## Algorithm

The main program has many steps:

1.- The algorithm gets the input data and digests it:

\* Every gene is a node in the network.

\* Links are every assay between three genes.

\* This links are tagged with 0 or 1 if there is interaction or not.

2.- Many data structures are initialized with the input:

\* 2D-Matrix where every gene has its vector of possibilities of behaving like one of the groups of genes.

\* 3D-Matrix where every group of genes has a matrix of possibilities of interact with the other two group of genes.

3.- The algorithm starts to iterate to get the maximization of the likelihood.

4.- If gets the covergement a message will be shown.

5.- All the data from the execution will be saved automatically for every sample even if it doesn't converge.

## Usage

To execute the code and speeding it up you'll use a \*\*pypy3\*\* virtual environment. There's one uploaded in the repository in the folder `src/`. To execute it open a terminal and situate it in the `src` folder. Then type, for executing the algorithm with the default options:

```

pypy3-v6.0.0-linux64/bin/pypy3 TrigenicInteractionPredictor.py

```

These default arguments are:

\* iterations = 1000

\* samples = 1

\* frequencyCheck = 1

\* filename = "Data\_S1.csv"

\* interactionType = "trigenic"

\* cutOffValue = -0.08

\* argk = 10

You can use other arguments, specifying \*\*\_all\_\*\* of them after the text that executes the program:

1.- iterations[positive integer]: Number of iterations done by algorithm.

2.- samples[positive integer]: Number of samples done by algorithm.

3.- frequencyCheck[positive integer] Number of iterations needed to check if likelihood has converged.

4.- filename[string]: Name of the dataset filename.

5.- interactionType[string:{Trigenic,Digenic,\\*}]: Type of interaction selected.

6.- cutOffValue[real]: Value used to determine if an interaction is positive or negative.

7.- argk[integer]: Number of groups to use in the algorithm (Increases lineally the computation cost).

For example:

```

pypy3-v6.0.0-linux64/bin/pypy3 TrigenicInteractionPredictor.py 1000 5 10 Data\_S1.csv Trigenic -0.08

```

## Code Health

[![Codacy Badge](https://api.codacy.com/project/badge/Grade/51cacbf196634b1f81521e09bfdc9617)](https://www.codacy.com/app/AleixMT/TrigenicInteractionPredictor?utm\_source=github.com&amp;utm\_medium=referral&amp;utm\_content=AleixMT/TrigenicInteractionPredictor&amp;utm\_campaign=Badge\_Grade)

## Authors

\* \*\*Aleix Mariné\*\* - [AleixMT](https://github.com/AleixMT) [aleix.marine@estudiants.urv.cat](aleix.marine@estudiants.urv.cat)

\* \*\*Marta Sales-Pardo\*\* - [seeslab](https://github.com/seeslab) [marta.sales@urv.cat](marta.sales@urv.cat)

\* \*\*Roger Guimerà\*\* - [roger.guimera@urv.cat](roger.guimera@urv.cat)

Genetic interactions occur when two or more mutations in different genes combine to result in a phenotype that is different from the expected phenotype when these mutations are tested separately in different individuals.

For example, let’s take a genetic interaction of two genes as a base:

Take x0 and x1 as two different genes.

Let f be the transformation from genotype to phenotype.

A genetic interaction occurs when f(x0+x1) != f(x0)+f(x1).

*Types of genetic interaction*

We’ll consider two main types of interaction:

* **Negative genetic interaction:** Occurs when a combination of mutations leads to a fitness defect that is more exacerbated than expected.
  + **Synthetic lethality** occurs when two non-letal mutations generate a non-viable mutant when combined.
* **Positive genetic interaction:** Occurs when a combination of mutations leads to a fitness greater than expected.
  + Genetic suppression: Occurs when the mutations in the fitness defect of a query mutant is alleviated by a mutation in a second gene.

→ Phenotypic variation is caused by genetic determinants that act as modifiers. By the moment we know where these *loci* are but we don’t know how they interact. As always, we understand DNA in a static way, but genetics actually behave in a dynamic way.

*Results and Data*

In this experimented they tested these three key factor from the digenic network interaction.

1. Digenic interaction strength
2. Average mutations.
3. Digenic interaction profile similarity.

We used the colony size as measure for the fitness.

*Data Size*

In total, we tested ~400,000 double and ~200,000 triple mutants for fitness defects and identified ~9500 digenic and ~3200 trigenic negative interactions.

//~ About One third of the trigenic interactions were enriched for functional relationship.

* One third of the trigenic interactions were “novel” (not observed in digenic control network).
* The rest (two thirds) of trigenic interactions modified an existing digenic interaction.
* Only ~1000 of the ~6000 total yeast genes are individually essential and cause lethality when deleted.
* ~550,000 different yeast gene pairs display a combinatorial negative genetic interaction, including a subset of ~10,000 extreme synthetic lethal interactions involving nonessential gene pairs.

→ The set of digenic interactions **(genetic interaction profile)** from a query gene provides a quantitative measure of function.

→ Genes with similar roles have overlapping profiles.

→ Genes in the same biological pathway have similar genetic interaction profiles.

→ A global network based on digenic interactio profile shows an hierarchical model.

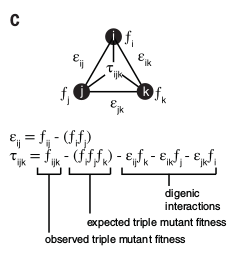
cellular compartments

bioprocesses

functional modules

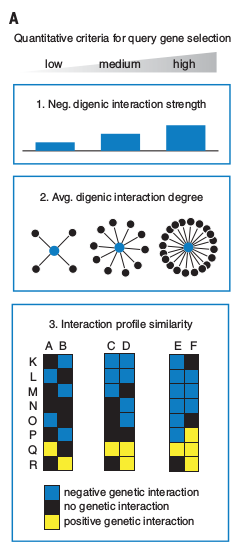
*Scoring method*

Third, we developed a scoring method, the t-SGA score, which combines double- and triple- mutant fitness estimates derived from colony size measurements to identify trigenic interactions quantitatively (Fig. 1C). The t-SGA score differs from the MinDC score reported previously (18), because it accounts for all cases in which two of the genes are not independent, resulting in an expectation that contains digenic interaction effects scaled by the fitness of the noninteracting genes (fig. S4) (16). The final trigenic t-SGA interaction score then accounts for digenic effects but also enables detection of trigenic interactions in which digenic effects of insufficient explanatory power can be found.



Triple-mutant

SGA quantitative scoring strategy. The top equation in C shows the quantification of a digenic interaction, where e ij is the digenic interaction score, ƒ ij is the observed double-mutant fitness, and the expected double-mutant fitness is expressed as the product of single-mutant fitness estimates ƒ i ƒ j . In the bottom equation, the trigenic interaction score (t ijk ) is derived from the digenic interaction score, where ƒ ijk is the observed triple-mutant fitness and ƒ i ƒ j ƒ k is the triple-mutant fitness expectation expressed as the product of three single-mutant fitness estimates. The influence of digenic interactions is subtracted from the expectation, and each digenic interaction is scaled by the fitness of the third mutation



Criteria for selecting query strains for sampling trigenic interaction landscape of singleton genes in yeast. The gene pairs were grouped into three general categories based on a range of features:

* (1) Digenic interaction strength. Gene pairs were directly connected by zero to very weak (digenic interaction score: 0 to –0.08, n = 74 strains), weak (–0.08 to –0.1, n = 32), or moderate (<–0.1, n = 45) negative digenic interactions.
* (2) Number of digenic interactions. Gene pairs had a low (10 to 45 interactions, n = 50), intermediate (46 to 70, n = 53), or high (>71, n = 48) average digenic interaction degree (denoted by the number of black edges of each node).
* (3) Digenic interaction profile similarity. Gene pairs had low (score: –0.02 to 0.03, n = 46; represented by genes A and B, which show a relatively low overlap of genetic interactions with genes K to R), intermediate (0.03 to 0.1, n = 59; represented by genes C and D, which display an intermediate overlap of genetic interactions), or high (>0.1, n = 46, represented by genes E and F, which display a relatively high level of overlap of genetic interactions) functional similarity, as measured by digenic interaction profile similarity and coannotation to the same term(s). Query mutant genes were either nonessential deletion mutant alleles (D) or conditional temperature-sensitive (ts) alleles of essential genes.

We focused exclusively on the analysis of del-

eterious negative trigenic interactions for two

reasons. First, quantitative scoring of negative genetic interactions is often more accurate than  
that for positive interactions because there is a  
greater signal-to-noise ratio for negative genetic  
interactions. Hence, negative genetic interac-  
tions are associated with lower false-positive  
and false-negative rates than positive interactions  
(8), a feature that is important for the robust  
statistical analysis necessary to differentiate true  
trigenic interactions from the extensive back-  
ground digenic network. Second, negative digenic  
interactions are generally more functionally in-  
formative than positive digenic interactions (8),  
and thus the large-scale mapping of a negative  
trigenic interaction network is expected to pro-  
vide the most mechanistic insight into gene  
function and pathway wiring.

**Datasets**

### **Data File S1. Raw genetic interaction dataset.**

This file contains digenic interaction scores as well as raw and adjusted trigenic interaction scores in a tab-delimited format with 12 columns: 1) Query Strain ID, 2) Query allele name, 3) Array strain ID, 4) Array allele name, 5) Combined mutant type, 6) Raw genetic interaction score (ε), 7) Adjusted trigenic interaction score (τ), 8) p-value, 9) Query fitness, 10) Array single mutant fitness, 11) Combined mutant fitness relative to wild-type, 12) Combined mutant fitness standard deviation

Column description for each of the new supplemental files

1. Query Strain ID
2. Query Allele name
3. Array Strain ID
4. Array Allele name
5. Combined mutant type
   * ‘digenic’ for double mutants resulting from a cross between a single mutant control query and a single mutant array strain
   * ‘trigenic’ for triple mutants resulting from a cross between a double mutant query and a single mutant array strain
6. Raw genetic interaction score (ε)
7. Final genetic interaction score (trigenic τ / digenic ε). Trigenic scores are adjusted according to the τ-SGA model shown in Fig. S4A. Digenic scored receive no further adjustment and the epsilon value is repeated.
8. Interaction *p* value
9. Query fitness
   * single mutant fitness for single mutant queries
   * double mutant fitness for double mutant queries
10. Array single mutant fitness
11. Combined mutant fitness relative to wild-type.
    * Double mutant fitness for the resulting combined digenic mutants
    * Triple mutant fitness for the resulting combined trigenic mutants
12. Combined mutant fitness standard deviation

### **Data File S2. Digenic and adjusted trigenic interaction dataset.**

This file contains digenic and trigenic interaction scores at an established interaction magnitude cut-off for digenic interactions (*p* < 0.05, |ε| > 0.08) and trigenic interactions (*p* < 0.05, τ < -0.08) in a tab-delimited format with 8 columns: 1) Query Strain ID, 2) Query allele name, 3) Array strain ID, 4) Array allele name, 5) Combined mutant type, 6) Final genetic interaction score (tau), 7) p-value, 8) Interaction type

Column description for each of the new supplemental files

1. Query Strain ID
2. Query Allele name
3. Array Strain ID
4. Array Allele name
5. Combined mutant type
   * ‘digenic’ for double mutants resulting from a cross between a single mutant control query and a single mutant array strain
   * trigenic’ for triple mutants resulting from a cross between a double mutant query and a single mutant array strain
6. Final genetic interaction score (trigenic tau / digenic epsilon). Trigenic scores are adjusted according to the τ-SGA model shown in Fig. S4A and account for any digenic interactions. Digenic scores receive no further adjustment and the epsilon value is repeated from Additional Table S1.
7. Interaction *p* value
8. Interaction types:
   * **Digenic** is digenic
   * **Novel** is novel trigenic
   * **Unclassified** apparently novel but with unknown query-query interaction score thus cannot be distinguished from modified and novel
   * **Modified Q-, Modified Q-A-, Modified Q-A+, Modified Q-A+-, Modified A-, Modified A+** or **Modified A-+** are modified trigenic interactions are further broken down by where the overlapping digenic interaction is found: Q- for a negative interaction between query genes, A- for a negative interaction between one or both of the query genes and the array gene, A+ for a positive interaction between one or both of the query genes and the array gene, A+- if both query genes have a digenic interaction with the array but of opposing signs.

**Data Structures**

**Interaction Types**

Classifying trigenic interactions into novel vs. modified

Our model for trigenic interactions allows for a trigenic interaction involving two

genes connected by a digenic interaction, providing the triple mutant demonstrates a

significant deviation from the expected fitness of the double mutant when combined with

the third perturbation. We term cases where such an overlap exists as “modified” trigenic

interactions because the third perturbation exacerbates or alleviates a previously known

digenic interaction leading to a more extreme phenotype than expected, and thus can be

said to modify an existing interaction. Alternatively, we observe a “novel” trigenic

interaction in cases where none of the two gene-gene connections within the triad

overlaps with a previously known digenic interaction. In these cases, we have gained

novel functional information for genes that were not previously observed to interact in

digenic space. In practice, a trigenic interaction (τ ijk ) between a double mutant query (Q ij )

and an array (A k ) is called novel if there is no significant interaction between either single

mutant control query (Q i or Q j ) and the array (A k ), and also no interaction between querygene pair itself. Digenic interactions between Q i -A k or Q j -A k were measured using our

single mutant control queries. Query pair interactions (Q i -Q j ) were measured using the

single and double mutant fitness standard (Additional Data S4) and applying the

multiplicative model to derive the genetic interaction between the two query genes. Each

query mutant fitness score has an associated standard deviation, and these were combined

to calculate the expected variance of the double mutant fitness under the product. As

epsilon scores are approximately normally distributed, this expected variance can be used

to calculate a p-value. If any such digenic interaction exists, either positive or negative,

the trigenic interaction is called modified (Fig. S8A). We classified negative trigenic

interactions into 1,859 modified and 1,024 novel at the following thresholds for digenic

interactions (p < 0.05, |e| > 0.08) and trigenic interactions (p < 0.05, t < -0.08). A further

313 trigenic interactions have no overlapping digenic interaction but stem from double

mutant queries for which double mutant fitness estimates have not been generated in our

standard due to quality control and have corresponding NaN values. Due to this

uncertainty, these interactions have been withheld from the novel class in Fig. S8A-C.

Furthermore, 1,508 modified trigenic interactions overlap a significant negative digenic

interaction, 243 modified trigenic interactions overlap a significant positive digenic

interaction, and 108 modified trigenic interactions overlap both significant positive and

negative digenic interactions (Fig. S8A, Additional Data S2). To simplify the analysis

depicted in Fig. S8B and due to the relatively small class size of 8 modified trigenic

interactions that overlapped a query-query interaction and two digenic interactions (one

positive and one negative) involving query-array pair, this group was collapsed with two

other groups and half of these interactions were summed with (1) a class containing

modified trigenic interactions that overlapped a query-query interaction and a positive

query-array interaction, and half were summed with (2) a class containing modified

trigenic interactions that overlapped a query-query interaction and a negative query-array

interaction. The same was applied to 17 modified trigenic interactions that overlapped

two digenic interactions (one positive and one negative) involving query-array pair but

did not show a query-query interaction. This group was collapsed with two other groups

and half of these interactions were summed with a class containing modified trigenic

interactions that overlapped a positive query-array interaction and half were summed

with a class containing modified trigenic interactions that overlapped a negative query-

array interaction.

**Strain construction**

Dear Mr. Myers,

My name is Aleix Mariné, an intern in the sees:lab research group, in the University Rovira i Virgili in Tarragona (Spain).

I read your article “Systematic analysis of complex genetic interactions" (DOI: 10.1126/science.aao1729). I found this article really interesting, so as a project we decided to model your results as a graph and use inference techniques in order to make predictions of interaction between genes.

Unfortunately, we have doubts when we try to interpret your datasets, concretely, datasets S1 and S2. I was hoping you could help me solve them.

Every row in dataset S1 or S2 represents an interaction between two or three genes. The fifth column represent the type of interaction between pairs (digenic) or triplets (trigenic) of genes.

Especifically I am interested in obtaining a network in which each node is a gene.

If I am interpreting the table correctly, for trigenic interactions if I find xxx+yyy in the query allele column and zzz in the array allele column then the nodes(genes) for this interaction are xxx, yyy and zzz.

The problem occurs in the digenic interactions. In this case, I should obtain a single allele (mutated gene) in the query allele column of digenic interactions ( because we have one allele coming from the query strain and another one coming from the array strain). However, we have TWO alleles in this column, that is the format is yyy+xxx as for the trigenic interactions. Which allele xxx or yyy is the one I should consider for the graph representation of the digenic interaction?

For example, given these two interactions from dataset S2:

<pastedImage.png>

Which is the deleted gene in the query strain in row 789, nup60delta or hodelta? Knowing that is digenic just two alleles should be implicated in this interaction.

Also, I assume that given the query alleles "a" and the array allele "b" in a digenic interaction, the interaction score "epsilon" should be the same as if alleles "a" and "b" swap their positions. In other words, I assume that interaction between genes is non-directional and the consideration of "query alleles" and "array allele" is just a decision design of this assay. Is that correct?

Thank you for your patience and time.

Best regards,

Aleix Mariné i Tena

Dear Aleix,

The single mutant queries that were used to derive digenic interactions carry a deletion or a temperature sensitive allele in the gene of interest and a deletion in a benign locus HO as a control. HO deletion is marked with the same genetic marker as the 2nd gene in the corresponding double mutant query that is used for trigenic interactions. This is described in the supplement pg. 3-4 of the PDF. Digenic interactions are thus specific only for non-HO genes (see bottom of pg. 5 and top of pg. 12 of the supplement PDF).

Your interpretations for trigenic interactions are correct. For digenic interactions, the nodes are non-HO query gene and the array gene. The deleted gene in the query strain in row 789 is nup60delta and not hodelta.

As for AB and BA scenario, you can definitely do what you are suggesting and in cases where you see differences you may set the strongest interaction as your AB, BA pair. We had three different selectable markers so we didn’t want to introduce noise into our trigenic model that relied on subtraction but for your purposes and for the sake of simplicity it should be ok. Ben, do you have any comments on this?

Just to clarify S1 has raw and adjusted scores and S2 has adjusted scores at an intermediate cut-off only. If you’d like to use a lenient cut-off then don’t threshold on the adjusted score but only threshold on the pvalue <0.05 and remember to use the adjusted scores because they represent the trigenic interaction scores according to our model shown in Fig 1 of the paper and described on pg. 2, 6 and 7 of the supplement PDF.

I hope this helps. Please, let us know if you have any more questions.

Elena

--

* Nodes in digenic interactions are array allele and the non-HO query allele (hoDelta alleles are discarded).
* Interaction between two alleles is non-directional taking in account query and array alleles.
* We can be more permissive and change the cut-off values for tau and epsilon in order to get more interaction from raw dataset S1, always maintaining P<0.05.