

1-20-2007

EDGEdb: a transcription factor-DNA interaction database for the analysis of *C. elegans* differential gene expression

M. Inmaculada Barrasa
University of Massachusetts Medical School

Philippe Vaglio
Modul-Bio

Fabien Cavasino
Modul-Bio

See next page for additional authors

Follow this and additional works at: <http://escholarship.umassmed.edu/oapubs>

 Part of the [Life Sciences Commons](#), and the [Medicine and Health Sciences Commons](#)

Repository Citation

Barrasa, M. Inmaculada; Vaglio, Philippe; Cavasino, Fabien; Jacotot, Laurent; and Walhout, Albertha J. M., "EDGEdb: a transcription factor-DNA interaction database for the analysis of *C. elegans* differential gene expression" (2007). *Open Access Articles*. 1243.
<http://escholarship.umassmed.edu/oapubs/1243>

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Open Access Articles by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

EDGEdb: a transcription factor-DNA interaction database for the analysis of *C. elegans* differential gene expression

Authors

M. Inmaculada Barrasa, Philippe Vaglio, Fabien Cavasino, Laurent Jacotot, and Albertha J. M. Walhout

Database

Open Access

EDGEdb: a transcription factor-DNA Interaction database for the analysis of *C. elegans* differential gene expression

M Inmaculada Barrasa^{†1}, Philippe Vaglio^{†2}, Fabien Cavasino², Laurent Jacotot² and Albertha JM Walhout^{*1}

Address: ¹Program in Gene Function and Expression and Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA and ²Modul-Bio, 232 Boulevard Sainte-Marguerite, 13009 Marseille, France

Email: M Inmaculada Barrasa - inmaculada.barrasa@umassmed.edu; Philippe Vaglio - p.vaglio@modul-bio.com; Fabien Cavasino - f.cavasino@modul-bio.com; Laurent Jacotot - l.jacotot@modul-bio.com; Albertha JM Walhout* - marian.walhout@umassmed.edu

* Corresponding author †Equal contributors

Published: 18 January 2007

Received: 10 November 2006

BMC Genomics 2007, 8:21 doi:10.1186/1471-2164-8-21

Accepted: 18 January 2007

This article is available from: <http://www.biomedcentral.com/1471-2164/8/21>

© 2007 Barrasa et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Transcription regulatory networks are composed of protein-DNA interactions between transcription factors and their target genes. A long-term goal in genome biology is to map protein-DNA interaction networks of all regulatory regions in a genome of interest. Both transcription factor -and gene-centered methods can be used to systematically identify such interactions. We use high-throughput yeast one-hybrid assays as a gene-centered method to identify protein-DNA interactions between regulatory sequences (e.g. gene promoters) and transcription factors in the nematode *Caenorhabditis elegans*. We have already mapped several hundred protein-DNA interactions and analyzed the transcriptional consequences of some by examining differential gene expression of targets in the presence or absence of an upstream regulator. The rapidly increasing amount of protein-DNA interaction data at a genome scale requires a database that facilitates efficient data storage, retrieval and integration.

Description: Here, we report the implementation of a *C. elegans* differential gene expression database (EDGEdb). This database enables the storage and retrieval of protein-DNA interactions and other data that relate to differential gene expression. Specifically, EDGEdb contains: i) sequence information of regulatory elements, including gene promoters, ii) sequence information of all 934 predicted transcription factors, their DNA binding domains, and, where available, their dimerization partners and consensus DNA binding sites, iii) protein-DNA interactions between regulatory elements and transcription factors, and iv) expression patterns conferred by regulatory elements, and how such patterns are affected by interacting transcription factors.

Conclusion: EDGEdb provides a protein-DNA -and protein-protein interaction resource for *C. elegans* transcription factors and a framework for similar databases for other organisms. The database is available at <http://edgedb.umassmed.edu>.

Background

Differential gene expression is governed, at least in part, by protein-DNA interactions between transcription factors (TFs) and their target genes. Together, such protein-DNA interactions can be modeled into transcription regulatory networks that describe the logic underlying the development, function, and pathology of a system of interest [1,2]. Two complementary strategies are currently being used to identify protein-DNA interactions: TF-centered approaches, where the DNA sequences that interact with a TF or set of TFs of interest are identified; and gene-centered methods that identify the TFs that interact with a regulatory DNA sequence or set of DNA sequences of interest (e.g. gene promoters)[2]. Previously, we developed a high-throughput yeast one-hybrid system for the gene-centered mapping of protein-DNA interactions between gene promoters or small *cis*-regulatory elements and TFs in the nematode *Caenorhabditis elegans* [3,4]. So far, we have identified 605 protein-DNA interactions between 115 gene promoters and 176 TFs [3,5-7]. In addition, we identified protein-DNA interactions between several small *cis*-regulatory DNA elements and TFs (i.e. to identify consensus TF binding sites). Several TFs bind DNA as dimers and such TF-TF dimers are being systematically identified by high-throughput protein-protein interaction mapping efforts [8-10]. Longer term, TF dimer information needs to be incorporated in transcription regulatory network models. We ultimately aim to generate a protein-DNA interaction map between all regulatory DNA elements and TFs in the *C. elegans* genome [2].

Most *C. elegans* research data are collected and maintained in the database WormBase [11,12]. This database provides an indispensable resource for *C. elegans* researchers and contains information about, for instance, the genome sequence and annotation, bioinformatic protein domain annotation, mutants, phenotypes, *etcetera*. However, this database is not as convenient for the retrieval and manipulation of comprehensive protein-protein and protein-DNA interaction datasets, as well as for the storage of manually curated annotations of families of genes (e.g. TFs). Several databases have been developed for the storage of data related to gene regulation, including Transfac [13,14] and Jaspar [15,16] that contain information about TF binding sites. Oreganno is a database that contains collections of regulatory sequences and TF binding sites for a variety of organisms [17,18]. The data in this database are not extensively curated as researchers are free to enter and manipulate data themselves. Hence, this database contains many types of experimental data, which does have clear advantages. However, it is not convenient to navigate and download protein-DNA interactions involving TFs and precisely defined genomic sequences.

The rapidly increasing amount of gene-centered protein-DNA interaction data and how such interactions affect differential gene expression requires a frequently updated and curated database for optimal data storage, retrieval and integration. Here, we report the implementation of a *C. elegans* differential gene expression database, or EDGEDb, a database that is specifically tailored for the storage, retrieval and integration of physical interactions between *C. elegans* TFs and precisely defined genomic regulatory DNA sequences.

Construction and content

In gene-centered yeast one-hybrid assays, two types of "DNA baits" are used to identify interacting TFs: single copy *C. elegans* genomic sequences such as gene promoters, and artificial baits such as (putative) *cis*-regulatory DNA elements [5]. EDGEDb contains information about i) DNA bait sequences and genomic coordinates; ii) all 934 predicted *C. elegans* TFs [19], i.e. their DNA binding domain, and, where available, dimerization partners and consensus binding sites; iii) protein-DNA interactions between DNA baits and TFs; and iv) where available, the transcriptional consequences of such protein-DNA interactions (see below). In total, the database contains 605 protein-DNA interactions between 115 *C. elegans* gene promoters and 176 TFs. In addition, the database contains protein-DNA interactions for 3 short DNA sequences that were either found by us or by other groups (referred to as "artificial baits", see e.g. ZTF-2 or DAF-12). Finally, the database contains 24 TF protein-protein dimer interactions. The regulation of several *C. elegans* genes by specific TFs has been documented in the literature, and we have included some of these in EDGEDb. However, we have not included cases where regulation is reported but where the genomic sequences involved have not been precisely mapped (i.e. sequences other than the promoter could be involved, the promoter sequence is not available, or the regulatory interaction is indirect). In the future, we aim to incorporate additional interaction data obtained by other laboratories as they become available. We encourage researchers to send us their data on the EDGEDb homepage. In addition, we will continue to incorporate TF dimerization data, both obtained by our own yeast two-hybrid assays and obtained by other groups.

EDGEDb is implemented on a Jboss application server using an Oracle Database. We used the Bio::DB::GFF [20] database schema to handle genomic sequence information. This schema is loaded with a filtered version of the GFF annotations and DNA sequences in FASTA format from WormBase [21]. Filtered GFF annotations only include gene structure information (5'UTR, exons, introns, 3'UTR), and operon and gene locations. We have also included a WormBase geneID file [22] that provides relationships between WormBase gene name, locus/CGC

name and the sequence name. Since only the WormBase ID is always maintained (*i.e.* the names and sequence may change as gene models and functional annotations improve), we use it as a key for the annotations. In addition to the Bio::DB::GFF schema, the database is composed of 26 tables that contain information about the protein-DNA interaction experiments, DNA baits, TFs, TF dimers and consensus binding sites, expression data, and publications. The interfaces have been developed using J2EE/JSP technology, except for the genome display that uses a modified version of GBrowse [23,24]. The genome sequence can be updated using new GFF annotations from WormBase. A series of Perl scripts can then "remap" the DNA baits to their new genomic position. EDGEdb will be updated with every WormBase freeze release (that is every 10 versions). For optimal utility, EDGEdb is linked to both WormBase [11,12] and Worfdb, the database that contains information regarding the collection of *C. elegans* cloned open reading frames (ORFs) in the ORFeome project [25,26].

Utility and discussion

EDGEdb description

The EDGEdb home page allows four types of queries: i) individual genes or lists of "space separated" genes can be searched (*e.g.* *daf-3*, Figure 1A); ii) a second text field is available for DNA bait names (*e.g.* *Pdaf-3*, the promoter of the *daf-3* gene, Figure 1B); iii) a scroll down menu in the DNA binding domain text field allows the retrieval of TFs by DNA binding domain (*e.g.* AP-2, Figure 1C), and iv) a search by publication links to experiments and associated excel files of individual studies (Figure 1D, F). An "interaction browser" lists DNA baits and TFs for which interactions are available and facilitates the retrieval of stored interactions (Figure 1G). The "interaction export" interface allows the query and download of interactions from different publications for a set of either DNA baits or TFs (Figure 1H). The "export all interactions" tool allows the download of all available interactions in the database (Figure 1I). These features are designed to facilitate the retrieval and integration of interactions from the different available datasets included in EDGEdb. They allow the user to quickly identify which DNA baits have been assayed so far, to identify the TFs retrieved, and to specifically obtain and download interactions for all or a set of DNA baits or TFs of interest from any of the datasets. Access to the complete collection of predicted *C. elegans* (worm) TFs (wTF2.0)[19] is available through the home page (Figure 1E). This resource will be updated as new TFs are identified and as gene predictions are updated [7]. To facilitate outsider data submission, we have implemented a page for data upload (Figure 1J). After manual curation, the data will be incorporated into EDGEdb.

EDGEdb gene page

A gene page (Figure 2A) can be accessed by searching for a sequence name (*e.g.* F25E2.5), CGC (*Caenorhabditis* Genetics Center) name (*e.g.* *daf-3*) or WormBase identifier (*e.g.* WBGene00000899). The page displays alternative names used for the gene (Figure 2A, top), the genomic coordinates, and a genome view of gene models and available DNA baits (Figure 2A, middle). All elements in the genome view link to their respective gene or DNA bait page (see below). The next section (Figure 2A, third section) displays a list of DNA baits associated with the gene (*e.g.* gene promoter sequences). DNA bait names link to the respective DNA bait pages (see below). Additional features are available for genes that encode TFs, including a link to the TF page (see below) (Figure 2A top, red circle), and a list of DNA targets the TF interacts with (Figure 2A bottom, blue circle). Finally, the detailed experiment for each interaction is retrieved through the "view experiment" button (Figures 2A bottom).

EDGEdb transcription factor page

There is a page for each of the 934 predicted TFs describing its DNA binding domain, dimerization partners, and consensus binding site, where available [19](data not shown). In addition, links to Jaspar [15,16], Transfac [13,14] and WormBook [27], that may contain additional TF binding site information, are provided. The experiment pages for the protein-DNA interactions are accessed through the "view experiment" button. References to the publications where protein-protein interactions and consensus binding sites were reported are included. The DNA baits bound by the TF are listed at the bottom of the page, and links to the corresponding experiments and export tools (see below) are included.

EDGEdb DNA bait Page

In EDGEdb, genomic sequences that correspond to DNA baits are specified with their chromosome coordinates, and the corresponding sequence is displayed on the page (Figure 2B). The DNA bait name relates to the downstream gene (*e.g.* *Pmdl-1*, the promoter of the gene *mdl-1*, Figure 2B). The DNA bait position with respect to the gene model may change as gene models are updated. TFs that interact with a DNA bait are listed at the bottom of the page (Figure 2B, bottom) and links to the respective gene page (Figure 2B, blue circle) and relevant experiments are provided. Finally, where available, expression patterns conferred by DNA baits and changes in these patterns in the absence of an interacting TF can also be accessed through this page (Figure 2B green circle, Figure 2C). Several artificial DNA baits are also included in the database (*e.g.* P2_multimer).

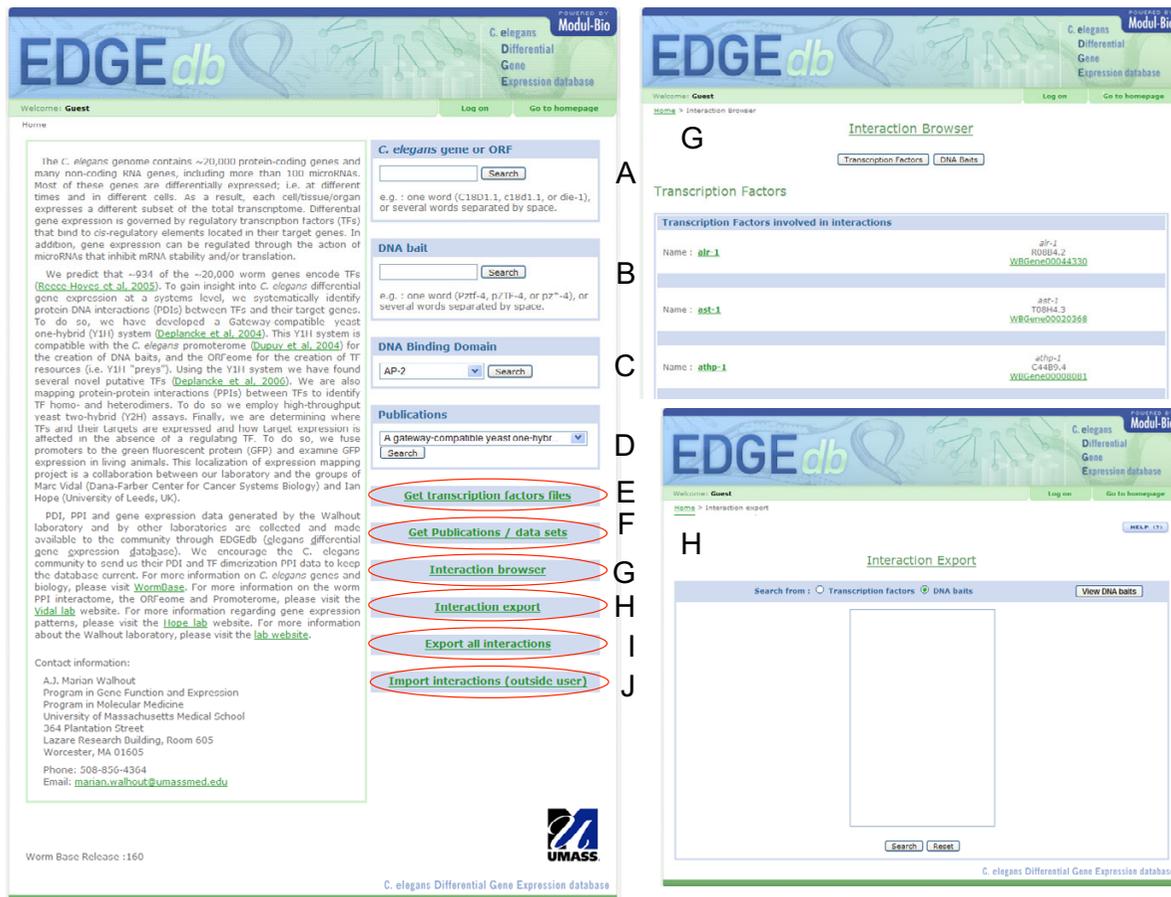


Figure 1
EDGEdb home page. This page contains an introduction to the database (left). The home page also provides links to several tools to explore and download the data (right): (A) Text field to search for *C. elegans* genes. (B) Text field to search for DNA baits (e.g. a promoter). (C) Scroll down menu to retrieve TFs by DNA binding domain. (D) Scroll down menu to retrieve publications and associated experiments. (E) Link to the collection of *C. elegans* TFs. (F) Yeast one -and two-hybrid datasets and pdf files of corresponding publications. (G) Interaction browser. (H, I) Export tools. (J) Import interaction tool for outside users. The WormBase release used is specified at the bottom left corner of the page.

EDGEdb experiment page

Each DNA bait used is linked to at least one experiment page that details the information about the type of experiment that was carried out and the list of interactors found. For interactions reported in the literature, "literature" is specified as the source of the interaction, and the PubMed ID is provided (e.g. the MDL-1/MLX-1 dimerizing protein-protein interaction reported on the corresponding TF pages is referred to as PMID: 9764821).

Export of protein-DNA interaction data

A user can download excel files containing protein-DNA interaction information from the TF, DNA bait and exper-

iment pages. The user can also download text files containing the sequence of DNA baits bound by a TF from the TF pages. Additionally, the home page links to an "interaction export" page and to an "export all interactions" tool that allow query and download of all or a subset of interactions.

EDGEdb significance

Several features make EDGEdb a valuable resource for both the *C. elegans* and systems biology of gene expression/transcription communities. First, as more interaction data become available, extracting information for the user's favorite set of genes or TFs will become time con-

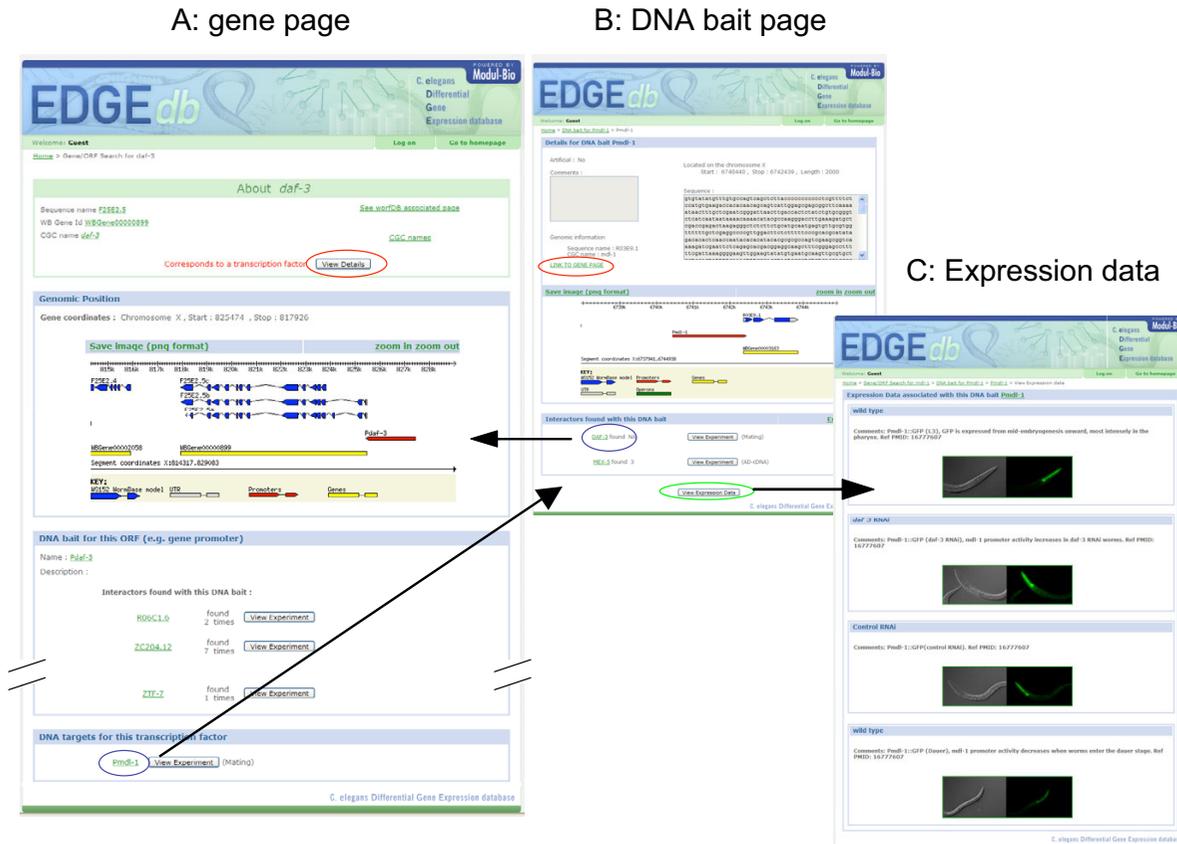


Figure 2
EDGEdb gene, DNA bait and expression pages. (A) EDGEdb *daf-3* gene page. The top section contains the different *daf-3* gene identifiers, each of which is linked to WormBase. In addition, it contains a link to Wormdb, and the detailed TF page (red circle). The middle section contains the genome view of the *daf-3* gene. The third section lists the DNA baits associated with *daf-3* (i.e. the *daf-3* promoter, *Pdaf-3*) and their corresponding yeast one-hybrid interactors (i.e. TFs that bind to *Pdaf-3*). The bottom section includes the DNA baits bound by the TF DAF-3 (blue circle). By clicking this one goes to the *Pmdl-1* DNA bait page. (B) EDGEdb DNA bait page for *Pmdl-1*. This page displays the DNA bait sequence and genomic coordinates; IDs of the gene corresponding to the DNA bait and link to the gene page (red circle). The middle section contains the genome browser. In the bottom section, the list of the interactors found with the DNA bait and links to the relevant experiments and gene pages are shown (blue circle). By clicking this link one goes back to the *daf-3* gene page. A link to the page containing the expression patterns driven by the DNA bait is included (green circle). By clicking this one goes to the expression page for *Pmdl-1*. (C) Expression data page for *Pmdl-1*. From top to bottom: GFP expression driven by *Pmdl-1* in wild type animals, in *daf-3(RNAi)* animals, in control RNAi animals, and in dauer animals. See Deplancke et al. [5] for details.

suming. EDGEdb allows the efficient retrieval and integration of different protein-DNA interaction and TF-TF dimerization datasets, either through the export tools (i.e. protein-DNA interactions) or through the TF page (i.e. TF-TF dimerization). Moreover, EDGEdb enables the export of interaction information for multiple genes or TFs at a time. Second, specific sequences bound by a TF or a set of

TFs can be retrieved. This is very important for users that aim to analyze sequences that interact with a particular TF (e.g. to computationally infer consensus TF binding sites). Third, sequence and gene identifiers in EDGEdb will be updated with every WormBase freeze. This is important as gene identifiers and coordinates may change, and because the conversion of old gene IDs to new gene IDs and coor-

dinates may be tedious. Thus, the data within EDGEdb will always be compatible with the latest WormBase freeze. Finally, EDGEdb includes TF consensus binding sites and TF dimerization data from the literature where available. We will continue to include protein-DNA interaction data from other sources in the future, as long as precise DNA sequence information is available for interacting genomic regulatory sequences. We encourage researchers to send us their interaction data that involve *C. elegans* TFs.

Conclusion

EDGEdb provides a database for the storage, retrieval and integration of gene-centered protein-DNA interactions between *C. elegans* regulatory genome sequences and TFs; information regarding TFs; and differential gene expression data. This database will provide a valuable resource and tool for both the *C. elegans* and the systems biology community, and provides a framework to create similar databases for other organisms, including humans.

Availability and requirements

Project name: EDGEdb a transcription factor-DNA Interaction database for the analysis of *C. elegans* differential gene expression

Project home page: <http://edgedb.umassmed.edu>

Operating system(s): platform independent

Programming language: Java

Licence: the database is freely available to academic and non-academic users

Abbreviations

Transcription factor (TF), open reading frame (ORF), untranslated region (UTR), green fluorescent protein (GFP), *C. elegans* differential gene expression database (EDGEdb).

Authors' contributions

M.I.B, P.V. and F.C. created the database, under supervision of L.J. and A.J.M.W. The manuscript was written by M.I.B. and A.J.M.W.

Acknowledgements

We thank the members of the Walhout laboratory and Job Dekker for discussion and critical reading of the manuscript, and NIDDK for funding (R01DK068429). We also thank D. Lapointe, C. Hull, R. Shulka, J. Schneider and D. Jones at UMass Medical School for administration, networking and backup of our server.

References

- Blais A, Dynlacht BD: **Constructing transcriptional regulatory networks.** *Genes Dev* 2005, **19**:1499-1511.

- Walhout AJM: **Unraveling Transcription Regulatory Networks by Protein-DNA and Protein-Protein Interaction Mapping.** *Genome Res* 2006, **16**:1445-1454.
- Deplancke B, Dupuy D, Vidal M, Walhout AJM: **A Gateway-compatible yeast one-hybrid system.** *Genome Res* 2004, **14**:2093-2101.
- Deplancke B, Vermeirssen V, Arda HE, Martinez NJ, Walhout AJM: **Gateway-compatible yeast one-hybrid screens.** *CSH Protocols* 2006, doi:10.1101/pdb.prot4590.
- Deplancke B, Mukhopadhyay A, Ao W, Elewa AM, Grove CA, Martinez NJ, Sequerra R, Doucette-Stam L, Reece-Hoyes JS, Hope IA, Tisenbaum HA, Mango SE, Walhout AJM: **A gene-centered *C. elegans* protein-DNA interaction network.** *Cell* 2006, **125**:1193-1205.
- Vermeirssen V, Barrasa MI, Hidalgo C, Babon JAB, Sequerra R, Doucette-Stam L, Barabasi AL, Walhout AJM: **Transcription Factor Modularity in a Gene-Centered *C. elegans* Core Neuronal Protein-DNA Interaction Network.** *submitted* 2007.
- Vermeirssen V, et al.: **A *C. elegans* transcription factor resource for smart pooling-based yeast one- and two-hybrid assays. in preparation.**
- Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JDJ, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Bauemgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, van den Heuvel S, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stam L, Gunsalus K, Harper JW, Cusick ME, Roth FP, Hill D, Vidal M: **A map of the interactome network of the metazoan *C. elegans*.** *Science* 2004, **303**:540-543.
- Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M: **Towards a proteome-scale map of the human protein-protein interaction network.** *Nature* 2005, **437**:1173-1178.
- Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanesen N, Carroll S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stan-yan CA, Finley RL Jr., White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM: **A protein interaction map of *Drosophila melanogaster*.** *Science* 2003, **302**:1727-1736.
- Schwarz EM, Antoshechkin I, Bastiani C, Bieri T, Blasiar D, Canaran P, Chan J, Chen N, Chen WJ, Davis P, Fiedler TJ, Girard L, Harris TW, Kenny EE, Kishore R, Lawson D, Lee R, Muller HM, Nakamura C, Ozersky P, Petcherski A, Rogers A, Spooner W, Tuli MA, Van Auken K, Wang D, Durbin R, Spieth J, Stein LD, Sternberg PW: **WormBase: better software, richer content.** *Nucleic Acids Res* 2006, **34**:D475-8.
- Wormbase [<http://www.wormbase.org>]
- Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I, Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel AE, Wingender E: **TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes.** *Nucleic Acids Res* 2006, **34**:D108-10.
- Transfac: [<http://www.gene-regulation.com>].
- Vlieghe D, Sandelin A, De Bleser PJ, Vleminckx K, Wasserman WW, van Roy F, Lenhard B: **A new generation of JASPAR, the open-access repository for transcription factor binding site profiles.** *Nucleic Acids Res* 2006, **34**:D95-7.
- Jaspar [<http://www.jaspar.cgb.ki.se>]
- Montgomery SB, Griffith OL, Sleumer MC, Bergman CM, Bilenyk M, Pleasance ED, Prychyna Y, Zhang X, Jones SJ: **OREGAnno: an open access database and curation system for literature-derived promoters, transcription factor binding sites and regulatory variation.** *Bioinformatics* 2006, **22**:637-640.
- oreganno [<http://www.oreganno.org>]

19. Reece-Hoyes JS, Deplancke B, Shingles J, Grove CA, Hope IA, Walhout AJM: **A compendium of C. elegans regulatory transcription factors: a resource for mapping transcription regulatory networks.** *Genome Biol* 2005, **6**:R110.
20. **Bioperl** [<http://www.bioperl.org>]
21. **E chromosomes** [ftp://ftp.sanger.ac.uk/pub/wormbase/development_release/CHROMOSOMES]
22. **EG IDs** [ftp://ftp.wormbase.org/pub/wormbase/genomes/elegans/annotations/gene_IDs/current.gz]
23. **Gbrowse** [<http://www.gmod.org>]
24. Stein LD, Mungall C, Shu S, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S: **The generic genome browser: a building block for a model organism system database.** *Genome Res* 2002, **12**:1599-1610.
25. **Worfdb.** [<http://worfdb.dfci.harvard.edu>].
26. Vaglio P, Lamesch P, Reboul J, Rual JF, Martinez M, Hill D, Vidal M: **Worfdb: the Caenorhabditis elegans ORFeome Database.** *Nucleic Acids Res* 2003, **31**:237-240.
27. **Wormbook** [<http://www.wormbook.org/>]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

