

SOFTWARE

Open Access

Genecentric: a package to uncover graph-theoretic structure in high-throughput epistasis data

Andrew Gallant¹, Mark DM Leiserson², Maxim Kachalov¹, Lenore J Cowen¹ and Benjamin J Hescott^{1*}

Abstract

Background: New technology has resulted in high-throughput screens for pairwise genetic interactions in yeast and other model organisms. For each pair in a collection of non-essential genes, an epistasis score is obtained, representing how much sicker (or healthier) the double-knockout organism will be compared to what would be expected from the sickness of the component single knockouts. Recent algorithmic work has identified graph-theoretic patterns in this data that can indicate functional modules, and even sets of genes that may occur in compensatory pathways, such as a BPM-type schema first introduced by Kelley and Ideker. However, to date, any algorithms for finding such patterns in the data were implemented internally, with no software being made publically available.

Results: Genecentric is a new package that implements a parallelized version of the Leiserson et al. algorithm (*J Comput Biol* 18:1399-1409, 2011) for generating generalized BPMs from high-throughput genetic interaction data. Given a matrix of weighted epistasis values for a set of double knock-outs, Genecentric returns a list of generalized BPMs that may represent compensatory pathways. Genecentric also has an extension, GenecentricGO, to query FuncAssociate (*Bioinformatics* 25:3043-3044, 2009) to retrieve GO enrichment statistics on generated BPMs. Python is the only dependency, and our web site provides working examples and documentation.

Conclusion: We find that Genecentric can be used to find coherent functional and perhaps compensatory gene sets from high throughput genetic interaction data. Genecentric is made freely available for download under the GPLv2 from <http://bcb.cs.tufts.edu/genecentric>.

Background

When two non-essential genes are simultaneously deleted, sometimes a surprising phenotype emerges compared to the phenotype of the deletion mutants of the single genes. When studying the yeast genome, often this can be quantified in terms of the growth rate of the double deletion mutant, compared to the growth rate of its component single deletion mutants, termed epistasis [1]. Recent SGA [2], dSLAM [3] and E-MAP [4] technology produces high throughput weighted epistasis values for large collections of double knockouts.

A variety of algorithmic methods have been proposed to infer functionally meaningful relationships between genes based on the structure of their epistatic genetic

interactions [5-10]. In particular, we consider the generalized “Between Pathway Model” (or BPM) as studied by Leiserson et al [5]. As discussed in [5], this involves finding pairs of sets of genes, where the sum of the epistasis values between genes in different sets, minus the sum of the epistasis values between genes in the same set, is as negative as possible. This is a generalization of an unweighted BPM model studied by Ma et al. [11] and Brady et al. [12], which was a simplification of the original BPM model introduced by Kelley and Ideker [13]. Other versions of BPMs, in different settings or criteria, have been studied [9,14,15].

Leiserson et al. [5] presented a randomized algorithm based on maximal graph cuts to generate these putative generalized BPMs from weighted epistasis data. They showed that the BPMs produced by their method were biologically enriched when their method was run on several different yeast E-MAP and SGA data sets [8,16]. We

*Correspondence: Benjamin.Hescott@tufts.edu

¹Department of Computer Science, Tufts University, Medford, MA 02155, USA
Full list of author information is available at the end of the article

now present a full implementation, Genecentric, of this algorithm that is fast, easy to use, well documented and open source. In addition, Genecentric has an extension that performs GO enrichment analysis using FuncAssociate's web service API [17].

Implementation

Genecentric implements a randomized algorithm to generate putative generalized BPMs as described in Leiserson et al [5].

Algorithm

Genecentric takes as input a file with genetic interaction data and outputs a list of pairs of sets of genes corresponding to the putative generalized BPMs. The genetic interaction data must be in a tab-delimited format where each row corresponds to a gene pairing and that pairing's interaction score. Specifically, the first column contains the first gene identifier, the second column contains the second gene identifier, and the third column contains the interaction score.

We'll now briefly recap the algorithm described in Leiserson et al [5]. (Figure 1 provides a graphical representation of the algorithm.) First, the input genetic interaction data is used to create M random bipartitions, where the vertices correspond to genes and an edge between genes corresponds to its interaction score. Second, each random bipartition is transformed into a LocalMaxCut by iteratively modifying it until every vertex in the graph satisfies the following property: the sum of its edges to other vertices in the same partition is greater than the sum of its edges to other vertices in the other partition. Third, every unique gene g in the interaction data generates a bipartite subgraph (called a BPM) from this set of M bipartitions. Namely, one partition of the BPM includes g and every gene from the M bipartitions that is in the same partition as g at least $C\%$ of the time. The other partition of the BPM includes every gene from the M bipartitions that is in the opposing partition as g at least $C\%$ of the time. Finally, the set of all BPMs from all the genes is pruned, either to

remove BPMs whose partite sets are too small or too large (see parameter settings, below), or to remove substantially overlapping BPMs generated from different genes g from the set.

Since a couple of the steps in the aforementioned algorithm involve many independent calculations, Genecentric parallelizes those computations automatically using the *multiprocessing* module included in the Python standard library. Therefore, large data sets can be used to generate BPMs in reasonable time. Genecentric was able to produce BPMs for a data set containing 1,685,210 genetic interactions in a few hours on an AMD 48 core Linux machine with default parameters. On a smaller data set (220,116 genetic interactions), Genecentric produces BPMs in under a minute on the same 48 core machine.

Parameters

There are several parameters that can be set to customize Genecentric. Firstly, Genecentric can modify the underlying weights on the edges of the genetic interaction data. In particular, Leiserson et al. [5] left SGA weights unchanged, but squared E-MAP weights (retaining the sign) to speed convergence.

The user may also set M , the number of randomized bipartitions (default is 250), and C , the proportion of the time a gene must be on the same or opposite side as gene g in order to be included in g 's BPM (default is 0.9). (Recall that Leiserson's algorithm generates a BPM for every gene in the input and prunes redundant BPMs.) Note that because Genecentric is a randomized algorithm, the set of BPMs produced will not be the same for every run; however, as M increases, the results will converge. Leiserson et al. [5] showed that with values of at least 250 and 0.9 for M and C respectively, different runs of the algorithm will produce similar results. M and C can be customized, but making M or C much smaller is not recommended since there could be too much variability between different runs of the randomized algorithm. Alternatively, increasing M can further decrease variability, but at a runtime performance cost.

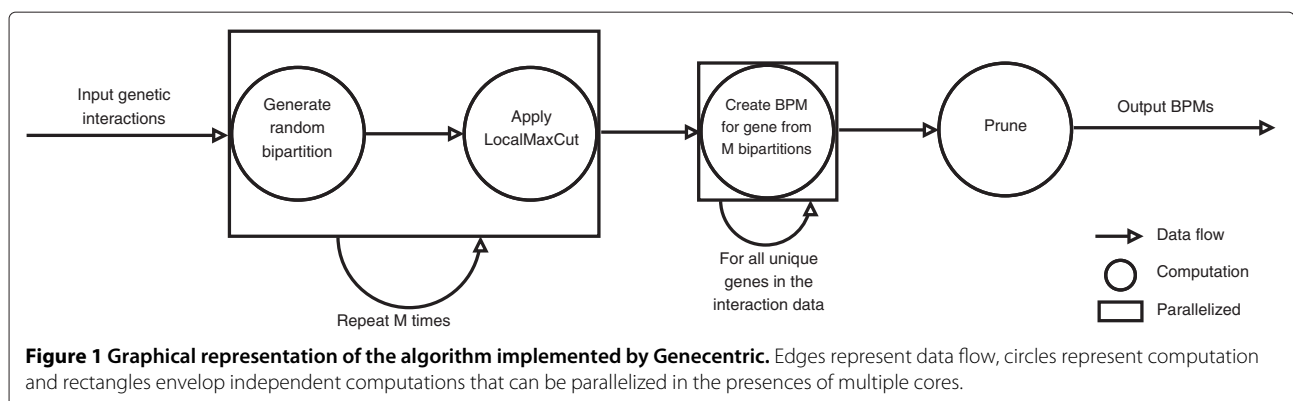


Figure 1 Graphical representation of the algorithm implemented by Genecentric. Edges represent data flow, circles represent computation and rectangles envelop independent computations that can be parallelized in the presences of multiple cores.

By default, Genecentric will prune the BPM set returned to avoid repeating many overlapping, similar sets. But there are also several pruning options that are completely configurable in Genecentric. The first is the Jaccard index, J (default is 0.66), which specifies the similarity threshold of the resulting BPMs. Namely, every pairing of BPMs in the output is guaranteed to have a Jaccard index less than J (where the Jaccard index between two BPMs is defined as the size of the intersection between both BPMs divided by the size of the union of both BPMs). The final two pruning options, “minimum-size” and “maximum-size” (defaults are 3 and 25, respectively), filter the resulting set of BPMs so that no BPM has a module with fewer than “minimum-size” genes and no BPM has a module with more than “maximum-size” genes.

Recall that Genecentric outputs a tab-delimited file which contains a list of pairs of sets of genes corresponding to its putative generalized BPMs. This file can be used as input to GenecentricGO which performs GO enrichment analysis on the BPMs using FuncAssociate [17]. GenecentricGO can also be configured with several parameters. Of them, the most notable are the p-value cutoff (default 0.05) and the genespace. The p-value cutoff corresponds to the p-value cutoff in FuncAssociate: only sets of genes whose p-value is less than or equal to this cutoff will be returned in the results. The genespace is by default set to only the genes in the input genetic interaction data, but can be toggled to use all genes recorded for that species by FuncAssociate. GenecentricGO also provides an interface to use different species and namespaces with FuncAssociate.

Finally, GenecentricGO by default automatically employs the FuncAssociate multiple-testing correction by setting the “simulations” parameter to the greater of 1000 and the number of BPM modules, where 1000 is the default value set by FuncAssociate. This value is also capped at 10000 by FuncAssociate. (While in theory, we should be able to discover more enrichment by just setting the parameter to the number of modules in the input BPM file, we found that FuncAssociate’s stochastic simulations to estimate p-values were too variable when this parameter was set much below their default of 1000.)

Results and discussion

We ran Genecentric on the same E-MAP dataset as discussed in the Leiserson et al. paper [5], and showed that it produced comparable results (though note that results will not be 100% identical, because, as discussed above and in their paper, results wobble slightly over different randomized runs of the algorithm). We set out next to show that Genecentric with default or nearly default parameters can be run “out of the box” to produce meaningful results that correlate well with the biological literature

on other datasets as well. In particular, a recent study [18] produced an E-MAP dataset of 374 genes involved in various aspects of plasma-membrane biology, including endocytosis, signaling, lipid metabolism and eisome function. We ran Genecentric on this E-MAP dataset using all default parameters except the weights were squared, as Leiserson et al. [5] recommends for E-MAP data, and C was lowered from .9 to .8 in order to produce more BPMs (22 instead of 6). Since default parameters were used, GO enrichment was computed by FuncAssociate with a genespace consisting of only genes from the E-MAP data set. Of these 22 BPMs, 7 exhibited GO enrichment in both gene sets, according to the Genecentric FuncAssociate GO module, whereas an additional 8 exhibited GO enrichment in one gene set of the BPM. These enrichment rates, while good, are still somewhat below the percentage enrichment reported for the ChromBio E-MAP dataset in [5], perhaps because the component genes in this plasma-membrane E-MAP data set are not all as well studied, and thus fewer functional annotations are yet known.

Of the 7 BPMs that exhibit enrichment in both gene sets, 6 have one set that contains the genes coding for proteins COG 5, 6, 7, and 8, and there is some additional overlap in some of the genes in these BPMs (recall that Genecentric with default pruning options will permit overlapping BPMs provided their Jaccard index is less than 0.66). Table 1 shows a list of these BPMs and corresponding GO enrichment terms. Figure 2 depicts one of these BPMs by showing edges for physical and genetic interactions between each pair of genes.

The proteins encoded by COG 5, 6, 7 and 8 constitute a structural component known of the conserved oligomeric Golgi (COG) complex, an important peripheral Golgi apparatus protein structure that has been most significantly implicated in retrograde trafficking [19] and has additional roles in supporting Golgi apparatus structure [20] and glycosylation [21]. It is a hetero-octamer consisting of two lobes, A and B, that are both comprised of four proteins, COG 1-4 for lobe A and COG 5-8 for lobe B. Recent studies have also identified COG 1 as the linking unit between the two lobes of the complex [22]. Four out of the eight COG proteins have significant homology with other organisms, including mammals, and studies have shown that mutation of COG proteins in humans can lead to a group of serious conditions known as congenital disorders of glycosylation [21,23]. Deletions in COG 1-4 (Lobe A) in *Saccharomyces cerevisiae* cause severe growth defects, but COG 5-8 have been shown to be non-essential [22]. Additionally, the mutations that cause these severe growth defects are phenotypically distinct, indicating different roles for the different subunits of the COG complex [24,25]. Other genes that occur multiple times in the same set as the COG proteins in the BPMs

Table 1 BPMs where 1 module includes the COG complex

BPM	Selected enriched GO terms
COG6 COG5 COG8 PIB2 COG7	intra-Golgi vesicle-mediated transport (4/5) protein targeting to vacuole (4/5)
ARL1 VPS35 GET3 ARL3 SYS1 GOT1 PEP8 SFT2 MNN1 VPS17	protein transport (9/10) Golgi apparatus (7/10) endosome transport (6/10) vesicle-mediated transport (9/10)
COG6 RIC1 MRP8 COG5 COG7 SNC1 COG8 GCS1 PIB2 SRO7	intra-Golgi vesicle-mediated transport (4/10) Golgi vesicle transport (7/10) vesicle-mediated transport (9/10)
ARL1 VPS35 GET3 ARL3 SYS1 SRO77 PEP8 SFT2 GOT1 SNC2 VPS13 VPS17	post-Golgi vesicle-mediated transport (7/12) Golgi vesicle transport (8/12) establishment of protein localization (10/12) retromer complex (3/12)
COG6 YPT7 COG7 MVB12 YPT52 COG5 CCZ1 ARF3 COG8 ENT3 YPT53	intra-Golgi vesicle-mediated transport (4/11) establishment of protein localization to vacuole (8/11) vacuolar transport (10/11) GTP binding (4/11)
ARL1 VPS35 GET3 ARL3 YPT31 BOR1 SYS1 GOT1 PEP8 SFT2 VPS17 YOL019W	endosome transport (7/12) Golgi vesicle transport (7/12) retrograde transport, endosome to Golgi (4/12) establishment of protein localization (11/12) retromer complex (3/12)
COG6 COG5 COG8 COG7	intra-Golgi vesicle-mediated transport (4/4) protein targeting to vacuole (4/4) establishment of protein localization (4/4)
ARL1 VPS35 GSF2 GET3 ARE2 SFT2 GOT1 SCS7 PEP8 VPS17	retromer complex (3/10)
COG6 RIC1 MRP8 EIS1 ARF1 COG5 PIB2 COG7 COG8 GCS1	intra-Golgi vesicle mediated transport (5/10) Golgi vesicle transport (6/10)
ARL1 VPS35 GSF2 ARL3 SYS1 APL5 GET3 PEP8 SFT2 GOT1 ARE1 SCS7 VPS13 VPS17	protein transport (11/14) Golgi apparatus (8/14) endosome transport (7/14) Golgi vesicle transport (7/14) establishment of protein localization (11/14)
COG6 RIC1 MRP8 VPS41 EIS1 ARF1 ARF3 COG5 YPT52 COG7 COG8 GCS1 SNC1 PIB2 ENT3	intra-Golgi vesicle-mediated transport (5/15) Golgi vesicle transport (8/15) vesicle-mediated transport (13/15) establishment of protein localization (11/15)
ARL1 VPS35 GSF2 ARL3 YPT31 SYS1 VPS17 GET3 PEP8 SFT2 GOT1 SNC2 SCS7 VPS13 MNN1 VPS29 ARE2	endosome transport (9/17) Golgi vesicle transport (8/17) retrograde transport, endosome to Golgi (5/17) establishment of protein localization (12/17) retromer complex (3/17)

This table lists several BPMs generated where one module contains genes in the COG complex. Selected enriched GO terms are included, and bolded genes have been enriched for at least one of the GO terms listed. For complete GO enrichment results, please see the Additional file 1.

include PIB2, a protein that binds phosphatidylinositol 3-phosphate, involved in vesicle-mediated transport. Phosphatidylinositol-3-phosphate, together with small GTPases, is also an important factor for sorting in the endocytic pathway [26]. Other proteins were CCZ1, a protein involved in vacuolar transport and vesicle docking, ARF1, ARF3 and GCS1, where ARFs are GTPases of the Ras superfamily that regulate the formation of coated vesicles in intracellular trafficking [27,28] are regulated by the GTPase activating protein GCS1 [29]. RIC1, a GTPase involved in localization of trans-Golgi membrane proteins [30] and MRP8, whose function is not known, also appear.

On the opposite module of the BPMs containing COG proteins, frequently occurring genes include VPS35 and PEP8, components of the retromer complex needed for retrograde transport, as well as VPS17, a protein associated with proper vesicle formation. Additionally, the protein GOT1, which plays a role in secretory transport [31] was found, which further hints at the variable function of the COG complex and its subunits. The second module also contained the genes encoding for the ARL1, SYS1, and ARL3 proteins—which are involved in vesicle tethering at the Golgi apparatus [32]. It is interesting that, different from the datasets discussed in the Leiser et al paper [5], a large percentage of the enriched BPMs contain the same COG complex. We don't know if the popularity of the COG complex in our results is simply because of what annotation is known, something about the interaction of the algorithm with the distribution of the E-MAP weights, or multiple roles for the COG complex that make it especially able to compensate for multiple different biological processes.

In addition to the BPMs containing the COG complex, there are two dually enriched and seven singly enriched BPMs (see Tables 2 and 3). We note that some of the BPM modules—that do not show up as GO enriched—support functional coherence when we dig more deeply into the literature. For example, a module in one of our BPMs, RTG3 ALP1 PEP12 XRN1 BCH2 RTG2 SIW14, is not flagged as GO-enriched by FuncAssociate, but we found that RTG3 and RTG2 are known to be involved in the retrograde signalling pathway. PEP12 is a multifunctional yeast syntaxin that controls entry of biosynthetic, endocytic and retrograde traffic into the prevacuolar compartment [33]. BCH2 is a member of a complex that mediates the export of specific cargo proteins from the Golgi to the plasma membrane [34].

A full list of Genecentric BPMs is provided in the Additional file 1.

Conclusions

We have introduced Genecentric, a package that can automatically accept a set of high-throughput genetic interaction data, and output generalized BPMs along with their

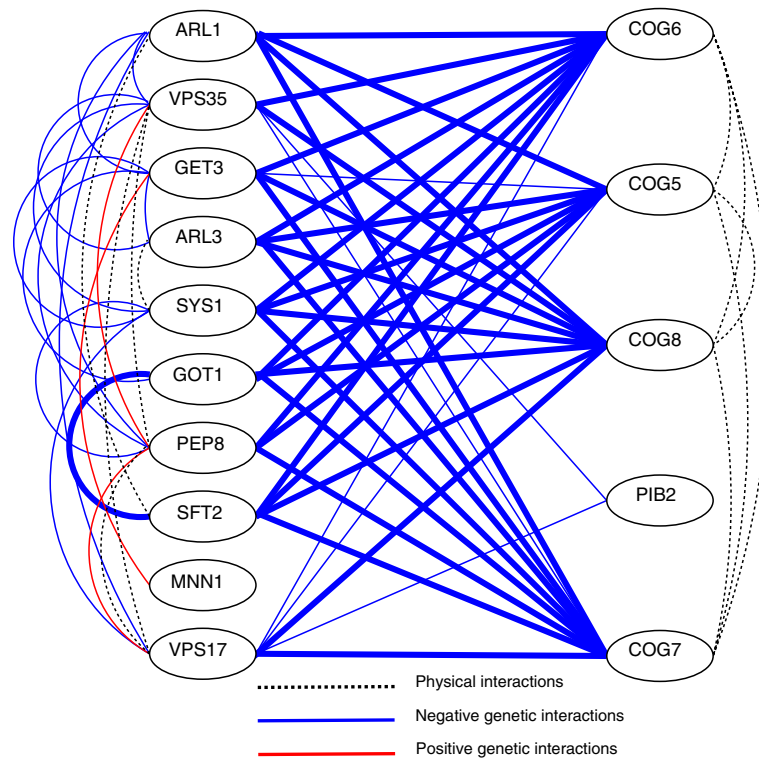


Figure 2 Sample of physical and genetic interactions of a BPM. Genetic interactions appear as solid lines. Thicker lines correspond to more negative weights. Dashed lines correspond to known protein-protein interactions. (Protein-protein interactions were taken from BioGRID's data, with the "low-throughput" filter active).

Table 2 Dually enriched BPMs

BPM	Selected enriched GO terms
SLT2 BCK1 CLC1	endoplasmic reticulum unfolded protein response (2/3)
PEX1 PEX6 EDE1 SKN7 ERG4 ADH1 PEX15 ARC18 ECM33	protein import into peroxisome matrix (3/9) receptor recycling (3/9)
COG6 YPT7 MVB12 YPT52 COG5 CCZ1 ARF3 YPT53 ENT3	vacuolar transport (8/9) protein targeting to vacuole (6/9) establishment of protein localization (9/9) GTP binding (4/9)
VPS35 YPT31 BOR1 YOL019W PEP8	retrograde transport, endosome to Golgi (3/5)

This table lists BPMs which have both modules enriched and do not include all of the COG 5, 6, 7 and 8 genes. Selected enriched GO terms are included, and bolded genes have been enriched for at least one of the GO terms listed. For complete GO enrichment results, please see the Additional file 1.

Table 3 Singly enriched BPMs

BPM module	Selected enriched GO terms
VPS35 GET3 ARL3 SYS1 GOT1 PEP8 SFT2 ARE1	protein transport (7/8) endosome transport (5/8)
PEX6 BZZ1 SKN7 BLS1 PEX1 TWF1 PEX15 ARC18 JSN1 ECM33	protein import into peroxisome matrix, receptor recycling (3/10)
SNC1 GCS1 SRO7	Golgi to plasma membrane transport (3/3)
YPT7 MVB12 ARF3 CCZ1 YPT52 VPS41 STD1 ENT3	establishment of protein localization (7/8)
VPS27 YOL019W VPS4	protein retention in Golgi apparatus (2/3) maintenance of protein location in cell (2/3)
RVS161 RVS167 SKN1	lipid tube assembly (2/3)
RVS161 PKH3 ASG7	conjugation with cellular fusion (2/3)

This table lists enriched modules from BPMs which have only one module enriched for GO terms. Selected enriched GO terms are included, and bolded genes have been enriched for at least one of the GO terms listed. For complete GO enrichment results, please see the Additional file 1.

enrichment values. Genecentric has several features that make using it both easy and fast:

- Its only dependency is Python—no compilation steps or third party libraries are required. Any system capable of running Python should also be able to run Genecentric.
- All of the parameters of the algorithm are easily configurable using Genecentric's command line interface. (Namely, M , C , and J as used in the aforementioned algorithm description.) The parameters are by default set to those used in [5]: $M = 250$, $C = 0.9$ and $J = 0.66$.
- Genecentric automatically takes advantage of multiple CPUs.
- Input data can be read directly from genetic interaction data files. The user may specify an additional set of genes to *exclude* from BPM generation (i.e., a list of essential genes).
- Genecentric is species and gene identifier agnostic. (i.e., Genecentric does not care which kind of gene identifiers are used.)
- Code is well documented and could be extended easily.

Finally, we provided an extension to Genecentric that can perform Gene Ontology (GO) enrichment on a set of BPMs. Genecentric uses FuncAssociate's web API to achieve this, and parameters like genespace, namespace and p-value are configurable on the command line.

Availability and requirements

Project name: Genecentric

Project home page: <http://bcf.cs.tufts.edu/genecentric>

Operating system(s): Platform independent

Programming language: Python

Other requirements: Python 2.6 or higher

License: GNU General Public License Version 2

Any restrictions to use by non-academics: No

Additional file

Additional file 1: All BPMs generated with GO enrichment. A full list of BPMs generated by Genecentric from E-MAP data with 374 genes involved in various aspects of plasma-membrane biology, including endocytosis, signaling, lipid metabolism and eisome function.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceived and designed the project: AG and BH. Algorithmic development: ML, LC and BH. Implemented the software: AG. Analyzed the data: AG, MK, LC and BH. Wrote the paper: AG, LC and BH. All authors read and approved the final manuscript.

Acknowledgements

Thanks to Noah Daniels for many helpful discussions. We would also like to thank Kenny Chua and the rest of the FuncAssociate development team for incredibly quick and helpful responses to our questions. Finally, we extend our thanks to the anonymous reviewers who provided feedback that made both our code and paper better.

Funding

This work was partially funded by the National Institutes of Health (grant 1R01GM080330 to LC).

Author details

¹Department of Computer Science, Tufts University, Medford, MA 02155, USA.

²Department of Computer Science, Brown University, Providence, RI 02912, USA.

Received: 15 June 2012 Accepted: 27 December 2012

Published: 18 January 2013

References

1. Cordell HJ: **Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans.** *Human Mol Genet* 2002, **11**(20):2463–8. <http://www.ncbi.nlm.nih.gov/pubmed/12351582>.
2. Tong A, Lesage G, Bader G, Ding H: **Global mapping of the yeast genetic interaction network.** *Science's* 2004, **303**(February):808–813. <http://stke.sciencemag.org/cgi/content/abstract/sci;303/5659/808>.
3. Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD: **A DNA integrity network in the yeast *Saccharomyces cerevisiae*.** *Cell* 2006, **124**(5):1069–81. <http://www.ncbi.nlm.nih.gov/pubmed/16487579>.
4. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, Punna T, Ihmels J, Andrews B, Boone C, Greenblatt JF, Weissman JS, Krogan NJ: **Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile.** *Cell* 2005, **123**(3):507–19. <http://www.ncbi.nlm.nih.gov/pubmed/16269340>.
5. Leiserson MDM, Tatar D, Cowen LJ, Hescott BJ: **Inferring Mechanisms of Compensation from E-MAP and SGA Data Using Local Search Algorithms for Max Cut.** *J Comput Biol* 2011, **18**(11):1399–409. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3216108&tool=pmcentrez&rendertype=abstract>.
6. Bandyopadhyay S, Kelley R, Krogan NJ, Ideker T: **Functional maps of protein complexes from quantitative genetic interaction data.** *PLoS Comput Biol* 2008, **4**(4):e1000065. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2289880&tool=pmcentrez&rendertype=abstract>.
7. Bellay J, Atluri G, Sing TL, Toufighi K, Costanzo M, Ribeiro PSM, Pandey G, Baller J, VanderSluis B, Michaut M, Han S, Kim P, Brown GW, Andrews BJ, Boone C, Kumar V, Myers CL: **Putting genetic interactions in context through a global modular decomposition.** *Genome Res* 2011, **21**(8):1375–87. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3149503&tool=pmcentrez&rendertype=abstract>.
8. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JLY, Toufighi K, Mostafavi S, Prinz J, St Onge RP, VanderSluis B, Makhnevych T, Vizeacoumar FJ, Alizadeh S, Bahr S, Brost RL, Chen Y, Cokol M, Deshpande R, Li Z, Lin ZY, Liang W, Marback M, Paw J, San Luis BJ, Shuteriqi E, Tong AHY, van Dyk N, Wallace IM, Whitney Ja, Weirauch MT, Zhong G, Zhu H, Houry Wa, Brudno M, Ragibizadeh S, Papp B, Pál C, Roth FP, Giaever G, Nislow C, Troyanskaya OG, Bussey H, Bader GD, Gingras AC, Morris QD, Kim PM, Kaiser Ca, Myers CL, Andrews BJ, Boone C: **The genetic landscape of a cell.** *Science (New York, N.Y.)* 2010, **327**(5964):425–31. <http://www.ncbi.nlm.nih.gov/pubmed/21945515>.
9. Kelley DR, Kingsford C: **Extracting between-pathway models from E-MAP interactions using expected graph compression.** *J Comput Biol* 2011, **18**(3):379–90. <http://www.ncbi.nlm.nih.gov/pubmed/21385041>.
10. Ulitsky I, Shlomi T, Kupiec M, Shamir R: **From E-MAPs to module maps: dissecting quantitative genetic interactions using physical interactions.** *Mol Syst Biol* 2008, **4**(209):209. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2516364&tool=pmcentrez&rendertype=abstract>.
11. Ma X, Tarone AM, Li W: **Mapping genetically compensatory pathways from synthetic lethal interactions in yeast.** *PLoS one* 2008, **3**(4):e1922. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2275788&tool=pmcentrez&rendertype=abstract>.

12. Brady A, Maxwell K, Daniels N, Cowen LJ: **Fault tolerance in protein interaction networks: stable bipartite subgraphs and redundant pathways.** *PLoS one* 2009, **4**(4):e5364. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2670499&tool=pmcentrez&rendertype=abstract>.
13. Kelley R, Ideker T: **Systematic interpretation of genetic interactions using protein networks.** *Nat Biotechnol* 2005, **23**(5):1–14. <http://www.nature.com/nbt/journal/v23/n5/abs/nbt1096.html>.
14. Ulitsky I, Shamir R: **Pathway redundancy and protein essentiality revealed in the Saccharomyces cerevisiae interaction networks.** *Mol Syst Biol* 2007, **3**(104):104. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1865586&tool=pmcentrez&rendertype=abstract>.
15. Hescott BJ, Leiserson MDM, Cowen LJ, Slonim DK: **Evaluating between-pathway models with expression data.** *J Comput Biology* 2010, **17**(3):477–87. <http://www.ncbi.nlm.nih.gov/pubmed/20377458>.
16. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, Schuldiner M, Gebbia M, Recht J, Shales M, Ding H, Xu H, Han J, Ingvarsdottir K, Cheng B, Andrews B, Boone C, Berger SL, Hieter P, Zhang Z, Brown GW, Ingles CJ, Emili A, Allis CD, Toczyski DP, Weissman JS, Greenblatt JF, Krogan NJ: **Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map.** *Nature* 2007, **446**(7137):806–10. <http://www.ncbi.nlm.nih.gov/pubmed/17314980>.
17. Berriz GF, Beaver JE, Cenik C, Tasan M, Roth FP: **Next generation software for functional trend analysis.** *Bioinformatics (Oxford, England)* 2009, **25**(22):3043–4. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2800365&tool=pmcentrez&rendertype=abstract>.
18. Aguilar PS, Fröhlich F, Rehman M, Shales M, Ulitsky I, Olivera-Couto A, Braberg H, Shamir R, Walter P, Mann M, Ejsing CS, Krogan NJ, Walther TC: **A plasma-membrane E-MAP reveals links of the eisosome with sphingolipid metabolism and endosomal trafficking.** *Nat Struct & Mol Biol* 2010, **17**(7):901–908. <http://www.nature.com/doi/10.1038/nsmb.1829>.
19. Smith R, Willett R, Kudlyk T, Pokrovskaya I: **The COG complex, Rab6 and COPI define a novel Golgi retrograde trafficking pathway that is exploited by SubAB toxin.** *Traffic* 2009, **10**(10):1502–1517. <http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0854.2009.00965.x/full>.
20. Ungar D, Oka T, Brittle EE, Vasile E, Lupashin VV, Chatterton JE, Heuser JE, Krieger M, Waters MG: **Characterization of a mammalian, Golgi-localized protein complex, COG, that is required for normal Golgi morphology and function.** *J Cell Biol* 2002, **157**(3):405–15. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2173297&tool=pmcentrez&rendertype=abstract>.
21. Steet R: **COG-7-deficient human fibroblasts exhibit altered recycling of Golgi proteins.** *Mol Biol Cell* 2006, **17**(May):2312–2321. <http://www.molbiolcell.org/content/17/5/2312.short>.
22. Fotso P, Koryakina Y, Pavliv O, Tsiomenko AB, Lupashin VV: **Cog1p plays a central role in the organization of the yeast conserved oligomeric Golgi complex.** *J Biol Chem* 2005, **280**(30):27613–23. <http://www.ncbi.nlm.nih.gov/pubmed/15932880>.
23. Ng BG, Sharma V, Sun L, Loh E, Hong W, Tay SKH, Freeze HH: **Identification of the first COG-CDG patient of Indian origin.** *Mol Genet Metab* 2011, **102**(3):364–7. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3058693&tool=pmcentrez&rendertype=abstract>.
24. Ram R, Li B: **Identification of Sec36p, Sec37p, and Sec38p: components of yeast complex that contains Sec34p and Sec35p.** *Mol Biol Cell* 2002, **13**(May):1484–1500. <http://www.molbiolcell.org/content/13/5/1484.short>.
25. Whyte JR, Munro S: **The Sec34/35 Golgi transport complex is related to the exocyst, defining a family of complexes involved in multiple steps of membrane traffic.** *Dev Cell* 2001, **1**(4):527–37. <http://www.ncbi.nlm.nih.gov/pubmed/11703943>.
26. Santiago-Tirado FH, Bretscher A: **Membrane-trafficking sorting hubs: cooperation between PI4P and small GTPases at the trans-Golgi network.** *Trends Cell Biol* 2011, **21**(9):515–25. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3164296&tool=pmcentrez&rendertype=abstract>.
27. Moss J, Vaughan M: **Molecules in the ARF orbit.** *J Biol Chem* 1998, **273**(34):21431–4. <http://www.ncbi.nlm.nih.gov/pubmed/9705267>.
28. Moss J, Vaughan M: **Activation of toxin ADP-ribosyltransferases by eukaryotic ADP-ribosylation factors.** *Mol Cell Biochem* 1999, **193**(1-2):153–7. <http://www.ncbi.nlm.nih.gov/pubmed/10331652>.
29. Poon P, Wang X, Rotman M: **Saccharomyces cerevisiae Gcs1 is an ADP-ribosylation factor GTPase-activating protein.** *Proc Natl Acad Sci USA* 1996, **93**(September):10074–10077. <http://www.pnas.org/content/93/19/10074.short>.
30. Bensen ES, Yeung BG, Payne GS: **Ric1p and the Ypt6p GTPase function in a common pathway required for localization of trans-Golgi network membrane proteins.** *Mol Biol Cell* 2001, **12**:13–26. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=30564&tool=pmcentrez&rendertype=abstract>.
31. Lorente-Rodríguez A, Heidtman M, Barlowe C: **Multicopy suppressor analysis of thermosensitive YIP1 alleles implicates GOT1 in transport from the ER.** *J Cell Sci* 2009, **122**(Pt 10):1540–50. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2680100&tool=pmcentrez&rendertype=abstract>.
32. Panic B, Whyte JRC, Munro S: **The ARF-like GTPases Arl1p and Arl3p act in a pathway that interacts with vesicle-tethering factors at the Golgi apparatus.** *Curr Biol: CB* 2003, **13**(5):405–10. <http://www.ncbi.nlm.nih.gov/pubmed/12620189>.
33. Gerrard SR, Levi BP, Stevens TH: **Pep12p is a multifunctional yeast syntaxin that controls entry of biosynthetic, endocytic and retrograde traffic into the prevacuolar compartment.** *Traffic (Copenhagen, Denmark)* 2000, **1**(3):259–69. <http://www.ncbi.nlm.nih.gov/pubmed/11208109>.
34. Trautwein M, Schindler C, Gauss R, Dengjel J, Hartmann E, Spang A: **Arf1p, Chs5p and the ChAPs are required for export of specialized cargo from the Golgi.** *EMBO J* 2006, **25**(5):943–54. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1409733&tool=pmcentrez&rendertype=abstract>.

doi:10.1186/1471-2105-14-23

Cite this article as: Gallant *et al.*: Genecentric: a package to uncover graph-theoretic structure in high-throughput epistasis data. *BMC Bioinformatics* 2013 **14**:23.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

