Published online 17 November 2015 *Nucleic Acids Research, 2016, Vol. 44, Database issue* ***D567–D573***
doi: 10.1093/nar/gkv1228

WeGET: predicting new genes for molecular systems by weighted co-expression

Radek Szklarczyk[1,2,*,†], Wout Megchelenbrink[2,3,4,†], Pavel Cizek[2], Marie Ledent[1], Gonny Velemans[2], Damian Szklarczyk[5] and Martijn A. Huynen[2,4,*]

[1]Maastricht University Medical Centre, P. Debyelaan 25, 6226XM Maastricht, The Netherlands, [2]Centre for Molecular and Biomolecular Informatics (CMBI), Radboud University Medical Centre, Geert Grooteplein Zuid 26–28, 6525 GA Nijmegen, Nijmegen, The Netherlands, [3]Institute for Computing and Information Sciences (ICIS), Radboud University, Toernooiveld 212, 6525 EC Nijmegen, The Netherlands, [4]Centre for Systems Biology and Bioenergetics (CSBB), Radboud University Medical Centre, Geert Grooteplein Zuid 10, 6500 HB Nijmegen, Nijmegen, The Netherlands and [5]Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, 8057 Zurich, Switzerland

Received August 24, 2015; Revised October 12, 2015; Accepted October 30, 2015

ABSTRACT

We have developed the Weighted Gene Expression Tool and database (WeGET, http://weget.cmbi.umcn.nl) for the prediction of new genes of a molecular system by correlated gene expression. WeGET utilizes a compendium of 465 human and 560 murine gene expression datasets that have been collected from multiple tissues under a wide range of experimental conditions. It exploits this abundance of expression data by assigning a high weight to datasets in which the known genes of a molecular system are harmoniously up- and down-regulated. WeGET ranks new candidate genes by calculating their weighted co-expression with that system. A weighted rank is calculated for human genes and their mouse orthologs. Then, an integrated gene rank and p-value is computed using a rank-order statistic. We applied our method to predict novel genes that have a high degree of co-expression with Gene Ontology terms and pathways from KEGG and Reactome. For each query set we provide a list of predicted novel genes, computed weights for transcription datasets used and cell and tissue types that contributed to the final predictions. The performance for each query set is assessed by 10-fold cross-validation. Finally, users can use the WeGET to predict novel genes that co-express with a custom query set.

INTRODUCTION

Ever since the publication of the first gene expression arrays, the correlated expression of genes involved in a related molecular process has been used to predict functional relations between gene pairs (1). Large amounts of microarray and RNA-seq transcript expression, measured under a plethora of conditions enable mining for concordantly expressed genes. Indeed, this concept has been successfully employed in databases such as COEXPRESSdb, GeneFriends, GeneMANIA and STARNET 2 (2–5). Nevertheless, relative to other types of genomics data, co-expression has lower sensitivity and selectivity (6). To improve the quality of the predictions, various strategies have been applied, like exploiting the conservation of co-expression between species (7), combining many gene expression datasets (8,9) or biclustering datasets to identify groups of genes that co-express within a subset of the experiments (see (10) for a review). Expression screening, an extension of biclustering methods (11), weighs gene expression datasets based on the co-expression of genes within a molecular system and uses those weights to predict new genes involved in that system. It has been successfully applied to predict new mitochondrial proteins essential for the organelle (11) and to discover new players in heme biosynthesis (12). The principle behind this method is appealing: it systematically exploits the available gene expression data and, via its weighting scheme, implicitly solves the question facing many researchers: which gene expression data to use to predict new genes for a pathway? Nevertheless, it is computationally costly, as the weighting has to be recalculated for each pathway separately and additional cross validation requires multiple runs per pathway. We have therefore developed and implemented a fast expression screening algorithm that includes a dataset

*To whom correspondence should be addressed. Tel: +31 43 38 82 918; Fax: +31 43 388 4573; Email: radek.szklarczyk@maastrichtuniversity.nl
Correspondence may also be addressed to Martijn Huynen. Tel: +31 24 36 19543; Fax: +31 24 36 19395; Email: Martijn.Huijnen@radboudumc.nl
†These authors contributed equally to this work as first authors.

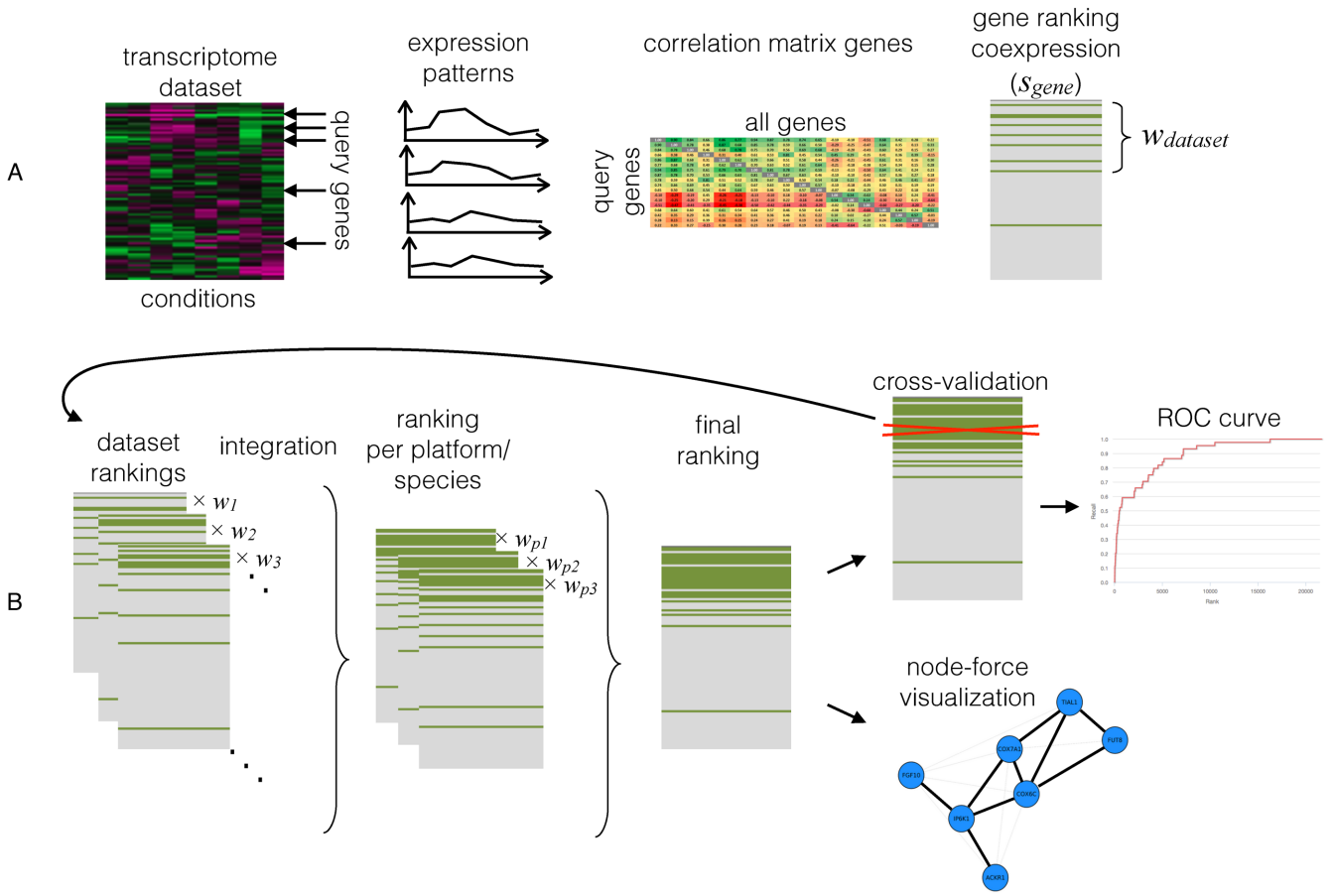

Figure 1. WeGET computational pipeline used to create the database. (**A**) Determining the dataset weight $w_{dataset}$. The transcriptome measurements are converted into a correlation matrix. The average correlation with the query set (s_{gene}) is used for gene ranking and the dataset weight calculation $w_{dataset}$. (**B**) Data integration across datasets, platforms and species. Gene scores s_{gene} from all datasets are combined taking into account the precomputed weights. Subsequently different transcriptome platforms and species data are integrated to arrive at the final ranking. The process is repeated after excluding each query gene to construct a receiver operating characteristic (ROC) curve that visualizes predictive power of the method for a specific query set of genes.

weighting and allows for the rapid computation of genes that co-regulate with a query gene set. Our algorithm was employed to compile a weighted co-expression database for all Gene Ontology (GO) terms and human pathways annotated in the KEGG and Reactome databases. Furthermore, we provide information regarding the original experimental setup of the highly weighted datasets. In particular, WeGet reports the cell and tissue types in which the query genes are consistently up- and down-regulated with each other. Finally, the robustness of the predicted results is assessed by 10-fold cross-validation and reported as the receiver operating characteristic (ROC) curve.

We compared WeGET with five popular web tools and databases that predict novel genes based on their co-expression with specified query gene sets, using two query gene sets published by Baughman *et al.* (ref 11) and show that indeed, weighting the datasets results in improved precision, in particular at low recall rates (the top 100 genes). The complete WeGET database, together with a custom query submission system, is available through the WeGET website.

THE WeGET ANALYSIS PIPELINE

WeGET uses a compendium of 465 human and 560 murine gene expression datasets ranging from 6 to 192 samples per dataset. In total, ∼30 000 samples from multiple mammalian platforms were collected from the Gene Expression Omnibus (GEO) (13).

The WeGET computational pipeline starts with selecting the normalized expression values for all probes associated with the query genes. For genes with multiple probes, the probe with the highest average Pearson correlation coefficient with all other query probes is selected. Subsequently, the pipeline calculates the average Pearson correlation between each gene and the set of query genes in every dataset (Figure 1). Then, all probes are ranked based on their average correlation with the query probes and mapped back to their associated gene. Each gene *i* obtains a score s_i depending on the fraction of the query set that has been ranked above that gene. These calculations are then repeated four times for the same query set and gene expression dataset, where the expression values have been randomly permuted between the genes in every measurement. This step estimates the number of genes that are expected to highly correlate with the query set in a random model. To calculate the

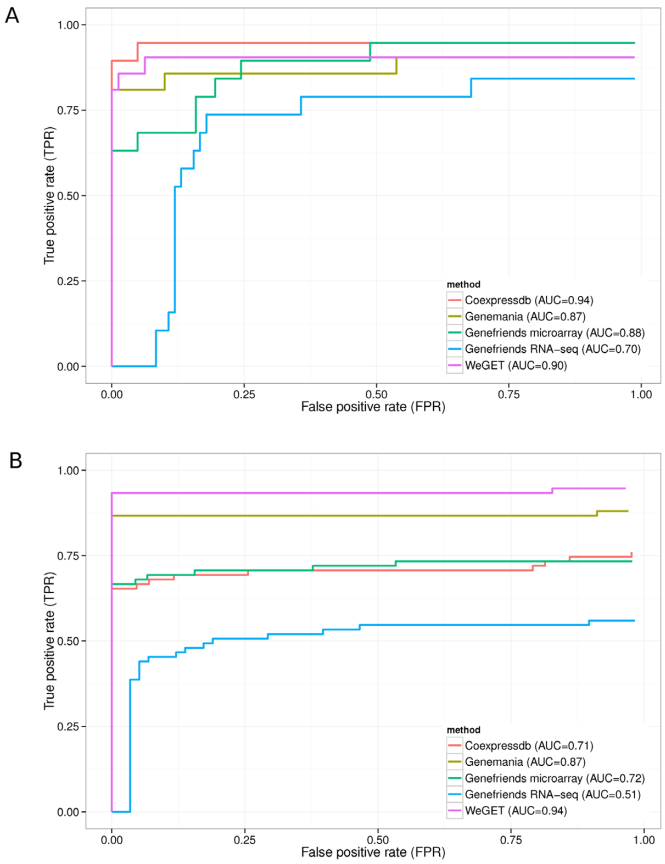

Figure 2. ROC performance curves for online co-expression tools (see Supplementary Table SI1). Performance measured by multiple cross-validation runs is indicated by the area under the curve (AUC) for the top 100 genes corresponding to a typical use case scenario. (**A**) Results for 19 genes in the cholesterol pathway using leave-one-out cross-validation. (**B**) Results for 10-fold cross-validation in the oxidative phosphorylation (OXPHOS) query set.

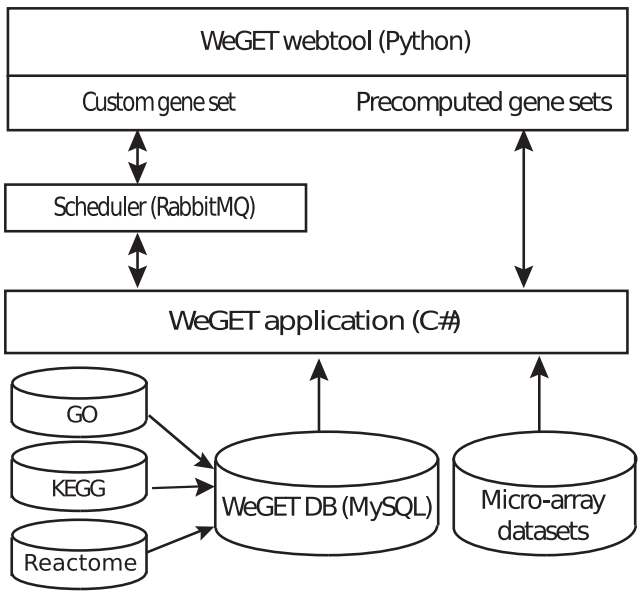

Figure 3. The WeGET system architecture. Results for predefined pathways (GO, KEGG and Reactome) are precomputed and exposed through the WeGET webtool. Custom defined gene sets can be analyzed by submitting the gene ids or gene symbols to the webserver.

dataset weight, an N100 value is calculated that is the fraction of query genes found among the top 100 genes with highest average Pearson correlation. The ratio between the N100 from the original dataset and the average N100 value for the randomized datasets constitutes the weight of the experiments. A species score is the weighted average of all its datasets. The final ranked gene list is obtained by integrating the ranked human and mouse list (mouse genes that are unambiguous human one-to-one orthologs). This is performed using the RobustRankAggreg R-package (14) that computes the final gene rankings using a rank-order statistic (15,16).

Thus for each set of expression data the pipeline measures whether genes in a given pathway are co-expressed with each other better than expected and uses that to assign weights to that expression dataset. These weights are subsequently used in determining the (weighted) co-expression of all genes with that pathway. The source of the variation in the weights between the datasets can be technical, e.g. variation in the probes that have been used, or biological, e.g. variation in the tissues in which gene expression has been measured. The important assumption behind the method is that new genes for a pathway are significantly co-expressed with the majority of the genes of a pathway that they belong to, rather than only with some of its members. This, in turn, depends on the pathway definition. To aid in finding the genes from a pathway that co-express with each other, the results include a visualization of co-expression between query genes displayed as a network. This allows the user to select a subset of co-expressed genes from that pathway to repeat the procedure.

WeGET VALIDATION AND COMPARISON TO OTHER CO-EXPRESSION DATABASES

To assess and compare the predictive power of different co-expression methods (see Supplementary Table SI1), we used two query gene sets (11): 19 query genes in the cholesterol biosynthesis pathway and 76 genes involved in oxidative phosphorylation (OXPHOS). We manually performed leave-one-out or 10-fold cross validation by multiple submissions (see Supplementary Methods for details). We took into account the top 100 ranked genes as a likely use case scenario. Figure 2 shows the WeGET results for the cholesterol biosynthesis pathway and OXPHOS system compared to other online tools employing the co-expression analysis.

Baughman *et al.* (11) carried out one-time computations for cholesterol and OXPHOS datasets. The Weget webserver achieves identical (cholesterol) or marginally better performance (OXPHOS, 86.4% sensitivity at 99.8% specificity, compared to 85% and 99.4%, respectively, see Supplementary Figure SI1).

THE WeGET DATABASE AND WEB ACCESS

Figure 3 depicts the architecture of the WeGET database. Human pathways and their associated genes from GO and KEGG are stored in a central database. The WeGET parallel algorithm that calculates each dataset on a separate thread precomputes the co-expressed genes and dataset weights for all pathways using the transcriptome compendium. The results are presented to the user using the

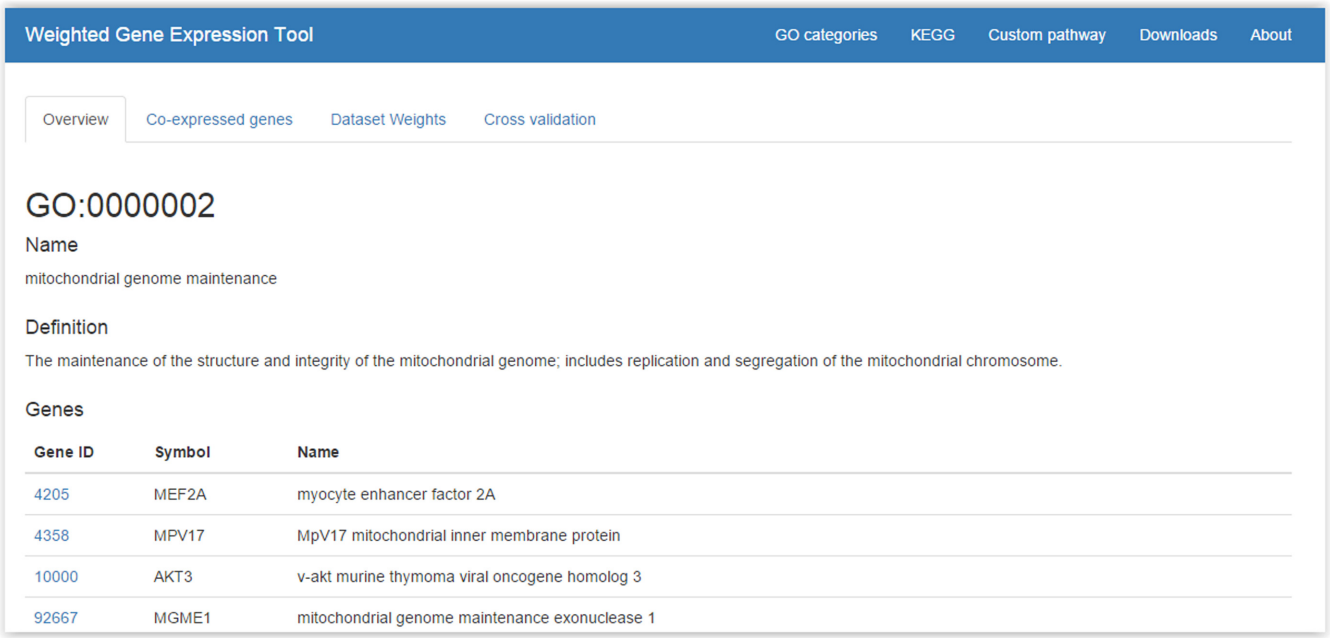

Figure 4. The GO data grid. Users can browse or search the list of Gene Ontology (GO) terms. When a row is clicked, detailed information is provided (see Figure 5). KEGG and Reactome pathways can be accessed in a similar fashion.

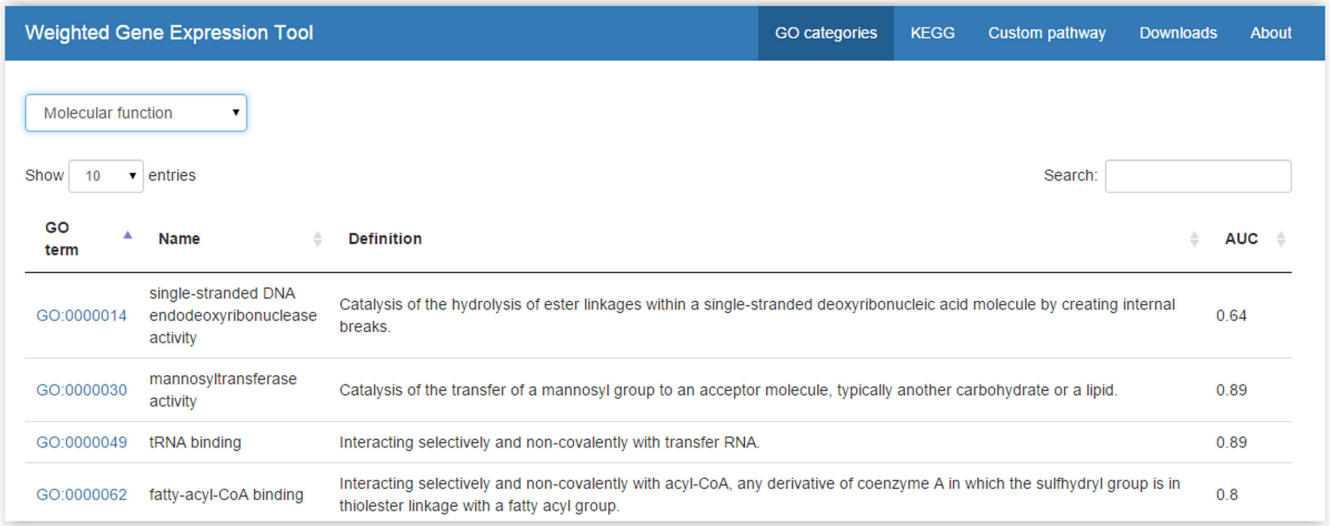

Figure 5. Detailed information for a precomputed term/pathway. The query genes, co-expressed genes, dataset weights and cross-validation results are shown.

WeGET webtool (implemented in Python Flask) and can additionally be downloaded.

On the WeGET website, pathways are shown in a data grid (Figure 4), which can be sorted and searched. Detailed information (Figure 5), such as the best scoring genes, dataset weights, cell and tissue types in which the genes highly co-express (see also Supplementary Figures SI2 and SI3) and cross-validation results are shown when a row entry is selected. A pathway can be accessed directly as *weget.cmbi.umcn.nl/pathwaydb/identifier* where *pathwaydb* denotes the pathway database (one of: GO, KEGG or Reactome) and *identifier* the category identification (e.g. http://weget.cmbi.umcn.nl/GO/GO:0000398). User queries (different than the predefined sets) can be entered using the '*Custom pathway*' tab, specifying genes as Entrez ID or HUGO gene symbol. The query is then scheduled for analysis. After the analysis, the user receives an email with results, including the cross-validation and a network that displays the co-regulation of the query genes within the datasets (see below), in a spreadsheet.

The website provides an opportunity to learn more about the experimental conditions in which the concordant expression of the query molecular system has been observed. The tab 'Dataset Weights' accessible for each precomputed

Table 1. Genes implicated in neuropathic pain collected from the literature.

No.	Gene Symbol	Sub-system	No.	Gene Symbol	Sub-system
1	ACTG1	CORE	11	DPYSL2	TAG
2	ANK3	CORE	12	KCNK3	TAG
3	SCN10A	CORE	13	NRCAM	TAG
4	SCN11A	CORE	14	ANXA2	TAG
5	SCN1B	CORE	15	PRKACA	TAG
6	SCN2B	CORE	16	PRKCB	TAG
7	SCN3A	CORE	17	SYN2	TAG
8	SCN3B	CORE	18	TNR	TAG
9	SCN8A	CORE	19	MSN	PI
10	SPTBN4	CORE	20	NEDD4L	PI

Core molecular sub-system associated with voltage-gated sodium channels (CORE), trafficking-associated genes (TAG) and the peripheral involvement (PI) classes are indicated (23–25).

Table 2. Results from custom molecular system as received by the user. Top 40 genes prioritized for their involvement in neuropathic pain are shown. Genes that were part of the query set are shown on shaded background.

Position	p-value	Entrez id	Symbol	Description	Is query gene
0	0.00000	6854	SYN2	synapsin II	1
1	0.00005	6327	SCN2B	sodium channel, voltage-gated, type II, beta subunit	1
2	0.00010	23705	CADM1	cell adhesion molecule 1	0
3	0.00018	60	ACTB	actin, beta	0
4	0.00018	730102	LOC730102	quinone oxidoreductase-like protein 2 pseudogene	0
5	0.00020	7532	YWHAG	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, ga	0
6	0.00025	283209	PGM2L1	phosphoglucomutase 2-like 1	0
7	0.00030	64207	IRF2BPL	interferon regulatory factor 2 binding protein-like	0
8	0.00037	71	ACTG1	actin, gamma 1	1
9	0.00040	288	ANK3	ankyrin 3, node of Ranvier (ankyrin G)	1
10	0.00049	8455	ATRN	attractin	0
11	0.00050	5579	PRKCB	protein kinase C, beta	1
12	0.00054	3785	KCNQ2	potassium voltage-gated channel, KQT-like subfamily, member 2	0
13	0.00054	3745	KCNB1	potassium voltage-gated channel, Shab-related subfamily, member 1	0
14	0.00055	6191	RPS4X	ribosomal protein S4, X-linked	0
15	0.00064	84502	JPH4	junctophilin 4	0
16	0.00064	6907	TBL1X	transducin (beta)-like 1X-linked	0
17	0.00066	2917	GRM7	glutamate receptor, metabotropic 7	0
18	0.00076	221336	BEND6	BEN domain containing 6	0
19	0.00081	26508	HEYL	hes-related family bHLH transcription factor with YRPW motif-like	0
20	0.00083	6235	RPS29	ribosomal protein S29	0
21	0.00088	2597	GAPDH	glyceraldehyde-3-phosphate dehydrogenase	0
22	0.00097	84677	DSCR8	Down syndrome critical region gene 8	0
23	0.00101	8825	LIN7A	lin-7 homolog A (C. elegans)	0
24	0.00101	6134	RPL10	ribosomal protein L10	0
25	0.00108	4155	MBP	myelin basic protein	0
26	0.00111	11280	SCN11A	sodium channel, voltage-gated, type XI, alpha subunit	1
27	0.00114	3778	KCNMA1	potassium large conductance calcium-activated channel, subfamily M, alpha me	0
28	0.00115	399947	C11orf87	chromosome 11 open reading frame 87	0
29	0.00115	6171	RPL41	ribosomal protein L41	0
30	0.00117	63895	PIEZO2	piezo-type mechanosensitive ion channel component 2	0
31	0.00122	222865	TMEM130	transmembrane protein 130	0
32	0.00125	2828	GPR4	G protein-coupled receptor 4	0
33	0.00126	23201	FAM168A	family with sequence similarity 168, member A	0
34	0.00129	57181	SLC39A10	solute carrier family 39 (zinc transporter), member 10	0
35	0.00129	1816	DRD5	dopamine receptor D5	0
36	0.00132	9378	NRXN1	neurexin 1	0
37	0.00135	4130	MAP1A	microtubule-associated protein 1A	0
38	0.00138	7314	UBB	ubiquitin B	0
39	0.00148	302	ANXA2	annexin A2	1
40	0.00149	115827	RAB3C	RAB3C, member RAS oncogene family	0

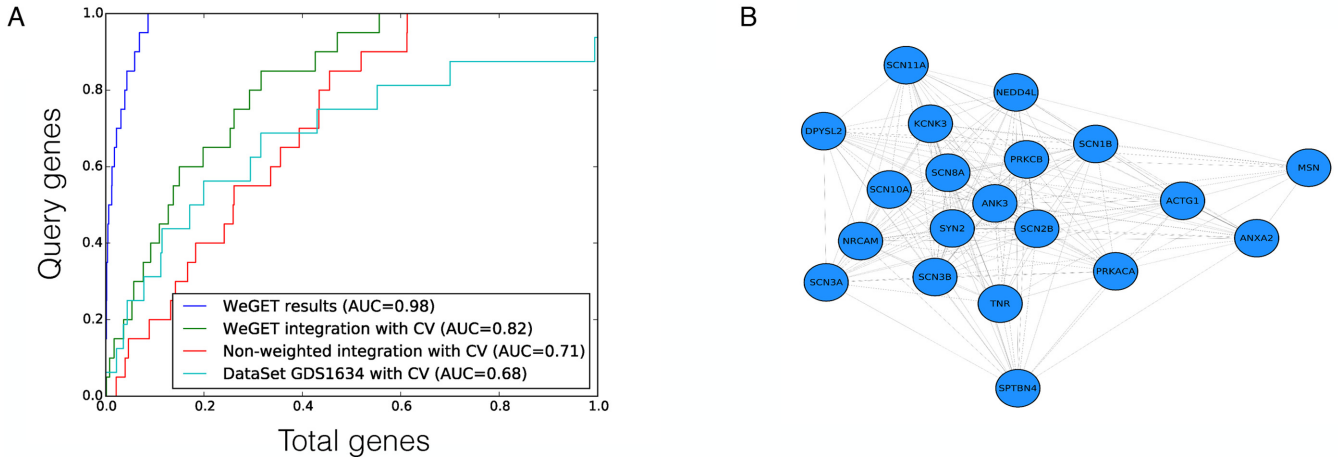

Figure 6. Visualization of performance of the WeGET results for neuropathic pain genes. (**A**) ROC curves for the neuropathic pain query set (Table 1). The X axis represents fraction of human genes, the Y axis the fraction of the neuropathic pain molecular system. Shown are ROC curves for final results (blue), the cross-validation (CV) of integrated datasets (green), the average co-expression across all datasets (integration with equal contribution of each dataset) with CV (red) and results of co-expression within a single high-weight dataset (GDS1634, a nodose and dorsal root ganglia comparison, cyan) (22). (**B**) Network visualization of the co-expression allows identification of genes less co-expressed with the core of the query set.

query set lists GEO dataset records with concordant expression patterns of the query system, indicating congruent co-expression of the gene components of the molecular system. The dataset identifier is directly hyperlinked with the GEO entry description (both online and in Excel output file) such that users can read details of the experiment that lead to harmonious expression of the query set.

EVALUATION OF WeGET RESULTS FOR A QUERY SET

The robustness of the results is tested by k-fold cross-validation and graphically displayed with a ROC curve. The curve illustrates the performance of the WeGET method, by plotting the true positive rate (successfully cross-validated query genes) versus all human genes (Figure 6). The curve is plotted for every molecular system stored in the database (GO, KEGG and Reactome pathways) separately. The area under the curve (AUC) is a measure of the prediction quality and robustness for that pathway. The average AUC computed for all pathways is around 0.7. Well-studied and clearly defined cellular components such as mitochondria and biological processes such as cilium movement and assembly have a higher AUC (average 0.83 and 0.84, respectively) reflecting their concordant expression patterns. For pathways with less than 50 genes we use leave-one-out cross-validation, for larger pathways 10-fold cross-validation is carried out.

Finally, the cohesion of the query gene set is displayed as a network using a node-force algorithm (Figure 6B). Query genes that consistently co-express perform a large attractive force and therefore cluster together. In contrast, genes that show little evidence of co-regulation exhibit a smaller force and do not cluster with the other query genes. Using this visualization, the user can resubmit the query gene set to omit genes that do not show evidence of co-regulation.

USING WeGET TO PREDICT GENES INVOLVED IN NEUROPATHIC PAIN

Previous studies indicate that mutations in genes coding for voltage-gated sodium channels and related processes may impair the nociceptive pathway and influence response to pain stimuli (17). From the literature we collected genes implicated in neuropathic pain (Table 1) and used the WeGET database to predict novel candidate genes for this pathway. Table 2 shows genes co-expressing with the neuropathic pain molecular system as calculated by WeGET. Next to sodium channels and its regulators (PRMT8, UNC80 rank 41 and 74, respectively) also genes of the voltage-gated potassium system are strongly represented among top co-expressing genes (MAP1A, PPP2R2C, KCNH3, KCNQ2 rank 6, 7, 66 and 87, respectively) consistent with their involvement in nociceptive processing (18), their expression in dorsal root ganglion neurons analogous to voltage-gated sodium channels (19) and with recently discovered genetic variants that modulate neuropathic pain (20). The PIEZO2 gene, a nociceptive component mechanically activated in nerve endings (21) ranks 74th among all genes. Additional poorly characterized genes such as SERP2, TMEM130 and CCDC155 (ranks 5, 9 and 20, respectively) are also present among genes co-expressing with the system and constitute novel candidate genes for nociceptive pathway. Figure 6A shows the higher performance of WeGET integration of all datasets (cross-validated AUC = 0.82) compared to integration of all datasets with equal weights (average co-expression across all experiments, AUC = 0.71) and a high-weight individual dataset GDS1634 of dorsal root ganglia neurons (AUC = 0.68). Weights assigned to GEO datasets reveal a high contribution of transcriptome measurements related to neurons: a murine nodose and dorsal root ganglia study (GDS1634, weight 3.0) (22), gene expression in human neurofibrillary tangles (GDS2795, weight 2.5) and DNA methylation effect on neural stem cells (GDS538, weight 3.0). The peripheral roles of DPYSL2 (trafficking subset, Table 1), MSN and NEDD4L proteins (peripheral

subclass) are visualized in the query gene network (Figure 6B). Currently we screen patients with a familial form of neuropathic pain for genetic variants that may impact the function of the candidate genes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

European Union 7th Framework Programme [PROPANE Study, 602273]; Metakids Foundation (to R.S.); Centre for Systems Biology and Bioenergetics [CSBR09/013V to W.M.]; European Union's FP7 large scale integrated network Gencodys [http://www.gencodys.eu/, HEALTH-241995 to P.C.]; SYSCILIA [241955]; Centre for Systems Biology and Bioenergetics. Funding for open access charge: Grant funding.
Conflict of interest statement. None declared.

REFERENCES

1. Chu,S., DeRisi,J., Eisen,M., Mulholland,J., Botstein,D., Brown,P.O. and Herskowitz,I. (1998) The transcriptional program of sporulation in budding yeast. *Science*, **282**, 699–705.
2. Jupiter,D., Chen,H. and VanBuren,V. (2009) STARNET 2: a web-based tool for accelerating discovery of gene regulatory networks using microarray co-expression data. *BMC Bioinformatics*, **10**, 332.
3. Obayashi,T., Okamura,Y., Ito,S., Tadaka,S., Motoike,I.N. and Kinoshita,K. (2013) COXPRESdb: a database of comparative gene coexpression networks of eleven species for mammals. *Nucleic Acids Res.*, **41**, D1014–D1020.
4. van Dam,S., Craig,T. and de Magalhães,J.P. (2015) GeneFriends: a human RNA-seq-based gene and transcript co-expression database. *Nucleic Acids Res.*, **43**, D1124–D1132.
5. Zuberi,K., Franz,M., Rodriguez,H., Montojo,J., Lopes,C.T., Bader,G.D. and Morris,Q. (2013) GeneMANIA prediction server 2013 update. *Nucleic Acids Res.*, **41**, W115–W122.
6. von Mering,C., Krause,R., Snel,B., Cornell,M., Oliver,S.G., Fields,S. and Bork,P. (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, **417**, 399–403.
7. Van Noort,V., Snel,B. and Huynen,M.A. (2003) Predicting gene function by conserved co-expression. *Trends in Genetics*, **19**, 238–242.
8. Lee,H.K., Hsu,A.K., Sajdak,J., Qin,J. and Pavlidis,P. (2004) Coexpression analysis of human genes across many microarray data sets. *Genome Res.*, **14**, 1085–1094.
9. Zhou,X.J., Kao,M.-C.J., Huang,H., Wong,A., Nunez-Iglesias,J., Primig,M., Aparicio,O.M., Finch,C.E., Morgan,T.E. and Wong,W.H. (2005) Functional annotation and network reconstruction through cross-platform integration of microarray data. *Nat. Biotechnol.*, **23**, 238–243.
10. Madeira,S.C. and Oliveira,A.L. (2004) Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **1**, 24–45.
11. Baughman,J.M., Nilsson,R., Gohil,V.M., Arlow,D.H., Gauhar,Z. and Mootha,V.K. (2009) A computational screen for regulators of oxidative phosphorylation implicates SLIRP in mitochondrial RNA homeostasis. *PLoS Genet.*, **5**, e1000590.
12. Nilsson,R., Schultz,I.J., Pierce,E.L., Soltis,K.A., Naranuntarat,A., Ward,D.M., Baughman,J.M., Paradkar,P.N., Kingsley,P.D., Culotta,V.C. *et al.* (2009) Discovery of genes essential for heme biosynthesis through large-scale gene expression analysis. *Cell Metab.*, **10**, 119–130.
13. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
14. Kolde,R., Laur,S., Adler,P. and Vilo,J. (2012) Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*, **28**, 573–580.
15. Aerts,S., Lambrechts,D., Maity,S., Van Loo,P., Coessens,B., De Smet,F., Tranchevent,L.-C., De Moor,B., Marynen,P., Hassan,B. *et al.* (2006) Gene prioritization through genomic data fusion. *Nat. Biotechnol.*, **24**, 537–544.
16. Stuart,J.M., Segal,E., Koller,D. and Kim,S.K. (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–255.
17. von Hehn,C.A., Baron,R. and Woolf,C.J. (2012) Deconstructing the neuropathic pain phenotype to reveal neural mechanisms. *Neuron*, **73**, 638–652.
18. Tsantoulas,C. and McMahon,S.B. (2014) Opening paths to novel analgesics: the role of potassium channels in chronic pain. *Trends Neurosci.*, **37**, 146–158.
19. Pollema-Mays,S.L., Centeno,M.V., Ashford,C.J., Apkarian,A.V. and Martina,M. (2013) Expression of background potassium channels in rat DRG is cell-specific and down-regulated in a neuropathic pain model. *Mol Cell. Neurosci.*, **57**, 1–9.
20. Costigan,M., Belfer,I., Griffin,R.S., Dai,F., Barrett,L.B., Coppola,G., Wu,T., Kiselycznyk,C., Poddar,M., Lu,Y. *et al.* (2010) Multiple chronic pain states are associated with a common amino acid-changing allele in KCNS1. *Brain*, **133**, 2519–2527.
21. Delmas,P., Hao,J. and Rodat-Despoix,L. (2011) Molecular mechanisms of mechanotransduction in mammalian sensory neurons. *Nat. Rev. Neurosci.*, **12**, 139–153.
22. Peeters,P.J., Aerssens,J., de Hoogt,R., Stanisz,A., Gohlmann,H.W., Hillsley,K., Meulemans,A., Grundy,D., Stead,R.H. and Coulie,B. (2006) Molecular profiling of murine sensory neurons in the nodose and dorsal root ganglia labeled from the peritoneal cavity. *Physiol. Genomics*, **24**, 252–263.
23. Porter,J.D., Merriam,A.P., Leahy,P., Gong,B. and Khanna,S. (2003) Dissection of temporal gene expression signatures of affected and spared muscle groups in dystrophin-deficient (mdx) mice. *Hum. Mol. Genet.*, **12**, 1813–1821.
24. Eijkelkamp,N., Linley,J.E., Baker,M.D., Minett,M.S., Cregg,R., Werdehausen,R., Rugiero,F. and Wood,J.N. (2012) Neurological perspectives on voltage-gated sodium channels. *Brain*, **135**, 2585–2612.
25. Vacher,H. and Trimmer,J.S. (2012) Trafficking mechanisms underlying neuronal voltage-gated ion channel localization at the axon initial segment. *Epilepsia*, **53**(Suppl. 9), 21–31.