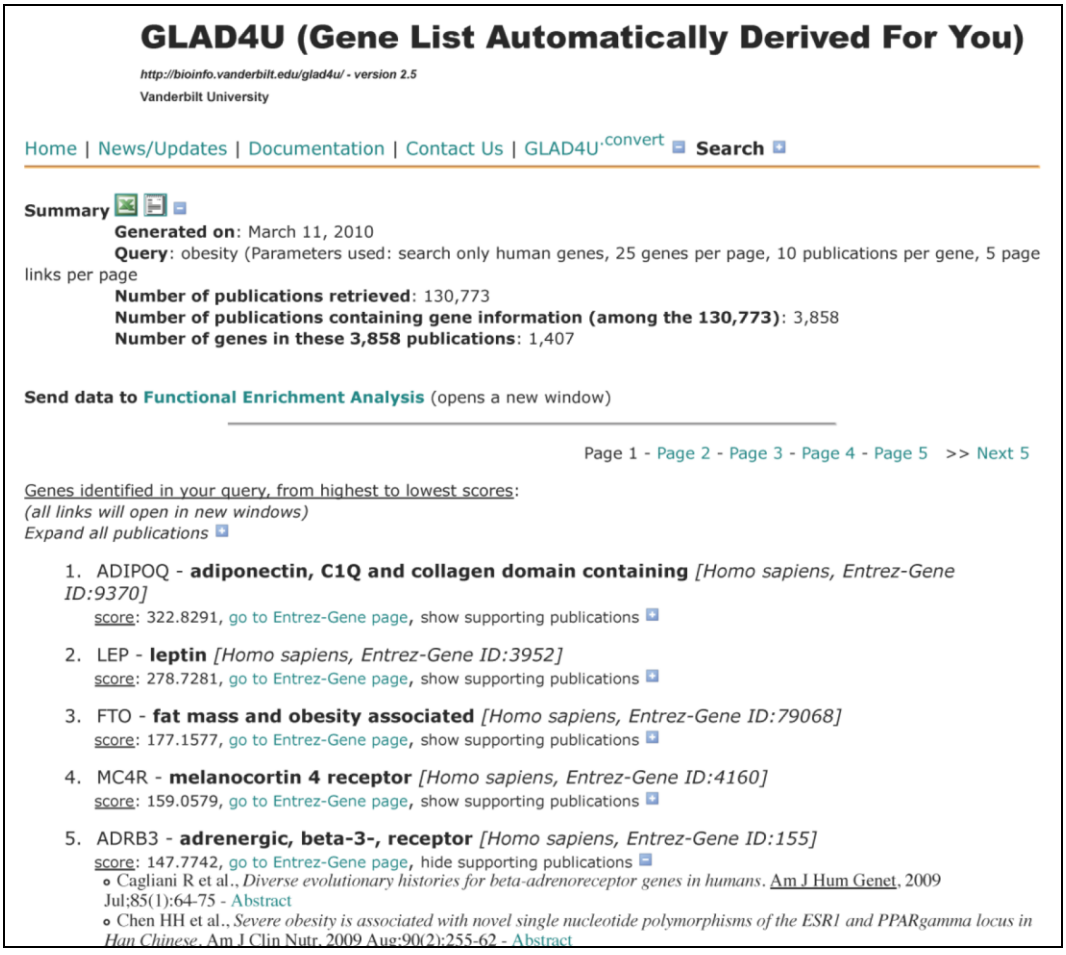

Figure 7: GLAD4U user interface – result page.

The summary section presents the main statistics for the query, along with two icons to download the results as a CSV or a text file. Below the summary, a link sends the results for functional enrichment analysis. In the main result section, user can click the "+" to show/hide the supporting publications. ADRB3 gene is presented with its supporting publications as an example.

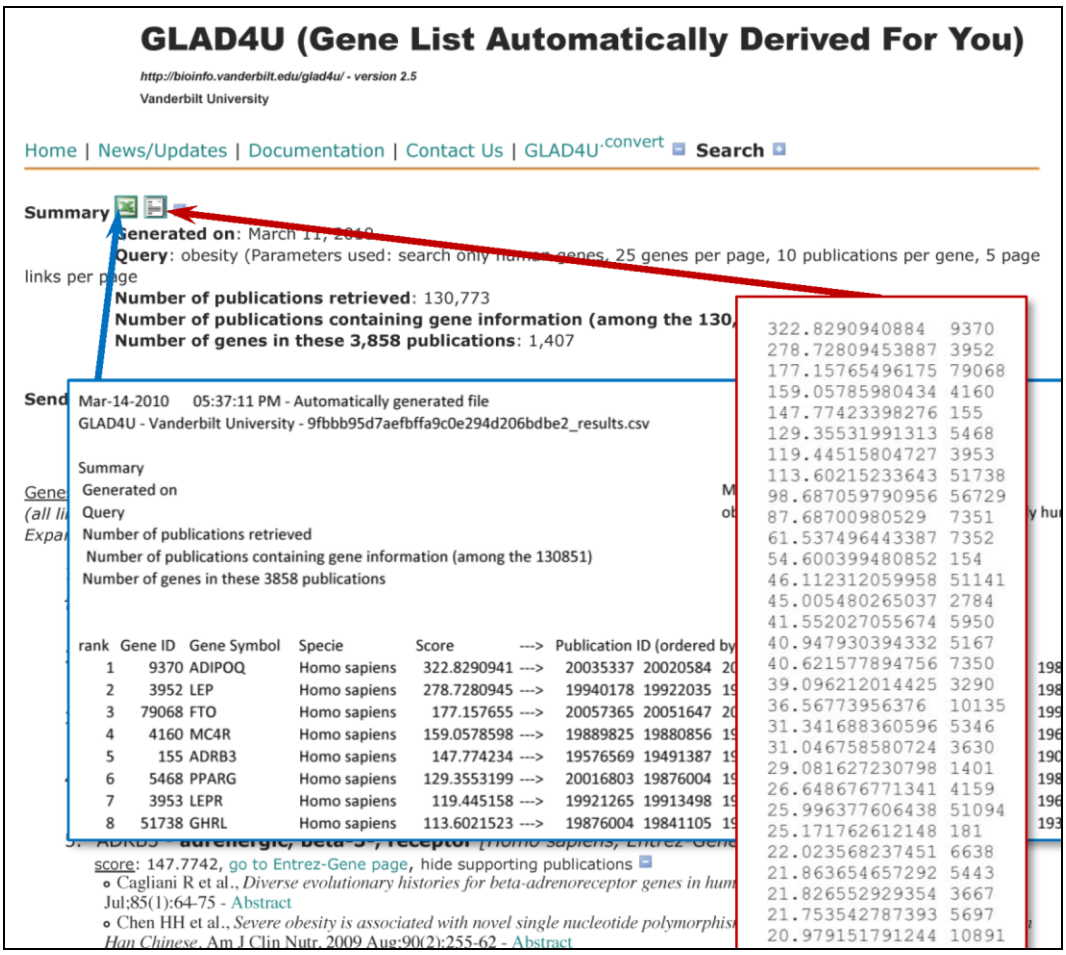

Figure 8: GLAD4U output 3 - export files.

Clicking the Excel icon will retrieve the results as an Excel spreadsheet (blue arrow and insert). Clicking the text-file icon will show the results as score/Entrez-Gene ID pair, one per line (red arrow and insert).

The output page displays the ranked gene list and information associated with each gene (Figure 7). As each gene is identified by an Entrez-Gene ID, we use eSummary, another one of NCBI's eUtilities [59], to fetch annotations for the gene including name, symbol, and species. Publications supporting the relationship between a gene and the query term are listed under the gene. The publications are ordered based on

their PubMed IDs so that the most recent publication is listed first (see Figure 7, under the "LAMC2" gene description). As for genes, we use eSummary to fetch information relevant to the publication such as title, authors and journal name. Genes and publications are hyperlinked to the corresponding NCBI pages, which will—by design—open in a new window to avoid disrupting the result page.

At the top of the output page, a summary of the run is also given, including: the date of the query, the query term(s) and options chosen, the number of genes and publications processed, as well as a hyperlink to download the complete results in the comma-separated values (CSV) format or as text file (Figure 8). Although this file may be difficult to interpret by experimentalists, it can be easily used as input for other computational analysis tools. For example, we have implemented a "send data to Functional Enrichment Analysis" link in the result page (Figure 7) of GLAD4U for submitting a gene list to the functional enrichment analysis tool GOTM [89, 90]. This function is particularly useful for the functional interpretation of a gene list, i.e., a list returned by a disease term query. That is, it could help reveal biological processes associated with the disease. As an example, enrichment analysis on the first 100 genes returned by the "obesity" query linked this disease to biological processes such as "fat cell differentiation" (20 genes, multiple-test adjusted enrichment p-value (adjp) = 8.64E-28), "lipid metabolic process" (40 genes, adjp=7.33E-23), and "response to insulin stimulus" (17 genes, adjp=6.65E-18) (Figure 9).

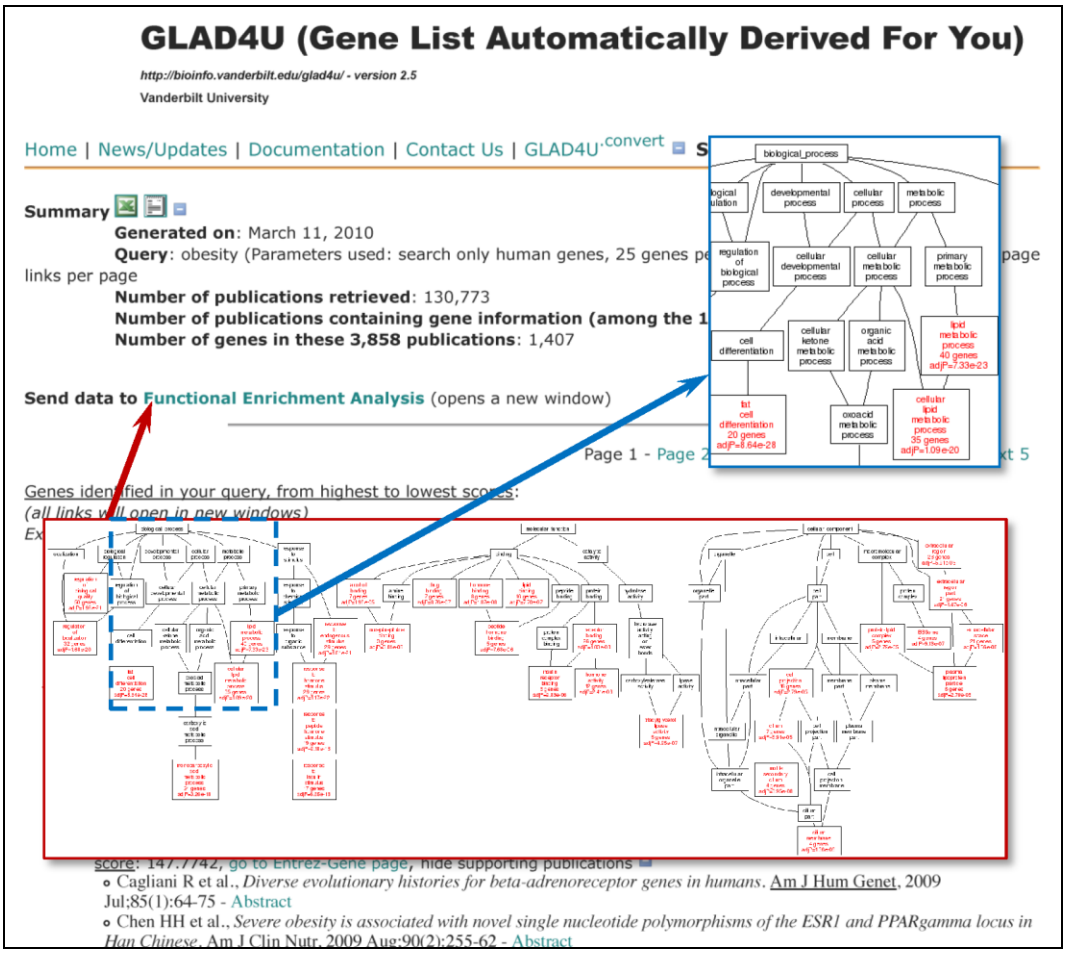

Figure 9: GLAD4U output 4 – functional enrichment analysis.

The functional enrichment analysis is performed by the GOTM [89, 90]. The resulting DAG is presented (red arrow and insert), as well as a magnified section for better readability (blue arrow and insert). The red boxes in the DAG correspond to enriched GO terms, based on the gene list returned by GLAD4U and submitted for analysis.

Limitations

Implementing GLAD4U as a web user-interface is solely dependent on one web server. If this server were down, GLAD4U would become unavailable. One alternative to overcome this would be to have GLAD4U web implementation iteratively installed on multiple web servers. In this case, if one server went down, GLAD4U Uniform Resource Locator (URL) would be redirected to one of the working web servers. Unfortunately,

this alternative would be resource-intensive relative to the inherent complexity of the Information Technology (IT) infrastructure of an Institution like Vanderbilt University. Another alternative is to make GLAD4U implementation available for download and third-party implementation. This solution would empower users with the possibility to have their own GLAD4U instance, but also with the responsibility of keeping their GLAD4U implementation updated. For that purpose, we would add another script to render the update of GLAD4U automatic.

CHAPTER V

DISCUSSION

Reading through all relevant literature to manually generate a gene list is time consuming [10, 27, 32, 91], a common and significant concern in all interviews of experimentalists that we performed (see Preliminary results). GLAD4U addresses this problem by automatically creating a ranked list of genes following a user's input query.

One important feature of GLAD4U is its information processing. Based on our survey among experimentalists, GLAD4U follows the exact steps that an experimentalist would follow (Figure 10A): gather literature, extract gene information, and create an expert list [92]. Whether a user queries a disease, a non-disease phenotype, a biological process, or a gene, GLAD4U will fetch corresponding biomedical publications using NCBI's eUtilities API, retrieve relevant gene information, rank them, and send them back to the user (Figure 10B).

Another important feature of GLAD4U is its simplicity. Researchers will be at ease using GLAD4U because its search engine is powered by PubMed's API [32, 59], and behaves similarly to Entrez-PubMed [93]. GLAD4U outputs a clean result page where the user can easily find and search for genes relevant to the concept queried and supporting publications. Additionally, the use of PubMed's API makes GLAD4U almost maintenance-free. GLAD4U will update itself along with the MEDLINE library update. This process will ensure that GLAD4U's results will always be up-to-date with the current literature.

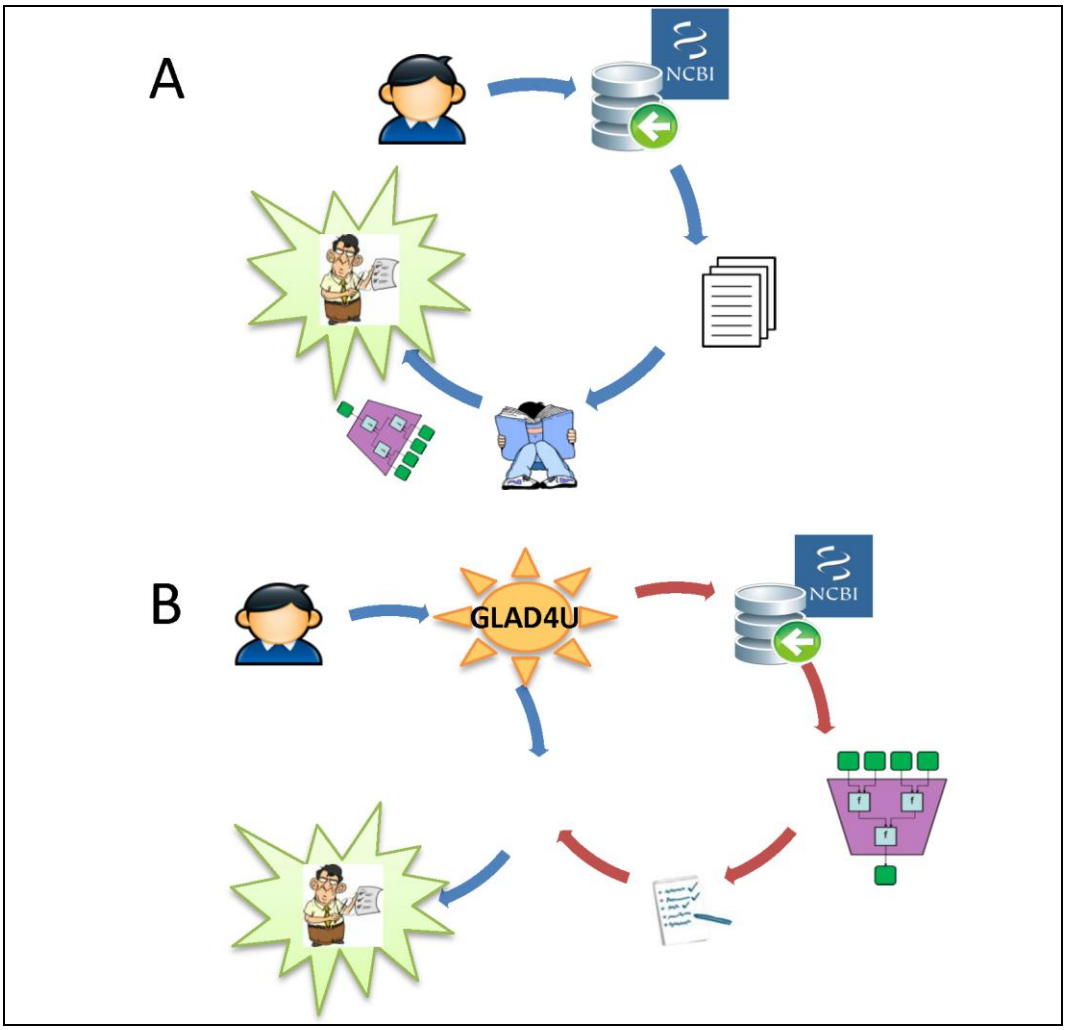

Figure 10: Traditional and GLAD4U workflows.

(A) Traditional workflow for manually creating gene lists. Users query PubMed, retrieve and read the relevant publications to create the gene list. (B) GLAD4U workflow. Users query GLAD4U, which automatically sends the query to PubMed, retrieve the publications, corresponding genes, and ranks them. The resulting gene lists is returned to users.

Several tools rely on PubMed to build disease candidate genes lists [5, 8, 12, 32, 34]. EBIMed [5] and FACTA [7] are recent, concept-oriented applications for mining existing biomedical literature. Similar to GLAD4U, these two tools mine biomedical literature and present ranked tables of query-associated concepts. They use co-occurrence

statistics to rank and organize results. For computational efficiency, EBIMed only processes a maximum of 10,000 publications. This limitation may explain why EBIMed's performance was surprisingly low, as compared to GLAD4U (Table 3). Nonetheless, EBIMed still ranks relevant genes at the top of the list. Unlike GLAD4U, FACTA is primarily focused on the speed of execution as it runs with local copies of resources. By comparison, GLAD4U uses live retrieval of publications, allowing for up-to-date information retrieval. Moreover, GLAD4U is gene-centered and will solely organize its results based on the genes.

Although using the biomedical literature as a knowledge source seems intuitive [27, 44, 94], certain limitations exist: the literature is indexed based on titles, abstracts and keywords, instead of on full-text [11, 22]. Thus, a set of publications retrieved may be incomplete (i.e., some publications relevant to the concept queried will not be retrieved because they do not contain the necessary keywords in their titles or abstracts) [95]. There is a possible bias in using the biomedical literature and ontology [93], as the most studied genes (those with the most publications) will have more weight [27, 54] at the expense of more relevant genes that might only be featured in few papers [96]. Thus, we use the hypergeometric test to rank genes based on how likely it would be to retrieve them by chance alone, based on the number of publications retrieved for this gene among the total number of publications linked to this gene. The less likely it is—the smaller the p value—the higher the score will be for the gene. Thus, even if GLAD4U is solely retrieving its data from the biomedical literature, it prioritizes following a statistical analysis of the retrieved data.

Nevertheless, even if our retrieval method performs well (high recall), there is room for improving our prioritization algorithm, as neither the k=50 nor the k=100 are as good as the recall. The improvement can come by reprioritizing GLAD4U-generated gene lists by using additional information, such as gene-to-gene networks based on co-occurrence [7], functional similarity [97], expression [98], or protein-protein interactions [55]. These alternatives will be addressed in future works.

The most obvious usage of GLAD4U is to generate a gene list for an input concept, as demonstrated in this thesis. This ability can be extremely useful for the design of targeted high-throughput experiments. If one needs to create a custom array or select proteins for targeted quantitative proteomic analysis using the selected reaction monitoring (SRM) assay, one can use GLAD4U and review the ranked list of genes that should likely be included in the experimental design. Besides generating gene lists for individual concepts, GLAD4U is very flexible and allows production of gene lists related to multiple concepts, which cannot be produced by searching only GO or OMIM databases. For example, a query of "smoking AND cancer" can generate a gene list that could potentially help exploring gene-environment interactions in cancer. GLAD4U also holds the potential to assist in improvement of the functional annotation of genes. Although GO contains more than 17,000 terms [4, 53] and is regularly used in the bioinformatics field as a standard [4, 99], it is not complete [27, 100]. Through manual checking of the top genes returned by GLAD4U that were not part of the gold standard lists, we easily found evidence that these genes were indeed linked to the query, and probably should have been included in the gold standard.

CHAPTER VI

FUTURE WORK

GLAD4U "Similarity"

We will explore additional approaches to integrate new information in GLAD4U to increase its performance. Considering that we are retrieving most genes relevant to our queries (high recall), it is important to focus on implementing prioritization methods that will increase precision. For example, we can build a co-occurrence network based on the scientific literature, GO, OMIM, or protein-protein interactions. Co-occurrence networks have been extensively used to prioritize and organize literature-based information (see the Literature review section and [93] for review). With these networks, we can derive gene pair similarities and integrate this new information to the original scores to reprioritize the gene lists, as applied in Ma *et al.* [101].

GLAD4U web implementations

Even if NCBI files are updated daily, we do not feel that it is necessary to update GLAD4U as often. Rather, we believe that a monthly update is enough to reasonably capture the changes in the number of publications and the mapping of publications-genes. Currently, the web implementation of GLAD4U is updated manually. Upon the download of updated files from the NCBI, the install Perl script is run to update all the files at the core of GLAD4U. To completely automate GLAD4U, we will create a "cron" job—a Unix time-based job scheduler—that will fetch and download the necessary NCBI

files, and run the install script to update GLAD4U. This approach will require writing additional scripts that will take GLAD4U offline for the duration of the process, update the necessary files and return GLAD4U back online. In view of our current update process, the turnover should be 30 minutes at most and would be processed at night, when both the NCBI and GLAD4U usage will be at their lowest.

We will also work on creating an API for GLAD4U to allow other applications to directly query GLAD4U and use its gene prioritizations in further computations. Additionally, we will integrate other resources from within the GLAD4U results page. Currently, we link GLAD4U results to the functional enrichment analysis performed by GOTM. Soon the enrichment analysis will incorporate GO, the Kyoto Encyclopedia of Genes and Genomes (KEGG, [102]), and WikiPathways [103], allowing the users to extend GLAD4U results to additional resources and facilitate the interpretation of their results.

CHAPTER VI

CONCLUSIONS

GLAD4U is a freely available web-application for creating prioritized gene lists tailored to a user's query. It follows the same steps that the experimentalist would follow: gather literature, extract gene information and create a gene list. The simple interface of GLAD4U ensures easy usage and interpretation. Because GLAD4U relies on existing biomedical literature, it has an immediate credibility with experimentalists, who use this resource as a primary means for enhancing their knowledge and expertise. Although the gene list directly returned from a PubMed query is usually lengthy and noisy, the prioritization method implemented in GLAD4U successfully ranks truly relevant genes at the top of the list and effectively facilitates efficient browsing of that list.

APPENDIX

A. APPENDIX 1: SURVEY TEMPLATE ADAPTED FROM THE FREE-STORY INTERVIEWS

Table 7: Survey template adapted from the free-story interviews.

Questions	*Possible answers*
Part 1: Tell us about yourself and your job.	
1. What is your age (select one):	a. Under 25 b. 25 to 34 c. 35 to 44 d. 45 and above
2. Which is best describing your position (select one):	a. Laboratory technician (such as up to Research Assistant 2) b. Medical or graduate student c. Post-doctoral fellow or research staff (such as Research Assistant 3 or more) d. Junior faculty e. Clinician or senior faculty
3. Which is best describing your workplace (select one):	a. Academia b. Hospital c. Government d. Health or life science industry e. Other industry
4. In which department do you work:	OPEN-ENDED QUESTION (one-line text box)
5. When using an online search engine, how satisfied are you with the speed at which you are getting the results (select one):	a. Satisfied b. Somewhat satisfied c. Somewhat unsatisfied d. Unsatisfied
6. When using an online search engine, how satisfied are you with the relevance of the results which you are getting (select one):	a. Satisfied b. Somewhat satisfied c. Somewhat unsatisfied d. Unsatisfied
Part 2: Tell us about the methods and type of data that you used in your work.	
7. Which of the following online science-oriented search engines have you used (check all that apply):	a. Entrez Pubmed b. Google Scholar c. Ensembl d. BioMart e. Other online search engines f. None of the above
8. If you used any of the online search engines mentioned in question 7 or others, how often do you use them on average a week (select one):	a. Once a week or less b. Up to five times a week c. Five to ten times a week d. More than ten times a week e. Never

Table 7, continued.

Questions	*Possible answers*
9. How much time a week do you approximately spend searching and/or reading scientific papers (abstracts or full publications) (select one):	a. None, I don't search publications online b. Less than 1hr to 5hrs c. 5hrs to 10hrs d. More than 10hrs
10. High-Throughput Screening uses robotics, data processing and control software, liquid handling devices, and sensitive detectors to quickly conduct millions of biochemical, genetic or pharmacological tests. Have you dealt with high-throughput screening data (select one):	a. Yes b. No c. I don't know
Part 3: Few questions about the method underlying GLAD4U.	
11. Do you think that any tool should provide instructions on how to evaluate the results by yourself (select one):	a. Yes b. No c. Don't have an opinion
12. Which of the following filters would you like to be able to use to constraint your gene results (check all that apply):	a. Species b. Methods used in the papers c. Diseases d. Pathways e. None of the above
13. Considering the prioritization method used to produce the rank of the hits, would you say that (select one):	a. No interest in the prioritization method, complete trust in the tool results b. Prefer a link to be provided to the details of the prioritization method c. Prefer the description of the prioritization method on the query and/or result pages d. Prefer to be able to choose the prioritization method to use
14. Do you think that supporting publications should incorporate the journal impact factors (select one):	a. Yes b. No c. Don't have an opinion
15. It is important to know if supporting publications are reviews or research papers (select one):	a. Agree b. Disagree c. Don't have an opinion
Part 4: Few questions about the results provided by GLAD4U.	
16. If a result page contains 50 hits, how many pages do you typically look at (select one):	a. All pages of results b. The first 10-15 pages of results (500-750 hits) c. The first 5 pages of results (250 hits) d. More than 1 but less than 5 pages of results (50-250 hits) e. Only the first page of results (50 hits)
17. When searching online, how important is it for you to be able to customize the result page (customization is defined as the ability "to build, fit, or alter according to individual specifications") (select one):	a. Very important b. Important c. Somewhat important d. Not important
18. Which of the following options would you like to be able to use to modify the look of the results page (check all that apply):	a. Number of genes per page b. Number of supporting publications per gene c. Number of pages created at once d. None of the above

Table 7, continued.

Questions	*Possible answers*
19. Considering the large number of online scientific information searching tools and their large spectrum of type of information retrieved, do you think that any online tool needs to provide links to other complimentary online tools (select one):	a. Yes b. No c. Don't have an opinion
20. Because GLAD4U relies on Pubmed to retrieve supporting publications, do you think that the tool needs to link back to Pubmed more than any other sources (select one):	a. Yes b. No c. Don't have an opinion
21. Do you think that GLAD4U should expand the search space by offering links to more general information such as online search engine (Google, Yahoo, etc.) (select one):	a. Yes b. No c. Don't have an opinion
Part 5: Few questions about the usability ("term used to denote the ease with which people can employ a particular tool") of GLAD4U.	
22. In general, do you prefer links to open in new windows (select one):	a. Yes b. No c. Don't have an opinion
23. Considering information gathered from the result page, do you prefer the screen to be (select one):	a. Compact (all information fits on one screen) b. Semi-compact (more than one screen, information solely from the tool) c. Semi-complete (more than one screen, links towards extra information) d. Complete (more than one screen, external information integrated in additional boxes)
24. Assuming that a tool provides optimal results, would you use it if it were not free:	a. Yes b. No c. Don't have an opinion
25. Assuming that a tool provides optimal results, would you use it if it were free but required users to register before use:	a. Yes b. No c. Don't have an opinion
26. Do you feel more confident in using tools that are backed up by renown Academic Research Facilities or Corporations (select one):	a. Yes b. No c. Don't have an opinion

B. APPENDIX 2: GOLD STANDARDS

GO-derived gold standards

Apoptosis

Gene Symbol|Entrez-Gene ID - GRIK2|2898, NPM1|4869, SELS|55829, DAPL1|92196, CARD10|29775, CIAPIN1|57019, TXNDC5|81567, CDKN2C|1031, MAEA|10296, DNM2|1785, PPP2R2B|5521, TIAM1|7074, POU3F3|5455, BAK1|578, USP17L5|728386, NRAS|4893, ZBTB16|7704, DCC|1630, BTK|695, TNFSF10|8743, SHISA5|51246, DIDO1|11083, SOX2|6657, ABR|29, APOH|350, EPHA2|1969, RPS27A|6233, MYO18A|399687, NOD2|64127, HDAC1|3065, GSK3B|2932, BBC3|27113, ERCC5|2073, RPS3|6188, HSPB1|3315, SMNDC1|10285, PDCD4|27250, LALBA|3906, TRIO|7204, CD3G|917, ROBO1|6091, AIFM1|9131, GIMAP1|170575, PDCD7|10081, NAIP|4671, APH1B|83464, TSPO|706, KCNH8|131096, MAPK7|5598, USP17L1P|401447, CASP1|834, GZMH|2999, IL7|3574, DPF1|8193, SHARPIN|81858, HIP1|3092, SFRP5|6425, BCLAF1|9774, SRA1|10011, RBM5|10181, FNTA|2339, DEDD|9191, EPHA7|2045, TMEM173|340061, GAL|51083, EP300|2033, SPATA3|130560, MNT|4335, NQO1|1728, INPP5D|3635, TNFRSF8|943, FKBP8|23770, NOX5|79400, DCUN1D3|123879, GDF5|8200, MPO|4353, EIF2AK2|5610, USP17L3|645836, CDH1|999, PEG10|23089, RELA|5970, PDCL3|79031, LYST|1130, MX1|4599, RIPK2|8767, HTATIP2|10553, ACTN1|87, MEF2D|4209, CRTAM|56253, MAGI3|260425, TWIST2|117581, PPP2CA|5515, IKBKG|8517, NME1|4830, ALX4|60529, WFS1|7466, UNC5A|90249, TUBB|203068, BAG1|573, HSP90B1|7184, CKAP2|26586, TRAF2|7186, BIK|638, SST|6750, NLRP12|91662, IP6K2|51447, SGPL1|8879, HBXIP|10542, PLEKHG5|57449, PDCD10|11235, NAIF1|203245, CFDP1|10428, GHR|2690, AIPL1|23746, BAX|581, ANXA5|308, PLG|5340, RAG1|5896, AIMP2|7965, WDR92|116143, CACNA1A|773, PIM3|415116, ANXA1|301, TNFRSF1B|7133, SLC5A8|160728, BLID|414899, AEN|64782, STAT5A|6776, OSR1|130497, IL17A|3605, DEDD2|162989, SPN|6693, DFFB|1677, CDKN1A|1026, IL1A|3552, GZMB|3002, PRNP|5621, CTNNB1|1499, MITF|4286, TGM2|7052, NKX2-6|137814, RYBP|23429, IL3|3562, KCNIP3|30818, COL4A3|1285, APH1A|51107, ING4|51147, NTRK1|4914, HSPA5|3309, POU4F1|5457, EAF2|55840, VEGFA|7422, ITM2B|9445, TNS4|84951, ERN1|2081, BCL7C|9274, ADAM9|8754, PPP3CC|5533, KLF11|8462, IL2RA|3559, CLN3|1201, GFRAL|389400, BOK|666, ZDHHC16|84287, FOXO1|2308, DNASE1L3|1776, NCKAP1|10787, ACIN1|22985, GNRH1|2796, APLP1|333, APBB2|323, XPA|7507, PIM1|5292, NTN1|9423, PF4|5196, RIPK3|11035, STRADB|55437, OBSCN|84033, SCRIB|23513, BCL2L13|23786, CIDEC|63924, TCF7L2|6934, CDK5R1|8851, CIDEA|1149, USP17L2|377630, AIMP1|9255, SEMA3A|10371, ESR1|2099,

RAD21|5885, SART1|9092, TRAF7|84231, DOCK1|1793, TP53|7157, RRAGC|64121, RIPK1|8737, P2RX7|5027, AZU1|566, UCN|7349, SEMA6A|57556, BCAP31|10134, IER3|8870, VAV1|7409, GML|2765, BCL6|604, BCAP29|55973, TDGF1|6997, IGFBP3|3486, PPIF|10105, IFNG|3458, POLR2G|5436, AKT1|207, BIRC3|330, PPT1|5538, FGF4|2249, VDAC1|7416, BNIP2|663, PTGS2|5743, FASTK|10922, HGF|3082, ARHGAP4|393, NCSTN|23385, BCL2L12|83596, PAK7|57144, CTSB|1508, TNFRSF25|8718, CDK11B|984, CUL5|8065, TMEM161A|54929, SGK1|6446, DUSP1|1843, P2RX4|5025, RHOA|387, PRKDC|5591, ARHGEF12|23365, API5|8539, BARHL1|56751, SPIN2B|474343, MGMT|4255, ALX3|257, BCL10|8915, PDCD6|10016, HSPA9|3313, CXCR4|7852, SCG2|7857, TBX3|6926, E2F1|1869, CSRNP1|64651, EPO|2056, ASAH2|56624, HIPK3|10114, GPX1|2876, ARHGEF9|23229, IFI16|3428, ARHGEF4|50649, TCTN3|26123, P2RX1|5023, PDCD6IP|10015, CUL4A|8451, CCL2|6347, CSRNP2|81566, KNG1|3827, CUL1|8454, CAMK1D|57118, TEX261|113419, 40425|5414, PRKCA|5578, LGALS7|3963, BCL2L14|79370, IL12A|3592, RPS6|6194, PIK3CA|5290, IGF1|3479, SPATA4|132851, ACTN3|89, SMPD2|6610, PCSK9|255738, KLF10|7071, ACVR1C|130399, CLCF1|23529, IL2|3558, BNIP3L|665, KIAA1967|57805, KALRN|8997, SON|6651, CSTB|1476, BNIPL|149428, DDAH2|23564, SH3RF1|57630, LUC7L3|51747, TNFRSF10D|8793, ZMAT3|64393, CASP12|120329, FGD1|2245, DNAJC5|80331, CARD9|64170, VAV2|7410, DHCR24|1718, GADD45G|10912, SULF1|23213, B4GALT1|2683, RNF34|80196, CYCS|54205, CD27|939, TNFRSF19|55504, PSEN1|5663, PDCD5|9141, GAS1|2619, THBS1|7057, RTKN|6242, CALR|811, MSH2|4436, DAPK1|1612, BCL2L11|10018, RARB|5915, EIF5A|1984, PRODH|5625, CARD18|59082, PIGT|51604, ANXA4|307, ADRB2|154, FXR1|8087, ANGPTL4|51129, XRCC5|7520, CAT|847, SOX7|83595, RRM2B|50484, CECR2|27443, CEBPG|1054, IFIH1|64135, CDKN2A|1029, TRAF6|7189, BIRC6|57448, TM2D1|83941, RAF1|5894, BLCAP|10904, IL19|29949, BCL2L1|598, NTF3|4908, TIAM2|26230, BLOC1S2|282991, ASNS|440, SIAH1|6477, ALDOC|230, TXNIP|10628, SHF|90525, RXFP2|122042, FAM82A2|55177, TP53BP2|7159, BCL3|602, CD5L|922, ARHGEF17|9828, CASP8|841, COMP|1311, CASP8AP2|9994, CFL1|1072, ACVR1B|91, IFNB1|3456, PREX1|57580, BUB1B|701, ATOH1|474, UBQLN1|29979, INS|3630, SNCA|6622, MSX2|4488, HELLS|3070, EDAR|10913, NGFRAP1|27018, LHX4|89884, ERCC6|2074, TRAF3|7187, XRCC4|7518, EYA2|2139, NDUFS1|4719, BRAF|673, GRAMD4|23151, PAX7|5081, OPA1|4976, TPD52L1|7164, PHLDA2|7262, ABL1|25, DRAM1|55332, FEM1B|10116, ERBB3|2065, CITED2|10370, SMAD6|4091, MSH6|2956, VAV3|10451, SOS1|6654, LCK|3932, ALOX12|239, INHA|3623, NGF|4803, SH3GLB1|51100, FADD|8772, TERT|7015, ATM|472, CAPN10|11132, PIM2|11040, PAX3|5077, APAF1|317, ESPL1|9700, DDIT3|1649, DDX19B|11269, ERCC3|2071, F3|2152, SIVA1|10572, ERN2|10595, COL2A1|1280, NME2|4831, TNFRSF4|7293, YARS|8565, MAP3K10|4294, OSM|5008, DNASE1|1773, RPS27L|51065, RNF144B|255488, VHL|7428, RASSF5|83593, ELMO3|79767, GSN|2934, CD24|100133941, ZNF443|10224, MADD|8567, FASTKD5|60493, PTK2B|2185, CLU|1191, CHEK2|11200, ZC3H8|84524, MTCH1|23787, ALOX15B|247, C6|729, ADAMTSL4|54507, AGTR2|186, SERPINB2|5055, LY86|9450, STK3|6788, PRKCI|5584, EI24|9538, PRF1|5551, ADIPOQ|9370, C9|735, GREM1|26585, NME6|10201, VNN1|8876, LIG4|3981,

ING3|54556, SLTM|79811, PRDX2|7001, DAPK3|1613, HERPUD1|9709, PLAGL1|5325, PPP1R15A|23645, SIAH2|6478, YWHAZ|7534, RARG|5916, MUL1|79594, BAK1P1|600, SOD2|6648, COL18A1|80781, PHLDA1|22822, SCIN|85477, NACC1|112939, ECE1|1889, EBAG9|9166, PRAME|23532, PTCRA|171558, PTPN6|5777, SNCB|6620, GLO1|2739, CADM1|23705, CASP4|837, SOCS3|9021, TIAL1|7073, PHF17|79960, ARHGEF6|9459, RNF130|55819, MAL|4118, SPHK2|56848, MAPK1|5594, GCLM|2730, ARHGEF11|9826, ARHGEF7|8874, SERPINB9|5272, TNFSF13B|10673, TAX1BP1|8887, ENDOG|2021, TGFBR1|7046, PAWR|5074, TOP2A|7153, FASTKD3|79072, TMEM85|51234, CARD8|22900, BMP7|655, ARHGEF2|9181, HSPD1|3329, TGFB1|7040, CARD6|84674, FGF2|2247, RNF7|9616, RFFL|117584, RUNX3|864, CDK11A|728642, CASP7|840, GADD45A|1647, TRIM69|140691, BCL2A1|597, ELMO1|9844, GRID2|2895, IHH|3549, UACA|55075, TERF1|7013, AHR|196, HRAS|3265, NGEF|25791, NELL1|4745, NLRP1|22861, PDE1B|5153, CD40LG|959, FGD2|221472, PPP2CB|5516, XIAP|331, CARD16|114769, STAT5B|6777, CHAC1|79094, BARD1|580, ATF5|22809, CDK5|1020, DAPK2|23604, TP73|7161, BAG3|9531, NEK6|10783, PMAIP1|5366, BCL2L15|440603, ALDH1A3|220, AVEN|57099, PGAP2|27315, AATF|26574, SMPD1|6609, MCL1|4170, TNFRSF10A|8797, TOPORS|10210, HSPE1|3336, BAD|572, BRCA2|675, DNM1L|10059, AXIN1|8312, PRDX5|25824, MAP3K11|4296, NDUFS3|4722, FIS1|51024, TNFRSF11B|4982, ARF6|382, E2F2|1870, SERINC3|10955, MEF2C|4208, HMGB1|3146, SLK|9748, SOX9|6662, TRADD|8717, GJB6|10804, JUN|3725, SYCP2|10388, BNIP1|662, SQSTM1|8878, MAPK9|5601, ADORA2A|135, APOE|348, INHBA|3624, PEA15|8682, BDNF|627, RB1CC1|9821, RYR2|6262, ILK|3611, RHOB|388, FOXC2|2303, PUF60|22827, STK17A|9263, PTPRF|5792, HTT|3064, SEMA4D|10507, PPP1R13L|10848, PIK3CG|5294, PTH|5741, MAPK8|5599, TBX5|6910, BTC|685, TUBB2C|10383, CNTF|1270, SOCS2|8835, EIF2AK3|9451, APC|324, BID|637, HMOX1|3162, APP|351, HSPA1A|3303, TNFSF9|8744, MCF2|4168, SSTR3|6753, CDH13|1012, CASP2|835, CDKN2D|1032, NAE1|8883, PSENEN|55851, PERP|64065, C1D|10438, STAMBP|10617, NKX3-2|579, ECT2|1894, LGALS1|3956, GIMAP5|55340, IL10|3586, TRIM39|56658, HIPK2|28996, PRKCZ|5590, PROC|5624, CASP14|23581, TICAM1|148022, TNFSF15|9966, TAF9|6880, GRK1|6011, MAEL|84944, TAOK2|9344, SGMS1|259230, IL4|3565, PDIA3|2923, JMJD6|23210, AKT1S1|84335, TGFB2|7042, ETS1|2113, GSTP1|2950, ADA|100, CD2|914, IRAK1|3654, MCF2L|23263, NEFL|4747, PLEKHF1|79156, TMBIM6|7009, TMX1|81542, MAP3K1|4214, TNFAIP3|7128, NEUROD1|4760, IL6R|3570, FOXC1|2296, PTPRC|5788, BTG1|694, ID3|3399, TRIB3|57761, NOD1|10392, ALB|213, PML|5371, GDNF|2668, PSME3|10197, FAF1|11124, TPT1|7178, NMNAT3|349565, THOC1|9984, TIAF1|9220, CD3E|916, RXRA|6256, HDAC6|10013, BIRC2|329, FAM188A|80013, CARD17|440068, DBH|1621, DLG5|9231, TNFRSF14|8764, XRCC2|7516, TNFRSF18|8784, TLR4|7099, CCAR1|55749, F2|2147, FAIM3|9214, BIRC8|112401, DNAJB13|374407, BRCA1|672, PRKRA|8575, CDK1|983, VCP|7415, BFAR|51283, ITGB2|3689, PAFAH2|5051, TNFRSF1A|7132, ALMS1|7840, PLEKHG2|64857, BMF|90427, DFFA|1676, PHB|5245, LTBR|4055, TNFRSF6B|8771, C3ORF38|285237, IGF1R|3480, PAK2|5062, BNIP3|664, EDNRB|1910, AMIGO2|347902, PAK1|5058, UNC5C|8633, CRADD|8738, CARD14|79092, HDAC3|8841, MALT1|10892, KRAS|3845, UBE4B|10277,

ESR2|2100, G2E3|55632, ITGB3BP|23421, CLN8|2055, IFI6|2537, SFN|2810, DYNLL1|8655, ERBB2|2064, DUSP22|56940, IL1B|3553, NDUFA13|51079, RPL11|6135, AKTIP|64400, GHRL|51738, TNFAIP8|25816, DNAJB6|10049, IKBKB|3551, FGD3|89846, PDCD2|5134, EGLN3|112399, SCN2A|6326, CASP10|843, RNF216|54476, SIRT1|23411, SOS2|6655, FAM176A|84141, DDX41|51428, BTG2|7832, GCM2|9247, FASTKD2|22868, AKAP13|11214, GSDMA|284110, IL12B|3593, SMAD3|4088, TRAF1|7185, LGALS12|85329, CRYAA|1409, AIFM3|150209, EEF1E1|9521, ZC3HC1|51530, SLC25A6|293, MMP9|4318, CARD11|84433, CLPTM1L|81037, DAP|1611, APBB1|322, MAP1S|55201, JAG2|3714, ZNF346|23567, GLRX2|51022, CD44|960, FAIM|55179, PROP1|5626, TNF|7124, ADAMTS20|80070, PPM1F|9647, TNFRSF10B|8795, LITAF|9516, NUAK2|81788, DNASE2|1777, TRIM35|23087, CSF2|1437, HIPK1|204851, BCAR1|9564, RHOT2|89941, MAPK8IP1|9479, RPS3A|6189, PPARD|5467, BIRC5|332, PHLDA3|23612, POLB|5423, RAD9A|5883, ACVR1|90, CRYAB|1410, TNFRSF21|27242, CSE1L|1434, MDM4|4194, DLC1|10395, YWHAE|7531, CD5|921, IDO1|3620, ELMO2|63916, IL31RA|133396, CASP9|842, DEPDC6|64798, DDX20|11218, SAP30BP|29115, TP53I3|9540, SGPP1|81537, SKIL|6498, GGCT|79017, NUPR1|26471, RASGRF2|5924, ASCL1|429, UNC5B|219699, LTA|4049, TRAF5|7188, RTN3|10313, IFT57|55081, TP53INP1|94241, CRH|1392, NOTCH2|4853, TNFRSF10C|8794, BAG2|9532, TNFSF8|944, JMY|133746, CFLAR|8837, PRKCE|5581, DAD1|1603, TP63|8626, CD70|970, TNFRSF12A|51330, PTPRH|5794, CYFIP2|26999, RASA1|5921, FAS|355, TLR2|7097, PGP|283871, GCH1|2643, PPP3R1|5534, GSPT1|2935, NR4A2|4929, MAP3K5|4217, PROK2|60675, KRT18|3875, PLA2G4A|5321, ATP7A|538, PPP1R13B|23368, GPR65|8477, NME5|8382, MAGED1|9500, RAC1|5879, AIFM2|84883, PDE3A|5139, MOAP1|64112, CHD8|57680, IL24|11009, CDKN1B|1027, MRPL41|64975, MLL|4297, UNC5D|137970, FKSG2|59347, SLC5A11|115584, CUL3|8452, SLC11A2|4891, PEG3|5178, NLRC4|58484, VDR|7421, DNAJA3|9093, TRAIP|10293, FOXL2|668, STAT1|6772, NRG1|3084, ATG5|9474, ARHGEF3|50650, RTN4|57142, MLH1|4292, NOL3|8996, FOXO3|2309, CASP5|838, GLI3|2737, PTEN|5728, NISCH|11188, TRIAP1|51499, IAPP|3375, SGK3|23678, SFRP1|6422, ERCC2|2068, ITGA1|3672, GRIN2A|2903, CSDA|8531, CREB1|1385, MAP3K7|6885, C8ORF4|56892, PACS2|23241, NOTCH1|4851, RASGRF1|5923, IFNA2|3440, NET1|10276, KCNMA1|3778, TBRG4|9238, PECR|55825, RHOT1|55288, PRDX1|5052, XAF1|54739, CEBPB|1051, AGT|183, PRDX3|10935, WWOX|51741, IL6|3569, PCGF2|7703, GRM4|2914, MYD88|4615, TNFSF14|8740, F7|2155, MTP18|51537, RABEP1|9135, STK4|6789, CASP3|836, NF1|4763, SLAMF7|57823, UBE2Z|65264, BECN1|8678, BCL2L2|599, SH3KBP1|30011, MBD4|8930, MRPS30|10884, TNFSF12|8742, NFKB1|4790, PSEN2|5664, HRK|8739, ARNT2|9915, ADRA1A|148, ITSN1|6453, DYRK2|8445, SOD1|6647, GZMA|3001, SOX4|6659, STK17B|9262, UTP11L|51118, TAF9B|51616, GATA6|2627, FASTKD1|79675, NLRP2|55655, ACTN4|81, ROCK1|6093, FOSL1|8061, TRAF4|9618, BCL2L10|10017, BDKRB2|624, HOXA13|3209, BCL11B|64919, SHB|6461, CHST11|50515, DPF2|5977, C16ORF5|29965, CIB1|10519, TIA1|7072, MEF2A|4205, PHLPP1|23239, NKX2-5|1482, CCK|885, DDIT4|54541, PDCD1|5133, POU4F3|5459, PYCARD|29108, TGFB3|7043, DLX1|1745, CSRNP3|80034, BCL2|596, AATK|9625, NOS3|4846,

IGF2R|3482, ARHGEF18|23370, KRT20|54474, NMNAT1|64802, CGB|1082, DIABLO|56616, PSMG2|56984, CD38|952, WRN|7486, ATP2A1|487, NGFR|4804, EGFR|1956, CIDEB|27141, MNAT1|4331, TNFRSF9|3604, NFKBIA|4792, RRAGA|10670, FGD4|121512, PTRH2|51651, NUP62|23636, TP53AIP1|63970, LRDD|55367, NLRP3|114548, PLAGL2|5326, HAND2|9464, ARHGEF16|27237, TMEM102|284114, BRE|9577, BIRC7|79444, UNC13B|10497, PPP2R1A|5518, SMO|6608, STEAP3|55240, TWIST1|7291, GULP1|51454, APIP|51074, F2R|2149, HTRA2|27429, BAG5|9529, BAG4|9530, CD14|929, NME3|4832, ZC3H12A|80149, DAP3|7818, DAXX|1616, ADNP|23394, GADD45B|4616, ARHGDIA|396, GAS2|2620, CBX4|8535, TNFSF18|8995, MIF|4282, FAIM2|23017, MAGEH1|28986, SORT1|6272, ACTC1|70, CASP6|839, EYA1|2138, SRGN|5552, SPHK1|8877, ACTN2|88, MFSD10|10227, CTNNBL1|56259, ADORA1|134, EEF1A2|1917, PRLR|5618, JAK2|3717, ADAM17|6868, MSX1|4487, GCLC|2729, NR4A1|3164, IL2RB|3560, MEN1|4221, TRAF3IP2|10758, NUDT2|318, CD74|972, TEX11|56159, PDIA2|64714, DHRS2|10202, MAP2K6|5608, CUL2|8453, SHH|6469, FASLG|356, YWHAB|7529.

Cell adhesion

Gene Symbol|Entrez-Gene ID - TEK|7010, CRNN|49860, ADAMDEC1|27299, DSG4|147409, TSC2|7249, CNTNAP3|79937, ITGA8|8516, CD40LG|959, RET|5979, GP1BB|2812, IL18|3606, STAT5B|6777, PCDHA7|56141, BCAN|63827, PRSS2|5645, PCDH7|5099, EPDR1|54749, CDK5|1020, ZYX|7791, PDZD2|23037, VWC2|375567, MUC4|4585, MAEA|10296, PPFIA2|8499, PVRL1|5818, CADM3|57863, CLDN4|1364, CDK6|1021, CEACAM1|634, BOC|91653, CNTNAP3B|728577, CNTNAP4|85445, AATF|26574, PTPRM|5797, COL8A2|1296, MOG|4340, FEZ1|9638, LOXL2|4017, CDON|50937, CD36|948, COL12A1|1303, STXBP3|6814, PARVG|64098, TPM1|7168, ARF6|382, APOA4|337, LPP|4026, CD93|22918, CLDN15|24146, TRIP6|7205, SELP|6403, PCDHB8|56128, ITGAM|3684, ITGA6|3655, SOX9|6662, ROBO1|6091, CHRD|8646, BTBD9|114781, ITGB5|3693, SCARF1|8578, COL6A3|1293, ALX1|8092, SIPA1|6494, FLRT3|23767, CD22|933, CCL4|6351, C22ORF28|51493, LAMB2|3913, CNTN3|5067, PDPN|10630, NCAM1|4684, STXBP1|6812, JAM2|58494, PCDHA2|56146, AGGF1|55109, COL17A1|1308, PCDHB15|56121, CDHR5|53841, RAPH1|65059, GTPBP4|23560, COL29A1|256076, ILK|3611, TGFBI|7045, HABP2|3026, ONECUT2|9480, RHOB|388, PTPRF|5792, SCARB2|950, SEMA4D|10507, MMRN1|22915, EMILIN2|84034, CDH17|1015, DEFB118|117285, VWF|7450, CD97|976, CDH2|1000, EGFL6|25975, ACHE|43, ECM2|1842, SIGLEC8|27181, PKD1|5310, DDR2|4921, ASTN1|460, TESK2|10420, CD99L2|83692, MYH9|4627, PKP2|5318, APC|324, EFS|10278, LY6D|8581, NRXN1|9378, APP|351, ESAM|90952, CD96|10225, MUC16|94025, CD164|8763, CDH1|999, GP1BA|2811, ATP2A2|488, CCR3|1232, AMIGO3|386724, CDH13|1012, MMP14|4323, GPR98|84059, ITGA11|22801, ITGA10|8515, PERP|64065, PCDHB10|56126, SDK2|54549, COL6A2|1292, ACTN1|87, MAG|4099, COL21A1|81578, EGFLAM|133584, NID2|22795, PPP2CA|5515, NRXN3|9369, ENTPD1|953, LAMB1|3912, PRKCZ|5590, MUC5B|727897, HAS1|3036, ISLR|3671, LAMA5|3911, COL20A1|57642, AMTN|401138, CLDN1|9076, CDH4|1002, TAOK2|9344,

ROBO2|6092, CTNNA3|29119, PCDH11X|27328, CPXM1|56265, CLDN22|53842, ATP5B|506, TGFB2|7042, F5|2153, NRCAM|4897, COL27A1|85301, ADA|100, ICAM1|3383, CD2|914, NEO1|4756, LPXN|9404, CD58|965, PCDHGC5|56097, ITGB8|3696, NLGN2|57555, PVRL3|25945, CFDP1|10428, PCDHGB1|56104, MYBPC1|4604, ADAM15|8751, HAPLN4|404037, CLDN3|1365, PCDHA3|56145, PDPK1|5170, TDGF3|6998, PCDHB16|57717, PVRL4|81607, NID1|4811, ROR2|4920, CTNNAL1|8727, ITGA4|3676, PCDHB2|56133, PDE3B|5140, DGCR6|8214, PCDHGB2|56103, CNTNAP1|8506, PTPRC|5788, PCDHB7|56129, IGSF11|152404, ACVRL1|94, DMP1|1758, STAT5A|6776, AMELX|265, ITGAV|3685, TRO|7216, TPBG|7162, FAF1|11124, DCBLD1|285761, SPN|6693, BCAM|4059, MADCAM1|8174, MFGE8|4240, CALCA|796, STAB1|23166, F11R|50848, CCL5|6352, ABL2|27, DLG5|9231, LAMA3|3909, DLL1|28514, CTNNB1|1499, NPNT|255743, SPON2|10417, SIGLEC12|89858, ITGB3|3690, COL19A1|1310, TGM2|7052, CHL1|10752, TLN2|83660, PCDHGA2|56113, NPHS1|4868, MTSS1|9788, PTPRK|5796, DSC3|1825, ITGAL|3683, ITGB7|3695, PCDHA1|56147, CELSR2|1952, COL4A3|1285, SELE|6401, EZR|7430, LAMA2|3908, CLDN6|9074, SEMA5A|9037, CXADR|1525, ITGB2|3689, ITGAE|3682, PCDHB9|56127, JUB|84962, DPP4|1803, ADAM9|8754, CTNND1|1500, MAGI1|9223, SIGLEC1|6614, TYRO3|7301, TRPM7|54822, AOC3|8639, AMICA1|120425, SMAD7|4092, COL8A1|1295, SORBS3|10174, BARX2|8538, CNTNAP2|26047, THBS3|7059, PPFIA1|8500, SSX2IP|117178, MYBPH|4608, APLP1|333, HEPACAM|220296, GP5|2814, ICAM5|7087, TMEM8A|58986, SIGLEC5|8778, GMDS|2762, HAPLN2|60484, AMIGO2|347902, CYTH1|9267, CLSTN3|9746, TNFAIP6|7130, PCDH1|5097, LAMC1|3915, RGMB|285704, CDH19|28513, WISP1|8840, SCRIB|23513, SDK1|221935, ICAM2|3384, CDH8|1006, CDK5R1|8851, NEGR1|257194, SERPINI2|5276, PCDH11Y|83259, ITGB3BP|23421, PCDHGA11|56105, PVRL2|5819, COL13A1|1305, CSF1|1435, ITGA7|3679, CD47|961, CLDN20|49861, ACAN|176, CASS4|57091, AIMP1|9255, CD34|947, COL4A6|1288, DSC1|1823, TESC|54997, CD4|920, NF2|4771, MGP|4256, CNTNAP5|129684, MAP2K1|5604, CD151|977, ERBB2|2064, AZU1|566, RADIL|55698, S1PR1|1901, IL1B|3553, TNC|3371, CTNNA2|1496, PCDHB12|56124, VAV1|7409, ENG|2022, IZUMO1|284359, EMILIN1|11117, SRPX|8406, BCL6|604, CLDN17|26285, CDH3|1001, LYVE1|10894, CORO1A|11151, CNTN2|6900, PCDHB3|56132, ITGB6|3694, SIRPG|55423, CDH16|1014, TDGF1|6997, PCDHB18|54660, KIRREL2|84063, FGF4|2249, BGLAP|632, ARHGAP5|394, REG3A|5068, MIA|8190, IL12B|3593, PCDHGB7|56099, RPSA|3921, FGF1|2246, MLLT4|4301, SMAD3|4088, FERMT1|55612, OLFM4|10562, CLDN2|9075, PTK7|5754, CLCA2|9635, ITGA2|3673, PTPRS|5802, KAL1|3730, NFASC|23114, ANXA9|8416, SIGLEC14|100049587, PSTPIP1|9051, CTNND2|1501, TGFB1I1|7041, CNTN1|1272, BMPR1B|658, PCDH18|54510, CDH22|64405, IL32|9235, FXC1|26515, GPR56|9289, F8|2157, ANGPTL3|27329, GNE|10020, MSLN|10232, SELL|6402, JAG2|3714, BAI1|575, LAMA1|284217, CXCR3|2833, SIGLEC6|946, SIRPA|140885, NRP2|8828, LAMB3|3914, CDH7|1005, SIGLEC11|114132, COL5A3|50509, TNN|63923, CD44|960, SYK|6850, HSD17B12|51144, TNF|7124, HES1|3280, CNTN5|53942, FBLN5|10516, RHOA|387, RAB13|5872, COL11A2|1302, PCDHA8|56140, CD209|30835, OMG|4974, FERMT2|10979, ITGA5|3678, COL7A1|1294, PCDHAC1|56135, ADAM12|8038,

FNDC3A|22862, GPNMB|10457, BCAR1|9564, PIK3CB|5291, CNTN6|27255, PPARD|5467, CDH24|64403, THY1|7070, FAT4|79633, DSC2|1824, DAB1|1600, PKD2|5311, DSCAML1|57453, ITGB4|3691, PCDHA12|56137, HSPG2|3339, ADAM10|102, CCL2|6347, CX3CL1|6376, DLC1|10395, KNG1|3827, RS1|6247, CDH6|1004, IGFALS|3483, MEGF10|84466, LGALS7|3963, HAPLN1|1404, MYF5|4617, IL12A|3592, CDH15|1013, AMBP|259, CTGF|1490, CD2AP|23607, PCDHA9|9752, ARHGAP6|395, ACTN3|89, CDHR4|389118, SGCE|8910, CLDN23|137075, IL2|3558, PARVA|55742, LGALS4|3960, PCDH8|5100, COL15A1|1306, ITGB1|3688, BMP1|649, KITLG|4254, SERPINI1|5274, TMEM8B|51754, NRP1|8829, MYBPC2|4606, CADM4|199731, NINJ1|4814, INPPL1|3636, OLR1|4973, JUP|3728, CELSR1|9620, PCDH15|65217, B4GALT1|2683, CDHR2|54825, SSPN|8082, SLURP1|57152, APBA1|320, PKP3|11187, PSEN1|5663, COL3A1|1281, CD226|10666, THBS1|7057, SIGLEC7|27036, TECTA|7007, CNTN4|152330, ITGA3|3675, LYPD3|27076, CYR61|3491, SCARF2|91179, CPXM2|119587, TNFRSF12A|51330, BCL2L11|10018, ITGA9|3680, EDIL3|10085, CDH23|64072, EMB|133418, PGM5|5239, DCHS1|8642, ICAM3|3385, PCDHA6|56142, CCR8|1237, NCAN|1463, CD72|971, CYFIP2|26999, ANTXR1|84168, RASA1|5921, PCDHGA10|56106, CUZD1|50624, PARVB|29780, CHAD|1101, PKD1L1|168507, PGP|283871, VCAN|1462, THBS4|7060, CDH5|1003, COL24A1|255631, PCDH19|57526, PKP1|5317, SVEP1|79987, MSN|4478, CDKN2A|1029, CYTIP|9595, CLDN5|7122, PCDHAC2|56134, THRA|7067, CERCAM|51148, PCDH17|27253, RAC1|5879, COL14A1|7373, DDR1|780, CLDN8|9073, IGFBP7|3490, CSF3R|1441, COL6A6|131873, PLEK|5341, NELL2|4753, PCDHB5|26167, CDH10|1008, COL5A1|1289, CD300A|11314, COMP|1311, FLOT2|2319, PCDHGB6|56100, DSCAM|1826, TTYH1|57348, COL1A1|1277, FGF6|2251, CLSTN1|22883, GP9|2815, NLGN4X|57502, PNN|5411, PCDHA13|56136, ALCAM|214, DCBLD2|131566, PCDHGA3|56112, AZGP1|563, C4ORF31|79625, ITGAD|3681, NRG1|3084, PCDHGC4|56098, PCDHB1|29930, PCDH9|5101, ADAM23|8745, PKP4|8502, HAPLN3|145864, PCDHGC3|5098, CBLL1|79872, EFNB1|1947, PRPH2|5961, LGALS3BP|3959, RND1|27289, WISP2|8839, ABL1|25, CLEC4A|50856, PTEN|5728, CD9|928, FREM2|341640, ERBB3|2065, NLGN3|54413, CITED2|10370, ITGA1|3672, CLDN18|51208, NTM|50863, CD84|8832, VAV3|10451, PXN|5829, CASK|8573, TNR|7143, ALOX12|239, NINJ2|4815, THBS2|7058, MIA3|375056, PKHD1|5314, PCDHGB4|8641, SUSD5|26032, FREM1|158326, EMID2|136227, PCDHB11|56125, CCDC80|151887, DGCR2|9993, AGT|183, CLDN12|9069, PCDHGA9|56107, ICAM4|3386, PCDHGA7|56108, BYSL|705, MPZL3|196264, ARHGDIG|398, SSPO|23145, PCDHGB5|56101, ADAMTS13|11093, NPHP4|261734, PLXNC1|10154, COL28A1|340267, NF1|4763, SLAMF7|57823, CLDN14|23562, CHST10|9486, CDH9|1007, DST|667, VTN|7448, ZAN|7455, COL2A1|1280, CD6|923, NME2|4831, DLG1|1739, MSLNL|401827, LAMA4|3910, CLDN9|9080, AJAP1|55966, ARHGDIB|397, CDH12|1010, NRXN2|9379, GSN|2934, HSPB11|51668, LEF1|51176, CD24|100133941, PCDHGA4|56111, ITGBL1|9358, NCAM2|4685, OPCML|4978, MFAP4|4239, IL8|3576, PTK2B|2185, VCAM1|7412, ROCK1|6093, SPON1|10418, LAMC2|3918, PCDHGA1|56114, SIGLEC10|89790, PCDHA4|56144, FLRT1|23769, CD99|4267, RELL2|285613, FN1|2335, AEBP1|165, CDSN|1041, CTNNA1|1495, NEDD9|4739, CIB1|10519, LAMC3|10319, FAT3|120114,

CLDN11|5010, PCDHB14|56122, CLDN16|10686, SPP1|6696, MCAM|4162, LRRN2|10446, COL6A1|1291, FXYD5|53827, BCL2|596, KIFAP3|22920, NPTN|27020, COL9A1|1297, LSAMP|4045, ADIPOQ|9370, DCHS2|54798, CLDN7|1366, SNED1|25992, FREM3|166752, PCDH10|57575, CDH26|60437, PODXL|5420, PODXL2|50512, TSTA3|7264, VNN1|8876, SDC3|9672, RELN|5649, UNCX|340260, PCDHGA6|56109, EGFR|1956, MYBPC3|4607, CLEC4M|10332, ROPN1B|152015, TROAP|10024, TSC1|7248, ONECUT1|3175, POSTN|10631, FAT1|2195, CG**REF**1|10669, B4GALNT2|124872, COL16A1|1307, CCL11|6356, NR2C2AP|126382, IBSP|3381, LMO7|4008, SORBS1|10580, PVR|5817, CELSR3|1951, TNXB|7148, COL18A1|80781, COL11A1|1301, SAA1|6288, PCDH20|64881, MUC5AC|4586, COL22A1|169044, PCDHGB3|56102, SPACA4|171169, PPP2R1A|5518, TINAG|27283, OTOR|56914, PPFIBP1|8496, PTPRT|11122, PCDHA10|56139, ITGAX|3687, VCL|7414, PCDHGA5|56110, DSG3|1830, DPT|1805, FER|2241, PTPRU|10076, MPZL2|10205, SELPLG|6404, ROM1|6094, SPAM1|6677, ERBB2IP|55914, AMIGO1|57463, CADM1|23705, CDHR1|92211, EMR1|2015, LAMB4|22798, PCDHA11|56138, LMLN|89782, ADAM2|2515, AMBN|258, ARHGDIA|396, HES5|388585, DSG1|1828, SCARB1|949, ADAM22|53616, FBLN7|129804, CLDN10|9071, CLDN19|149461, STAB2|55576, FBLIM1|54751, CX3CR1|1524, SPOCK1|6695, CLSTN2|64084, ACTN2|88, CDH11|1009, PECAM1|5175, PCDHB13|56123, ITGA2B|3674, FERMT3|83706, PCDHGA8|9708, LY9|4063, CCR1|1230, PCDHA5|56143, PRLR|5618, DSG2|1829, CHST4|10164, ARVCF|421, JAK2|3717, CXCL12|6387, FAT2|2196, PCDHB6|56130, FPR2|2358, TGFB1|7040, ADAM17|6868, CLEC7A|64581, ATP2C1|27032, CYTH3|9265, FLRT2|23768, NLGN1|22871, ITGB1BP1|9270, OMD|4958, CDH18|1016, SYMPK|8189, SIGLEC9|27180, EDA|1896, RND3|390, PCDH12|51294, CDH20|28316, NLGN4Y|22829, NPHP1|4867, L1CAM|3897, PCDHGA12|26025, PCDHB4|56131, NELL1|4745, CD33|945.

DNA repair

Gene Symbol|Entrez-Gene ID - RAD9A|5883, POLB|5423, CUL4A|8451, CCNO|10309, KAT5|10524, MORF4L2|9643, GTF2H5|404672, BARD1|580, DCLRE1A|9937, EYA4|2070, NBN|4683, TP73|7161, RAD23B|5887, POLL|27343, MRE11A|4361, CRY2|1408, ERCC1|2067, RAD1|5810, NONO|4841, PIPSL|266971, CSNK1E|1454, EYA3|2140, C1ORF124|83932, MSH5|4439, TTRAP|51567, RFC2|5982, RPA3|6119, BRCA2|675, RAD18|56852, NSMCE1|197370, EME2|197342, TYMS|7298, ERCC5|2073, RPS3|6188, TNP1|7141, SSRP1|6749, RECQL|5965, HMGB1|3146, NTHL1|4913, CLSPN|63967, SLK|9748, POLG|5428, GADD45G|10912, KIN|22944, PNKP|11284, GIYD1|548593, JMY|133746, SMC3|9126, APLF|200558, MSH2|4436, USP1|7398, POLD4|57804, UHRF1|29128, ASF1A|25842, GTF2H3|2967, DDB2|1643, TOPBP1|11073, POLA1|5422, CHAF1B|8208, MPG|4350, FANCA|2175, FANCF|2188, XRCC5|7520, TEX15|56154, ALKBH3|221120, RRM2B|50484, RAD51|5888, CEBPG|1054, WDR33|55339, ATXN3|4287, FANCC|2176, BRIP1|83990, GTF2H2C|728340, PTTG1|9232, CHD1L|9557, IGHMBP2|3508, MLH3|27030, PMS2L12|392713, MMS19|64210, TDG|6996, BTBD12|84464, ZSWIM7|125150,

NCOA6|23054, DCLRE1B|64858, OBFC2A|64859, ATRIP|84126, RECQL5|9400, MSH4|4438, CDKN2D|1032, FEN1|2237, SFPQ|6421, ASTE1|28990, SMUG1|23583, RFC4|5984, PMS1|5378, RFC3|5983, CCNH|902, TREX1|11277, RAD17|5884, CDK7|1022, XRCC6BP1|91419, FANCG|2189, POLG2|11232, SUPT16H|11198, RAD51C|5889, RDM1|201299, ERCC6|2074, NSMCE2|286053, XRCC6|2547, SMC1A|8243, H2AFX|3014, UIMC1|51720, ALKBH1|8846, BRCC3|79184, XRCC4|7518, GEN1|348654, UBE2N|7334, EYA2|2139, CSNK1D|1453, SETX|23064, POLQ|10721, FAM175A|84142, UVRAG|7405, PMS2L5|5383, XPC|7508, RAD54L|8438, FBXO6|26270, MLH1|4292, BLM|641, NEIL2|252969, SUMO1|7341, RPA2|6118, ABL1|25, ERCC2|2068, CRY1|1407, GTF2H2|2966, REV1|51455, MSH6|2956, APEX1|328, PAPD7|11044, POLK|51426, FBXO18|84893, UBE2A|7319, PARP4|143, EPC2|26122, RNF8|9025, APTX|54840, PSMD4|5710, FANCM|57697, RPA1|6117, FANCB|2187, POLE2|5427, POLE|5426, UNG|7374, ATM|472, HMGB2|3148, NHEJ1|79840, DDB1|1642, ATRX|546, ERCC3|2071, CUL4B|8450, FANCD2|2177, PRMT6|55170, FANCI|55215, MBD4|8930, FANCL|55120, PCNA|5111, RBM14|10432, XRCC2|7516, KIF22|3835, OGG1|4968, TRIP13|9319, HINFP|25988, TDP1|55775, ERCC8|1161, PMS2CL|441194, SMC5|23137, SOD1|6647, RPS27L|51065, RBX1|9978, RAD51AP1|10635, INTS3|65123, UBE2V1|7335, BRCA1|672, POLN|353497, TTC5|91875, VCP|7415, HUS1|3364, CHAF1A|10036, HSPA1L|3305, NEIL1|79661, PARP3|10039, TP53BP1|7158, RAD51L1|5890, UBE2B|7320, RNF168|165918, EME1|146956, CIB1|10519, MTMR15|22909, LIG1|3978, ERCC4|2072, CHEK1|1111, LIG3|3980, CEP164|22897, RECQL4|9401, XPA|7507, DNA2|1763, APEX2|27301, GTF2H1|2965, ATR|545, PMS2L2|5380, SETMAR|6419, LIG4|3981, RAD52|5893, WRN|7486, XRCC1|7515, PMS2L11|441263, MNAT1|4331, REV3L|5980, CHRNA4|1137, MUS81|80198, RUVBL2|10856, POLD2|5425, RAD21|5885, BCCIP|56647, XAB2|56949, PMS2|5395, SOD2|6648, TP53|7157, ESCO2|157570, MDC1|9656, PMS2L1|5379, POLD1|5424, RPAIN|84268, RAD23A|5886, BRE|9577, PRKCG|5582, GTF2H4|2968, ALKBH2|121642, MC1R|4157, POLI|11201, EEPD1|80820, PRPF19|27339, SIRT1|23411, PARP1|142, BTG2|7832, PARP2|10038, DCLRE1C|64421, FANCE|2178, RAD54B|25788, EXO1|9156, NUDT1|4521, RBBP8|5932, UBE2V2|7336, XRCC3|7517, RFC1|5981, MUTYH|4595, POLH|5429, RAD51L3|5892, SMG1|23049, ESCO1|114799, EYA1|2138, SHPRH|257218, RAD9B|144715, SMC6|79677, WRNIP1|56897, MORF4L1|10933, TOP2A|7153, TREX2|11219, TMEM161A|54929, OBFC2B|79035, RTEL1|51750, PRKDC|5591, MEN1|4221, POLD3|10714, MGMT|4255, UPF1|5976, GADD45A|1647, RFC5|5985, NEIL3|55247, RAD50|10111, MSH3|4437.

Disease-centered gold standards

Hypertension

Gene Symbol|Entrez-Gene ID - ACVRL1|94, ADD1|118, ADD2|119, ADD3|120, ADM|133, ADRA2B|151, ADRB1|153, ADRB2|154, AGT|183, AGTR1|185,

AGTR2|186, APOB|338, APOE|348, ATP1B1|481, BDKRB1|623, BDKRB2|624, BMPR2|659, CAT|847, CLU|1191, CPS1|1373, CYBA|1535, CYP2C8|1558, CYP2C9|1559, CYP2J2|1573, CYP3A5|1577, CYP4A11|1579, CYP11B2|1585, ACE|1636, DRD1|1812, |1889, EDN1|1906, EDN2|1907, EDNRA|1909, EPHX1|2052, FGB|2244, FOXF1|2294, GCK|2645, GNB3|2784, GNB3|2784, GRK4|2868, HGF|3082, KCNMB1|3779, KRT8|3856, KRT18|3875, SMAD9|4093, NR3C2|4306, NOS2|4843, NOS3|4846, SERPINE1|5054, PEE1|5177, ABCB1|5243, PTGIS|5740, SCN7A|6332, SCNN1B|6338, SCNN1G|6340, SELE|6401, SLC6A2|6530, SLC8A1|6546, SLC12A1|6557, SLC12A3|6559, SOD2|6648, SOD3|6649, TH|7054, TNF|7124, PHA2A|7830, HTNB|8080, RGS5|8490, CART|9607, MBOAT5|10162, CORIN|10699, CAPN10|11132, HYT2|50986, MEX3C|51320, RETN|56729, RFH1|59331, WNK1|65125, WNK4|65266, ACSM1|116285, HYT1|117191, STOX1|219736, HYT3|387575, HYT4|444980, HYT8|100188321, HYT5|100188807, HYT6|100188808, HYT7|100188825, CTEPH1|100302516.

Obesity

Gene Symbol|Entrez-Gene ID - A2M|2, ACP1|52, ADRA2A|150, ADRA2B|151, ADRB1|153, ADRB2|154, ADRB3|155, AGRP|181, AGT|183, AHSG|197, APOA4|337, APOB|338, APOC3|345, APOE|348, FAS|355, AQP7|364, AR|367, BBS4|585, BDNF|627, CAPN5|726, SERPINA6|866, CETP|1071, CIDEA|1149, CNR1|1268, CPT1A|1374, CPT1B|1375, DBP|1628, ACE|1636, DRD4|1815, F7|2155, FAAH|2166, FABP2|2169, FOXC2|2303, GABRA6|2559, GAD2|2572, GFPT1|2673, GNB3|2784, GPR10|2834, MCHR1|2847, NR3C1|2908, HSD11B1|3290, HSPA1B|3304, HTR2A|3356, HTR2C|3358, IGF2|3481, IL6|3569, IL6R|3570, INS|3630, IRS1|3667, LDLR|3949, LEP|3952, LEPR|3953, LIPC|3990, LIPE|3991, LMNA|4000, LPL|4023, MAOA|4128, MC3R|4159, MC4R|4160, MIF|4282, MTTP|4547, NMB|4828, NPY|4852, NPY5R|4889, SERPINE1|5054, PCSK1|5122, ENPP1|5167, PLIN|5346, PLTP|5360, POMC|5443, PPARA|5465, PPARD|5467, PPARG|5468, PYY|5697, PTPN1|5770, PTPRF|5792, ACSM3|6296, SIM1|6492, SREBF1|6720, STCH|6782, TH|7054, TNF|7124, TNFRSF1B|7133, UCP1|7350, UCP2|7351, UCP3|7352, VDR|7421, WTS|7492, MEHMO|8422, NR0B2|8431, IRS2|8660, ADIPOQ|9370, CART|9607, SDC3|9672, PPARGC1A|10891, CAPN10|11132, SLC6A14|11254, GHRL|51738, INPP5E|56623, OB10P|56694, RETN|56729, AOMS1|65076, AOMS2|65077, FTO|79068, APOA5|116519, PPARGC1B|133522, VPS13B|157680, BMIQ4|338026, OB10Q|353126, OB4|404683, WAGRO|100233159.

Schizophrenia

Gene Symbol|Entrez-Gene ID - ADRA1A|148, AKT1|207, ALDH3A1|218, APOE|348, CCK|885, CCKAR|886, CHGB|1114, CHI3L1|1116, CHRM1|1128, CNR1|1268, COMT|1312, CTLA4|1493, CYP1A2|1544, DAO|1610, DRD2|1813, DRD3|1814, DRD4|1815, DRD5|1816, DRP2|1821, ATN1|1822, EGF|1950, ACSL4|2182, GABBR1|2550, GABRA5|2558, GDNF|2668, GRIA4|2893, GRIK3|2899,

GRIN1|2902, GRIN2A|2903, GRIN2B|2904, GRM3|2913, GRM5|2915, GRM8|2918, GSTM1|2944, NRG1|3084, NRG1|3084, DRB1|3123, HTR2A|3356, MTHFR|4524, NOS1|4842, NPY|4852, NOTCH4|4855, NTF3|4908, NR4A2|4929, OPRS1|4989, PDYN|5173, ABCB1|5243, PIK3C3|5289, PLA2G1B|5319, PON1|5444, PPP3CC|5533, PRODH|5625, RGS4|5999, RXRB|6257, S100B|6285, SCA1|6310, SCZD3|6365, SCZD1|6377, SCZD2|6378, SCZD4|6379, SLC1A2|6506, SLC6A4|6532, SNAP25|6616, SOD2|6648, SYN2|6854, TH|7054, TNF|7124, FZD3|7976, SYN3|8224, SCZD6|8400, SCZD7|8401, APOL1|8542, SCZD8|8806, SYNGR1|9145, SNAP29|9342, CART|9607, FEZ1|9638, CLINT1|9685, SLC12A6|9990, PDLIM5|10611, CHL1|10752, APOL2|23780, DISC2|27184, DISC1|27185, RTN4|57142, SCZD10|63944, RTN4R|65078, APOL4|80832, DTNBP1|84062, DAOA|267012, TAAR6|319100, SCZD11|404686, SCZD12|619488, SCZD13|100196913.

C. APPENDIX 3: EBIMED-EXTRACTED PRIORITIZED GENE LISTS

Apoptosis

Gene Symbol|Entrez-Gene ID:Score - TNFSF10|8743:2370, TNF|7124:1289, TP53|7157:1287, ANXA5|308:1179, CYCS|54205:1116, AIFM1|9131:633, XIAP|331:632, ALPI|248:615, CD47|961:615, MAGT1|84061:615, FASLG|356:547, ERVK6|64006:501, HLA-DRB1|3123:491, TNFRSF10B|8795:491, TNFRSF10A|8797:363, CFLAR|8837:340, CD4|920:323, FADD|8772:315, PCNA|5111:312, EPHB2|2048:300, CRK|1398:276, GRAP2|9402:276, AHSA1|10598:276, POLDIP2|26073:276, CDKN1A|1026:222, IFNG|3458:202, CASP8|841:194, DIABLO|56616:186, APAF1|317:174, TNFRSF10D|8793:160, TNFRSF10C|8794:160, CSF2|1437:157, PCBD1|5092:147, NLRP1|22861:142, C9ORF3|84909:140, PAWR|5074:135, BAX|581:129, IL1B|3553:124, PIK3CA|5290:124, PIK3CB|5291:124, PIK3CD|5293:124, PIK3CG|5294:124, PIK3R1|5295:124, CASP1|834:121, NOS2|4843:119, BIRC2|329:108, DNTT|1791:106, NFKBIA|4792:104, CASP3|836:102, IL2|3558:102, FAS|355:101, FASN|2194:101, BIRC3|330:100, IL10|3586:99, TGFB1|7040:96, IL6|3569:94, DDIT3|1649:93, DFFA|1676:93, CAT|847:90, IL8|3576:88, MAPK3|5595:88, E2F1|1869:87, CES2|8824:83, ANXA1|301:81, EGFR|1956:81, GCHFR|2644:81, MAP3K5|4217:81, MNAT1|4331:81, CDK5R1|8851:81, UPK3B|80761:81, CDCA5|113130:81, EGF|1950:80, BBC3|27113:78, AKT1|207:77, PML|5371:76, UBC|7316:76, TNFRSF1A|7132:73, STS|412:72, CALM3|808:71, ATM|472:70, AGT|183:69, ESR1|2099:69, SMPD1|6609:67, BCR|613:66, PTGS2|5743:66, BAD|572:65, BCL2|596:64, IFI27|3429:64, SYT1|6857:64, WNK1|65125:64, IL5|3567:62, HTRA2|27429:61, GLI2|2736:59, IL3|3562:59, UMOD|7369:59, EIF2AK2|5610:57, SOAT1|6646:56, NR3C1|2908:55, IL4|3565:55, ELF4|2000:54, MEFV|4210:54, VEGFA|7422:54, PTEN|5728:53, APC|324:52, JUN|3725:52, MMRN1|22915:52, TP73|7161:51, NAIP|4671:48, CLEC4D|338339:48, CALML3|810:47, MDM2|4193:47, TIMP1|7076:47, CSRP3|8048:47, COTL1|23406:47, CD34|947:46, CTNNB1|1499:46, ITGB2|3689:46, PLEKHF1|79156:46, BIRC7|79444:46, CD40|958:45, MTX1|4580:45, PLAT|5327:45, KRT18|3875:44, TERT|7015:44, BAK1|578:43, MTOR|2475:43, BID|637:42, MS4A1|931:42, PPARG|5468:42, CPOX|1371:41, PTK2|5747:41, DFFB|1677:40, HSPA5|3309:40, ST3GAL4|6484:40, BTG2|7832:40, CBX8|57332:40, STAT1|6772:39, STAT3|6774:39, NUP43|348995:39, CAST|831:38, GLB1|2720:38, SLC25A3|5250:38, MAPK8|5599:38, REG1A|5967:38, EBP|10682:38, SNAP25|6616:37, PARP1|142:36, APP|351:36, TNFRSF11B|4982:35, DAP|1611:34, IAPP|3375:34, VIM|7431:34, AR|367:33, CEL|1056:33, EIF2S1|1965:33, PTPRC|5788:33, SLC27A5|10998:33, PARP9|83666:33, HIF1A|3091:32, PGR|5241:32, PI3|5266:32, BECN1|8678:32, AATF|26574:32, CD40LG|959:31, CDK2|1017:31, KITLG|4254:31, NTF3|4908:31, PDCD5|9141:31, SCYL1|57410:31, ADM|133:30, LOX|4015:30, RENBP|5973:30, MBTPS1|8720:30, AIFM2|84883:29, CDK4|1019:28,

EPHA3|2042:28, ABL1|25:27, CA1|759:27, CD14|929:27, DAXX|1616:27, IL15|3600:27, TFPI|7035:27, SIRT1|23411:27, CREB1|1385:26, CREBBP|1387:26, EDN1|1906:26, IKBKB|3551:26, PAG1|55824:26, CSF1|1435:25, MAP3K1|4214:25, PLAU|5328:25, FBXO8|26269:25, FBRS|64319:25, BIRC5|332:24, CD28|940:24, CD68|968:24, CDKN1B|1027:24, PLAUR|5329:24, MAP2K1|5604:24, TNFSF11|8600:24, CAV1|857:23, CD69|969:23, FGF2|2247:23, HPD|3242:23, MMP9|4318:23, SMPD2|6610:23, TLR4|7099:23, TNFRSF6B|8771:23, EIF2C2|27161:23, CGN|57530:23, ALOX5|240:22, BRCA1|672:22, GSK3B|2932:22, HMGCR|3156:22, HSPD1|3329:22, ODC1|4953:22, PKLR|5313:22, ERVK5|60358:22, CDK1|983:21, CORT|1325:21, DLD|1738:21, EP300|2033:21, ERV3|2086:21, HGF|3082:21, IGF2|3481:21, PAEP|5047:21, PRIM1|5557:21, MSC|9242:21, ERVWE1|30816:21, SLC25A21|89874:21, LDHD|197257:21, HERV-FRD|405754:21, ARCN1|372:20, AGFG1|3267:20, LGALS3|3958:20, MPO|4353:20, RASA1|5921:20, RIPK1|8737:20, BCL2L11|10018:20, RPAIN|84268:20, CPE|1363:19, DCC|1630:19, EPO|2056:19, ABCB1|5243:19, PRKAA2|5563:19, PRKAB1|5564:19, MAPK1|5594:19, TGM2|7052:19, TLR1|7096:19, KDM5D|8284:19, EPX|8288:19, CDC123|8872:19, STYK1|55359:19, MRS2|57380:19, APCS|325:18, ARSA|410:18, ATF4|468:18, BDNF|627:18, CALB1|793:18, CCNA2|890:18, CSF3|1440:18, IFNB1|3456:18, SH2D1A|4068:18, PRKD1|5587:18, STAT5A|6776:18, TGFB2|7042:18, TRAF2|7186:18, AOC3|8639:18, EIF2AK3|9451:18, NES|10763:18, SETBP1|26040:18, CCAR1|55749:18, RNF34|80196:18, DST|667:17, TSPO|706:17, CASR|846:17, CD86|942:17, DAPK1|1612:17, DAPK3|1613:17, INS|3630:17, ABCC1|4363:17, PAX2|5076:17, CCL2|6347:17, TLR2|7097:17, TTF2|8458:17, EIF3F|8665:17, MARCKSL1|65108:17, TRIM63|84676:17, CD44|960:16, SFN|2810:16, H2AFX|3014:16, ITK|3702:16, SMAD7|4092:16, P2RX7|5027:16, SERPINE1|5054:16, SERPINB5|5268:16, RARA|5914:16, RPS6KB1|6198:16, SARS|6301:16, MAP2K4|6416:16, SLC22A3|6581:16, XDH|7498:16, TRADD|8717:16, F2RL3|9002:16, MAGED1|9500:16, REXO2|25996:16, BNIP3|664:15, RUNX1T1|862:15, CCNH|902:15, CDKN2A|1029:15, COL15A1|1306:15, CSE1L|1434:15, CTNND1|1500:15, HADH|3033:15, LGALS9|3965:15, MMP2|4313:15, MTTP|4547:15, MYCN|4613:15, NCAM1|4684:15, RAF1|5894:15, RPA2|6118:15, XBP1|7494:15, DAP3|7818:15, CXCR4|7852:15, TNFSF13|8741:15, TMPRSS11D|9407:15, BCAR1|9564:15, AATK|9625:15, ANP32B|10541:15, MGEA5|10724:15, DLGAP3|58512:15, COL18A1|80781:15, SNHG3-RCC1|751867:15, AMD1|262:14, ATR|545:14, BRCA2|675:14, TNFRSF8|943:14, CCR5|1234:14, GADD45A|1647:14, TIMM8A|1678:14, EDA|1896:14, HPR|3250:14, IL1A|3552:14, IL11|3589:14, MVD|4597:14, PPARA|5465:14, SAA2|6289:14, SFRS1|6426:14, TGFA|7039:14, TNFSF14|8740:14, SMC2|10592:14, KHDRBS1|10657:14, ANTXR1|84168:14, SERPINA2|390502:14, ACHE|43:13, AGA|175:13, CA3|761:13, CD3D|915:13, CD19|930:13, CD38|952:13, ELANE|1991:13, FPR1|2357:13, GAST|2520:13, HMOX1|3162:13, ICAM1|3383:13, IL7|3574:13, TNPO1|3842:13, MIP|4284:13, MIPEP|4285:13, RPE|6120:13, SFRP1|6422:13, SFRS2|6427:13, SON|6651:13, PRDX2|7001:13, TXN|7295:13, HGS|9146:13, TXNRD2|10587:13, MALT1|10892:13, ACD|65057:13, TAS1R3|83756:13, POLDIP3|84271:13, CMTM8|152189:13, AMH|268:12, ENDOG|2021:12, ERBB2|2064:12, HTT|3064:12, IGFALS|3483:12, CXCR2|3579:12, LGALS1|3956:12, RAB8A|4218:12,

NR3C2|4306:12, NOVA2|4858:12, VDAC1|7416:12, WT1|7490:12,
TNFSF13B|10673:12, LAT|27040:12, SPESP1|246777:12, ARHGAP4|393:11,
ATF3|467:11, C5AR1|728:11, CAPN2|824:11, PLK3|1263:11, E2F4|1874:11,
PTK2B|2185:11, GDNF|2668:11, GFAP|2670:11, GLA|2717:11, HK2|3099:11,
KCNA5|3741:11, KIF2A|3796:11, PBX1|5087:11, MAP2K3|5606:11, PRL|5617:11,
RNASE3|6037:11, SAG|6295:11, TNFAIP1|7126:11, UGCG|7357:11, MIA|8190:11,
RIPK2|8767:11, TNFRSF11A|8792:11, RNF7|9616:11, IFI44|10561:11,
CHEK2|11200:11, GCA|25801:11, HUNK|30811:11, GMCL1|64395:11,
GMCL1L|64396:11, AGTR2|186:10, BIK|638:10, BLM|641:10, CD27|939:10,
CD80|941:10, CPA1|1357:10, CRP|1401:10, CSRP1|1465:10, DDT|1652:10,
HBEGF|1839:10, HDAC2|3066:10, IL2RB|3560:10, IL18|3606:10, KCNJ6|3763:10,
KIT|3815:10, LAMP2|3920:10, LPA|4018:10, NFATC1|4772:10, PRKCA|5578:10,
PTH|5741:10, RNH1|6050:10, STAT4|6775:10, TF|7018:10, TNFRSF1B|7133:10,
TRAF1|7185:10, TRPV1|7442:10, STK17B|9262:10, PPP1R13L|10848:10,
CABIN1|23523:10, NGFRAP1|27018:10, HIPK2|28996:10, PYCARD|29108:10,
MAP9|79884:10, KIAA1804|84451:10, PTRH1|138428:10, NRK|203447:10,
ADRB2|154:9, ABCD1|215:9, ALOX12|239:9, ARSF|416:9, CEACAM1|634:9,
CALCA|796:9, CASP2|835:9, CD5|921:9, CD59|966:9, CPB1|1360:9, CTGF|1490:9,
CD55|1604:9, NQO1|1728:9, ELK3|2004:9, EPHB1|2047:9, FCGR3A|2214:9,
FCGR3B|2215:9, FGF1|2246:9, FGF12|2257:9, FHIT|2272:9, GAPDH|2597:9,
GPI|2821:9, HARS|3035:9, HCLS1|3059:9, NRG1|3084:9, HLA-DMA|3108:9,
HRG|3273:9, HSD11B1|3290:9, IVL|3713:9, LSS|4047:9, LTA|4049:9, LTF|4057:9,
MAP3K3|4215:9, MMP7|4316:9, MYC|4609:9, NOS3|4846:9, PEPD|5184:9,
SERPINF2|5345:9, PSMD4|5710:9, PTGS1|5742:9, RNASEL|6041:9, SFRS5|6430:9,
SLC6A2|6530:9, SRM|6723:9, TAT|6898:9, TFDP1|7027:9, KLF10|7071:9,
TOP1|7150:9, REEP5|7905:9, KHSRP|8570:9, DYNLL1|8655:9, CRADD|8738:9,
AIP|9049:9, PCSK7|9159:9, CD83|9308:9, POM121|9883:9, SDS|10993:9,
SIGLEC7|27036:9, RTEL1|51750:9, AURKAIP1|54998:9, CCL28|56477:9,
SFXN1|94081:9, UNC5B|219699:9, AFP|174:8, CHEK1|1111:8, CRAT|1384:8,
CYBB|1536:8, DAD1|1603:8, DBP|1628:8, EIF4E|1977:8, EPHA4|2043:8,
ERN1|2081:8, F9|2158:8, GC|2638:8, GDF2|2658:8, MSH6|2956:8, HRH2|3274:8,
IGF1R|3480:8, PDX1|3651:8, SMAD4|4089:8, MAFG|4097:8, MIF|4282:8,
MLH1|4292:8, MLN|4295:8, MNDA|4332:8, MSH2|4436:8, MST1|4485:8,
MUC2|4583:8, NOTCH1|4851:8, NOTCH2|4853:8, NPPB|4879:8, NT5E|4907:8,
PGC|5225:8, PGD|5226:8, PHEX|5251:8, PLG|5340:8, PTCH1|5727:8, PVR|5817:8,
SNCA|6622:8, STK4|6789:8, SULT2A1|6822:8, PHLDA2|7262:8, XRCC5|7520:8,
TRIM26|7726:8, ST8SIA4|7903:8, PEA15|8682:8, TNFRSF25|8718:8, IL1RL1|9173:8,
DEDD|9191:8, MAGI1|9223:8, BAG3|9531:8, HDAC9|9734:8, CASP8AP2|9994:8,
HAX1|10456:8, RNPS1|10921:8, MAP4K1|11184:8, ATF6|22926:8, RRS1|23212:8,
KIAA0664|23277:8, PLA2G15|23659:8, SETD2|29072:8, C18ORF8|29919:8,
SERTAD1|29950:8, NTM|50863:8, HDAC7|51564:8, ATP8A2|51761:8,
ACSS2|55902:8, MAVS|57506:8, EIF2AK4|440275:8, NCF1|653361:8,
ALOX5AP|241:7, APOE|348:7, KLK3|354:7, BMP2|650:7, KLF5|688:7, CASP9|842:7,
CD36|948:7, CHKA|1119:7, CPM|1368:7, ATF2|1386:7, DES|1674:7,
SERPINB1|1992:7, ERCC1|2067:7, ERCC2|2068:7, GIF|2694:7, PDIA3|2923:7,
ICAM3|3385:7, CYR61|3491:7, IL2RA|3559:7, IDO1|3620:7, KRT1|3848:7,

KRT14|3861:7, MAP2|4133:7, MEN1|4221:7, MPP1|4354:7, MRC1|4360:7, 40423|4735:7, NUCB2|4925:7, P4HB|5034:7, REG3A|5068:7, SLC26A4|5172:7, PSEN1|5663:7, PTGDS|5730:7, RARB|5915:7, RBBP4|5928:7, RBBP6|5930:7, S100A2|6273:7, SAT1|6303:7, SLC9A1|6548:7, FSCN1|6624:7, SRC|6714:7, ST13|6767:7, SYP|6855:7, TAGLN|6876:7, TP53BP2|7159:7, HSP90B1|7184:7, XRCC1|7515:7, BAT3|7917:7, TFPI2|7980:7, IKBKG|8517:7, PRKRA|8575:7, NPEPPS|9520:7, CHAF1A|10036:7, RBM5|10181:7, NDC80|10403:7, SH2B2|10603:7, HPSE|10855:7, ERP29|10961:7, CDC37|11140:7, AKAP13|11214:7, POLG2|11232:7, PSAT1|29968:7, CKLF|51192:7, PBRM1|55193:7, IL21|59067:7, HHIP|64399:7, DYNLRB1|83658:7, UNC5A|90249:7, ABP1|26:6, ACAT1|38:6, CRISP1|167:6, AGER|177:6, APEH|327:6, RERE|473:6, CCK|885:6, CD33|945:6, CHUK|1147:6, CR1|1378:6, CYP19A1|1588:6, DBT|1629:6, CFD|1675:6, DMD|1756:6, DNMT3A|1788:6, DNMT3B|1789:6, ENPEP|2028:6, FDX1|2230:6, FDXR|2232:6, DARC|2532:6, GTF3A|2971:6, GUK1|2987:6, GZMB|3002:6, HLCS|3141:6, NR4A1|3164:6, HSPB1|3315:6, IGF1|3479:6, IL6ST|3572:6, IL13|3596:6, ING1|3621:6, ITPR1|3708:6, KCNK3|3777:6, STMN1|3925:6, LIPE|3991:6, LPL|4023:6, LTBR|4055:6, SMAD6|4091:6, MAT2A|4144:6, MMP1|4312:6, NCF2|4688:6, PAH|5053:6, CFP|5199:6, PLA2G4A|5321:6, PRIM2|5558:6, PRKCG|5582:6, PTHLH|5744:6, RAD23B|5887:6, RAGE|5891:6, RARS|5917:6, RB1|5925:6, PRPH2|5961:6, REN|5972:6, RPS6KA3|6197:6, RRAD|6236:6, S100A8|6279:6, S100A9|6280:6, S100A10|6281:6, S100A12|6283:6, CCL3|6348:6, SDHB|6390:6, SHBG|6462:6, SLPI|6590:6, SPTA1|6708:6, TEF|7008:6, TMBIM6|7009:6, TERF2|7014:6, TFAP2A|7020:6, TRAF6|7189:6, UCP1|7350:6, ZBTB16|7704:6, MANF|7873:6, DYRK2|8445:6, MAPKAPK2|9261:6, MAPK8IP1|9479:6, TBPL1|9519:6, CCL4L1|9560:6, MFN2|9927:6, NR1H4|9971:6, PDCD6|10016:6, ARFRP1|10139:6, AASS|10157:6, PRG3|10394:6, UBD|10537:6, RIPK3|11035:6, MPRIP|23164:6, SIRT3|23410:6, PDLIM3|27295:6, REM1|28954:6, NXT1|29107:6, IL21R|50615:6, BFAR|51283:6, UBR5|51366:6, CRLS1|54675:6, NAT10|55226:6, AKR1B10|57016:6, LRRC4|64101:6, UCN3|114131:6, KCTD11|147040:6, C1ORF52|148423:6, NME1-NME2|654364:6, ACTB|60:5, ADA|100:5, ADRB1|153:5, AGTR1|185:5, ANG|283:5, SLC25A4|291:5, SLC25A6|293:5, ABCC6|368:5, BCL6|604:5, BMPR2|659:5, CASP7|840:5, RUNX1|861:5, CD2|914:5, CD7|924:5, CD22|933:5, CD70|970:5, CDK11B|984:5, CEACAM5|1048:5, KLF6|1316:5, CYLD|1540:5, DLAT|1737:5, DNASE1|1773:5, DSG3|1830:5, DUSP1|1843:5, ESR2|2100:5, EVPL|2125:5, F2R|2149:5, FOXO3|2309:5, FLNB|2317:5, GGT1|2678:5, GJB3|2707:5, GP1BA|2811:5, GSS|2937:5, HMGB1|3146:5, IGFBP5|3488:5, IHH|3549:5, KRT8|3856:5, LTB|4050:5, MDK|4192:5, MKI67|4288:5, FOXO4|4303:5, MMP14|4323:5, MSH3|4437:5, MT2A|4502:5, MYBL1|4603:5, NBN|4683:5, NF2|4771:5, NPPA|4878:5, NR4A2|4929:5, OGG1|4968:5, SLC22A18|5002:5, PEBP1|5037:5, PAK1|5058:5, PDC|5132:5, PLA2G1B|5319:5, PLCG1|5335:5, PLCG2|5336:5, PLD1|5337:5, PPT1|5538:5, MAPK6|5597:5, PRPH|5630:5, PXN|5829:5, RBP3|5949:5, RCVRN|5957:5, REL|5966:5, RPA1|6117:5, RPA3|6119:5, SORT1|6272:5, ATXN2|6311:5, CCL20|6364:5, SHH|6469:5, SKI|6497:5, SMN2|6607:5, SNRPN|6638:5, SPI1|6688:5, TACR3|6870:5, TAP1|6890:5, TAPBP|6892:5, TBP|6908:5, TEP1|7011:5, TYRP1|7306:5, SUMO1|7341:5, UCN|7349:5, UVRAG|7405:5, WARS|7453:5, DEK|7913:5, ADAM12|8038:5,

CDC45L|8318:5, USO1|8615:5, TP63|8626:5, UNC5C|8633:5, RNGTT|8732:5, INA|9118:5, SLC33A1|9197:5, PDLIM7|9260:5, TAOK2|9344:5, IKBKE|9641:5, BCAP31|10134:5, NOD1|10392:5, NXF1|10482:5, STK25|10494:5, SIVA1|10572:5, GMEB1|10691:5, MAP3K2|10746:5, SPIN1|10927:5, CKAP4|10970:5, ACIN1|22985:5, CDK19|23097:5, ITGB3BP|23421:5, SEC14L2|23541:5, DDX58|23586:5, BCL2L13|23786:5, CCRN4L|25819:5, RPA4|29935:5, CMPK1|51727:5, TLR9|54106:5, ING3|54556:5, CHDH|55349:5, SMPD3|55512:5, TRERF1|55809:5, AICDA|57379:5, ALS2CR8|79800:5, MCHR2|84539:5, OPN4|94233:5, BIRC8|112401:5, CMTM5|116173:5, TAAR9|134860:5, BNIPL|149428:5, NANOS2|339345:5, CDK11A|728642:5, CCR2|729230:5, JAG1|182:4, ALB|213:4, SLC25A5|292:4, ANXA4|307:4, APEX1|328:4, ARG2|384:4, AVP|551:4, BCL2L1|598:4, BNIP1|662:4, CA4|762:4, CAMP|820:4, CDKN1C|1028:4, CGA|1081:4, ERCC8|1161:4, CLCN3|1182:4, CLCN6|1185:4, CNR2|1269:4, COL2A1|1280:4, MAP3K8|1326:4, CPD|1362:4, CTBS|1486:4, CYB5A|1528:4, CYP11A1|1583:4, CYP17A1|1586:4, ACE|1636:4, DNASE1L3|1776:4, DSP|1832:4, E2F3|1871:4, EMD|2010:4, EPAS1|2034:4, ERBB3|2065:4, ETS1|2113:4, FCGR2B|2213:4, FGF4|2249:4, FGF7|2252:4, FGR|2268:4, FKBP5|2289:4, FUS|2521:4, FUT1|2523:4, GLS|2744:4, GPX4|2879:4, GYPC|2995:4, HCFC1|3054:4, HDAC1|3065:4, HLF|3131:4, HMOX2|3163:4, HNRNPU|3192:4, HRAS|3265:4, HSPG2|3339:4, HTN1|3346:4, ID1|3397:4, ID3|3399:4, IFIT3|3437:4, IK|3550:4, TNFRSF9|3604:4, IMPA1|3612:4, IRF1|3659:4, IRS1|3667:4, JAK2|3717:4, KNG1|3827:4, KRT10|3858:4, LRP1|4035:4, LY6E|4061:4, MAG|4099:4, MAT1A|4143:4, MBP|4155:4, SMCP|4184:4, MMP3|4314:4, MPZ|4359:4, MYB|4602:4, NHLH1|4807:4, NOTCH3|4854:4, OCLN|4950:4, PDK1|5163:4, PDK2|5164:4, PDPK1|5170:4, PECAM1|5175:4, PGF|5228:4, PIP|5304:4, PLD2|5338:4, PLN|5350:4, PRG2|5553:4, PRH2|5555:4, PRKDC|5591:4, MAPK7|5598:4, PSAP|5660:4, RBP4|5950:4, RELA|5970:4, RET|5979:4, MAPK12|6300:4, SATB1|6304:4, SCD|6319:4, SDC2|6383:4, SLC6A3|6531:4, SPTAN1|6709:4, TRIM21|6737:4, TROVE2|6738:4, SSB|6741:4, C2ORF3|6936:4, TRPS1|7227:4, ZNF23|7571:4, ZNF143|7702:4, BRAP|8315:4, NSMAF|8439:4, HRK|8739:4, TNFSF12|8742:4, IER3|8870:4, CDK5R2|8941:4, USP14|9097:4, ATP6V0D1|9114:4, EBAG9|9166:4, STK17A|9263:4, SFRS11|9295:4, KIAA0101|9768:4, MVP|9961:4, BPNT1|10380:4, SPAG11B|10407:4, LYPLA1|10434:4, ENOX2|10495:4, CIB2|10518:4, CCL27|10850:4, BRD8|10902:4, IKZF2|22807:4, PHLDA1|22822:4, LMTK2|22853:4, ICMT|23463:4, RHBDD3|25807:4, GLS2|27165:4, SIGLEC8|27181:4, UHRF1|29128:4, TAS2R13|50838:4, F11R|50848:4, SMOX|54498:4, CPVL|54504:4, TRIB3|57761:4, PERP|64065:4, MOAP1|64112:4, QTRT1|81890:4, PLVAP|83483:4, MYST1|84148:4, AKT1S1|84335:4, TP53INP1|94241:4, EGLN2|112398:4, SESN3|143686:4, DEDD2|162989:4, IL28RA|163702:4, SDK1|221935:4, REXO1L1|254958:4, NPS|594857:4, ACACA|31:3, APLNR|187:3, AHR|196:3, BAG1|573:3, BRAF|673:3, CA9|768:3, CAPG|822:3, CDH1|999:3, CENPB|1059:3, CCR1|1230:3, CCR7|1236:3, CNP|1267:3, CRH|1392:3, CSF1R|1436:3, CSH2|1443:3, CYP2E1|1571:3, DDC|1644:3, DEFB1|1672:3, DPT|1805:3, DUSP4|1846:3, DYRK1A|1859:3, ELAVL3|1995:3, ERBB4|2066:3, FCGR2A|2212:3, FLG|2312:3, FLT1|2321:3, FLT3|2322:3, FPR2|2358:3, XRCC6|2547:3, GEM|2669:3, GPR3|2827:3, GRB10|2887:3, GRIN2A|2903:3, GRIN2B|2904:3, IGFBP2|3485:3, CXCR1|3577:3, IRF3|3661:3,

LAG3|3902:3, RPSA|3921:3, LDLR|3949:3, EPCAM|4072:3, MARCKS|4082:3, CD46|4179:3, MGP|4256:3, MLL|4297:3, MOG|4340:3, MX2|4600:3, NAP1L1|4673:3, NDUFA6|4700:3, NDUFAB1|4706:3, NFATC4|4776:3, NGFR|4804:3, NOS1|4842:3, SERPINA5|5104:3, SERPINA1|5265:3, PLCB2|5330:3, PLEK|5341:3, POLG|5428:3, SRGN|5552:3, MAPK9|5601:3, MAP2K6|5608:3, MAP2K7|5609:3, PSEN2|5664:3, PTPN13|5783:3, RHD|6007:3, RHO|6010:3, SAFB|6294:3, CLEC11A|6320:3, CCL7|6354:3, SDC4|6385:3, SLC25A1|6576:3, SLCO2A1|6578:3, SLCO1A2|6579:3, SMARCA2|6595:3, SP2|6668:3, SRF|6722:3, SULT1E1|6783:3, STK11|6794:3, SULT1A1|6817:3, ADAM17|6868:3, TFDP2|7029:3, TFF1|7031:3, TNNT3|7140:3, TPO|7173:3, TTR|7276:3, VDR|7421:3, VIP|7432:3, VWF|7450:3, NR4A3|8013:3, SLC25A16|8034:3, PDHX|8050:3, GDF5|8200:3, RBM10|8241:3, PIP5K1A|8394:3, PDE5A|8654:3, BANF1|8815:3, NRP1|8829:3, NOL3|8996:3, HAP1|9001:3, FCGR2C|9103:3, ZBED1|9189:3, TRIP10|9322:3, NUP153|9972:3, DNM1L|10059:3, OPTN|10133:3, RBM6|10180:3, SAP18|10284:3, LANCL1|10314:3, PITRM1|10531:3, LEPREL2|10536:3, CCT4|10575:3, C1QL1|10882:3, CCNI|10983:3, LIAS|11019:3, CHP|11261:3, ICK|22858:3, RYBP|23429:3, MAPK8IP2|23542:3, SHC2|25759:3, PTPN22|26191:3, SLC17A5|26503:3, C1ORF107|27042:3, NAAA|27163:3, IL17B|27190:3, GPR77|27202:3, MAGEH1|28986:3, A1CF|29974:3, IL22|50616:3, FIS1|51024:3, SH3GLB1|51100:3, FZR1|51343:3, ZAK|51776:3, CHIC1|53344:3, FLVCR2|55640:3, NHP2|55651:3, C21ORF62|56245:3, MIB1|57534:3, CHD8|57680:3, GORASP1|64689:3, FKRP|79147:3, CHRFAM7A|89832:3, MYOCD|93649:3, FBXO32|114907:3, UHRF2|115426:3, NANOS3|342977:3, SFTPA1|653509:3, SFTPA2|729238:3, AFM|173:2, AIF1|199:2, ALK|238:2, ANXA6|309:2, AIRE|326:2, BARD1|580:2, BNIP3L|665:2, C4BPA|722:2, CALB2|794:2, CANX|821:2, CARS|833:2, CCNG1|900:2, CD5L|922:2, CD6|923:2, CD9|928:2, CD81|975:2, CDC34|997:2, CPT1A|1374:2, CPT1B|1375:2, CRHR2|1395:2, DNM1|1759:2, DPEP1|1800:2, DSPP|1834:2, DTNB|1838:2, ENO2|2026:2, ERCC4|2072:2, ERG|2078:2, ESD|2098:2, EWSR1|2130:2, F12|2161:2, ACSL4|2182:2, FOXO1|2308:2, FOS|2353:2, GAS1|2619:2, GATA4|2626:2, GJA1|2697:2, CFHR1|3078:2, UBE2K|3093:2, HPS1|3257:2, HSD11B2|3291:2, HSF1|3297:2, IL17A|3605:2, ITIH4|3700:2, STT3A|3703:2, KCNH2|3757:2, KRT19|3880:2, LCN2|3934:2, LGALS4|3960:2, MAD2L1|4085:2, MB|4151:2, MBL2|4153:2, ADAM11|4185:2, MEF2D|4209:2, MLLT3|4300:2, MMP12|4321:2, ND4|4538:2, MYBL2|4605:2, NAGLU|4669:2, NEFM|4741:2, NGF|4803:2, NTRK1|4914:2, OPA1|4976:2, ORM1|5004:2, PAM|5066:2, PFDN5|5204:2, PNN|5411:2, PPIA|5478:2, PRF1|5551:2, PRKCB|5579:2, PRNP|5621:2, PRTN3|5657:2, MAP4K2|5871:2, RBL2|5934:2, RFX1|5989:2, CCL21|6366:2, CCL22|6367:2, SNRNP70|6625:2, SNRPF|6636:2, SORD|6652:2, SP1|6667:2, SPINK1|6690:2, SPN|6693:2, SPP1|6696:2, STK3|6788:2, TCF7L2|6934:2, TFRC|7037:2, TGFB3|7043:2, THBS1|7057:2, TNR|7143:2, TPR|7175:2, TUFM|7284:2, UCHL1|7345:2, UCP2|7351:2, UTRN|7402:2, VCP|7415:2, VHL|7428:2, FOSL1|8061:2, NCOA3|8202:2, AXIN2|8313:2, HIST2H4A|8370:2, API5|8539:2, DENR|8562:2, DLK1|8788:2, PIAS2|9063:2, EXO1|9156:2, B4GALT6|9331:2, NCR2|9436:2, NCR1|9437:2, YAF2|10138:2, SPRY2|10253:2, EFS|10278:2, CLEC4M|10332:2, SEMA3A|10371:2, DEAF1|10522:2, KAT5|10524:2, PROCR|10544:2, ERN2|10595:2, TXNIP|10628:2, CSPG5|10675:2, TCFL5|10732:2, FASTK|10922:2, RALBP1|10928:2, TPPP|11076:2, CAPN10|11132:2, ECD|11319:2,

FAIM2|23017:2, MYCBP2|23077:2, JMJD6|23210:2, PLXNB2|23654:2, BRD1|23774:2, SYF2|25949:2, GRHL1|29841:2, GPR132|29933:2, MINK1|50488:2, SOST|50964:2, UTP11L|51118:2, CHMP5|51510:2, NLRP2|55655:2, BEX4|56271:2, SHD|56961:2, NUP37|79023:2, SLC25A22|79751:2, CD276|80381:2, SLC25A18|83733:2, ING5|84289:2, MT4|84560:2, BMF|90427:2, MOGAT1|116255:2, LRG1|116844:2, ACVR1C|130399:2, STT3B|201595:2, ZBTB38|253461:2, NCR3|259197:2, BEX5|340542:2, ACP2|53:1, ADAR|103:1, NR0B1|190:1, AKR1B1|231:1, ANXA13|312:1, APOH|350:1, ATP12A|479:1, ATP4A|495:1, CBL|867:1, CCNT1|904:1, ENTPD1|953:1, CD63|967:1, CDK9|1025:1, CHGA|1113:1, COMP|1311:1, CLDN4|1364:1, CR2|1380:1, CRMP1|1400:1, CTLA4|1493:1, CTSD|1509:1, CUX1|1523:1, CYP1B1|1545:1, CYP11B2|1585:1, DGUOK|1716:1, DTX1|1840:1, E2F5|1875:1, EEF1A2|1917:1, ETS2|2114:1, NR5A1|2516:1, FYN|2534:1, GAD2|2572:1, GAP43|2596:1, GPX1|2876:1, HGD|3081:1, HLA-A|3105:1, HNRNPA1|3178:1, IARS|3376:1, IDH1|3417:1, IDH2|3418:1, IRF4|3662:1, IVD|3712:1, JAG2|3714:1, LAMP1|3916:1, LMNA|4000:1, LMNB1|4001:1, LPO|4025:1, MFGE8|4240:1, MGMT|4255:1, MUC1|4582:1, OMP|4975:1, OXT|5020:1, P2RX1|5023:1, P2RX3|5024:1, P2RX4|5025:1, P2RX5|5026:1, P2RY1|5028:1, P2RY2|5029:1, PCYT1A|5130:1, PIGA|5277:1, PKD1|5310:1, PLAGL2|5326:1, PNOC|5368:1, PPP1CA|5499:1, PTBP1|5725:1, PTPRF|5792:1, TRIM27|5987:1, RIT1|6016:1, RNF5|6048:1, RXRG|6258:1, RYR1|6261:1, RYR2|6262:1, CCL5|6352:1, SLA|6503:1, SLC2A4|6517:1, SLC8A1|6546:1, SSFA2|6744:1, SSTR2|6752:1, SSTR3|6753:1, SSTR4|6754:1, SSTR5|6755:1, TSN|7247:1, TTN|7273:1, SLMAP|7871:1, CDK2AP1|8099:1, PIR|8544:1, PPAP2B|8613:1, HDAC3|8841:1, PROM1|8842:1, ARHGEF7|8874:1, BCL10|8915:1, P2RX6|9127:1, CBFA2T2|9139:1, AIMP1|9255:1, NTN1|9423:1, BAG4|9530:1, SLK|9748:1, NAMPT|10135:1, CEBPZ|10153:1, YAP1|10413:1, MERTK|10461:1, CUGBP1|10658:1, BTG3|10950:1, KLK11|11012:1, DIDO1|11083:1, CIT|11113:1, CARD8|22900:1, SIRT2|22933:1, P2RX2|22953:1, DAPK2|23604:1, RNF19A|25897:1, TES|26136:1, LAP3|51056:1, TH1L|51497:1, FAIM|55179:1, ARFGAP1|55738:1, CENPJ|55835:1, SLC2A4RG|56731:1, RGMA|56963:1, RTN4|57142:1, PTBP2|58155:1, CERK|64781:1, RTN4R|65078:1, COASY|80347:1, ADC|113451:1, CPO|130749:1, OR2AG1|144125:1, PAOX|196743:1.

Cell adhesion

Gene Symbol|Entrez-Gene ID:Score - ICAM1|3383:5018, VCAM1|7412:3856, PTK2|5747:2655, TNF|7124:1804, CALM3|808:1753, NCAM1|4684:1418, MMRN1|22915:1106, ITGB2|3689:783, PXN|5829:772, ITGA4|3676:756, CD44|960:719, CTNNB1|1499:561, ERVK5|60358:531, ERVK6|64006:531, SELE|6401:511, IL1B|3553:499, PECAM1|5175:466, VCL|7414:457, EPHB2|2048:384, CD34|947:369, KLK3|354:367, NPEPPS|9520:367, PSAT1|29968:367, IL8|3576:366, IL6|3569:346, IFNG|3458:335, CD4|920:309, GLI2|2736:294, UMOD|7369:294, CEACAM5|1048:292, CCL2|6347:252, CD2|914:240, MAPK3|5595:225, CEACAM1|634:222, PIK3CA|5290:213, PIK3CB|5291:213, PIK3CD|5293:213, PIK3CG|5294:213, PIK3R1|5295:213, CRK|1398:212, GRAP2|9402:212,

AHSA1|10598:212, POLDIP2|26073:212, SRC|6714:193, IL4|3565:184, EGF|1950:183, CSF2|1437:179, ICAM2|3384:172, CXCL12|6387:169, F11R|50848:163, TGFB1|7040:159, VEGFA|7422:158, CD58|965:157, ERV3|2086:148, ERVWE1|30816:148, HERV-FRD|405754:148, HGF|3082:143, MADCAM1|8174:142, ICAM3|3385:139, THBS1|7057:138, IL1A|3552:136, PLAUR|5329:135, SELPLG|6404:132, ILK|3611:127, CD14|929:126, CXCR4|7852:121, PTK2B|2185:115, ITGA5|3678:115, KIAA0101|9768:115, MLLT4|4301:113, MMP2|4313:109, IL10|3586:108, VWF|7450:104, CD36|948:102, CD40|958:99, SERPINE1|5054:99, PLAT|5327:99, VIM|7431:98, TNC|3371:97, PAEP|5047:96, EGFR|1956:95, CTNND1|1500:94, MMP9|4318:94, CEACAM6|4680:93, IL2|3558:92, MSC|9242:91, SELL|6402:90, CDH5|1003:89, SELP|6403:89, JAM3|83700:88, SPP1|6696:87, ITGA1|3672:86, AKT1|207:84, LGALS3|3958:83, EZR|7430:83, AOC3|8639:81, SDS|10993:81, ALCAM|214:80, FGF2|2247:80, MUC1|4582:80, RPE|6120:80, EDN1|1906:77, ANPEP|290:75, CRP|1401:75, CSF3|1440:75, CSRP1|1465:75, RBL2|5934:75, GPI|2821:73, PVRL1|5818:73, CDC123|8872:73, CD47|961:72, ITGA2|3673:72, PCNA|5111:72, KITLG|4254:71, IL5|3567:70, BCR|613:69, IL3|3562:69, SYT1|6857:69, WNK1|65125:69, ALPK1|80216:69, AFM|173:68, IL13|3596:68, PVR|5817:67, SOAT1|6646:67, ABL1|25:65, HNRNPU|3192:65, PLAU|5328:65, IKBKAP|8518:65, BRD8|10902:65, CD38|952:64, MAG|4099:64, SPN|6693:64, STAT3|6774:64, APC|324:62, ITGA2B|3674:61, NFKBIA|4792:61, TNPO1|3842:60, MCAM|4162:60, MIP|4284:60, MIPEP|4285:60, XPC|7508:60, SEC23IP|11196:60, ANXA5|308:59, LTA|4049:59, VASP|7408:59, CD28|940:58, CD19|930:57, SSFA2|6744:57, CDKN1A|1026:55, PTPRC|5788:53, RASA1|5921:53, ELF4|2000:52, ESR1|2099:52, MEFV|4210:52, NRCAM|4897:52, NDC80|10403:52, CD151|977:51, RAC2|5880:51, CD209|30835:51, LGALS8|3964:50, PTEN|5728:50, IQGAP1|8826:50, CCL4L1|9560:50, DLD|1738:49, JUN|3725:49, TP53|7157:49, ZYX|7791:49, CDH1|999:48, CXCL1|2919:48, PVRL3|25945:48, AGT|183:47, OCLN|4950:47, TGFBI|7045:47, CEACAM8|1088:46, CD9|928:45, CD86|942:45, IL6R|3570:45, PRKCA|5578:45, TLR1|7096:45, RAB8A|4218:44, MPO|4353:43, RENBP|5973:43, LGALS1|3956:42, CD99|4267:42, PI3|5266:42, PPARG|5468:41, MAPK1|5594:41, TNFRSF6B|8771:41, CAV1|857:40, F10|2159:40, JAK2|3717:40, MMP14|4323:40, MSN|4478:40, MAP2K1|5604:40, GPR56|9289:40, IL18|3606:39, BCAM|4059:39, CXCL9|4283:39, TIMP1|7076:39, BTG2|7832:39, MBTPS1|8720:39, CBX8|57332:39, CAPG|822:38, ICAM4|3386:38, CD46|4179:38, DSP|1832:37, LPA|4018:37, MYLK|4638:37, OPCML|4978:37, ITK|3702:36, RDX|5962:36, SLC22A3|6581:36, MS4A1|931:35, CD97|976:35, DDX11|1663:35, HSPG2|3339:35, IFI27|3429:35, ITGA6|3655:35, SDC2|6383:35, TF|7018:35, IFI44|10561:35, CHL1|10752:35, APOE|348:34, POMC|5443:34, CX3CL1|6376:34, DDR1|780:33, NOS2|4843:33, RNGTT|8732:33, CA3|761:32, CTGF|1490:32, NOS3|4846:32, PKLR|5313:32, CSF1|1435:31, EPHA3|2042:31, ERBB2|2064:31, CCL3|6348:31, CCL21|6366:31, CX3CR1|1524:30, CXCR3|2833:30, CXCL10|3627:30, KIT|3815:30, MMP1|4312:30, GPR126|57211:30, CD276|80381:30, GPR116|221395:30, CD55|1604:29, DSG3|1830:29, MTOR|2475:29, GDNF|2668:29, IL11|3589:29, FSCN1|6624:29, ADAM9|8754:29, LPHN2|23266:29, LPHN3|23284:29, GPR123|84435:29, GPR112|139378:29, GPR114|221188:29, GPR110|266977:29, ERCC8|1161:28, CCR1|1230:28, DCC|1630:28, LAG3|3902:28, MET|4233:28,

SERPINB5|5268:28, PTGS2|5743:28, SCD|6319:28, UBC|7316:28, CNTNAP1|8506:28, TNFSF11|8600:28, CD226|10666:28, FBXO8|26269:28, CHDH|55349:28, SLC2A4RG|56731:28, FBRS|64319:28, CD40LG|959:27, CEL|1056:27, COL15A1|1306:27, CSE1L|1434:27, HBD|3045:27, L1CAM|3897:27, LOX|4015:27, MTX1|4580:27, SLC25A3|5250:27, REG1A|5967:27, TIMP2|7077:27, TTN|7273:27, PKP4|8502:27, BCAR1|9564:27, SLC27A5|10998:27, COL18A1|80781:27, PARP9|83666:27, ACTB|60:26, ARVCF|421:26, CAT|847:26, CD80|941:26, GSK3B|2932:26, CCL5|6352:26, CCL25|6370:26, CD33|945:25, CD81|975:25, DMD|1756:25, GP1BA|2811:25, IL7|3574:25, LTF|4057:25, NID1|4811:25, NTF3|4908:25, PPARA|5465:25, RRM1|6240:25, CCL28|56477:25, NUP43|348995:25, CALML3|810:24, CD69|969:24, CPB1|1360:24, DSCAM|1826:24, HBEGF|1839:24, ELANE|1991:24, FAP|2191:24, GAST|2520:24, GJA1|2697:24, PTPN1|5770:24, CSRP3|8048:24, GLMN|11146:24, COTL1|23406:24, ADAM10|102:23, CA1|759:23, CSHL1|1444:23, DES|1674:23, FGF7|2252:23, GFAP|2670:23, IFNB1|3456:23, CXCR2|3579:23, LDLR|3949:23, MBP|4155:23, MAP3K1|4214:23, OLR1|4973:23, PRG2|5553:23, PTPN11|5781:23, TAT|6898:23, CNTN2|6900:23, ST8SIA4|7903:23, ST8SIA2|8128:23, F2R|2149:22, HBE1|3046:22, CD82|3732:22, NOV|4856:22, PLXNA1|5361:22, PVRL2|5819:22, SPARC|6678:22, LPXN|9404:22, CGN|57530:22, APOC3|345:21, CD7|924:21, CRAT|1384:21, DPP4|1803:21, GIF|2694:21, LPP|4026:21, RPS6KB1|6198:21, SLPI|6590:21, STAT1|6772:21, CXCL13|10563:21, PDLIM3|27295:21, NAT10|55226:21, IL33|90865:21, NRK|203447:21, CD24|100133941:21, ANXA2|302:20, CHUK|1147:20, FGFR1|2260:20, NEDD9|4739:20, PAH|5053:20, RET|5979:20, TAC1|6863:20, MAGI1|9223:20, RND1|27289:20, BGLAP|632:19, BMP2|650:19, MRPL49|740:19, CD22|933:19, CLU|1191:19, F2RL2|2151:19, GFRA1|2674:19, HMGCR|3156:19, MARCKS|4082:19, PKP1|5317:19, CCL20|6364:19, TGFB3|7043:19, TGM2|7052:19, TRO|7216:19, XDH|7498:19, ZP3|7784:19, CYCS|54205:19, PARVA|55742:19, PARD3|56288:19, GPR125|166647:19, ANXA1|301:18, AQP4|361:18, CR1|1378:18, CTBS|1486:18, CTNND2|1501:18, FGFR4|2264:18, GSK3A|2931:18, CXCR1|3577:18, IRS1|3667:18, TNFRSF11B|4982:18, PLA2G4A|5321:18, SEMA3F|6405:18, SNAP25|6616:18, PRDX2|7001:18, TLN1|7094:18, ARHGEF7|8874:18, CBFA2T2|9139:18, CLSTN1|22883:18, FLRT3|23767:18, GORASP2|26003:18, ICOS|29851:18, GP6|51206:18, RTEL1|51750:18, SLAMF7|57823:18, GRASP|160622:18, AGER|177:17, CD68|968:17, CEACAM3|1084:17, EVPL|2125:17, KRT18|3875:17, MATN1|4146:17, PRTN3|5657:17, PTBP1|5725:17, RAGE|5891:17, CXCL11|6373:17, TFPI|7035:17, TGFA|7039:17, TPO|7173:17, VTN|7448:17, IL32|9235:17, GNB2L1|10399:17, EMR2|30817:17, GHRL|51738:17, PTBP2|58155:17, SIRPA|140885:17, ARCN1|372:16, CCNA2|890:16, CREBBP|1387:16, DEFA3|1668:16, JAK1|3716:16, KRT12|3859:16, NR3C2|4306:16, PTPRM|5797:16, RNF5|6048:16, CCL17|6361:16, WASL|8976:16, CYTH1|9267:16, WASF2|10163:16, SH2B2|10603:16, LPHN1|22859:16, GPR124|25960:16, APBB1IP|54518:16, PAG1|55824:16, JAM2|58494:16, EMR3|84658:16, GPR128|84873:16, AMICA1|120425:16, GPR113|165082:16, GPR115|221393:16, GPR111|222611:16, OR10A4|283297:16, GPR133|283383:16, DEFA1B|728358:16, ADRB2|154:15, CDH2|1000:15, CDK5|1020:15, CXCL3|2921:15, LIF|3976:15, SERPINA5|5104:15, PGD|5226:15, PTPRB|5787:15, S100A12|6283:15, CCL19|6363:15, ADAM17|6868:15, TK1|7083:15,

BAG3|9531:15, CEBPZ|10153:15, FBLIM1|54751:15, APP|351:14, AR|367:14, AVP|551:14, BMP7|655:14, ENTPD1|953:14, CDH17|1015:14, CCR7|1236:14, DSPP|1834:14, EP300|2033:14, FCGR2A|2212:14, FCGR2B|2213:14, FHL1|2273:14, GIP|2695:14, GRIA1|2890:14, GRP|2922:14, HSPB1|3315:14, KRT14|3861:14, PBX1|5087:14, PGR|5241:14, PRL|5617:14, PTPRF|5792:14, STAT2|6773:14, ICAM5|7087:14, TLR2|7097:14, FCGR2C|9103:14, MAGED1|9500:14, TM9SF4|9777:14, GPR64|10149:14, LSM4|25804:14, CXCL16|58191:14, ESAM|90952:14, C1QTNF1|114897:14, OPN5|221391:14, ACTN4|81:13, ACTN1|87:13, ADM|133:13, ALOX5|240:13, CA9|768:13, CD59|966:13, CD63|967:13, CLDN7|1366:13, CPOX|1371:13, CSK|1445:13, VCAN|1462:13, ENO2|2026:13, EPHB6|2051:13, GALNS|2588:13, TNFRSF9|3604:13, ITGB3|3690:13, KRT19|3880:13, MITF|4286:13, MST1R|4486:13, MYBL2|4605:13, NOVA2|4858:13, NT5E|4907:13, P2RY2|5029:13, PRH2|5555:13, PRNP|5621:13, PSAP|5660:13, CCL7|6354:13, ST6GAL1|6480:13, ST3GAL4|6484:13, TCF19|6941:13, XRCC5|7520:13, RECK|8434:13, LY6D|8581:13, CLDN1|9076:13, VAMP3|9341:13, MAGI2|9863:13, WDR1|9948:13, PDCD6IP|10015:13, HPSE|10855:13, ADAM28|10863:13, PACSIN2|11252:13, ITGB1BP3|27231:13, LEF1|51176:13, ARFGAP1|55738:13, HM13|81502:13, RASSF5|83593:13, MYST1|84148:13, DAB2IP|153090:13, XIAP|331:12, FASLG|356:12, BRCA1|672:12, CALCA|796:12, CASP1|834:12, CASP4|837:12, RUNX1|861:12, CCND2|894:12, CCNG1|900:12, CCNH|902:12, CD5|921:12, CHEK1|1111:12, CCR6|1235:12, CSF3R|1441:12, CTLA4|1493:12, DSG2|1829:12, FOXO3|2309:12, FLG|2312:12, FLI1|2313:12, HCLS1|3059:12, IL2RB|3560:12, IL6ST|3572:12, ITGB4|3691:12, IVL|3713:12, KRT8|3856:12, LCN2|3934:12, MYO6|4646:12, MYOG|4656:12, PAM|5066:12, PLN|5350:12, PSEN1|5663:12, PTHLH|5744:12, RNASE3|6037:12, RRAD|6236:12, AURKA|6790:12, TNFRSF1B|7133:12, KHSRP|8570:12, ADAM15|8751:12, CCRL2|9034:12, GDF15|9518:12, TNFSF13B|10673:12, MYCBP2|23077:12, SIGLEC7|27036:12, GIT1|28964:12, PDLIM2|64236:12, NUP37|79023:12, PDCD1LG2|80380:12, MUC16|94025:12, CCR2|729230:12, AMH|268:11, FAS|355:11, ARSA|410:11, AZU1|566:11, CREB1|1385:11, CTSK|1513:11, CTSZ|1522:11, DAG1|1605:11, FUT4|2526:11, FUT5|2527:11, FUT6|2528:11, FUT7|2529:11, GCHFR|2644:11, GRPR|2925:11, GUK1|2987:11, KDR|3791:11, KRT7|3855:11, LGALS9|3965:11, EPCAM|4072:11, MIF|4282:11, MNAT1|4331:11, OMP|4975:11, PAK1|5058:11, PENK|5179:11, CFP|5199:11, PPARD|5467:11, RPA3|6119:11, S100A9|6280:11, CXCL6|6372:11, CXCL5|6374:11, SLC6A3|6531:11, SOD3|6649:11, TLR4|7099:11, BCAR3|8412:11, CDK5R1|8851:11, STK19|8859:11, MYL9|10398:11, FUT9|10690:11, SUB1|10923:11, ROBLD3|28956:11, UPK3B|80761:11, POLDIP3|84271:11, CDCA5|113130:11, AKR1B1|231:10, AMBN|258:10, BAD|572:10, SCARB1|949:10, CHAT|1103:10, CCR5|1234:10, DLG4|1742:10, DPYS|1807:10, F2RL1|2150:10, FGFR2|2263:10, FPR1|2357:10, GJB1|2705:10, NRG1|3084:10, IGF1R|3480:10, IK|3550:10, IL1RN|3557:10, IL15|3600:10, IMPA1|3612:10, KRT15|3866:10, KRT16|3868:10, PGC|5225:10, PLD2|5338:10, RAF1|5894:10, ST14|6768:10, TACR1|6869:10, TNNC1|7134:10, TUFM|7284:10, NCOA3|8202:10, BRAP|8315:10, PIAS1|8554:10, TNFSF10|8743:10, SH2D3C|10044:10, SH2D3A|10045:10, RAPGEF3|10411:10, NLRP1|22861:10, ANGPTL3|27329:10, CTNNBL1|56259:10, MARCKSL1|65108:10, MDGA1|266727:10, CLEC4D|338339:10,

ACP2|53:9, AFP|174:9, AKT2|208:9, ANK1|286:9, AQP1|358:9, STS|412:9,
BAG1|573:9, CD6|923:9, CD37|951:9, ENTPD2|954:9, CYP19A1|1588:9,
DAB2|1601:9, DPT|1805:9, F11|2160:9, FN1|2335:9, GAD2|2572:9, GFI1|2672:9,
GNRH1|2796:9, HIF1A|3091:9, HMOX1|3162:9, IGFBP3|3486:9, LHX1|3975:9,
LPL|4023:9, MATK|4145:9, MEN1|4221:9, MST1|4485:9, NPPA|4878:9, OXT|5020:9,
PLCG1|5335:9, PPBP|5473:9, PRKAA2|5563:9, PRKAB1|5564:9, RAP1GAP|5909:9,
RARB|5915:9, SAFB|6294:9, SS18|6760:9, STAT4|6775:9, SULT2A1|6822:9,
SYP|6855:9, TFAP2A|7020:9, UVRAG|7405:9, TRIM26|7726:9, SCG2|7857:9,
TP63|8626:9, DYNLL1|8655:9, TNFSF14|8740:9, PCSK7|9159:9, IL1RL1|9173:9,
ZBED1|9189:9, PDLIM7|9260:9, CD83|9308:9, PRG4|10216:9, CIB1|10519:9,
CXCR6|10663:9, 40430|10801:9, CKAP4|10970:9, ECD|11319:9, OSBPL3|26031:9,
INVS|27130:9, LSM1|27257:9, PYCARD|29108:9, LAP3|51056:9, NCKIPSD|51517:9,
TRPM7|54822:9, SMPD3|55512:9, CENPJ|55835:9, ANKH|56172:9, CXCR7|57007:9,
SCYL1|57410:9, DSCAML1|57453:9, TMPRSS13|84000:9, ABHD14B|84836:9,
DMBX1|127343:9, OPN1SW|611:8, BMP1|649:8, SERPINH1|871:8, CD48|962:8,
COL4A3|1285:8, COL8A1|1295:8, COL17A1|1308:8, CLDN3|1365:8, DOCK2|1794:8,
EPB41|2035:8, FLT3LG|2323:8, FLT4|2324:8, FYN|2534:8, GAP43|2596:8,
NR3C1|2908:8, HSPA8|3312:8, ITGAL|3683:8, LAD1|3898:8, LOXL2|4017:8,
LTB|4050:8, SMAD7|4092:8, MAP3K5|4217:8, MMP8|4317:8, MMP12|4321:8,
MYBPC1|4604:8, NCL|4691:8, NFIC|4782:8, PDE3B|5140:8, PDYN|5173:8,
PLEK|5341:8, SERPINF2|5345:8, PNMT|5409:8, PRCP|5547:8, PTN|5764:8,
ROCK1|6093:8, SHC1|6464:8, SON|6651:8, SOS2|6655:8, SPTAN1|6709:8,
SYK|6850:8, VEGFC|7424:8, TRIP10|9322:8, SH3BP5|9467:8, SDC3|9672:8,
SLK|9748:8, LRIG2|9860:8, TROAP|10024:8, CD96|10225:8, TACC3|10460:8,
CCT4|10575:8, NES|10763:8, FRS2|10818:8, LYVE1|10894:8, EHD1|10938:8,
CAPN10|11132:8, SEC14L2|23541:8, GCA|25801:8, ADAMDEC1|27299:8,
FXYD5|53827:8, SMOX|54498:8, STAP2|55620:8, SMYD3|64754:8, ACD|65057:8,
FERMT3|83706:8, UCN3|114131:8, EMB|133418:8, ACAT1|38:7, ADAM8|101:7,
APOA1|335:7, BDNF|627:7, CCKBR|887:7, CYP17A1|1586:7, E2F1|1869:7,
EVC|2121:7, FANCC|2176:7, FIGF|2277:7, GATA1|2623:7, MSH6|2956:7,
HAL|3034:7, HLA-G|3135:7, HOXB9|3219:7, IGFALS|3483:7, IL1R1|3554:7,
ITGAX|3687:7, STT3A|3703:7, KNG1|3827:7, LHCGR|3973:7, LMO7|4008:7,
MAP2|4133:7, SMCP|4184:7, MGAT5|4249:7, NEB|4703:7, NFIA|4774:7,
NFE2|4778:7, NFIB|4781:7, NFIX|4784:7, NME1|4830:7, ODC1|4953:7, PEBP1|5037:7,
PGF|5228:7, PHB|5245:7, KLK7|5650:7, PTH|5741:7, CLIP1|6249:7, CLEC11A|6320:7,
SIPA1|6494:7, SLA|6503:7, SOD1|6647:7, TBXAS1|6916:7, TFRC|7037:7,
TNFRSF1A|7132:7, TSC2|7249:7, TYRP1|7306:7, RNF112|7732:7, ADAM12|8038:7,
COIL|8161:7, ASAP2|8853:7, AIP|9049:7, PNMA1|9240:7, CYTIP|9595:7,
ABI1|10006:7, MYBBP1A|10514:7, UTS2|10911:7, ADRM1|11047:7,
AKAP13|11214:7, EPB41L3|23136:7, SH3KBP1|30011:7, SEPSECS|51091:7,
AURKAIP1|54998:7, WDR48|57599:7, GPR172A|79581:7, HDAC11|79885:7,
SFXN1|94081:7, LRIG3|121227:7, PTRH1|138428:7, ZFPM1|161882:7,
ACER2|340485:7, AGXT|189:6, ANXA6|309:6, APCS|325:6, ARRB2|409:6,
BMP4|652:6, BRCA2|675:6, BRS3|680:6, BTC|685:6, BYSL|705:6, CDC25A|993:6,
CDH16|1014:6, CEACAM7|1087:6, PLK3|1263:6, CPE|1363:6, CRH|1392:6,
CRKL|1399:6, CUX1|1523:6, DSC2|1824:6, DSC3|1825:6, DUSP1|1843:6,

CTTN|2017:6, EPO|2056:6, FBN1|2200:6, HNF4A|3172:6, HPD|3242:6, IBSP|3381:6, IGF2|3481:6, CYR61|3491:6, IKBKB|3551:6, IL4R|3566:6, ITGAM|3684:6, LOR|4014:6, LTK|4058:6, SH2D1A|4068:6, MMP19|4327:6, MTR|4548:6, NEFL|4747:6, PDE4D|5144:6, PF4|5196:6, PLD1|5337:6, PNN|5411:6, PSEN2|5664:6, RALGDS|5900:6, REN|5972:6, MAP2K4|6416:6, SHBG|6462:6, ST3GAL1|6482:6, SMPD2|6610:6, SSTR2|6752:6, SSTR3|6753:6, STK10|6793:6, TFF1|7031:6, TGFB2|7042:6, TNR|7143:6, TRAF2|7186:6, SLC35A2|7355:6, VIP|7432:6, VSNL1|7447:6, WHSC1|7468:6, WT1|7490:6, MIA|8190:6, EPX|8288:6, WISP3|8838:6, WISP2|8839:6, MAP3K13|9175:6, AKAP5|9495:6, GAL3ST1|9514:6, GNE|10020:6, PEMT|10400:6, STK25|10494:6, SPTLC1|10558:6, KHDRBS1|10657:6, AVIL|10677:6, PLK2|10769:6, CCR9|10803:6, C1QL1|10882:6, CORO1C|23603:6, EIF2C2|27161:6, STOML2|30968:6, UGT1A10|54575:6, UGT1A8|54576:6, UGT1A7|54577:6, UGT1A6|54578:6, UGT1A5|54579:6, UGT1A9|54600:6, UGT1A4|54657:6, UGT1A1|54658:6, UGT1A3|54659:6, CASZ1|54897:6, UGGT1|56886:6, PAK6|56924:6, GOPC|57120:6, PAK7|57144:6, MICAL1|64780:6, TLN2|83660:6, HOPX|84525:6, NT5C1A|84618:6, OR2D2|120776:6, KCTD11|147040:6, ASAH2B|653308:6, ACP1|52:5, ADCY6|112:5, AGTR2|186:5, ALOX5AP|241:5, APOC1|341:5, APOC2|344:5, ATP7B|540:5, BID|637:5, BLM|641:5, BMP5|653:5, BMP6|654:5, C4BPA|722:5, CA8|767:5, RUNX2|860:5, CDA|978:5, CDKN1B|1027:5, CDS1|1040:5, CFL1|1072:5, CORT|1325:5, CR2|1380:5, CTNNA1|1495:5, CYP26A1|1592:5, DAP|1611:5, DBH|1621:5, DEFB4|1673:5, CFD|1675:5, DLAT|1737:5, DLX4|1748:5, DOCK3|1795:5, EMD|2010:5, ENPEP|2028:5, EPB41L1|2036:5, STX2|2054:5, ERBB3|2065:5, F12|2161:5, FCGR1A|2209:5, FEN1|2237:5, DARC|2532:5, FYB|2533:5, GAD1|2571:5, GJB3|2707:5, CXCL2|2920:5, HAS2|3037:5, HAS3|3038:5, HMMR|3161:5, NR4A1|3164:5, HOXB8|3218:5, IAPP|3375:5, ITGA3|3675:5, LAMC2|3918:5, LOXL1|4016:5, LSAMP|4045:5, LTBP2|4053:5, MMP3|4314:5, MMP7|4316:5, MPG|4350:5, NAP1L1|4673:5, NELL1|4745:5, PITX1|5307:5, PKD1|5310:5, PMP22|5376:5, PPY|5539:5, PRKCB|5579:5, PRKCG|5582:5, PRKD1|5587:5, MAPK7|5598:5, PTAFR|5724:5, PTGDS|5730:5, PTGS1|5742:5, PXMP2|5827:5, RBP3|5949:5, RNASE2|6036:5, RPL22|6146:5, RREB1|6239:5, S100A8|6279:5, MAPK12|6300:5, SERPINB3|6317:5, SGCG|6445:5, SST|6750:5, VAMP2|6844:5, TAF1|6872:5, TGFBR2|7048:5, TXN|7295:5, UBE2I|7329:5, SUMO1|7341:5, UTRN|7402:5, VAV2|7410:5, PSCA|8000:5, SYMPK|8189:5, NCK2|8440:5, TTF2|8458:5, DENR|8562:5, TRADD|8717:5, NRP1|8829:5, CLDN2|9075:5, ITGB1BP1|9270:5, RBX1|9978:5, OPTN|10133:5, SORBS3|10174:5, SORBS1|10580:5, MYL12A|10627:5, IQGAP2|10788:5, CYSLTR1|10800:5, IMMT|10989:5, CORO1A|11151:5, PKP3|11187:5, KLF8|11279:5, FNDC3A|22862:5, SWAP70|23075:5, CABIN1|23523:5, LDLRAP1|26119:5, ANKRD1|27063:5, HPGDS|27306:5, CTNNA3|29119:5, NTM|50863:5, C2ORF28|51374:5, ADAM22|53616:5, SIRPG|55423:5, PPP1R9A|55607:5, FERMT1|55612:5, CCAR1|55749:5, C11ORF75|56935:5, AKTIP|64400:5, MMEL1|79258:5, RNF34|80196:5, EPC1|80314:5, ANTXR1|84168:5, NANOS2|339345:5, SFTPA1|653509:5, SFTPA2|729238:5, ACHE|43:4, ACTG1|71:4, ALPP|250:4, ANK3|288:4, BCL9|607:4, C1QBP|708:4, CAST|831:4, RUNX1T1|862:4, CBFB|865:4, CBR1|873:4, CD53|963:4, CD72|971:4, CD74|972:4, CDC42|998:4, CDH13|1012:4, CD52|1043:4, CHGA|1113:4, CHI3L1|1116:4, CIRBP|1153:4,

CNR1|1268:4, DAXX|1616:4, DEFB1|1672:4, DNM2|1785:4, DUSP2|1844:4,
DVL2|1856:4, DVL3|1857:4, EDA|1896:4, EIF4EBP1|1978:4, ETV6|2120:4,
FCGR3B|2215:4, FDPS|2224:4, FER|2241:4, FES|2242:4, FGA|2243:4, FGB|2244:4,
XRCC6|2547:4, GH1|2688:4, GRIN2B|2904:4, HEXA|3073:4, HEXB|3074:4,
HINT1|3094:4, HRAS|3265:4, AGFG1|3267:4, HSPA5|3309:4, KLC1|3831:4,
KRT5|3852:4, LCT|3938:4, LEPR|3953:4, MAOA|4128:4, MME|4311:4,
MMP16|4325:4, MT1E|4493:4, MT3|4504:4, MTRR|4552:4, MVD|4597:4,
MYB|4602:4, MYBL1|4603:4, MYH11|4629:4, NAGLU|4669:4, NBN|4683:4,
NCAM2|4685:4, NF1|4763:4, NME3|4832:4, NTS|4922:4, PIP|5304:4, POR|5447:4,
PSPN|5623:4, PSPH|5723:4, RBMS2|5939:4, RHCE|6006:4, S100A6|6277:4,
SRL|6345:4, CCL22|6367:4, ST3GAL3|6487:4, SLC25A1|6576:4, STXBP3|6814:4,
TBP|6908:4, NKX2-1|7080:4, TJP1|7082:4, TRAF3|7187:4, TTF1|7270:4,
UGCG|7357:4, WAS|7454:4, NR0B2|8431:4, IRS2|8660:4, RIPK1|8737:4,
ADAM23|8745:4, TNFRSF11A|8792:4, CDK5R2|8941:4, MAP3K14|9020:4,
PIAS2|9063:4, ATP6V0D1|9114:4, PTTG1|9232:4, PLAA|9373:4, MAP4K4|9448:4,
IQSEC1|9922:4, HDAC6|10013:4, BCL2L11|10018:4, HRSP12|10247:4,
CLEC4M|10332:4, BPNT1|10380:4, SPAG11B|10407:4, TM9SF1|10548:4,
MGEA5|10724:4, TUBGCP2|10844:4, TPPP|11076:4, PTPRT|11122:4,
KLRK1|22914:4, DDAH2|23564:4, DDAH1|23576:4, PDSS1|23590:4,
SLC17A5|26503:4, B3GAT1|27087:4, NAAA|27163:4, MCAT|27349:4,
ATAD2|29028:4, ASAP1|50807:4, EXOSC3|51010:4, GAL|51083:4, NT5C3|51251:4,
FZR1|51343:4, C22ORF28|51493:4, TTRAP|51567:4, SPA17|53340:4,
FGFRL1|53834:4, RCBTB1|55213:4, SLC35C1|55343:4, LRRC7|57554:4,
MAGEE1|57692:4, CENPK|64105:4, AHNAK|79026:4, FIP1L1|81608:4,
LOXL4|84171:4, RPAIN|84268:4, PPP1R9B|84687:4, LOXL3|84695:4,
STARD13|90627:4, HN1L|90861:4, ADC|113451:4, SSX2IP|117178:4,
B3GAT2|135152:4, CD109|135228:4, C20ORF70|140683:4, GCET2|257144:4,
ACTA1|58:3, AHSG|197:3, AMT|275:3, APOA2|336:3, AQP3|360:3, ARF1|375:3,
ATP12A|479:3, ATP4A|495:3, BAK1|578:3, BGN|633:3, BSG|682:3, CALD1|800:3,
CAMLG|819:3, CCK|885:3, CD79A|973:3, CDH3|1001:3, CDK2|1017:3, LYST|1130:3,
CNR2|1269:3, COMT|1312:3, CSN3|1448:3, CSNK2A1|1457:3, CSNK2A2|1459:3,
DBP|1628:3, DCN|1634:3, DIAPH2|1730:3, DNM1|1759:3, EIF4E|1977:3,
ELAVL3|1995:3, EPAS1|2034:3, ERBB4|2066:3, ESR2|2100:3, F9|2158:3,
FABP7|2173:3, FASN|2194:3, FGF1|2246:3, ADAM2|2515:3, GC|2638:3, GDF2|2658:3,
GCLC|2729:3, GCLM|2730:3, GP1BB|2812:3, HBB|3043:3, HK1|3098:3, HLA-
A|3105:3, HLA-DMB|3109:3, HLA-DQB1|3119:3, HLA-DRB1|3123:3, HLF|3131:3,
HLCS|3141:3, HSD11B2|3291:3, HTR1A|3350:3, HTR1B|3351:3, IARS|3376:3,
IL2RA|3559:3, INS|3630:3, IRF1|3659:3, KCNA4|3739:3, KLRB1|3820:3,
KPNA1|3836:3, LAMB2|3913:3, LAMP1|3916:3, LIMK1|3984:3, MC1R|4157:3,
MCF2|4168:3, ADAM11|4185:3, MMP15|4324:3, MMP17|4326:3, ABCC1|4363:3,
MUC2|4583:3, NELL2|4753:3, NEUROD1|4760:3, NEUROG1|4762:3,
NFATC1|4772:3, NUCB2|4925:3, OSBP|5007:3, P2RX7|5027:3, P2RY4|5030:3,
PAX2|5076:3, PDC|5132:3, PDGFB|5155:3, PDK1|5163:3, PDPK1|5170:3,
PHEX|5251:3, PKP2|5318:3, PPYR1|5540:3, MAPK11|5600:3, PTCH1|5727:3,
RIT2|6014:3, RNPEP|6051:3, RPS4X|6191:3, S100A10|6281:3, SAA2|6289:3,
SARS|6301:3, CCL13|6357:3, CCL24|6369:3, SDC4|6385:3, SFTPB|6439:3,

SLC19A1|6573:3, SNCA|6622:3, SP3|6670:3, TAGLN|6876:3, TAPBP|6892:3, C2ORF3|6936:3, TEP1|7011:3, TGFBR3|7049:3, TNXB|7148:3, TRIP6|7205:3, TSHR|7253:3, PLA2G7|7941:3, CUL5|8065:3, KDM5D|8284:3, CASK|8573:3, FADD|8772:3, TNFRSF10B|8795:3, NRP2|8828:3, HDAC3|8841:3, SQSTM1|8878:3, SELENBP1|8991:3, LIMD1|8994:3, SEMA5A|9037:3, COX7A2L|9167:3, CD163|9332:3, MAPK8IP1|9479:3, GTPBP1|9567:3, ARHGEF11|9826:3, PAK4|10298:3, TLR6|10333:3, CLEC10A|10462:3, FST|10468:3, PITRM1|10531:3, EDAR|10913:3, FERMT2|10979:3, CD93|22918:3, MAPRE1|22919:3, ACIN1|22985:3, PDZD2|23037:3, PLA2G15|23659:3, FAM155B|27112:3, ANAPC2|29882:3, GLTSCR2|29997:3, ST8SIA3|51046:3, CKLF|51192:3, POMP|51371:3, MAGEC2|51438:3, EVL|51466:3, LSR|51599:3, HECA|51696:3, CD244|51744:3, ATP8A2|51761:3, KRT20|54474:3, STYK1|55359:3, ACSS2|55902:3, MEPE|56955:3, LRRC4C|57689:3, GMCL1|64395:3, GMCL1L|64396:3, RTN4R|65078:3, PLVAP|83483:3, ANGPTL6|83854:3, C9ORF3|84909:3, OPN4|94233:3, NME2P1|283458:3, GPR144|347088:3, LOC100294189|100294189:3, LOC100294468|100294468:3, CRISP1|167:2, ALPI|248:2, AMBP|259:2, AMD1|262:2, AMELX|265:2, DST|667:2, C5AR1|728:2, SLC25A20|788:2, CASR|846:2, KRIT1|889:2, CD8A|925:2, CDH11|1009:2, CETP|1071:2, CFTR|1080:2, CLK1|1195:2, CCR4|1233:2, CRHR1|1394:2, CSF1R|1436:2, DOK1|1796:2, ERCC6|2074:2, EWSR1|2130:2, FLNA|2316:2, FLT3|2322:2, GJB2|2706:2, GRN|2896:2, GYPC|2995:2, HDLBP|3069:2, IGF1|3479:2, IGFBP1|3484:2, RBPJ|3516:2, IDO1|3620:2, PDX1|3651:2, KLRD1|3824:2, KRT1|3848:2, KRT13|3860:2, LAMA3|3909:2, LCK|3932:2, LEP|3952:2, LGALS4|3960:2, LIM2|3982:2, LRP1|4035:2, MAL|4118:2, MAP1A|4130:2, MGAT3|4248:2, MMP13|4322:2, MOG|4340:2, NMBR|4829:2, NTRK2|4915:2, P4HB|5034:2, PALM|5064:2, REG3A|5068:2, PDGFA|5154:2, PDZK1|5174:2, SERPINF1|5176:2, PIGR|5284:2, PKD2|5311:2, PLCD1|5333:2, PLCG2|5336:2, PML|5371:2, SRGN|5552:2, PTPRA|5786:2, RHEB|6009:2, RPA1|6117:2, RPA2|6118:2, SAG|6295:2, SAT1|6303:2, SLC9A1|6548:2, SLN|6588:2, SMARCA2|6595:2, SNCB|6620:2, SNX1|6642:2, SP1|6667:2, SRF|6722:2, SSTR4|6754:2, STAT5A|6776:2, TFF2|7032:2, TFF3|7033:2, THY1|7070:2, TIMP3|7078:2, TLR3|7098:2, CRISP2|7180:2, TSC1|7248:2, YES1|7525:2, SLBP|7884:2, TFEB|7942:2, PDHX|8050:2, SPARCL1|8404:2, SDPR|8436:2, PIR|8544:2, SOCS1|8651:2, MARCO|8685:2, MPDZ|8777:2, TNFRSF10A|8797:2, WISP1|8840:2, MGAM|8972:2, SOCS3|9021:2, PAPSS2|9060:2, PAPSS1|9061:2, PICK1|9463:2, RAPGEF2|9693:2, MVP|9961:2, G3BP1|10146:2, SIRPB1|10326:2, PAPOLA|10914:2, RALBP1|10928:2, BTG3|10950:2, LDB3|11155:2, CHP|11261:2, PDAP1|11333:2, ADNP|23394:2, BRD1|23774:2, BAMBI|25805:2, CCRN4L|25819:2, TPSG1|25823:2, LRIG1|26018:2, AATF|26574:2, LAT|27040:2, REM1|28954:2, NXT1|29107:2, RPA4|29935:2, HEBP1|50865:2, CLDN18|51208:2, TLR8|51311:2, DDX41|51428:2, TLR9|54106:2, TXNL4B|54957:2, MTPAP|55149:2, ZDHHC7|55625:2, RADIL|55698:2, ST6GALNAC1|55808:2, C10ORF2|56652:2, CTNNBIP1|56998:2, CNOT6|57472:2, MKL1|57591:2, LRRC4|64101:2, PDF|64146:2, MAGT1|84061:2, C12ORF57|113246:2, TIRAP|114609:2, RPTN|126638:2, CNTN4|152330:2, NANOS3|342977:2, TRIM67|440730:2, NME1-NME2|654364:2, SSX2B|727837:2, ADA|100:1, ALOX12|239:1, ANK2|287:1, AQP5|362:1, RHOA|387:1, SERPINC1|462:1, RERE|473:1, ATR|545:1, CA4|762:1, CALB1|793:1,

CASP3|836:1, CD27|939:1, CHM|1121:1, CPM|1368:1, HAPLN1|1404:1, DCTN1|1639:1, NQO1|1728:1, DR1|1810:1, SLC26A2|1836:1, TOR1A|1861:1, ELN|2006:1, STOM|2040:1, EPHB4|2050:1, ETV3|2117:1, FCGR3A|2214:1, FGFR3|2261:1, GAPDH|2597:1, GCG|2641:1, GCK|2645:1, GEM|2669:1, GPX1|2876:1, GSN|2934:1, HCK|3055:1, HLA-B|3106:1, HLA-C|3107:1, HLA-DMA|3108:1, HPRT1|3251:1, HSF1|3297:1, IGFBP7|3490:1, JUP|3728:1, KRT10|3858:1, LAMB1|3912:1, SMAD5|4090:1, MAOB|4129:1, MECP2|4204:1, MLN|4295:1, MLL|4297:1, MT2A|4502:1, MTHFR|4524:1, MYCN|4613:1, NINJ1|4814:1, NME4|4833:1, NRL|4901:1, OAS1|4938:1, CLDN11|5010:1, PCBD1|5092:1, SLC26A4|5172:1, PLA2G1B|5319:1, PPOX|5498:1, PRKCSH|5589:1, RARA|5914:1, RARS|5917:1, RASGRF1|5923:1, RP1|6101:1, RTN1|6252:1, SCN1B|6324:1, SFPQ|6421:1, SKP2|6502:1, SNAI2|6591:1, SRI|6717:1, SSTR1|6751:1, SSTR5|6755:1, STXBP1|6812:1, TACR2|6865:1, TBCA|6902:1, TCF3|6929:1, TP73|7161:1, TYROBP|7305:1, UCHL1|7345:1, SLMAP|7871:1, MANF|7873:1, GDF5|8200:1, TKTL1|8277:1, PDXK|8566:1, PEA15|8682:1, TNFRSF25|8718:1, C19ORF2|8725:1, CD164|8763:1, DLGAP1|9229:1, CYTH2|9266:1, KLF4|9314:1, ZFYVE9|9372:1, TM9SF2|9375:1, LIPG|9388:1, CLSTN3|9746:1, MTSS1|9788:1, CHAF1A|10036:1, ACTR2|10097:1, ARFRP1|10139:1, KLF2|10365:1, MERTK|10461:1, LILRB1|10859:1, MAPRE2|10982:1, METAP2|10988:1, DSTN|11034:1, ADAMTS5|11096:1, NUDT6|11162:1, SEC63|11231:1, SEPHS1|22929:1, TPX2|22974:1, RRS1|23212:1, ANGPTL2|23452:1, TNFRSF13B|23495:1, CD2AP|23607:1, LRIT1|26103:1, GREM1|26585:1, CHMP2A|27243:1, STK39|27347:1, SETD2|29072:1, PARD6A|50855:1, APH1A|51107:1, ZBTB7A|51341:1, PBRM1|55193:1, KIRREL|55243:1, DIABLO|56616:1, C8ORF4|56892:1, AKR1B10|57016:1, CD177|57126:1, RTN4|57142:1, HRH4|59340:1, CLSTN2|64084:1, DSG4|147409:1, SERPINA2|390502:1, EIF2AK4|440275:1, NCF1|653361:1.

DNA repair

Gene Symbol|Entrez-Gene ID:Score - NR1H2|7376:2723, MRC1|4360:1953, XRCC1|7515:1602, TP53|7157:1540, ERCC2|2068:1329, ERCC1|2067:987, XPA|7507:987, XPC|7508:967, ATM|472:901, XRCC3|7517:898, OGG1|4968:858, MLH1|4292:854, ERCC4|2072:853, POLB|5423:853, MSH2|4436:840, PCNA|5111:838, ERCC5|2073:754, BRCA1|672:707, RAD51|5888:590, MSH6|2956:568, PRKDC|5591:545, RPA1|6117:523, RPA2|6118:523, RPA3|6119:523, RPA4|29935:523, APEX1|328:511, NBN|4683:496, NLRP2|55655:496, XRCC6|2547:469, MGMT|4255:467, XRCC5|7520:461, ATR|545:453, ANTXR1|84168:453, SERPINA2|390502:453, BRCA2|675:442, H2AFX|3014:436, PMS2|5395:434, PARP1|142:428, XRCC2|7516:417, ERCC6|2074:405, XRCC4|7518:403, CPD|1362:393, ERCC3|2071:391, LIG4|3981:385, MRE11A|4361:353, SSB|6741:300, RAD50|10111:293, RAD52|5893:283, UNG|7374:278, MUTYH|4595:264, FEN1|2237:257, BLM|641:251, MPG|4350:231, UBC|7316:230, MSH3|4437:222, WRN|7486:219, LIG1|3978:214, CYP1A1|1543:194, RAD23B|5887:193, GSTT1|2952:191, PMS1|5378:190, CDKN1A|1026:174, FUS|2521:171, LIG3|3980:167,

AGT|183:160, AGXT|189:160, GSTM1|2944:160, FANCD2|2177:155, ELF4|2000:154, MEFV|4210:154, CYP2E1|1571:149, GSTP1|2950:149, DDB2|1643:147, EPHX1|2052:147, ERCC8|1161:140, EXO1|9156:136, LRIG3|121227:130, POLL|27343:127, NAT2|10:126, APC|324:126, NAA16|79612:126, ABL1|25:122, BCR|613:119, TP53BP1|7158:113, POLH|5429:110, POLQ|10721:110, CHEK2|11200:108, FANCG|2189:101, CCRL2|9034:101, EIF2AK1|27102:101, CCHCR1|54535:101, NQO1|1728:99, NTHL1|4913:96, NEIL1|79661:95, DHFR|1719:93, NHEJ1|79840:92, MLH3|27030:90, PAH|5053:86, DDB1|1642:85, CCNO|10309:84, RAD18|56852:83, RAD51C|5889:80, HPRT1|3251:78, TRIM21|6737:77, TROVE2|6738:77, RAD51L1|5890:76, ESR1|2099:75, REV3L|5980:75, RAD51L3|5892:73, TIMM8A|1678:72, TDG|6996:71, CHEK1|1111:69, MDC1|9656:68, LRIG1|26018:68, APTX|54840:68, MDM2|4193:67, NAT1|9:65, MPO|4353:65, SLC6A2|6530:65, SLC38A3|10991:65, CYP2C19|1557:64, CYP2C9|1559:64, SLC19A1|6573:64, AICDA|57379:64, TERF2|7014:63, TBPL1|9519:63, CYP1A2|1544:61, EP300|2033:61, RAD1|5810:61, RRAD|6236:61, RAD23A|5886:60, GADD45A|1647:57, EGFR|1956:57, SMUG1|23583:57, HMGB1|3146:56, KAT5|10524:55, MGEA5|10724:54, MRPL36|64979:52, BRIP1|83990:52, CDKN2A|1029:51, E2F1|1869:51, NUDT1|4521:51, DCLRE1A|9937:51, ZDHHC2|51201:51, TDP1|55775:51, MTHFR|4524:50, TP73|7161:50, CAT|847:49, CRAT|1384:49, CYP3A4|1576:49, PDXK|8566:49, MBD4|8930:47, KRT12|3859:46, SLC6A3|6531:46, POLM|27434:46, NEIL2|252969:46, CYP1B1|1545:45, CYP2D6|1565:45, TOP1|7150:45, ERVK6|64006:45, EPHB2|2048:44, HAP1|9001:44, FANCM|57697:44, COMT|1312:43, PIK3CA|5290:43, PIK3CB|5291:43, PIK3CD|5293:43, PIK3CG|5294:43, PIK3R1|5295:43, DNASE1|1773:42, FANCA|2175:42, SOD2|6648:42, ACTB|60:41, ADH1B|125:41, ALDH2|217:41, BARD1|580:41, CYP2A6|1548:41, DRD2|1813:41, DRD4|1815:41, FANCC|2176:41, GRPR|2925:41, GSTA4|2941:41, GSTM3|2947:41, SMC1A|8243:41, GSTT2B|653689:41, ETV6|2120:40, XAB2|56949:40, SCARA3|51435:39, DNMT1|1786:38, POLD1|5424:38, PTEN|5728:38, CREBBP|1387:37, NOTCH2|4853:37, DMC1|11144:37, PAG1|55824:37, C17ORF28|283987:37, CCNH|902:36, BAX|581:35, MUS81|80198:35, BRAF|673:34, RBBP8|5932:34, POLI|11201:34, OPN4|94233:33, HIF1A|3091:32, IL6|3569:32, YBX1|4904:32, PTGS2|5743:32, UBE2V2|7336:32, CALM3|808:31, CD4|920:31, JAK2|3717:31, AMD1|262:30, BACH1|571:30, CRK|1398:30, HMGA1|3159:30, ITK|3702:30, SLC22A3|6581:30, TGFB1|7040:30, RECQL4|9401:30, GRAP2|9402:30, AHSA1|10598:30, POLDIP2|26073:30, EGF|1950:29, MAPK3|5595:29, BCL2|596:28, CYP3A5|1577:28, NR3C1|2908:28, KRAS|3845:28, DCLRE1C|64421:28, CANT1|124583:28, CCND1|595:27, PML|5371:27, TNF|7124:27, PRPF19|27339:27, CD34|947:26, CYP17A1|1586:26, IGFALS|3483:26, RPE|6120:26, IFI44|10561:26, RRM2B|50484:26, FHIT|2272:25, PTTG1|9232:25, PARP2|10038:25, PARP3|10039:25, CLEC10A|10462:25, RTEL1|51750:25, C1QBP|708:24, CD74|972:24, SOD1|6647:24, UBE2A|7319:24, UBE2N|7334:24, TMPRSS11D|9407:24, REV1|51455:24, ZNF350|59348:24, CENPK|64105:24, CTNNB1|1499:23, CYP11B2|1585:23, IFI27|3429:23, IK|3550:23, PLAT|5327:23, UCN|7349:23, TRRAP|8295:23, SDS|10993:23, GAL|51083:23, GTF2H5|404672:23, CDK2|1017:22, ING1|3621:22, JUN|3725:22, NOS2|4843:22, TBP|6908:22, TGFA|7039:22, SHFM1|7979:22,

CUL4A|8451:22, ACAT1|38:21, DDR1|780:21, CDK7|1022:21, PAM|5066:21, POLR2B|5431:21, SFRS1|6426:21, RNF8|9025:21, CHAF1A|10036:21, MYCBP2|23077:21, FANCL|55120:21, MTX1|4580:20, PNKP|11284:20, PLEKHB1|58473:20, EME1|146956:20, ATF4|468:19, CD9|928:19, HMGN1|3150:19, LPO|4025:19, PNN|5411:19, ST3GAL4|6484:19, TGFBR2|7048:19, TOPBP1|11073:19, TPPP|11076:19, ATRIP|84126:19, ADA|100:18, AKT1|207:18, DUT|1854:18, GTF3A|2971:18, NTS|4922:18, SYCP1|6847:18, SMC3|9126:18, POLK|51426:18, ALK|238:17, APRT|353:17, MTOR|2475:17, IL1B|3553:17, ATXN3|4287:17, NPM1|4869:17, RAD17|5884:17, UBE2B|7320:17, TLK1|9874:17, MCPH1|79648:17, CA1|759:16, RECQL|5965:16, PIPOX|51268:16, CYCS|54205:16, CLSPN|63967:16, CETN2|1069:15, CPA1|1357:15, CSE1L|1434:15, CTNND1|1500:15, DNMT3A|1788:15, EZH2|2146:15, HLA-DMA|3108:15, IGFBP3|3486:15, PDC|5132:15, PGR|5241:15, REG1A|5967:15, SETMAR|6419:15, TXN|7295:15, PARG|8505:15, BCAR1|9564:15, CASP8AP2|9994:15, SUB1|10923:15, PCSK4|54760:15, TRIM39|56658:15, DCLRE1B|64858:15, PALB2|79728:15, PARP4|143:14, ANXA5|308:14, CCNA2|890:14, ATF2|1386:14, DNMT3B|1789:14, DNTT|1791:14, FLT3|2322:14, IL3|3562:14, IRS1|3667:14, NOTCH1|4851:14, PI3|5266:14, ZNF143|7702:14, MSC|9242:14, TNIP1|10318:14, SPO11|23626:14, BTBD12|84464:14, BIN1|274:13, CDKN2D|1032:13, CSF2|1437:13, F9|2158:13, FAP|2191:13, HSPE1|3336:13, IL10|3586:13, KITLG|4254:13, NR4A2|4929:13, PTCH1|5727:13, THRA|7067:13, TIA1|7072:13, WEE1|7465:13, MBD2|8932:13, GLMN|11146:13, RBM17|84991:13, SFXN1|94081:13, CREB1|1385:12, FGFR1|2260:12, GRK4|2868:12, HGF|3082:12, IFNG|3458:12, NF1|4763:12, PCYT1A|5130:12, PLK1|5347:12, CCL2|6347:12, TCEA1|6917:12, HMGA2|8091:12, AXIN2|8313:12, PAPD7|11044:12, TERF2IP|54386:12, SMOX|54498:12, N4BP2|55728:12, AHR|196:11, BID|637:11, CDK4|1019:11, DACH1|1602:11, DBP|1628:11, ESR2|2100:11, GAA|2548:11, GC|2638:11, SFN|2810:11, HRAS|3265:11, MAD2L1|4085:11, MCM2|4171:11, MCM7|4176:11, MEN1|4221:11, MYB|4602:11, NPY1R|4886:11, NTRK3|4916:11, PTPN1|5770:11, RAG1|5896:11, SRD5A2|6716:11, TFAM|7019:11, TNR|7143:11, SUMO1|7341:11, VHL|7428:11, YY1|7528:11, ZFP36|7538:11, HIST2H4A|8370:11, TTF2|8458:11, CSDA|8531:11, CDC123|8872:11, SGPL1|8879:11, DNAJA3|9093:11, ARL6IP5|10550:11, MORF4L1|10933:11, SYNM|23336:11, REXO2|25996:11, SYCP3|50511:11, DEFA1B|728358:11, APP|351:10, ARSE|415:10, ATP7A|538:10, COX11|1353:10, DYRK1A|1859:10, EPHA3|2042:10, HBB|3043:10, HFE|3077:10, MAG|4099:10, MYC|4609:10, POR|5447:10, RASGRF1|5923:10, RRM1|6240:10, SMARCA4|6597:10, TEP1|7011:10, VEGFA|7422:10, KIAA0101|9768:10, POLG2|11232:10, TREX1|11277:10, APEX2|27301:10, UIMC1|51720:10, KCNH7|90134:10, SYCE2|256126:10, ATRX|546:9, BAG1|573:9, RUNX1|861:9, CEBPG|1054:9, CFL1|1072:9, CLK1|1195:9, CTBS|1486:9, DLD|1738:9, FGF10|2255:9, GSTA1|2938:9, GYPA|2993:9, HDAC2|3066:9, NR4A1|3164:9, IRF4|3662:9, MLL|4297:9, MMP9|4318:9, PRH2|5555:9, PSAP|5660:9, RAG2|5897:9, RARB|5915:9, PRPH2|5961:9, RFC2|5982:9, ROS1|6098:9, SAFB|6294:9, SMARCA2|6595:9, SMARCC1|6599:9, PRDX2|7001:9, TERT|7015:9, TSC2|7249:9, UBTF|7343:9, UTRN|7402:9, RAD54L|8438:9, SMARCA5|8467:9, DENR|8562:9, STK19|8859:9, NR2E3|10002:9, CEBPZ|10153:9, TRIM28|10155:9, RPIA|22934:9, ZMYND8|23613:9, LDHD|197257:9,

ABP1|26:8, ANXA6|309:8, DST|667:8, CCNB1|891:8, CD40|958:8, CD44|960:8, CENPA|1058:8, CYP19A1|1588:8, E2F4|1874:8, GAPDH|2597:8, GGT1|2678:8, GZMB|3002:8, HSD17B4|3295:8, INSRR|3645:8, MSH4|4438:8, MVD|4597:8, POLE|5426:8, PPYR1|5540:8, MAP2K1|5604:8, RAD21|5885:8, RET|5979:8, RPS6KB1|6198:8, S100A9|6280:8, S100A11|6282:8, SHBG|6462:8, SOAT1|6646:8, TOP3A|7156:8, TYMS|7298:8, UBE3A|7337:8, XDH|7498:8, VAMP8|8673:8, CCNA1|8900:8, CCNB2|9133:8, INSL6|11172:8, SIRT1|23411:8, CABIN1|23523:8, RAD54B|25788:8, MCAT|27349:8, ROBLD3|28956:8, RIF1|55183:8, CHFR|55743:8, C1ORF103|55791:8, RMI1|80010:8, STS|412:7, CCNG1|900:7, CD40LG|959:7, CRYZ|1429:7, CSNK2A1|1457:7, CSNK2A2|1459:7, DDIT3|1649:7, EEF1A1|1915:7, EGR1|1958:7, PDIA3|2923:7, HLX|3142:7, LCK|3932:7, CXCL9|4283:7, MSH5|4439:7, MTR|4548:7, NCL|4691:7, NME1|4830:7, NOVA2|4858:7, PPARG|5468:7, PRIM2|5558:7, MAPK8|5599:7, RBL2|5934:7, RFC1|5981:7, RPS3|6188:7, CXCL11|6373:7, SLC1A5|6510:7, TAF12|6883:7, TGFB2|7042:7, DYSF|8291:7, RIPK1|8737:7, THOC4|10189:7, STUB1|10273:7, MYBBP1A|10514:7, POLD3|10714:7, NUPR1|26471:7, BBC3|27113:7, A1CF|29974:7, GMNN|51053:7, POMP|51371:7, SCYL1|57410:7, LIN9|286826:7, SNHG3-RCC1|751867:7, ACVR2A|92:6, AFM|173:6, AFP|174:6, APAF1|317:6, FAS|355:6, AR|367:6, CDKN2B|1030:6, CHKA|1119:6, LYST|1130:6, CKM|1158:6, CPOX|1371:6, DAP|1611:6, DCC|1630:6, ERBB2|2064:6, FASN|2194:6, FXN|2395:6, KAT2A|2648:6, GNAS|2778:6, HMOX1|3162:6, HNRNPK|3190:6, HNRNPU|3192:6, HOXB7|3217:6, AGFG1|3267:6, HSPA5|3309:6, IAPP|3375:6, IL6ST|3572:6, LHCGR|3973:6, NEFM|4741:6, NFYA|4800:6, NOV|4856:6, PKD1|5310:6, PLXNA1|5361:6, PRKD1|5587:6, PTPRC|5788:6, RB1|5925:6, RFC4|5984:6, RFX1|5989:6, RNH1|6050:6, SKIL|6498:6, AURKA|6790:6, TLR1|7096:6, UBA1|7317:6, WT1|7490:6, TRIM26|7726:6, ZMYM2|7750:6, PPM1D|8493:6, ESPL1|9700:6, NDC80|10403:6, PPP1R13L|10848:6, NLRP1|22861:6, MMRN1|22915:6, GRHL1|29841:6, TEX15|56154:6, AKR1B10|57016:6, THOC2|57187:6, RPAIN|84268:6, LRRC3B|116135:6, GCET2|257144:6, ALPP|250:5, APCS|325:5, ABCC6|368:5, OPN1SW|611:5, BLMH|642:5, C4BPA|722:5, CA3|761:5, CALML3|810:5, CD33|945:5, CDK1|983:5, CDC25A|993:5, CDKN1B|1027:5, CNP|1267:5, CPE|1363:5, CTH|1491:5, FDPS|2224:5, FES|2242:5, DARC|2532:5, GART|2618:5, GAS1|2619:5, GH1|2688:5, GPX1|2876:5, GTF2H1|2965:5, GYPC|2995:5, HBE1|3046:5, PRMT1|3276:5, IARS|3376:5, IGF2R|3482:5, IL1RN|3557:5, KARS|3735:5, SH2D1A|4068:5, MAPT|4137:5, MET|4233:5, MMP3|4314:5, MT2A|4502:5, MT3|4504:5, MUT|4594:5, NQO2|4835:5, NRAS|4893:5, PIP|5304:5, PRNP|5621:5, PTH|5741:5, RAF1|5894:5, SP1|6667:5, SYT1|6857:5, TAT|6898:5, NKX2-1|7080:5, TSNAX|7257:5, TTF1|7270:5, UBE2I|7329:5, USP1|7398:5, UVRAG|7405:5, CSRP3|8048:5, KDM5D|8284:5, TP63|8626:5, TNFSF10|8743:5, TNFSF9|8744:5, PLAA|9373:5, UBA2|10054:5, C4ORF6|10141:5, BPNT1|10380:5, CKAP4|10970:5, PHB2|11331:5, COTL1|23406:5, VSIG2|23584:5, NIPBL|25836:5, DLL1|28514:5, WRAP53|55135:5, NEIL3|55247:5, SIL1|64374:5, WNK1|65125:5, SYCE1|93426:5, PTRH1|138428:5, GPR180|160897:5, CLEC4D|338339:5, ADM|133:4, XIAP|331:4, APOBEC1|339:4, BDNF|627:4, CDC25C|995:4, CMA1|1215:4, CYC1|1537:4, DMXL1|1657:4, E2F3|1871:4, EDN1|1906:4, FABP7|2173:4, FOXO1|2308:4, GTF2B|2959:4, HBD|3045:4,

HDAC1|3065:4, UBE2K|3093:4, HLA-A|3105:4, HNMT|3176:4, HSF1|3297:4, KPNA2|3838:4, MBP|4155:4, MC1R|4157:4, MOS|4342:4, MUC1|4582:4, MUC2|4583:4, NPPA|4878:4, ODC1|4953:4, PAX3|5077:4, PAX7|5081:4, POMC|5443:4, PPARD|5467:4, PRG2|5553:4, PRSS1|5644:4, PURA|5813:4, RFC3|5983:4, RPL17|6139:4, RRAS|6237:4, ATXN7|6314:4, MAP2K4|6416:4, SFTPC|6440:4, SLC25A1|6576:4, SRY|6736:4, TERF1|7013:4, TNP2|7142:4, TPT1|7178:4, TRPC1|7220:4, TSN|7247:4, TYRP1|7306:4, PIR|8544:4, TNKS|8658:4, CDK5R1|8851:4, SMC4|10051:4, PTGES3|10728:4, ACIN1|22985:4, DDX58|23586:4, TFIP11|24144:4, SESN1|27244:4, HIPK2|28996:4, SETD2|29072:4, DONSON|29980:4, ZNRD1|30834:4, DHH|50846:4, SIRT6|51548:4, NGLY1|55768:4, ERVK5|60358:4, PLEKHF1|79156:4, SETD7|80854:4, SLC25A21|89874:4, C12ORF57|113246:4, MAS1L|116511:4, PAOX|196743:4, ZBTB12|221527:4, ADAM8|101:3, ADK|132:3, AGTR1|185:3, ARCN1|372:3, BAK1|578:3, BTC|685:3, CACNB2|783:3, CASP3|836:3, CASP5|838:3, CAV3|859:3, RUNX1T1|862:3, CD6|923:3, CDC34|997:3, CDK6|1021:3, CGA|1081:3, MAPK14|1432:3, CSF1R|1436:3, DAZ1|1617:3, DFFB|1677:3, DMD|1756:3, DPYS|1807:3, ELANE|1991:3, FGF4|2249:3, FGF9|2254:3, FKBP1A|2280:3, FOS|2353:3, GFRA1|2674:3, GLB1|2720:3, HARS|3035:3, HDGF|3068:3, HSPB1|3315:3, ICAM1|3383:3, ID2|3398:3, IGF1R|3480:3, IL8|3576:3, KRT10|3858:3, LOR|4014:3, MCM5|4174:3, MMP1|4312:3, NOS3|4846:3, NTRK1|4914:3, OXT|5020:3, ABCB1|5243:3, MAP2K2|5605:3, PSEN1|5663:3, RAC1|5879:3, SFRS5|6430:3, SHH|6469:3, SLC12A3|6559:3, SMARCC2|6601:3, SPINK2|6691:3, SPP1|6696:3, SULT2A1|6822:3, SYP|6855:3, TALDO1|6888:3, TCF7L2|6934:3, TOP2B|7155:3, TPI1|7167:3, TSC1|7248:3, UBE2E2|7325:3, ZBTB16|7704:3, AXIN1|8312:3, HAT1|8520:3, NCOA1|8648:3, DYNLL1|8655:3, MBTPS1|8720:3, DLEU2|8847:3, BTRC|8945:3, HGS|9146:3, TMSB10|9168:3, CARTPT|9607:3, MAD2L2|10459:3, SMC2|10592:3, NES|10763:3, DBF4|10926:3, PRSS21|10942:3, STIP1|10963:3, KIF2C|11004:3, CHP|11261:3, SIRT2|22933:3, SMG1|23049:3, SMG6|23293:3, MYST4|23522:3, EIF2C2|27161:3, GLTSCR1|29998:3, SMARCAL1|50485:3, NTM|50863:3, RHOF|54509:3, UGT1A7|54577:3, TESC|54997:3, PARL|55486:3, PPCS|79717:3, H2AFB3|83740:3, LRSAM1|90678:3, ARSK|153642:3, FAM84B|157638:3, NRK|203447:3, NANOS3|342977:3, ADARB1|104:2, CRISP1|167:2, AGA|175:2, RERE|473:2, BCL6|604:2, CALB1|793:2, CAPG|822:2, CAV1|857:2, RUNX3|864:2, CD19|930:2, CD68|968:2, CDA|978:2, CLU|1191:2, CRH|1392:2, DUSP1|1843:2, ELAVL3|1995:2, EMP1|2012:2, EPHB6|2051:2, ERV3|2086:2, F12|2161:2, FANCB|2187:2, FANCF|2188:2, FER|2241:2, FMR1|2332:2, FOLH1|2346:2, GAST|2520:2, FUT2|2524:2, GCLM|2730:2, GPI|2821:2, HES1|3280:2, IFNB1|3456:2, IL4|3565:2, IL5|3567:2, IRF3|3661:2, KCNN1|3780:2, TNPO1|3842:2, KRT14|3861:2, LTA|4049:2, MBNL1|4154:2, CD46|4179:2, ABCC1|4363:2, NF2|4771:2, OAZ1|4946:2, REG3A|5068:2, PBX1|5087:2, PEPD|5184:2, PGK1|5230:2, PLEK|5341:2, PRL|5617:2, PTBP1|5725:2, RENBP|5973:2, RNPEP|6051:2, SAT1|6303:2, CCL13|6357:2, TCEB1|6921:2, TCF3|6929:2, TF|7018:2, TGM2|7052:2, THBS1|7057:2, TOP2A|7153:2, VIM|7431:2, XPO1|7514:2, BTG2|7832:2, MANF|7873:2, SLC25A16|8034:2, C21ORF33|8209:2, BRAP|8315:2, IKBKG|8517:2, PIAS1|8554:2, SAP30|8819:2, HDAC3|8841:2, ASAP2|8853:2, PAPSS1|9061:2, PIAS2|9063:2, USP14|9097:2, NPIP|9284:2, BRE|9577:2, NCOR2|9612:2, GDA|9615:2,

THOC1|9984:2, ARFRP1|10139:2, PPIE|10450:2, MERTK|10461:2, CRTAP|10491:2, PAICS|10606:2, HPSE|10855:2, C1QL1|10882:2, PAPOLA|10914:2, PDAP1|11333:2, ARSG|22901:2, SWAP70|23075:2, DDAH2|23564:2, PDSS1|23590:2, BAMBI|25805:2, FBXO8|26269:2, ERVWE1|30816:2, TRAT1|50852:2, TTRAP|51567:2, MED15|51586:2, UGT1A1|54658:2, MTPAP|55149:2, CCL28|56477:2, DIABLO|56616:2, MUC13|56667:2, SPPL2B|56928:2, SLURP1|57152:2, CBX8|57332:2, CLK4|57396:2, PTBP2|58155:2, APOBEC3G|60489:2, FBRS|64319:2, MARCKSL1|65108:2, INTS3|65123:2, SLC25A22|79751:2, TLE6|79816:2, TNKS2|80351:2, QTRT1|81890:2, ZNF830|91603:2, IMP4|92856:2, RNF168|165918:2, CTAG1A|246100:2, ASPM|259266:2, TMED8|283578:2, POLN|353497:2, HERV-FRD|405754:2, JAG1|182:1, FASLG|356:1, SERPINC1|462:1, CACNB1|782:1, TPP1|1200:1, CPM|1368:1, CYP2B6|1555:1, DAPK1|1612:1, DCTD|1635:1, DDC|1644:1, DFFA|1676:1, DTYMK|1841:1, EIF4E|1977:1, EMD|2010:1, ENO1|2023:1, EPAS1|2034:1, F11|2160:1, FAH|2184:1, FLOT2|2319:1, FOSB|2354:1, NR5A2|2494:1, G6PD|2539:1, GALK1|2584:1, GIF|2694:1, GLI2|2736:1, HTT|3064:1, HLF|3131:1, HPD|3242:1, HSPD1|3329:1, HTN1|3346:1, IL2|3558:1, IL18|3606:1, ING2|3622:1, KCNJ5|3762:1, KRT5|3852:1, KRT8|3856:1, MECP2|4204:1, RAB8A|4218:1, MIP|4284:1, MIPEP|4285:1, MAP3K10|4294:1, MMP10|4319:1, MMP11|4320:1, MPST|4357:1, MTRR|4552:1, MYLK|4638:1, NFKBIA|4792:1, POLR2I|5438:1, POU5F1|5460:1, PRIM1|5557:1, PRTN3|5657:1, SLC17A1|6568:1, SPTAN1|6709:1, TAP1|6890:1, TAP2|6891:1, TNFRSF1A|7132:1, TTPA|7274:1, TYR|7299:1, TYRO3|7301:1, UBE2V1|7335:1, UGT8|7368:1, UMOD|7369:1, DEK|7913:1, IKBKAP|8518:1, FADD|8772:1, SQSTM1|8878:1, BAG3|9531:1, CIR1|9541:1, MATR3|9782:1, MTSS1|9788:1, KBTBD10|10324:1, CCL26|10344:1, FST|10468:1, KHDRBS3|10656:1, CPSF4|10898:1, STARD3|10948:1, DCTN3|11258:1, ECD|11319:1, KLRK1|22914:1, SETX|23064:1, CCRN4L|25819:1, POT1|25913:1, UBE2T|29089:1, GLTSCR2|29997:1, FN3K|64122:1, MMS19|64210:1, ACTR5|79913:1, ALPK1|80216:1, ANGPTL6|83854:1, REG4|83998:1, TP53INP1|94241:1, TMEM54|113452:1, FAM84A|151354:1, SEC14L3|266629:1, SCGB1D4|404552:1.

Hypertension

Gene Symbol|Entrez-Gene ID:Score - AGT|183:128, NR3C2|4306:44, EDN1|1906:41, AGTR1|185:37, KNG1|3827:37, SLC33A1|9197:37, NPPA|4878:31, ADM|133:27, SHBG|6462:26, SELENBP1|8991:26, CYP11B1|1584:25, DBP|1628:25, GC|2638:25, ADRB2|154:21, IL6|3569:20, DBH|1621:19, POMC|5443:19, S100A6|6277:19, CYP11B2|1585:16, HSD11B2|3291:16, IRS1|3667:16, PNMT|5409:16, S100A12|6283:16, SLC12A1|6557:16, PEMT|10400:16, YME1L1|10730:16, WNK4|65266:16, CHDH|55349:15, NPY|4852:11, TRH|7200:11, WNK1|65125:10, ADD1|118:9, CRH|1392:9, ICAM1|3383:9, CCL2|6347:9, TNF|7124:9, VIP|7432:9, AGXT|189:8, NR3C1|2908:8, NOS3|4846:7, ADRB1|153:6, VEGFA|7422:6, ADD2|119:5, ADD3|120:5, AVP|551:5, CRP|1401:5, CSRP1|1465:5, EPO|2056:5, GRK4|2868:5, VCAM1|7412:5, XDH|7498:5, EPX|8288:5, CBFA2T2|9139:5, CALM3|808:4, GUCA2A|2980:4, GUCA2B|2981:4, IMPA1|3612:4,

NTS|4922:4, PPARA|5465:4, PTGIS|5740:4, SCNN1B|6338:4, SCNN1G|6340:4, BRAP|8315:4, RAPGEF5|9771:4, C1QL1|10882:4, CABIN1|23523:4, AHSP|51327:4, ANXA13|312:3, EPHB6|2051:3, GH1|2688:3, GRB2|2885:3, PAH|5053:3, RENBP|5973:3, CCL21|6366:3, TPR|7175:3, HPSE|10855:3, ABP1|26:2, ADRBK1|156:2, ANG|283:2, ABCC6|368:2, SLC25A10|1468:2, CYBA|1535:2, EDN2|1907:2, EDN3|1908:2, EDNRB|1910:2, ELANE|1991:2, ENPEP|2028:2, GCK|2645:2, GNB3|2784:2, GRK5|2869:2, HMOX1|3162:2, HPS1|3257:2, HSD11B1|3290:2, ITGA2|3673:2, KLK1|3816:2, MBP|4155:2, OTC|5009:2, SERPINE1|5054:2, REG3A|5068:2, PPIA|5478:2, PRG2|5553:2, MAP4K2|5871:2, SLC12A3|6559:2, TAC1|6863:2, PDE5A|8654:2, ASAP2|8853:2, ADIPOQ|9370:2, PAPOLA|10914:2, PDAP1|11333:2, CIC|23152:2, NNT|23530:2, MTPAP|55149:2, CENPJ|55835:2, PNPLA3|80339:2, APEH|327:1, CPOX|1371:1, DCT|1638:1, DPYS|1807:1, FANCC|2176:1, HMGCR|3156:1, HTR2A|3356:1, KCNMA1|3778:1, MAP2|4133:1, MMP2|4313:1, MMP9|4318:1, PGF|5228:1, PLAT|5327:1, PMCH|5367:1, PTH|5741:1, GNE|10020:1, CEBPZ|10153:1, ARIH1|25820:1, LDLRAP1|26119:1, FLVCR2|55640:1, PTRH1|138428:1.

Obesity

Gene Symbol|Entrez-Gene ID:Score - POMC|5443:79, PPARG|5468:78, TNF|7124:74, NPY|4852:71, GH1|2688:66, GHRL|51738:65, PPARA|5465:61, MC4R|4160:57, IL6|3569:51, HPSE|10855:50, UCP1|7350:45, GCG|2641:34, LIPE|3991:34, ADRB3|155:30, PMCH|5367:30, PNLIP|5406:28, LPL|4023:27, ADRB2|154:25, CRH|1392:25, ESR1|2099:24, C1QL1|10882:24, CARTPT|9607:20, PYY|5697:19, CCL2|6347:19, AGT|183:18, PCSK1|5122:17, ENPP1|5167:17, SHBG|6462:17, MGAM|8972:17, SPAG8|26206:17, CCK|885:16, NR3C1|2908:16, SERPINE1|5054:16, CNR1|1268:15, UCP2|7351:15, NAMPT|10135:15, AGRP|181:14, CRP|1401:14, CSRP1|1465:14, RBP4|5950:14, UCP3|7352:14, APOB|338:11, APOC3|345:11, FASN|2194:11, SOCS3|9021:11, APOA1|335:10, ATM|472:10, CETP|1071:10, CSF2|1437:10, HCRT|3060:10, NOVA2|4858:10, ADD1|118:9, APOE|348:9, CAV1|857:9, CDKN1A|1026:9, CNTF|1270:9, COX7A1|1346:9, DBP|1628:9, EDN1|1906:9, FGF2|2247:9, GC|2638:9, IGFBP3|3486:9, PTEN|5728:9, VEGFA|7422:9, SOCS1|8651:9, BBS4|585:8, CYP19A1|1588:8, CFD|1675:8, IAPP|3375:8, NTS|4922:8, TNFRSF11B|4982:8, PPY|5539:8, SREBF1|6720:8, TNFSF11|8600:8, APOA4|337:7, KLK3|354:7, BBS1|582:7, BBS2|583:7, CSF1|1435:7, FABP2|2169:7, GCK|2645:7, HNF4A|3172:7, HSD11B1|3290:7, IL10|3586:7, PDX1|3651:7, MC3R|4159:7, NEUROD1|4760:7, PPARD|5467:7, HNF1A|6927:7, HNF1B|6928:7, SELENBP1|8991:7, NPEPPS|9520:7, PSAT1|29968:7, CHDH|55349:7, SERPINA12|145264:7, CAT|847:6, CCKAR|886:6, EGF|1950:6, GHSR|2693:6, GLB1|2720:6, LTA|4049:6, NR3C2|4306:6, PON3|5446:6, TIMP1|7076:6, UBC|7316:6, NR1H4|9971:6, EBP|10682:6, AR|367:5, CYP2E1|1571:5, CYP3A4|1576:5, HBEGF|1839:5, IL1B|3553:5, LEP|3952:5, MAOA|4128:5, PBX1|5087:5, SRGN|5552:5, PRL|5617:5, PTPRF|5792:5, TGFB1|7040:5, KIAA0101|9768:5, RAPGEF5|9771:5, SDS|10993:5, BBS5|129880:5, NPB|256933:5, NPW|283869:5, AQP7|364:4, DST|667:4, CALCA|796:4, CAPG|822:4, CD4|920:4, COMT|1312:4, CRAT|1384:4, CRHBP|1393:4,

EPHB2|2048:4, GIP|2695:4, HTR1B|3351:4, HTR2C|3358:4, HTR6|3362:4, LEPR|3953:4, LPA|4018:4, MAOB|4129:4, CD46|4179:4, NOS2|4843:4, NPPA|4878:4, NPY5R|4889:4, OAT|4942:4, PCYT1A|5130:4, RBP1|5947:4, SLC6A3|6531:4, STAT3|6774:4, IRS2|8660:4, PPARGC1A|10891:4, LEPROT|54741:4, ITLN1|55600:4, C1QTNF1|114897:4, PPARGC1B|133522:4, ABCA1|19:3, ACTB|60:3, FAS|355:3, ABCC6|368:3, BMP7|655:3, SCARB1|949:3, CES1|1066:3, CNR2|1269:3, CTSS|1520:3, CYP7A1|1581:3, DLG4|1742:3, GAST|2520:3, GAPDH|2597:3, GNB3|2784:3, HGF|3082:3, HTR1A|3350:3, ICAM1|3383:3, IL7|3574:3, LIPC|3990:3, MT2A|4502:3, NBN|4683:3, PHB|5245:3, PRPH2|5961:3, RHO|6010:3, SLC18A2|6571:3, TCF7L2|6934:3, WNT10B|7480:3, PNPLA4|8228:3, ARHGEF7|8874:3, CBFA2T2|9139:3, CCL4L1|9560:3, GAL|51083:3, MRAP|56246:3, CTNNBL1|56259:3, ERVK6|64006:3, PNPLA3|80339:3, LYPLAL1|127018:3, SIRPA|140885:3, ABP1|26:2, ADRB1|153:2, CRISP1|167:2, AGTR1|185:2, AGTR2|186:2, AHR|196:2, AKT1|207:2, ARCN1|372:2, ARG1|383:2, RERE|473:2, OPN1SW|611:2, BDNF|627:2, CANX|821:2, CCKBR|887:2, CTNNB1|1499:2, DRD2|1813:2, GHRH|2691:2, GZMB|3002:2, IKBKB|3551:2, IL17A|3605:2, LDLR|3949:2, NAGLU|4669:2, NDUFAB1|4706:2, OXT|5020:2, PFKFB3|5209:2, PGR|5241:2, PLAT|5327:2, PON1|5444:2, PON2|5445:2, PTH|5741:2, PTGS2|5743:2, PTPN1|5770:2, MAP4K2|5871:2, RENBP|5973:2, SKIV2L|6499:2, SLC10A2|6555:2, TFRC|7037:2, XDH|7498:2, MANF|7873:2, HMGA2|8091:2, MKKS|8195:2, PIAS1|8554:2, DGAT1|8694:2, TNFRSF11A|8792:2, SLC33A1|9197:2, ADIPOQ|9370:2, H6PD|9563:2, ARFRP1|10139:2, MLYCD|23417:2, MAGEL2|54551:2, CENPJ|55835:2, AKR1B10|57016:2, GOPC|57120:2, PRDM16|63976:2, PTRH1|138428:2, SGMS1|259230:2, ACO1|48:1, ACR|49:1, ACO2|50:1, AMT|275:1, ANPEP|290:1, ARRB1|408:1, ARRB2|409:1, AZU1|566:1, BCKDHA|593:1, BMP1|649:1, C1QBP|708:1, PTTG1IP|754:1, CD36|948:1, CD68|968:1, CD74|972:1, CHUK|1147:1, CPE|1363:1, EEF1A2|1917:1, EPHA1|2041:1, ESD|2098:1, F12|2161:1, FAAH|2166:1, FDFT1|2222:1, GABRA6|2559:1, GGT1|2678:1, GNAS|2778:1, SFN|2810:1, HCLS1|3059:1, HDLBP|3069:1, HMGCR|3156:1, HMGA1|3159:1, HRH1|3269:1, IGF1|3479:1, IHH|3549:1, IL1RN|3557:1, IL2RB|3560:1, IL15|3600:1, KNG1|3827:1, KIF22|3835:1, LBP|3929:1, LTB|4050:1, MAG|4099:1, MAP2|4133:1, MET|4233:1, MPG|4350:1, NOS3|4846:1, NTF3|4908:1, NUCB2|4925:1, OPRL1|4987:1, PCSK2|5126:1, PFKFB1|5207:1, PFKFB2|5208:1, PFKFB4|5210:1, ABCB1|5243:1, SLC25A3|5250:1, POR|5447:1, PRCP|5547:1, PRH2|5555:1, PRKAA2|5563:1, PRKAB1|5564:1, PSAP|5660:1, RAG2|5897:1, S100A6|6277:1, S100A12|6283:1, SAT1|6303:1, SGCA|6442:1, SLC2A4|6517:1, TAT|6898:1, TBCA|6902:1, TFPI|7035:1, TNFRSF1B|7133:1, TUB|7275:1, TYRP1|7306:1, SLBP|7884:1, KHSRP|8570:1, PAPSS2|9060:1, PAPSS1|9061:1, IKBKE|9641:1, MFN2|9927:1, NMU|10874:1, SIRT1|23411:1, PLA2G15|23659:1, BRD1|23774:1, REXO2|25996:1, SIGLEC7|27036:1, SCG3|29106:1, F11R|50848:1, HEBP1|50865:1, RTEL1|51750:1, TRIT1|54802:1, CNO|55330:1, ZNF395|55893:1, ACSS2|55902:1, RETN|56729:1, CENPK|64105:1, MTMR9|66036:1, DGAT2|84649:1, ASAH2B|653308:1, SFTPA1|653509:1, SFTPA2|729238:1.

Schizophrenia

Gene Symbol|Entrez-Gene ID:Score - NRG1|3084:102, COMT|1312:100, DISC1|27185:95, DTNBP1|84062:92, DAOA|267012:75, RGS4|5999:64, NNT|23530:57, BMP1|649:53, PRCP|5547:53, BDNF|627:51, DRD2|1813:50, CFP|5199:48, PRODH|5625:48, EBPL|84650:47, GRM3|2913:42, EP300|2033:35, IL10|3586:35, AKT1|207:32, CSF2|1437:31, TNF|7124:31, CHRNA7|1139:30, CNR1|1268:30, HTR2A|3356:25, NOTCH4|4855:25, DRD3|1814:24, DAO|1610:22, USH1G|124590:22, NOS1AP|9722:21, RTN4|57142:21, GAD1|2571:20, CCKAR|886:18, MYL4|4635:18, NR4A2|4929:18, RELN|5649:18, MLC1|23209:18, C6ORF15|29113:18, IL2|3558:17, ZDHHC8|29801:14, ABP1|26:13, EPHB2|2048:12, DRD4|1815:11, GRIA1|2890:11, GRM5|2915:11, IL6|3569:11, NTF3|4908:11, TPH1|7166:11, CA1|759:10, IL2RB|3560:10, PLA2G4A|5321:10, PPP3CC|5533:10, SLC6A3|6531:10, SLC6A9|6536:10, APOE|348:9, GRIN1|2902:9, GRIN2A|2903:9, IL1B|3553:9, PAH|5053:9, PLA2G1B|5319:9, SLC1A2|6506:9, SLC1A6|6511:9, NRXN1|9378:9, HPSE|10855:9, GPRIN1|114787:9, CREB1|1385:8, DNTT|1791:8, FGA|2243:8, GRIK3|2899:8, IL1RN|3557:8, SYP|6855:8, DGCR6|8214:8, DGCR2|9993:8, CPLX2|10814:8, CPLX1|10815:8, DST|667:7, CDKN1A|1026:7, IFNG|3458:7, IL4|3565:7, MAP2|4133:7, PTPN4|5775:7, CPZ|8532:7, FEZ1|9638:7, LZTS1|11178:7, DBH|1621:6, FGF9|2254:6, HTR2C|3358:6, HTR6|3362:6, IL3|3562:6, NRGN|4900:6, SLC26A4|5172:6, PTGDS|5730:6, PTGS2|5743:6, SLC6A4|6532:6, SYN2|6854:6, TCF4|6925:6, TCF7L2|6934:6, TIMP1|7076:6, CEBPZ|10153:6, DDN|23109:6, DNAJC5|80331:6, ADC|113451:6, APP|351:5, CA2|760:5, CA3|761:5, CNTF|1270:5, CSF2RA|1438:5, CSF2RB|1439:5, GRIK2|2898:5, GRIK5|2901:5, GSK3B|2932:5, IL3RA|3563:5, MAG|4099:5, MTHFR|4524:5, MTR|4548:5, TPT1|7178:5, SYN3|8224:5, PCSK7|9159:5, PPP1R1B|84152:5, PHOX2A|401:4, CCK|885:4, CYP2D6|1565:4, DLAT|1737:4, ERBB3|2065:4, ERF|2077:4, GRIK1|2897:4, GRIK4|2900:4, HAL|3034:4, CXCL10|3627:4, CXCL9|4283:4, NF2|4771:4, NPY|4852:4, OMP|4975:4, PCNT|5116:4, PSPN|5623:4, PSPH|5723:4, REG1A|5967:4, CCL2|6347:4, CCL3|6348:4, CCL24|6369:4, STXBP3|6814:4, TAT|6898:4, PHLDA2|7262:4, PHOX2B|8929:4, HRSP12|10247:4, FST|10468:4, SPDEF|25803:4, TPSG1|25823:4, SRR|63826:4, C20ORF70|140683:4, TAAR6|319100:4, ADCYAP1|116:3, ATF4|468:3, CALB1|793:3, CALB2|794:3, CAT|847:3, CBR1|873:3, CD9|928:3, CDH13|1012:3, CHI3L1|1116:3, CRAT|1384:3, CREBBP|1387:3, CTNNB1|1499:3, CTNND2|1501:3, EFNA5|1946:3, EPB49|2039:3, GAST|2520:3, GRM2|2912:3, HCRT|3060:3, HLA-A|3105:3, IL1A|3552:3, KRT7|3855:3, LTA|4049:3, NCAM1|4684:3, NEDD9|4739:3, PAFAH1B1|5048:3, PDYN|5173:3, PER1|5187:3, PLP1|5354:3, POMC|5443:3, RPA3|6119:3, S100A9|6280:3, SMARCA2|6595:3, SPTAN1|6709:3, TAL1|6886:3, TST|7263:3, RIPK1|8737:3, HDAC3|8841:3, HERC2|8924:3, NRXN3|9369:3, NRXN2|9379:3, AKAP5|9495:3, TRAF4|9618:3, MED12|9968:3, CACNG2|10369:3, CORIN|10699:3, SUB1|10923:3, TPPP|11076:3, MYT1L|23040:3, SIT1|27240:3, ROBLD3|28956:3, PLLP|51090:3, CRNKL1|51340:3, PAG1|55824:3, MUTED|63915:3, PORCN|64840:3, NDEL1|81565:3, ACO1|48:2, ACO2|50:2, ACTC1|70:2, ADORA1|134:2, ADRB3|155:2, NR0B1|190:2, AMH|268:2, APBA2|321:2, CA4|762:2, CD4|920:2, CYP2C19|1557:2, CYP2C9|1559:2, DBP|1628:2, TIMM8A|1678:2, DR1|1810:2,

DUSP2|1844:2, DUSP4|1846:2, EDA|1896:2, ELK3|2004:2, EPHB1|2047:2, ESR1|2099:2, GABRA1|2554:2, GABRA6|2559:2, GABRP|2568:2, GAP43|2596:2, GC|2638:2, GCG|2641:2, GH1|2688:2, GCLC|2729:2, CXCL2|2920:2, HLA-DRB1|3123:2, HPD|3242:2, HRH1|3269:2, HSPA9|3313:2, HTR1A|3350:2, IL6ST|3572:2, IDO1|3620:2, MPI|4351:2, MSRA|4482:2, NEFM|4741:2, NOVA2|4858:2, OGDH|4967:2, PAM|5066:2, PBX1|5087:2, PDGFB|5155:2, MAPK1|5594:2, MAPK3|5595:2, PRL|5617:2, PSEN1|5663:2, RASA1|5921:2, RBP4|5950:2, RHO|6010:2, RTN1|6252:2, S100A10|6281:2, CCL11|6356:2, SHBG|6462:2, SLC6A2|6530:2, SST|6750:2, SYN1|6853:2, TDO2|6999:2, TFCP2|7024:2, TRH|7200:2, TSNAX|7257:2, XDH|7498:2, SELENBP1|8991:2, SYNGR1|9145:2, ZFYVE9|9372:2, QKI|9444:2, PICK1|9463:2, CASP8AP2|9994:2, SDS|10993:2, CHP|11261:2, MYCBP2|23077:2, PLCB1|23236:2, SYNM|23336:2, SULT4A1|25830:2, LRIT1|26103:2, EIF2C2|27161:2, RHOF|54509:2, IL17RD|54756:2, TRIT1|54802:2, SLC25A32|81034:2, COL25A1|84570:2, LIN9|286826:2, ALDH3B1|221:1, FAS|355:1, ARCN1|372:1, STS|412:1, ASIP|434:1, ASPA|443:1, C4A|720:1, C4B|721:1, CACNA1A|773:1, CARS|833:1, CD5|921:1, CD19|930:1, CCR3|1232:1, CP|1356:1, DDC|1644:1, DLG4|1742:1, DPT|1805:1, DRD1|1812:1, EMD|2010:1, ENPEP|2028:1, F12|2161:1, FASN|2194:1, FES|2242:1, FLNB|2317:1, FN1|2335:1, GIF|2694:1, GLUD1|2746:1, GLUD2|2747:1, NR3C1|2908:1, GRM7|2917:1, GRM8|2918:1, HARS|3035:1, HP|3240:1, HTR3A|3359:1, HTR5A|3361:1, IL8|3576:1, IMPA1|3612:1, IPO5|3843:1, LRP1|4035:1, MXD1|4084:1, MBL2|4153:1, MUC1|4582:1, NEB|4703:1, NTS|4922:1, PBX2|5089:1, PDE4B|5142:1, SERPINF2|5345:1, PLXNB1|5364:1, PPIC|5480:1, RHD|6007:1, RRM1|6240:1, SORT1|6272:1, SFRS5|6430:1, SGCA|6442:1, SGTA|6449:1, SRI|6717:1, TAC1|6863:1, TCP1|6950:1, TF|7018:1, TFPI|7035:1, TNR|7143:1, TYR|7299:1, TYRP1|7306:1, XBP1|7494:1, BRAP|8315:1, APOL1|8542:1, USO1|8615:1, HGS|9146:1, ATG5|9474:1, H6PD|9563:1, MVP|9961:1, NXF1|10482:1, MASP2|10747:1, SLC27A5|10998:1, CIT|11113:1, GRIP1|23426:1, SEC14L2|23541:1, GCA|25801:1, GREM1|26585:1, GPSM2|29899:1, A1CF|29974:1, CSAD|51380:1, MED15|51586:1, CCAR1|55749:1, CCL28|56477:1, GJD2|57369:1, NPAS3|64067:1, GMCL1|64395:1, GMCL1L|64396:1, LGALS12|85329:1, CHRFAM7A|89832:1, UCN2|90226:1, RBM45|129831:1.

REFERENCES

1. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25:**25-29.

2. **The Gene Ontology** [http://www.geneontology.org/]

3. **Online Mendelian Inheritance in Man** [http: //www.ncbi.nlm.nih.gov/omim/]

4. Erhardt RA, Schneider R, Blaschke C: **Status of text-mining techniques applied to biomedical text.** *Drug Discov Today* 2006, **11:**315-325.

5. Rebholz-Schuhmann D, Kirsch H, Arregui M, Gaudan S, Riethoven M, Stoehr P: **EBIMed--text crunching to gather facts for proteins from Medline.** *Bioinformatics* 2007, **23:**e237-244.

6. **EBIMed** [http://www.ebi.ac.uk/Rebholz-srv/ebimed/index.jsp]

7. Tsuruoka Y, Tsujii J, Ananiadou S: **FACTA: a text search engine for finding associated biomedical concepts.** *Bioinformatics* 2008, **24:**2559-2560.

8. Hur J, Schuyler AD, States DJ, Feldman EL: **SciMiner: web-based literature mining tool for target identification and functional enrichment analysis.** *Bioinformatics* 2009, **25:**838-840.

9. **GAD** [http://geneticassociationdb.nih.gov/]

10. Becker KG, Hosack DA, Dennis G, Jr., Lempicki RA, Bright TJ, Cheadle C, Engel J: **PubMatrix: a tool for multiplex literature mining.** *BMC Bioinformatics* 2003, **4:**61.

11. Jenssen TK, Laegreid A, Komorowski J, Hovig E: **A literature network of human genes for high-throughput analysis of gene expression.** *Nat Genet* 2001, **28:**21-28.

12. Plake C, Schiemann T, Pankalla M, Hakenberg J, Leser U: **AliBaba: PubMed as a graph.** *Bioinformatics* 2006, **22:**2444-2445.

13. **AliBaba** [http://alibaba.informatik.hu-berlin.de/]

14. Rebholz-Schuhmann D, Kirsch H, Arregui M, Gaudan S, Rynbeek M, Stoehr P: **Protein annotation by EBIMed.** *Nat Biotechnol* 2006, **24:**902-903.

15. **FACTA** [http://text0.mib.man.ac.uk/software/facta/]

16. **PubGene** [http://www.pubgene.org/]

17. **Pubmatrix** [http://pubmatrix.grc.nia.nih.gov]

18. **SciMiner** [http://jdrf.neurology.med.umich.edu/SciMiner/]

19. Kelso J, Visagie J, Theiler G, Christoffels A, Bardien S, Smedley D, Otgaar D, Greyling G, Jongeneel CV, McCarthy MI, et al: **eVOC: a controlled vocabulary for unifying gene expression data.** *Genome Res* 2003, **13:**1222-1230.

20. **eVOC** [http://www.evocontology.org]

21. McKusick VA: *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders.* 12th edn. Baltimore: Johns Hopkins University Press; 1998.

22. Muller HM, Kenny EE, Sternberg PW: **Textpresso: an ontology-based information retrieval and extraction system for biological literature.** *PLoS Biol* 2004, **2:**e309.

23. **Textpresso** [http://www.textpresso.org]

24. Perez-Iratxeta C, Perez AJ, Bork P, Andrade MA: **Update on XplorMed: A web server for exploring scientific literature.** *Nucleic Acids Res* 2003, **31:**3866-3868.

25. **XplorMed** [http://www.ogic.ca/projects/xplormed/]

26. Gaulton KJ, Mohlke KL, Vision TJ: **A computational system to select candidate genes for complex human traits.** *Bioinformatics* 2007, **23:**1132-1140.

27. Aerts S, Lambrechts D, Maity S, Van Loo P, Coessens B, De Smet F, Tranchevent LC, De Moor B, Marynen P, Hassan B, et al: **Gene prioritization through genomic data fusion.** *Nat Biotechnol* 2006, **24:**537-544.

28. **Endeavour** [http://www.esat.kuleuven.be/endeavour]

29. van Driel MA, Cuelenaere K, Kemmeren PP, Leunissen JA, Brunner HG: **A new web-based data mining tool for the identification of candidate genes for human genetic disorders.** *Eur J Hum Genet* 2003, **11:**57-63.

30. van Driel MA, Cuelenaere K, Kemmeren PP, Leunissen JA, Brunner HG, Vriend G: **GeneSeeker: extraction and integration of human disease-related information from web-based genetic databases.** *Nucleic Acids Res* 2005, **33:**W758-761.

31. **GeneSeeker** [http://www.cmbi.ru.nl/GeneSeeker]

32. Cheng D, Knox C, Young N, Stothard P, Damaraju S, Wishart DS: **PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites.** *Nucleic Acids Res* 2008, **36:**W399-405.

33. **PolySearch** [http://wishart.biology.ualberta.ca/polysearch/index.htm]

34. Chen J, Xu H, Aronow BJ, Jegga AG: **Improved human disease candidate gene prioritization using mouse phenotype.** *BMC Bioinformatics* 2007, **8:**392.

35. **ToppGene** [http://toppgene.cchmc.org]

36. Jelier R, Schuemie MJ, Veldhoven A, Dorssers LC, Jenster G, Kors JA: **Anni 2.0: a multipurpose text-mining tool for the life sciences.** *Genome Biol* 2008, **9:**R96.

37. **Anni 2.1** [http://biosemantics.org/anni]

38. Doms A, Schroeder M: **GoPubMed: exploring PubMed with the Gene Ontology.** *Nucleic Acids Res* 2005, **33:**W783-786.

39. **GoPubMed** [http://www.gopubmed.com/web/gopubmed/]

40. Fernandez JM, Hoffmann R, Valencia A: **iHOP web services.** *Nucleic Acids Res* 2007, **35:**W21-26.

41. **iHOP** [http://www.ihop-net.org/UniPub/iHOP/]

42. States DJ, Ade AS, Wright ZC, Bookvich AV, Athey BD: **MiSearch adaptive pubMed search tool.** *Bioinformatics* 2009, **25:**974-976.

43. **MiSearch** [http://misearch.ncibi.org/]

44. Turner FS, Clutterbuck DR, Semple CA: **POCUS: mining genomic sequence annotation to predict disease genes.** *Genome Biol* 2003, **4:**R75.

45. Adie EA, Adams RR, Evans KL, Porteous DJ, Pickard BS: **Speeding disease gene discovery by sequence based candidate prioritization.** *BMC Bioinformatics* 2005, **6:**55.

46. **PROSPECTR** [http://www.genetics.med.ed.ac.uk/prospectr]

47. Adie EA, Adams RR, Evans KL, Porteous DJ, Pickard BS: **SUSPECTS: enabling fast and effective prioritization of positional candidates.** *Bioinformatics* 2006, **22:**773-774.

48. **SUSPECTS** [http://www.genetics.med.ed.ac.uk/suspects]

49. Hatcher E, Gospodnetic O: *Lucene In Action.* Manning Pubns Co; 2004.

50. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13:**2498-2504.

51. **AMIGO** [http://amigo.geneontology.org/cgi-bin/amigo/go.cgi]

52. **Gene Ontology OBO data** [http://geneontology.org/ontology/obo_format_1_2/]

53. Bada M, Stevens R, Goble C, Gil Y, Ashburner M, Blake JA, Cherry JM, Harris M, Lewis S: **A short study on the success of the Gene Ontology.** *Web Semantics: Science, Services and Agents on the World Wide Web* 2004, **1:**235-240.

54. Tiffin N, Adie E, Turner F, Brunner HG, van Driel MA, Oti M, Lopez-Bigas N, Ouzounis C, Perez-Iratxeta C, Andrade-Navarro MA, et al: **Computational disease gene identification: a concert of methods prioritizes type 2 diabetes and obesity candidate genes.** *Nucleic Acids Res* 2006, **34:**3067-3081.

55. Chen J, Aronow BJ, Jegga AG: **Disease candidate gene identification and prioritization using protein interaction networks.** *BMC Bioinformatics* 2009, **10:**73.

56. **Entrez Gene FTP** [ftp://ftp.ncbi.nih.gov/gene/DATA/]

57. **UniProt** [http://www.uniprot.org/]

58. Becker KG, Barnes KC, Bright TJ, Wang SA: **The genetic association database.** *Nat Genet* 2004, **36:**431-432.

59. Masys DR: **Linking microarray data to the literature.** *Nat Genet* 2001, **28:**9-10.

60. Buckley C, Voorhees EM: **Retrieval System Evaluation.** In *TREC: Experiment and Evaluation in Information Retrieval.* Edited by Voorhees EM, Harman DK: The MIT Press; 2006: 53-79

61. Tanabe L, Scherf U, Smith LH, Lee JK, Hunter L, Weinstein JN: **MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling.** *Biotechniques* 1999, **27:**1210-1214, 1216-1217.

62. Hwang YC, Lu TY, Huang DY, Kuo YS, Kao CF, Yeh NH, Wu HC, Lin CT: **NOLC1, an enhancer of nasopharyngeal carcinoma progression, is essential for TP53 to regulate MDM2 expression.** *Am J Pathol* 2009, **175:**342-354.

63. Han CT, Schoene NW, Lei KY: **Influence of zinc deficiency on Akt-Mdm2-p53 and Akt-p21 signaling axes in normal and malignant human prostate cells.** *Am J Physiol Cell Physiol* 2009, **297:**C1188-1199.

64. Huart AS, MacLaine NJ, Meek DW, Hupp TR: **CK1alpha plays a central role in mediating MDM2 control of p53 and E2F-1 protein stability.** *J Biol Chem* 2009, **284:**32384-32394.

65. Khor LY, Bae K, Paulus R, Al-Saleem T, Hammond ME, Grignon DJ, Che M, Venkatesan V, Byhardt RW, Rotman M, et al: **MDM2 and Ki-67 predict for distant metastasis and mortality in men treated with radiotherapy and androgen deprivation for prostate cancer: RTOG 92-02.** *J Clin Oncol* 2009, **27:**3177-3184.

66. Kim SH, Oh J, Choi JY, Jang JY, Kang MW, Lee CE: **Identification of human thioredoxin as a novel IFN-gamma-induced factor: mechanism of induction and its role in cytokine production.** *BMC Immunol* 2008, **9:**64.

67. Kurimoto C, Kawano S, Tsuji G, Hatachi S, Jikimoto T, Sugiyama D, Kasagi S, Komori T, Nakamura H, Yodoi J, Kumagai S: **Thioredoxin may exert a protective effect against tissue damage caused by oxidative stress in salivary glands of patients with Sjogren's syndrome.** *J Rheumatol* 2007, **34:**2035-2043.

68. Cha MK, Suh KH, Kim IH: **Overexpression of peroxiredoxin I and thioredoxin1 in human breast carcinoma.** *J Exp Clin Cancer Res* 2009, **28:**93.

69. Zhou F, Gomi M, Fujimoto M, Hayase M, Marumo T, Masutani H, Yodoi J, Hashimoto N, Nozaki K, Takagi Y: **Attenuation of neuronal degeneration in thioredoxin-1 overexpressing mice after mild focal ischemia.** *Brain Res* 2009, **1272:**62-70.

70. Billiet L, Furman C, Cuaz-Perolin C, Paumelle R, Raymondjean M, Simmet T, Rouis M: **Thioredoxin-1 and its natural inhibitor, vitamin D3 up-regulated protein 1, are differentially regulated by PPARalpha in human macrophages.** *J Mol Biol* 2008, **384:**564-576.

71. Ehret GB, O'Connor AA, Weder A, Cooper RS, Chakravarti A: **Follow-up of a major linkage peak on chromosome 1 reveals suggestive QTLs associated with essential hypertension: GenNet study.** *Eur J Hum Genet* 2009, **17:**1650-1657.

72. Tsyvian PB, Markova TV, Mikhailova SV, Hop WC, Wladimiroff JW: **Left ventricular isovolumic relaxation and renin-angiotensin system in the growth restricted fetus.** *Eur J Obstet Gynecol Reprod Biol* 2008, **140:**33-37.

73. Fan X, Wang Y, Sun K, Zhang W, Yang X, Wang S, Zhen Y, Wang J, Li W, Han Y, et al: **Polymorphisms of ACE2 gene are associated with essential hypertension and antihypertensive effects of Captopril in women.** *Clin Pharmacol Ther* 2007, **82:**187-196.

74. Stewart JM, Ocon AJ, Clarke D, Taneja I, Medow MS: **Defects in cutaneous angiotensin-converting enzyme 2 and angiotensin-(1-7) production in postural tachycardia syndrome.** *Hypertension* 2009, **53:**767-774.

75. Minami J, Ohno E, Furukata S, Ishimitsu T, Matsuoka H: **Pretreatment levels of plasma renin activity predict ambulatory blood pressure response to valsartan in essential hypertension.** *J Hum Hypertens* 2009, **23:**683-686.

76. Brugts JJ, de Maat MP, Boersma E, Witteman JC, van Duijn C, Uitterlinden AG, Bertrand M, Remme W, Fox K, Ferrari R, et al: **The rationale and design of the PERindopril GENEtic association study (PERGENE): a pharmacogenetic analysis of angiotensin-converting enzyme inhibitor therapy in patients with stable coronary artery disease.** *Cardiovasc Drugs Ther* 2009, **23:**171-181.

77. Feng Y, Yue X, Xia H, Bindom SM, Hickman PJ, Filipeanu CM, Wu G, Lazartigues E: **ACE2 over-expression in the subfornical organ prevents the Angiotensin-II-mediated pressor and drinking responses and is associated with AT1 receptor down-regulation.** *Circ Res* 2008, **102:**729-736.

78. Niu W, Qi Y, Hou S, Zhou W, Qiu C: **Correlation of angiotensin-converting enzyme 2 gene polymorphisms with stage 2 hypertension in Han Chinese.** *Transl Res* 2007, **150:**374-380.

79. Palmer BR, Jarvis MD, Pilbrow AP, Ellis KL, Frampton CM, Skelton L, Yandle TG, Doughty RN, Whalley GA, Ellis CJ, et al: **Angiotensin-converting enzyme 2 A1075G polymorphism is associated with survival in an acute coronary syndromes cohort.** *Am Heart J* 2008, **156:**752-758.

80. Tsezou A, Karayannis G, Giannatou E, Papanikolaou V, Triposkiadis F: **Association of renin-angiotensin system and natriuretic peptide receptor A gene polymorphisms with hypertension in a Hellenic population.** *J Renin Angiotensin Aldosterone Syst* 2008, **9:**202-207.

81. Wilke RA, Simpson RU, Mukesh BN, Bhupathi SV, Dart RA, Ghebranious NR, McCarty CA: **Genetic variation in CYP27B1 is associated with congestive heart failure in patients with hypertension.** *Pharmacogenomics* 2009, **10:**1789-1797.

82. Yan T, Feng Y, Zhai Q: **Axon degeneration: Mechanisms and implications of a distinct program from cell death.** *Neurochem Int* 2010, **56:**529-534.

83. Prohaska JR: **Role of copper transporters in copper homeostasis.** *Am J Clin Nutr* 2008, **88:**826S-829S.

84. Konstantakou EG, Voutsinas GE, Karkoulis PK, Aravantinos G, Margaritis LH, Stravopodis DJ: **Human bladder cancer cells undergo cisplatin-induced apoptosis that is associated with p53-dependent and p53-independent responses.** *Int J Oncol* 2009, **35:**401-416.

85. Mangala LS, Zuzel V, Schmandt R, Leshane ES, Halder JB, Armaiz-Pena GN, Spannuth WA, Tanaka T, Shahzad MM, Lin YG, et al: **Therapeutic Targeting of ATP7B in Ovarian Carcinoma.** *Clin Cancer Res* 2009, **15:**3770-3780.

86. Yu Z, Liu J, Guo S, Xing C, Fan X, Ning M, Yuan JC, Lo EH, Wang X: **Neuroglobin-overexpression alters hypoxic response gene expression in primary neuron culture following oxygen glucose deprivation.** *Neuroscience* 2009, **162:**396-403.

87. Jishage M, Fujino T, Yamazaki Y, Kuroda H, Nakamura T: **Identification of target genes for EWS/ATF-1 chimeric transcription factor.** *Oncogene* 2003, **22:**41-49.

88. Drutel G, Heron A, Kathmann M, Gros C, Mace S, Plotkine M, Schwartz JC, Arrang JM: **ARNT2, a transcription factor for brain neuron survival?** *Eur J Neurosci* 1999, **11:**1545-1553.

89. Zhang B, Schmoyer D, Kirov S, Snoddy J: **GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies.** *BMC Bioinformatics* 2004, **5:**16.

90. **GOTM** [http://bioinfo.vanderbilt.edu/gotm/]

91. Zhang B, Kirov S, Snoddy J: **WebGestalt: an integrated system for exploring gene sets in various biological contexts.** *Nucleic Acids Res* 2005, **33:**W741-748.

92. Khatri P, Bhavsar P, Bawa G, Draghici S: **Onto-Tools: an ensemble of web-accessible, ontology-based tools for the functional design and interpretation of high-throughput gene expression experiments.** *Nucleic Acids Res* 2004, **32:**W449-456.

93. Jensen LJ, Saric J, Bork P: **Literature mining for the biologist: from information retrieval to biological discovery.** *Nat Rev Genet* 2006, **7:**119-129.

94. de Bruijn DR, dos Santos NR, Kater-Baats E, Thijssen J, van den Berk L, Stap J, Balemans M, Schepens M, Merkx G, van Kessel AG: **The cancer-related protein SSX2 interacts with the human homologue of a Ras-like GTPase interactor, RAB3IP, and a novel nuclear protein, SSX2IP.** *Genes Chromosomes Cancer* 2002, **34:**285-298.

95. Grivell L: **Mining the bibliome: searching for a needle in a haystack? New computing tools are needed to effectively scan the growing amount of scientific literature for useful information.** *EMBO Rep* 2002, **3:**200-203.

96. Perez-Iratxeta C, Bork P, Andrade MA: **Association of genes to genetically inherited diseases using data mining.** *Nat Genet* 2002, **31:**316-319.

97. Zhang P, Zhang J, Sheng H, Russo JJ, Osborne B, Buetow K: **Gene functional similarity search tool (GFSST).** *BMC Bioinformatics* 2006, **7:**135.

98. Elo LL, Jarvenpaa H, Oresic M, Lahesmaa R, Aittokallio T: **Systematic construction of gene coexpression networks with applications to human T helper cell differentiation process.** *Bioinformatics* 2007, **23:**2096-2103.

99. Tiffin N, Kelso JF, Powell AR, Pan H, Bajic VB, Hide WA: **Integration of text- and data-mining using ontologies successfully selects disease gene candidates.** *Nucleic Acids Res* 2005, **33:**1544-1552.

100. Wren JD, Garner HR: **Shared relationship analysis: ranking set cohesion and commonalities within a literature-derived relationship network.** *Bioinformatics* 2004, **20:**191-198.

101. Ma X, Lee H, Wang L, Sun F: **CGI: a new approach for prioritizing genes by combining gene expression and protein-protein interaction data.** *Bioinformatics* 2007, **23:**215-221.

102. **KEGG** [http://www.genome.jp/kegg/]

103. **WikiPathways** [http://www.wikipathways.org/index.php/WikiPathways]