The genetic interaction ontology (GIO) incorporating the genetic interactions structured terminology (GIST)

Christian A. Grove[1], Rose W. Oughtred[2], Kara Dolinski[2], Mike Tyers[3,4], Paul Sternberg[1]
1. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
2. Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA
3. The Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada
4. Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Québec H3T 1J4 Canada

ABSTRACT

Genetic interactions have been studied for more than a century as a means to elucidate gene function. Built on prior work by the authors, a new unified genetic interaction ontology (GIO) incorporating the genetic interactions structured terminology (GIST) is proposed for inclusion into the Human Proteome Organization (HUPO) Protein Standards Initiative (PSI) Molecular Interactions (MI) controlled vocabulary (https://github.com/HUPO-PSI/psi-mi-CV). The updated, proposed ontology includes a restructuring of the upper level of the ontology, clear references to definitions and meanings described in the literature, direct incorporation of genetic interaction terms from the Biological General Repository for Interaction Datasets (BioGRID) database (https://thebiogrid.org/), relabeling of some terms, GIST synonyms for existing terms, obsoletion of outdated terms, and some new proposed terms to incorporate GIST and ensure consistent ontology structure. An important distinction is made between genetic interactions that are reported as modifications of an existing phenotype versus genetic interactions that defy expectation (i.e. deviate significantly from an inherent or mathematically determined expected double genetic perturbation phenotype), as both meanings have been used in the literature and in biological data repositories. The proposed ontology can accommodate many existing use cases for quantitative genetic interactions described in the model organism literature, but is not intended or purported to be exhaustive or comprehensive. Suggestions for edits or new terms from the genetics community are strongly encouraged.

CHALLENGE

Biological and genetics databases curate and catalog various genetic interactions from the literature. Currently, genetic interactions are curated using different classification schemes and vocabularies making it difficult to integrate these data and draw consistent conclusions from them. To best integrate genetic interactions across different databases, and consistently infer the correct biological conclusions from those interactions, it is important that databases use a common vocabulary and classification scheme so as to appropriately compare and contrast these interactions. A structured genetic interaction ontology (GIO) of interaction types is needed to help genetics researchers construct genetic pathway models by effectively drawing consistent conclusions from genetic interactions curated by different databases.

GOAL

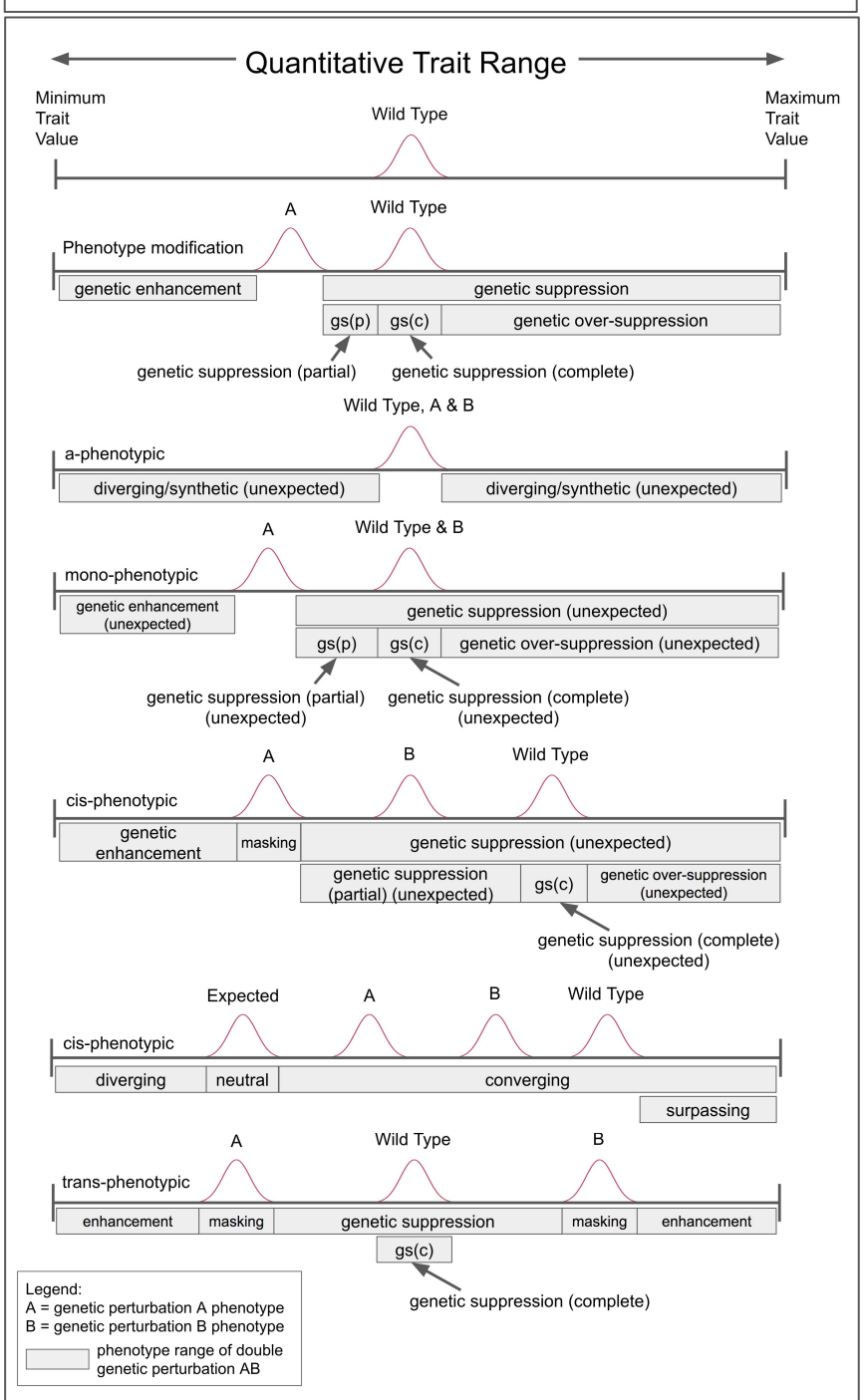
The goal of this work is to compare and contrast the different genetic interaction types used by different databases, draw equivalencies between appropriate terms/types, and develop a classification hierarchy within an ontology so as to understand, and to help compute over, the relationships between different genetic interaction types. To place existing genetic interaction types into a common ontology, some new terms will also need to be instantiated as grouping terms that act as a common superclass for various related genetic interaction types. Once the GIO is built, genetic interactions from different databases can be qualitatively compared and used to consistently construct genetic pathways for various model organisms.

Genetic Interaction: Phenotype Modification vs. Unexpected

Whereas many published definitions of "genetic interaction" assert that there must be a deviation from an expected double perturbation phenotype[1-4], many examples of genetic interactions in the literature and in biological databases[5-7] include phenotype modifications like suppression or enhancement that are not necessarily reported with an expected double perturbation trait value, precluding an analysis of expectation or deviation from it. For this reason, this genetic interaction ontology has two terms for genetic interaction distinguished by their meanings as either "sensu phenotype modification" or "sensu unexpected". Both types of interactions provide meaningful insights into the biology of the genes in question, yet they have distinct sets of conclusions that can be drawn from them. This will be important for future curation and interpretation of genetic interactions catalogued by genetic interaction databases.

Genetic Interactions Structured Terminology (GIST)

As presented previously[8], the GIST system aims to classify quantitative genetic interaction types on three semi-orthogonal attributes, termed GIST modules 1-3. **GIST module 1** classifies genetic interactions based on (if known) the single genetic perturbation phenotypes with respect to wild type (or control) and to each other. Terms from GIST module 1 are as follows: **a-phenotypic**, meaning that both single genetic perturbations lack an observable phenotype (difference from wild type/control) with respect to the phenotype/trait in question; **mono-phenotypic**, meaning that one single genetic perturbation results in an observable phenotype whereas the other single genetic perturbation lacks an observable phenotype; **cis-phenotypic**, meaning that both single genetic perturbations result in an observable phenotype, and both phenotypes deviate from wild type/control in the same direction (e.g. both increased trait quality or both decreased trait quality); **iso-phenotypic**, meaning a subclass of "cis-phenotypic" in which both single genetic perturbations result in the same quantitative phenotype (statistically indistinguishable from each other); and **trans-phenotypic**, meaning that both single genetic perturbations result in an observable phenotype, and both phenotypes deviate from wild type/control but in the opposite direction (e.g. one exhibits an increased trait quality and the other a decreased trait quality). **GIST module 2** classifies genetic interactions based on the double genetic perturbation phenotype with respect to the single genetic perturbation phenotypes and wild type/control. This module includes familiar genetic interaction type terms such as **enhancing/enhancement**, **suppressing/suppression**, **masking/epistasis**, **synthetic** and more novel terms such as **super-suppressing/over-suppression**, meaning that a phenotype is suppressed beyond wild type/control (resulting in an opposite quantitative phenotype). **GIST module 3** classifies genetic interactions based on the observed double genetic perturbation phenotype with respect to the expected (if stated or inferable) double genetic perturbation phenotype. Terms from GIST module 3 are as follows: **diverging**, meaning that the observed double genetic perturbation phenotype is further from wild type/control than the expected double genetic perturbation phenotype; **converging**, meaning that the observed double genetic perturbation phenotype is closer to (or in the direction of) wild type/control than the expected double genetic perturbation phenotype; **surpassing**, a subclass of converging meaning that the observed double genetic perturbation phenotype is beyond wild type/control compared to the expected double genetic perturbation phenotype; and **neutral**, meaning that the observed double genetic perturbation phenotype is equivalent to the expected double genetic perturbation phenotype. Together, these terms provide detailed expressivity about the nature of curated genetic interactions so that the appropriate interpretation may be applied in each case when constructing genetic pathways.



What is an ontology? Why an ontology?

Ontologies are structured controlled vocabularies, the core of which consists of a taxonomy relating all terms/classes with superclass/supertype, subclass/subtype relationships, probably the most widely used example in the biological sciences being the Gene Ontology[9,10]. Ontologies provide a logical structure with which databases and query tools can make useful inferences about which terms relate to other terms and how. For the GIO, the goal is that such a structure will eventually help enable automatic genetic pathway construction based on genetic interaction data.

GIO Structure (partial)

- multigenic phenotype result
  - deleterious multigenic phenotype result
    - dosage growth defect (sensu BioGRID)
    - negative genetic interaction (sensu BioGRID)
    - synthetic growth defect (sensu BioGRID)
    - synthetic haploinsufficiency (sensu BioGRID)
  - genetic interaction (sensu phenotype modification)
    - genetic enhancement
      - genetic enhancement (sensu unexpected)
        - mutual genetic enhancement (sensu unexpected)
        - monophenotypic genetic enhancement
      - mutual genetic enhancement (expected)
      - phenotypic enhancement (sensu BioGRID)
    - genetic suppression
      - dosage rescue (sensu BioGRID)
      - genetic over-suppression
      - genetic suppression (complete)
      - genetic suppression (partial)
      - genetic suppression (sensu unexpected)
      - mutual genetic suppression
      - phenotypic suppression (sensu BioGRID)
      - positive genetic interaction (sensu BioGRID)
      - synthetic rescue (sensu BioGRID)
  - genetic interaction (sensu unexpected)
    - converging genetic interaction
      - asynthetic genetic interaction
      - converging genetic epistasis
      - genetic suppression (sensu unexpected)
      - positive genetic interaction
    - diverging genetic interaction
      - genetic enhancement (sensu unexpected)
      - negative genetic interaction
      - synthetic genetic interaction
    - genetic epistasis (sensu Bateson)
      - qualitative genetic epistasis
      - quantitative genetic epistasis
    - surpassing genetic interaction
      - mutual genetic over-suppression
      - monophenotypic genetic over-suppression
    - transphenotypic enhancing genetic interaction
  - neutral multigenic phenotype result
    - aphenotypic neutral multigenic phenotype result
    - mutual genetic enhancement (expected)
    - transphenotypic genetic suppression (expected)

CONCLUSION

The entire genetic interaction ontology (GIO) proposal can be downloaded as part of the PSI-MI controlled vocabulary in OBO format here: https://bit.ly/2yoidCe. This file can be opened in the ontology editor Protege (https://protege.stanford.edu/). To navigate to the GIO within the file, navigate to "interaction type" > "phenotype result". Whenever possible, definitions provided in the literature[1-4,11-13] were cited in the ontology as cross-referenced definitions for terms. The genetic interaction type terms included in this ontology are directly used by the WormBase[5], BioGRID[6], and FlyBase[7] databases. Thus, the construction of the GIO has been an attempt to reconcile the different terminologies used and provide an overarching framework within which each of these terms can exist with its appropriate relations to all other such terms used. Future efforts will include assimilation of this ontology for use in the Alliance of Genome Resources[15] (www.alliancegenome.org) for all genetic interaction classification needs. Much of the development of this ontology was inspired by work done to derive all possible observable phenotype inequalities between the four genotypes in question (wild type, genetic perturbation A, genetic perturbation B, and double genetic perturbation AB), observe their occurrences in yeast genetic interaction networks and derive useful functional biological insights from them for genetic network analysis. We believe the use of this ontology with such approaches may not only uncover genetic network modularity but also causal linear genetic pathways that often lay the groundwork for detailed biochemical pathways that help elucidate the mechanisms and potential treatment options for human disease.

References

1. Eddy SR. Genetics. Total information awareness for worm genetics. Science. 2006 Mar 10;311(5766):1381-2. PubMed PMID: 16527956.
2. Mani R, St Onge RP, Hartman JL 4th, Giaever G, Roth FP. Defining genetic interaction. Proc Natl Acad Sci U S A. 2008 Mar 4;105(9):3461-6. doi: 10.1073/pnas.0712255105. Epub 2008 Feb 27. PubMed PMID: 18305163; PubMed Central PMCID: PMC2265146.
3. Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C. Systematic mapping of genetic interaction networks. Annu Rev Genet. 2009;43:601-25. doi: 10.1146/annurev.genet.39.073003.114751. Review. PubMed PMID: 19712041.
4. Baryshnikova A, Costanzo M, Myers CL, Andrews B, Boone C. Genetic interaction networks: toward an understanding of heritability. Annu Rev Genomics Hum Genet. 2013;14:111-33. doi: 10.1146/annurev-genom-082509-141730. Epub 2013 Jun 26. Review. PubMed PMID: 23808365.
5. Harris TW, Arnaboldi V, Cain S, Chan J, Chen WJ, Cho J, Davis P, Gao S, Grove CA, Kishore R, Lee RYN, Muller HM, Nakamura C, Nuin P, Paulini M, Raciti D, Rodgers FH, Russell M, Schindelman G, Auken KV, Wang Q, Williams G, Wright AJ, Yook K, Howe KL, Schedl T, Stein L, Sternberg PW. WormBase: a modern Model Organism Information Resource. Nucleic Acids Res. 2020 Jan 8;48(D1):D762-D767. doi: 10.1093/nar/gkz920. PubMed PMID: 31642470.
6. Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, Kolas N, O'Donnell L, Leung G, McAdam R, Zhang F, Dolma S, Willems A, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. The BioGRID interaction database: 2019 update. Nucleic Acids Res. 2019 Jan 8;47(D1):D529-D541. doi: 10.1093/nar/gky1079. PubMed PMID: 30476227; PubMed Central PMCID: PMC6324058.
7. Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, Kaufman TC, Calvi BR; FlyBase Consortium . FlyBase 2.0: the next generation. Nucleic Acids Res. 2019 Jan 8;47(D1):D759-D765. doi: 10.1093/nar/gky1003. PubMed PMID: 30364959; PubMed Central PMCID: PMC6323960.
8. Grove CA, Oughtred RW, Lee R, Dolinski K, Tyers M, Sternberg P, Baryshnikova A. Genetic Interactions Structured Terminology (GIST): A new standard for describing and annotating cross-species genetic interactions data. Biocuration 2017, Stanford University, Stanford CA.
9. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25-9. PubMed PMID: 10802651; PubMed Central PMCID: PMC3037419.
10. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019 Jan 8;47(D1):D330-D338. doi: 10.1093/nar/gky1055. PubMed PMID: 30395331; PubMed Central PMCID: PMC6323945.
11. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. Nat Rev Genet. 2007 Jun;8(6):437-49. Review. PubMed PMID: 17510664.
12. St Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, Nislow C, Roth FP, Giaever G. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet. 2007 Feb;39(2):199-206. Epub 2007 Jan 7. PubMed PMID: 17206143; PubMed Central PMCID: PMC2716756.
13. Drees BL, Thorsson V, Carter GW, Rives AW, Raymond MZ, Avila-Campillo I, Shannon P, Galitski T. Derivation of genetic interaction networks from quantitative phenotype data. Genome Biol. 2005;6(4):R38. Epub 2005 Mar 31. PubMed PMID: 15833125; PubMed Central PMCID: PMC1088966.
14. Carter GW, Galas DJ, Galitski T. Maximal extraction of biological information from genetic interaction data. PLoS Comput Biol. 2009 Apr;5(4):e1000347. doi: 10.1371/journal.pcbi.1000347. Epub 2009 Apr 3. PubMed PMID: 19343223; PubMed Central PMCID: PMC2659753.
15. Alliance of Genome Resources Consortium. The Alliance of Genome Resources: Building a Modern Data Ecosystem for Model Organism Databases. Genetics. 2019 Dec;213(4):1189-1196. doi: 10.1534/genetics.119.302523. PubMed PMID: 31796553; PubMed Central PMCID: PMC6893393.