Unravelling the molecular mechanisms of genetic interactions: **from yeast to cancer**

Saman Amini

Unravelling the molecular mechanisms of genetic interactions:
from yeast to cancer

ISBN: 978-94-92801-32-6
Cover and layout: Proefschrift-aio
Printed by: Proefschrift-aio

Unravelling the molecular mechanisms of genetic interactions: from yeast to cancer

Ontrafelen van de moleculaire mechanismen van genetische interacties: van gist tot kanker
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op donderdag 24 mei 2018 des middags te 12.45 uur

door

Seyedsaman Amini
geboren op 8 augustus 1985 te Baneh, Iran

Promotor: Prof. dr. F.C.P. Holstege

Copromotor: Dr. P. Kemmeren

"The universe is not only queerer than we imagine, but queerer than we can imagine."

John B S Haldane

Table of contents

General introduction: defining, identifying and understanding genetic interactions

Defining genetic interactions

DNA is the carrier of genetic information and can have a profound influence on health. A major challenge of biology and medicine is to understand how variation in DNA sequence can result in complex genetic diseases. Due to spectacular advances in genome sequencing technology, in particular during the last decade, many disease-causing variants have been linked to common human diseases. Despite these important discoveries we frequently still lack the ability to accurately predict the phenotype of an individual based on a combination of variants [1]. For many diseases, risk variants detected through genome-wide association studies (GWAS) can only explain 20-50% of heritability [1,2]. One of the possible reasons for this so-called 'missing heritability' could be genetic interactions or epistasis [1–5]. This concept is also more broadly known as the dependency of mutation on genetic background, in which the consequence of mutations varies across individuals [6–8]. A genetic interaction is an interaction between two alleles whereby the phenotypic outcome of the combination differs from the expected sum or product of the effects of the individual alleles. Currently, most genetic interactions are unpredictable, making accurate prediction of phenotype based on genotype very challenging. The general goal of the work described in this thesis is aimed at understanding the mechanisms of genetic interactions, in order to make such interactions more predictable. This introduction therefore begins by explaining genetic interactions in more detail, how they can be defined and identified, what is currently known about the underlying molecular mechanisms and how genetic interactions can be modeled. The introduction, as well as the work in this thesis, ends with a section about the involvement of genetic interactions in cancer, a disease that stems from changes in our DNA.

The most intuitive example of a genetic interaction is a redundancy relationship between two genes. Redundancy is the phenomenon whereby one gene can take over the function of another gene, for example if they both code for highly similar proteins. In the case of a complete redundancy relationship, inactivation of either gene on its own will result in no defect or in other words, the phenotype will be identical to the wild-type (WT) (Fig 1A). Assuming that there are no further redundancy relationships with other genes and that the gene products have an important role, simultaneous deletion of both genes together will lead to a significant defect (Fig 1A). This is clearly an example whereby the combined effect of two mutations is larger than the product or addition of the effects of the individual mutations. The combined effect cannot be predicted unless the redundancy relationship is already known. Numerous different types of genetic interactions have been described [9]. However, based on a simple phenotype such as growth for example, two major types of genetic interactions can be discerned: negative and positive genetic interactions. As in the example depicted in Fig 1A, negative genetic interactions occur when the combined effect of two mutations is greater than expected based on the phenotype of single mutations.

Conversely, positive genetic interactions occur when the combined effect is weaker than expected (Fig 1B).

The term epistasis was coined more than one hundred years ago by Bateson to describe a specific type of genetic interaction [11]. He used the term to describe cases of interactions where effects of one locus masks the allelic effect of another locus. The literal translation of epistasis is 'standing upon' which was meant to describe the masking effect of one locus on another one. In one of the most seminal papers in genetics, one hundred years ago, Fisher called any non-additive effect of two loci 'epistacy' [12]. This term quickly became epistasis. Thereafter, two distinct definitions of epistasis entered the field of genetics. This double meaning of the term frequently causes confusion. The limited definition of epistasis described by Bateson is still often used by classical geneticists, while the broader Fisher definition is frequently used by quantitative geneticists. In this thesis, both genetic interaction and the term epistasis with the broader Fisher definition are used interchangeably.

Early examples of epistasis in the 1930s included unusual segregation ratios that differed from simple Mendelian ratios [13–15]. Since then there was little interest in epistasis until the 1980s when interest in genetic interaction detection was increased due to advances in molecular biology. Those studies were mainly focused on using epistasis to order genes into linear pathways (reviewed by Avery and Wasserman [16]). As explained below, genes whereby the products function in linear signalling pathways frequently show positive genetic interactions.

In recent years, numerous studies have played crucial roles in better understanding the relationship between genotype and phenotype in model organisms by performing large-scale analyses of single mutations [17–21]. These efforts have been extended to identify genetic interactions between genes on a massive scale in budding yeast and on a smaller scale in fission yeast, *Caenorhabditis elegans* and *Drosophila melanogaster* cells [22–27]. One of the important messages from such studies is that genetic interactions are pervasive. In budding yeast, 900,000 genetic interactions have been detected from all possible pairwise interactions between the approximately 6000 protein-coding genes [23]. This global network of genetic interactions has revealed that negative and positive genetic interactions connect both within and between pathway modules which has enhanced our knowledge about the interplay between different cellular processes. Systematic pair-wise genetic interaction studies in yeast may have just exposed the tip of the iceberg since the majority of double mutants do not show a growth defect [23]. Despite the fact that yeast is arguably the most well studied model eukaryote, we still lack a comprehensive understanding of its genotype-phenotype relationship because most large-scale genetic interaction studies are conducted in one condition. Approaching phenotype-genotype relationships in multicellular organisms will be even more challenging considering the generally larger genome sizes and different cell types. This also highlights the importance of genetic interactions and their role in complex human diseases.

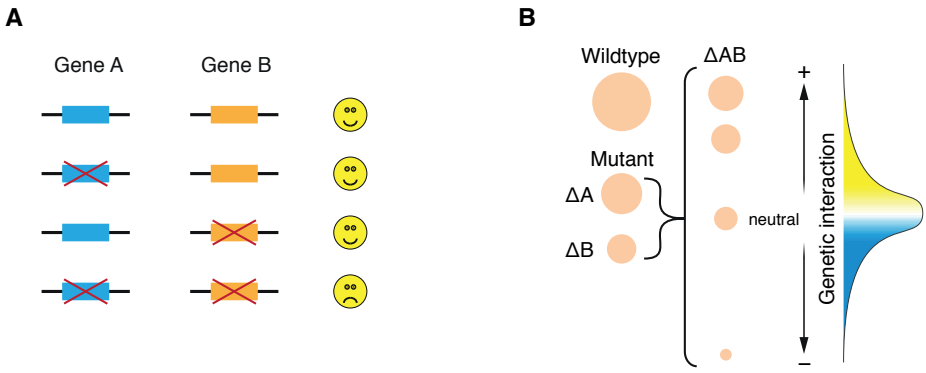


Fig 1. Defining genetic interactions. (A) One intuitive example of a genetic interaction is the presence of two functionally redundant genes (gene A and gene B). Individual deletion of redundant genes has no consequence but simultaneous deletion of both genes does result in a severe outcome. **(B)** Genetic interactions can be broadly subdivided into two types: negative and positive genetic interactions. The size of circles represents fitness of deletion mutants. The neutral situation is the expected fitness of double mutant based on the fitness of single mutant (A and B). Negative genetic interactions occur when the fitness of double mutant is lower than expected and positive genetic interactions occur when the fitness of double mutant is higher. The density plot represents an expected distribution of genetic interaction scores (adapted from Beltrao et al. [10]).

Genetic interaction detection in yeast

The ease of using model organisms coupled to advances in molecular biology have enabled the detection of genetic interactions in a high-throughput manner. The majority of large-scale genetic interaction assays have used yeast *Saccharomyces cerevisiae*. With its simple genome and straightforward genetics, yeast is a suitable model organism for studying fundamental properties of eukaryotic cells and has served a key role in the development of many functional genomic technologies including methods to detect genetic interactions [28]. Three different technologies are used to systematically construct double mutants and detect genetic interactions in yeast. The first method is synthetic genetic array (SGA) analysis which is capable of generating double mutants on a large scale by combining arrays of single mutants with a query mutant [29]. The SGA procedure consists of mating and multiple selection steps (Fig 2, left panel). Fitness of double mutants is quantified using imaging software by assessing the size of colonies grown on plates. Genetic interactions are scored as negative (aggravating) or positive (alleviating) when the double mutant fitness is compared with the expected value derived from a multiplicative model of the single

mutants fitness. A variation of the SGA method has been applied to detect all possible genetic interaction pairs for a smaller and defined subset of genes and is called epistatic miniarray profile (E-MAP) [30]. The second high throughput method for identification of genetic interactions in yeast is diploid synthetic lethal analysis by microarray (dSLAM) which takes advantage of strain-specific barcodes associated with each deletion mutant (Fig 2, right panel) [31]. The third method is genetic interaction mapping (GIM) which is a hybrid between SGA and dSLAM (Fig 2, right panel) [32]. The main requirement for genome-scale mapping of genetic interactions in model organisms is the availability of mutant strains on a large scale. Mapping of genetic interaction should also be high-throughput to enable rapid and comprehensive experiments. The employment of high-throughput technologies was pioneered in *S. cerevisiae* which has enabled early progress in quantitative genetic interactions detection in this organism.

When measuring genetic interactions in yeast, two important factors should be taken into account. The first one is a quantitative phenotype to measure [33]. Fitness has been the prime class of phenotype in the vast majority of genetic interaction studies [34]. Different measures of fitness have been also used [33]. One of the preferred measures is the relative growth rate compared to a WT strain under a given laboratory condition. This phenotype is an appropriate readout to measure genetic interactions between genes. In SGA, the relative growth rate of single and double mutants is obtained based on the colony size of strains [35]. In dSLAM and GIM the fitness of double mutants is measured by measuring the intensity of hybridized barcodes specific to mutants. This is indicative of the relative abundance of double mutants in a population [31,32]. Less frequently, fitness has been measured by an increase in mutant population or progeny of mutants relative to WT in one WT generation [36,37]. Finally, fitness can also be measured by optical density of a growing yeast culture over time and calculating its exponential growth rate [38]. Other more specific phenotypes such as morphological defects as identified by high-content screening have also been applied [39].

The second important factor to be taken into account when measuring genetic interactions is the neutral definition of double mutants [33]. Four mathematical definitions have been used for calculating the expected growth defect in a double mutant based on the fitness of singles. These models include the product, the additive effect, the log, and the min. A genetic interaction score is calculated for a pair of genes (A, B) if W_{AB}, the observed phenotype of the double mutant, is significantly different from $E(W_{AB})$, the expected phenotype of the double mutant based on the fitness of singles (W_A, W_B) if they were noninteracting. The expected phenotype of the double mutant under each definition is (W_A .W_B), (W_A + W_B), log2[(2 ^{W_A} -*1*)(2 ^{W_B}-*1*)], and min (W_A, W_B), for the product, the additive, the log, and the min, respectively. The predictive neutral fitness provided by these mathematical definitions differ which leads to different genetic interaction scores [33]. The most commonly used neutral mathematical model is the product whereby the expected fitness of the double

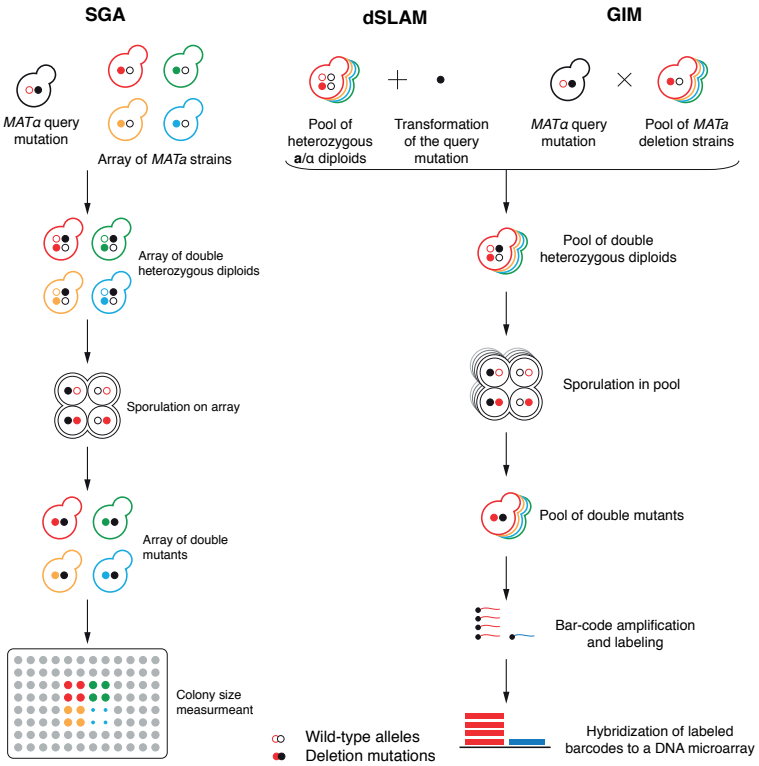


Fig 2. High-throughput genetic interaction detection methodologies in yeast. (Left panel) In synthetic genetic array (SGA) technology, a *MATα* mutant strain is crossed to an array of viable nonessential *MATα* mutant strains or conditional alleles of essential genes. A *MATα* deletion mutant is tagged with the drug-resistance marker gene *natMX4* (filled black circle) and the *MATα* deletion mutants are tagged with the drug-resistance *kanMX4* gene cassette (filled red, green, yellow, and blue circles). The process is continued by mating and sporulation. Next meiotic progeny is robotically pinned on solid media which can select for double-mutant cells. The double mutant growing on an array can be scored for fitness by using their colony size. **(Right panel)** Most steps in diploid synthetic lethal analysis by microarray (dSLAM) and genetic interaction mapping (GIM) are similar except for constructing a pool of double mutant heterozygous diploids. In dSLAM, using high-efficiency integrative transformation, a query mutation (filled black circle) marked with the *URA3* selection gene is transformed into a pool of haploid-convertible heterozygous diploids. In GIM, a *MATα* haploid deletion strain tagged with *natMX4* (filled black circle) is mated with a pool of *MATα* deletion mutants. The result in both methods is a pool of double heterozygous diploids. Then, after sporulation, haploid double mutants are selected. Subsequently, genomic DNA is isolated and used as a template for polymerase chain reaction amplification of the tags. During this process the amplified DNA that contains the tags are labeled with fluorescent dyes and hybridized to microarrays. Microarray hybridization of double mutants and control single mutants in the same pool enables quantitative measurement of genetic interactions. A synthetic lethal interaction is revealed by a high single/double ratio of hybridization signal intensity. (Figure adapted from Baryshnikova et al. [3]).

mutant in the absence of a genetic interaction, should be equivalent to the product of the corresponding single mutants fitness [22,23,38,40,41].

Molecular mechanisms of negative genetic interactions

Negative genetic interactions occur when a double mutant exhibits a more severe phenotype than expected based on corresponding single mutants. This type of interaction can be subdivided into 'synthetic sick' and 'synthetic lethal' (Fig 3A). As described earlier, the simplest explanation for a negative genetic interaction is a redundancy relationship between two highly functionally related genes [7]. If two genes perform a similar molecular function the loss of one can be compensated by the other (Fig 3B). Such functional redundancy relationships can arise through gene duplications. Gene duplication events are great source of adaptation and new genes [42–44]. They can be divided into two main categories: whole-genome duplication (WGD) and small-scale duplication (SSD). Whole-genome duplications are rare evolutionary events whereby the entire genome of organism doubles. High sequence identity between blocks of the genome and conserved synteny between other closely related species confirm a whole-genome duplication event for baker's yeast [45–47]. On the other hand, small-scale duplications only occur for one or few genes [48]. Two duplicated genes should initially have the same function. In the long-term, accumulation of mutations in both copies result in two functionally divergent copies [42]. In total 547 gene pairs in yeast *Saccharomyces cerevisiae* derived from whole-genome duplication which are known as ohnologs [49]. Some of these duplicated genes may still have the capability to compensate for each other's loss. This property of functional redundancy has been linked to non-responsiveness upon gene deletion [50–52]. In *S. cerevisiae* many genes can be deleted with little or no phenotypic effect on growth in rich media [17]. This is also confirmed using a more sensitive phenotype such as genome-wide gene expression changes where more than 50% of deletion mutants behave like WT [53]. The degree to which functional redundancy contributes to non-responsiveness behaviour is still unclear and is reported upon in detail in chapter 2 of this thesis.

A functional redundancy relationship can also be extended to parallel pathways. For instance, if two pathways perform the same molecular function. The disruption of either pathway may result in a minor defect, whereas inactivation of both pathways simultaneously may have major consequences (Fig 3C). Large-scale genetic interaction studies have confirmed extensive connections between different pathways [22,23,35,54]. Even though there are examples of redundancy between pathways [55], many genetic interactions are detected between seemingly unrelated processes [23]. To what extent negative genetic interactions between different molecular pathways always represent functional redundancy is still unresolved. This ambiguity highlights the importance of unravelling the molecular mechanisms underlying genetic interactions.

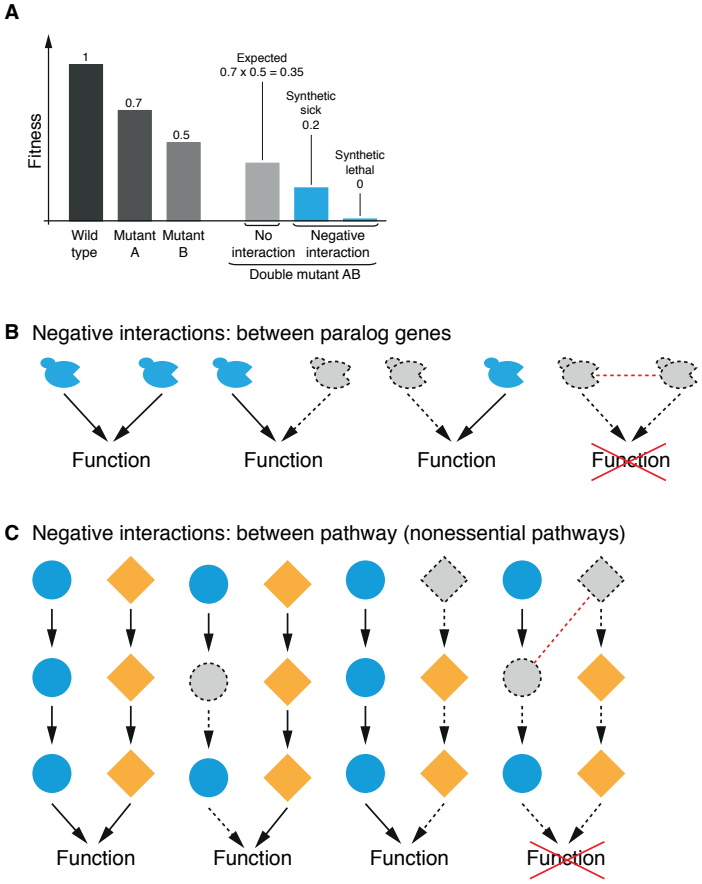


Fig 3. Negative genetic interactions and molecular mechanisms. (A) A cartoon representation of negative genetic interaction is depicted. No interaction is defined when the double mutant fitness is as expected based on the fitness of single mutants. The expected fitness of double mutant based on the product of two singles (A and B) is 0.35. Negative genetic interactions occur when the double mutant phenotype is more severe than expected and can be subdivided into synthetic sick and synthetic lethal. Synthetic lethality occurs when the double mutant is nonviable. **(B)** The most simplistic explanation for negative genetic interaction is depicted using two redundant proteins. Individual deletion of redundant genes has no consequences but simultaneous deletion of them will lead to disruption of a shared function. Blue symbols represent active genes. Grey symbols with dashed outline indicate a deleted gene. **(C)** Graphical representation of negative genetic interactions that can arise from genes in two parallel pathways with a similar function. Disruption of individual pathway can be compensated by the parallel pathway. Genetic interactions could potentially occur between all possible pair-wise combinations of genes functioning within each pathway. Red dashed lines indicate genetic interactions between two genes.

Molecular mechanisms of positive genetic interactions

Positive genetic interactions occur when double mutants exhibit a less severe phenotype than expected based on the phenotype of the corresponding single mutants (Fig 4A and 4B). Positive genetic interactions can be subdivided into different categories. For example, symmetric positive in which the phenotype of single and double mutants is indistinguishable (Fig 4A). Such a scenario can be explained by the interaction between members of the same essential linear pathway or protein complex (Fig 4C and 4D) [56]. The logic behind this proposed mechanism is that inactivating any individual member of an essential pathway or protein complex results in complete disruption of that pathway or protein complex. Therefore, additional inactivation of any other members will have no further consequences. Another subcategory of positive interaction is when the phenotypic effect varies between single and double mutants which is called asymmetric positive. Depending on the relative fitness of the double mutant to the sickest single mutant, asymmetric genetic interactions can further subdivided into masking and suppression (Fig 4B) [56]. The underlying molecular mechanisms of masking and suppression are not well understood.

Other proposed molecular mechanisms

There have been more specific proposed mechanisms for both negative and positive genetic interactions which do not easily fall into the simple explanations mentioned above. For example, it has been proposed that negative genetic interactions could also occur within the same essential pathway or protein complex [56]. This situation could occur with altered alleles of essential genes which result in a decreased flux through a metabolic pathway (Fig 5A). For example, members of the endoplasmic reticulum membrane protein complex show strong negative genetic interactions [23]. In this situation a pathway will only get fully disrupted when two essential genes are altered. On the contrary, the individual effect of single mutants is negligible.

Different varieties of positive genetic interactions have also been described. For example, it is possible that the deletion of one gene in a pathway results in a toxic state, whereby deletion of a second downstream gene could reverse the toxic situation (Fig 5B) [56]. Furthermore, a subset of genes can have many genetic interactions with other genes with diverse molecular functions (Fig 5C). These genes are known as ‘hubs’. Large-scale studies in baker’s yeast have shown that both negative and positive genetic hubs are abundant among both essential and nonessential genes [22,23]. The underlying mechanism for these hubs is yet to be discovered. Another interesting proposed mechanism is ‘induced essentiality’ that could occur due to network wiring and activation of other gene sets following a gene loss (Fig 5C) [57].

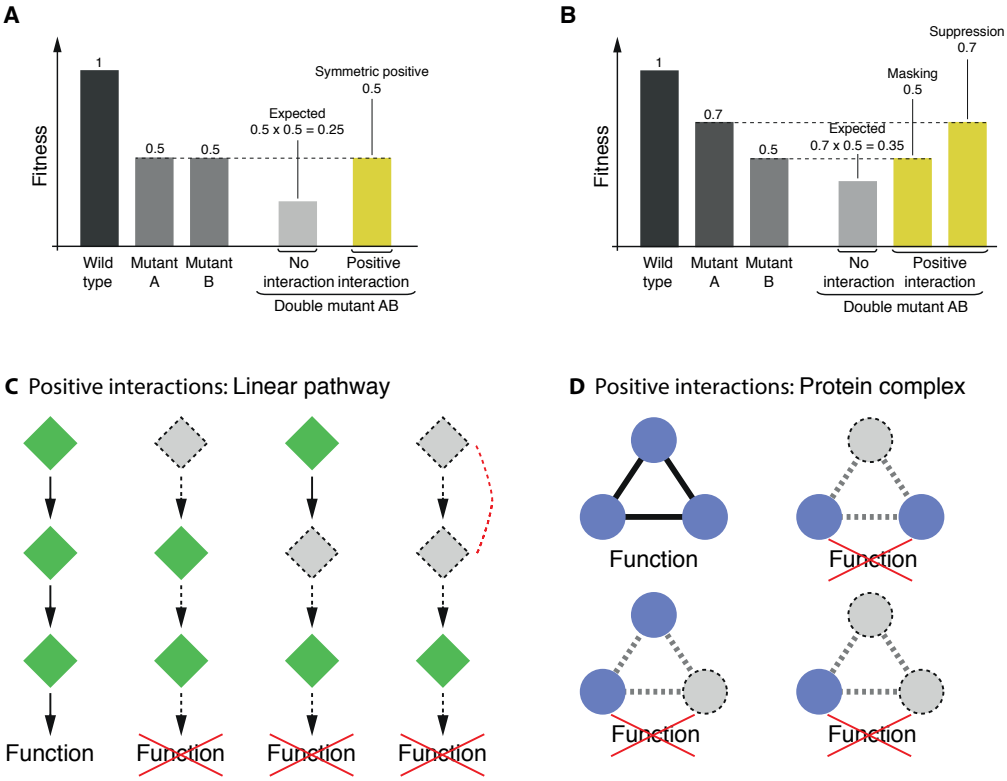


Fig 4. Positive genetic interaction and molecular mechanisms. (A) A cartoon representation of symmetric genetic interactions is depicted. The expected phenotype of double mutant (0.25) is defined based on single mutant fitnesses (0.5). Symmetric positive genetic interactions occur when the fitness of double mutant is better than expected and similar to the fitness of single mutants. **(B)** A cartoon representation of asymmetric genetic interaction is depicted. This type of positive interaction can be further subdivided into two different kinds of interactions: masking and suppression. In masking the fitness of the double mutant resembles the fitness of the sickest single mutant. Conversely, in suppression the fitness of double mutant is increased compared to the sickest single mutant. **(C)** The molecular mechanism of symmetrical positive genetic interaction can be explained by members of the same pathway or **(D)** protein complex. In both C and D deletion mutants are depicted using gray diamond/circles with dashed outline. Deletion of either genes in the same pathway or protein complex leads to inactivation the related function. Additional deletions will have no further consequences.

Genome-wide expression changes to unravel molecular mechanisms of genetic interactions

The mechanisms proposed above have been valuable in understanding negative and positive genetic interactions. However, more detailed phenotypes such as genome-wide gene expression changes can be instrumental in further unravelling the mechanisms of genetic interactions [38,58–60]. Unlike growth-based genetic interactions, expression-based genetic interaction profiling can deliver more detailed data. This has been recently demonstrated through investigating genetic interactions between yeast kinases and/or phosphatases [38]. This study revealed that a genetic interaction between two genes can have different types of epistatic influences on the other genes. Genetic interactions among other functional classes such as gene specific transcription factors (GSTFs) have been detected [23] and can be investigated by genome-wide expression changes as is demonstrated in chapter 3 of this thesis.

The number of patterns that emerge from expression-based profiling is higher compared to growth or fitness based genetic interactions. This requires more systematic approaches to understand the different patterns. As discussed below, computational modelling such as Boolean and Nested Effects Models (NEMs) have therefore been applied to understand specific types of genetic interactions [38,61].

General principals of modelling in biology

Innovations in experimental methods have enabled large-scale studies which has led to the development of large regulatory network models [53,62,63]. Consequently, understanding the underlying mechanisms of these complex networks requires the integration of a great quantity of data. One essential approach for understanding complex biological systems has been computational modelling. Various computational modelling approaches have been developed. Broadly speaking, these can be divided into three major classes: logical models, continuous models, and single molecule levels [64]. An overview of different methods is represented in fig 6.

A mathematical model is a representation of the essential features of a complex system [65]. The purpose of modelling is to represent simplified images of reality rather than replicating it. Simplification allows us to capture the crucial features of a complex process avoiding the overwhelming amount details. Mathematical models are amenable to simulation using computers. Models describing biological networks are generally too complex to be solved manually, therefore they are mostly solved using computers [66]. With the increase in computer power our ability to enumerate large numbers of models has increased in recent years. Due to the large numbers of models that can explain different genetic interaction patterns the speed of analysis for each model can be a limiting factor (Fig 6). Boolean and Petri net methods are known to be relatively fast in the speed of their analysis. Therefore, they can be suitable for enumerating large numbers of models.

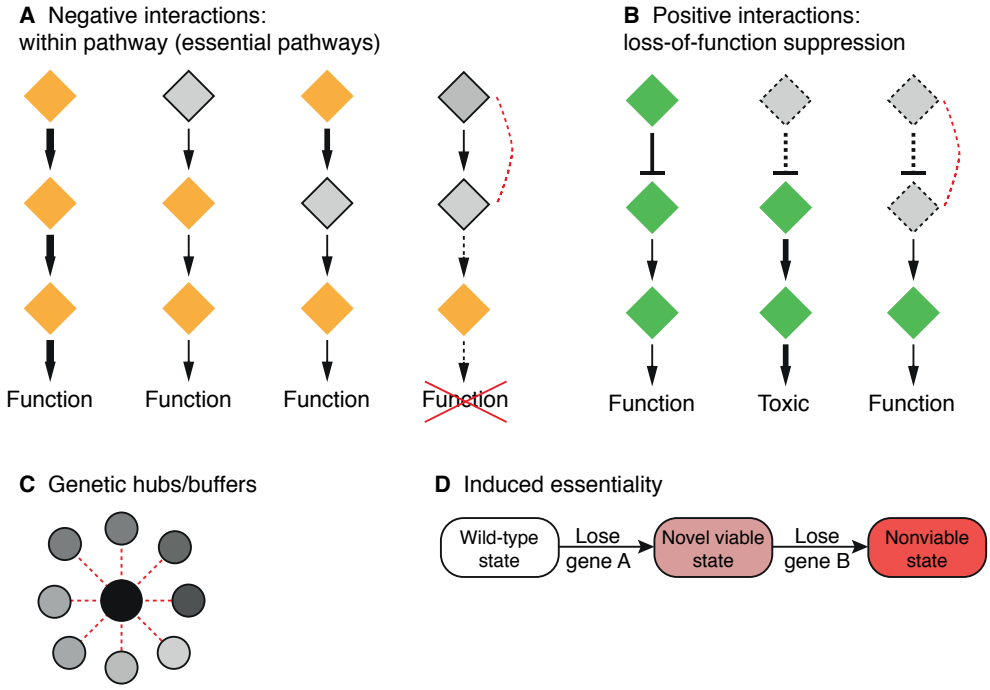


Fig 5. Miscellaneous proposed mechanisms of genetic interactions. (A) Altered alleles of two essential genes in an essential pathway (depicted in grey diamond with solid outline) can reduce the flux through the pathway which can give rise to negative genetic interactions. The phenotype of double knockdown mutant could not be predicted base on the phenotype of single knockdowns **(B)** Deletion of an activator in a pathway can increase the expression of a downstream gene which can lead to accumulation of toxic products in the cell. Subsequent deletion of downstream gene can reduce the toxic products and therefore suppress the effect of upstream gene deletion. **(C)** Genetic 'hubs' can have genetic interactions with many other genes with diverse functions. **(D)** Deletion of one gene can stimulate the activation another gene giving rise to a novel viable state. Deletion of both genes can have a much more severe consequence than expected.

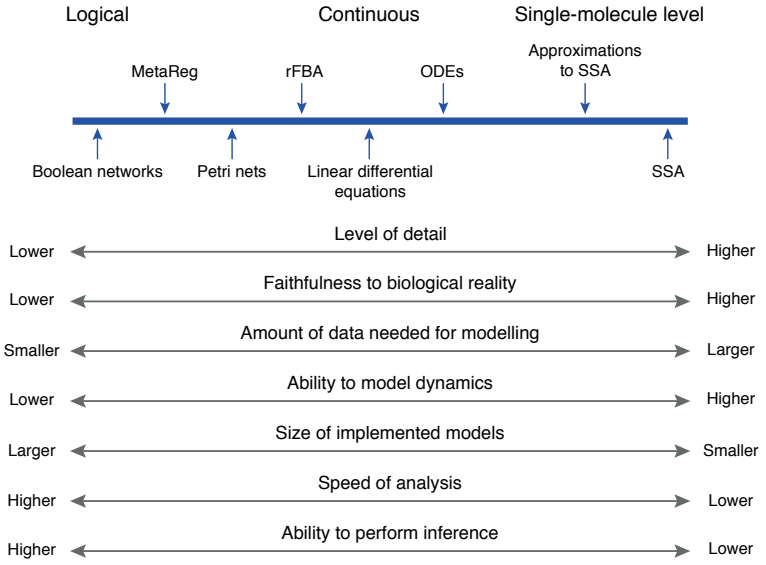


Fig 6. A schematic comparison of mathematical modelling of regulatory networks. Different mathematical modelling techniques are presented on an imaginary scale which represents the level of quantitativeness. The level of quantitativeness and details captured by each technique increases from left to right. Several features of different modelling techniques are indicated below the scale. Boolean networks are the simplest type of logical modelling. They are highly abstract, therefore the amount of data required is minimal. They can capture a qualitative dynamic of a biological network. MetaReg is more faithful to biological reality since it can capture intermediate states of a network. Petri nets can capture finer details of metabolic and signalling networks. It can also handle dynamics to a certain degree. Regulated flux balance analysis (rFBA) can handle metabolic predictions. This method requires more data and its analysis is challenging. Linear differential equations can better predict the dynamics of a regulatory network. General ordinary differential equations (ODEs) are more faithful to biological reality but computationally more expensive than the former. The most detailed type of modelling is single-molecule level models. This approach can capture stochasticity of single molecules but is computationally expensive. Stochastic simulation algorithms (SSAs) are a form of single-molecule approach which can perform better by sacrificing some details. (Adapted from Karlebach and Shamir [64]).

Boolean modelling and genetic interactions

Boolean regulatory networks were first introduced by Kauffman [67,68]. They are regularly used to approximate the dynamics of genetic regulatory networks. In Boolean notation, genes which are referred to as entities can be either active (true state) or inactive (false state) (Fig 7A). The state of each gene is updated according to levels of other input regulators, using a specific Boolean function. In spite of neglecting intermediate states Boolean techniques can introduce a rough approximation of the final state of a gene regulatory network. Boolean networks have been applied to analyze the stability of transcriptional network and cell-cycle network regulation [69,70]. Boolean modelling as one of the logical modelling techniques rely on purely qualitative knowledge. Despite its simplicity it has been also applied to predict certain genetic interaction patterns based on gene expression changes in double and single mutants [38]. However, due to its limitations it cannot capture the detail of all possible models and therefore all different genetic interaction patterns.

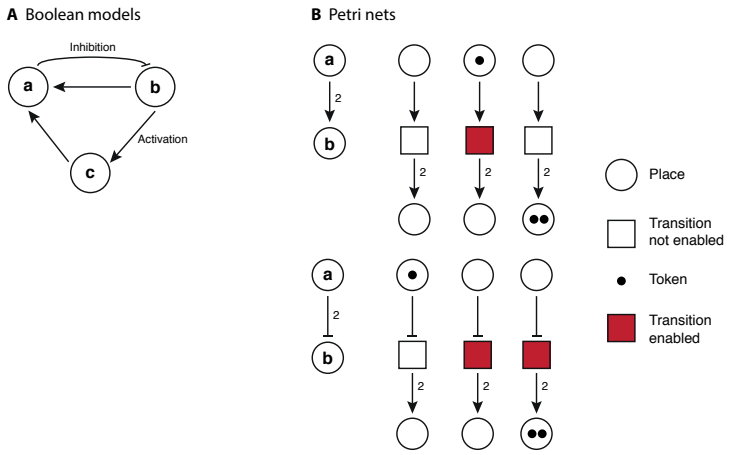


Fig 7. Graphical representation of Boolean and Petri net modelling. (A) In Boolean modelling, regulatory relationships between genes (nodes/entities) are drawn using arrows (edges). Both activation and inhibition edges can be incorporated. The state of one node can be either active or inactive depending on the incoming edges. **(B)** In Petri nets, genes (a and b) are depicted in circles (places). Regulatory relationships between two genes are depicted in squares (transitions) connecting two places via arrows (arcs). Tokens (black circles) represent resources that can be transferred from one place to another via a transition. Transitions are enabled when number of tokens in the input place match or exceed the corresponding arc weight. Edge weight represents the strength of activation or inhibition of genes on each other. Upper panel: A Petri net describing the activation of gene a on gene b. The edge weight, depicted by 2, represents the strength of activation. Lower panel: A Petri net describing the inhibition of gene a on gene b. The edge weight, depicted by 2, represents the strength of inhibition. The three Petri nets illustrate different states of the model.

Petri net modelling and genetic interactions

A petri net is a directed graph in which nodes can be divided into two disjoined sets (bipartite graph). Petri nets consist of two types of nodes, namely places and transitions. These nodes are connected by directed arrows which are called arcs. Graphically, places and transitions are depicted by circles and squares respectively. Tokens represent resources which can be transferred from one place to another depending on the direction of the arcs. An enabled transition is fired when tokens in the input place(s) match or exceed the weight(s) on the corresponding incoming arc(s) to the transition (Fig 7B). The state of a system is determined by the distribution of tokens in places in a certain time which is called marking. The dynamics of the model (starting from the initial marking) is caught by state changes due to the firing of transitions and by the consequent movement of tokens over the places. Commonly used descriptions in biology such as biochemical reactions or interactions between genes can be easily transformed to Petri nets. In Petri net notation, places represent genes/proteins/substrates and transitions can transformed to interactions.

Petri net modelling framework was first introduced by Carl Adam Petri in 1962 at the age of 13 and applied to describe chemical processes [71]. The method has been also intensively employed in computer science to specify and analyze concurrent and distributed systems [72,73]. The intuitiveness and graphical illustration of this method has made it appealing to computational systems biologists to describe biochemical reaction systems, where tokens are interpreted as single molecules of metabolites. Petri nets were first introduced to biology to describe metabolic pathways by Reddy et al. [74]. Since then, more studies have adapted the Petri net method to obtain quantitative information about the behaviour of biological systems [75–78].

The Petri net notation provides an accessible framework for construction and execution of biological networks. Owing to the integration of both qualitative (given by the topology of the Petri nets) and quantitative (given by the flow of the tokens) analysis, Petri nets can deal with the complexity of gene regulatory networks. Given the complexity of genetic interactions, a Petri net framework could be devised to predict the genetic interaction patterns detected using gene expression changes. This can provide clues about the molecular mechanisms of genetic interactions and is demonstrated in chapter 4.

Genetic interaction detection in cancer

Over the past decades, research has led to an understanding of many cellular processes that are involved in complex diseases such as cancer. It is clear that the phenotype of individuals does not arise from the individual action of alterations but is rather caused by the combination of them. The notion that cancer is the result of multiple rate-limiting events began as early as the 1960s [79,80]. This has indeed been confirmed by large-scale studies of DNA sequences of tumors in the past few years [81–84]. These projects have also shown the large heterogeneity of cancer driver mutations across different cancer types.

The principles underlying mutational heterogeneity and to what degree genetic interactions play a role in cancer predisposition and progression is still poorly understood.

The first instance of genetic interaction in cancer, named oncogene cooperation, was discovered in 1983 by Land and Ruley [85,86]. Cooperation between two mutations is reflected in co-occurrence of two mutations in the same individuals more often than expected by chance. Conversely, mutual exclusivity between two genes happen when simultaneous mutations in the same individuals are less frequent than it is expected by chance. There are at least three underlying explanations for mutual exclusivity. These explanations are not entirely intuitive and rely in part also on understanding that cancer cells arise and survive due to an evolutionary process of their own. A first explanation for mutual exclusivity is when genes are members of the same functional pathway. In such cases, mutations in either gene can be sufficient to disrupt the pathway and initiate cancer progression [87,88]. Mutation in the other gene is therefore not seen simultaneously because there is no added benefit to the tumor. A second explanation is when two genes are active under different developmental stages. Here too simultaneous mutation is rarely observed because this offers no growth advantage. Finally, mutual exclusivity can occur between genes from parallel pathways for example, whereby simultaneous mutation is lethal for cancer cells: synthetic lethality [89]. This type of genetic interaction has been recently exploited as a therapeutic strategy to selectively target tumor cells [90,91]. One of the clinically relevant examples of synthetic lethality is patients carrying mutations in *BRCA* tumor suppressor genes which are highly sensitive to PARP inhibition [92]. Although there are other examples of potential synthetic lethal interactions *BRCA-PARP* interaction is currently the most successful case in terms of treatment potential [93].

Statistical analyses of data from thousands of adult tumors have led to identification of co-occurrence and mutual exclusivity relationships between alterations [94]. Several methods have been developed to identify genetic interactions using adult tumors [87,94,95]. One of the important requirements to identify potential genetic interactions using statistical analysis is to have a large population of tumors [94]. In the majority of adult tumors sample size is not a limiting factor. Conversely, identification of genetic interactions in pediatric cancers may be more challenging due to their rarity. Therefore, other strategies may be required to identify genetic interactions in pediatric cancer, and this is also investigated in chapter 5.

Scope and aim of the work described in this thesis

The overall aim of the work described in this thesis is to understand genetic interactions at the molecular level. In order to investigate genetic interactions different experimental and computational approaches have been employed starting with the model organism *S. cerevisiae* and ending with an investigation of pediatric cancer. An important aspect of any variant in DNA, also for the purpose of studying genetic interactions, is whether or not there is an associated phenotype. There are many documented cases of complete gene loss with no apparent phenotypic consequence. **Chapter 2** investigates this non-responsiveness to genetic perturbation. It employs the very sensitive phenotype offered by genome-wide expression analysis to study the relative contribution of redundancy and condition dependency to non-responsiveness upon systematic deletion of yeast genes. **Chapter 3** focuses on understanding the genetic interactions found between gene specific transcription factors (GSTFs) by analysing genome-wide gene expression changes of 72 double mutants and comparing them with the respective single mutants. Both negative and positive genetic interactions are considered. As exemplified, the high-resolution gene expression atlas provides a great resource for investigating molecular mechanisms of genetic interactions. In **chapter 4**, differences in genetic interaction patterns are compared between two different functional classes of proteins (GSTFs and kinases/phosphatases). Petri net modelling is applied to understand the molecular mechanism of a specific genetic interaction pattern more dominant in GSTFs compared to kinases/phosphatases. Finally, **chapter 5** describes a preliminary effort to identify genetic interactions between somatic changes in pediatric cancers. Two different approaches are compared resulting in an initial map of genetic interactions in pediatric cancer. **Chapter 6** discusses all these findings as well as future research directions.

References

1. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci. 2012;109: 1193–1198. doi:10.1073/pnas.1119675109
2. Lander ES. Initial impact of the sequencing of the human genome. Nature. 2011;470: 187. doi:10.1038/nature09792
3. 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: 111–133. doi:10.1146/annurev-genom-082509-141730
4. Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet. 2014;15: 22–33. doi:10.1038/nrg3627
5. Lehner B. Modelling genotype–phenotype relationships and human disease with genetic interaction networks. J Exp Biol. 2007;210: 1559–1566. doi:10.1242/jeb.002311
6. Chandler CH, Chari S, Dworkin I. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet TIG. 2013;29: 358–366. doi:10.1016/j.tig.2013.01.009
7. Lehner B. Molecular mechanisms of epistasis within and between genes. Trends Genet. 2011;27: 323–331. doi:10.1016/j.tig.2011.05.007
8. Phillips PC. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9: nrg2452. doi:10.1038/nrg2452
9. Drees BL, Thorsson V, Carter GW, Rives AW, Raymond MZ, Avila-Campillo I, et al. Derivation of genetic interaction networks from quantitative phenotype data. Genome Biol. 2005;6: R38. doi:10.1186/gb-2005-6-4-r38
10. Beltrao P, Cagney G, Krogan NJ. Quantitative Genetic Interactions Reveal Biological Modularity. Cell. 2010;141: 739–745. doi:10.1016/j.cell.2010.05.019
11. Bateson W. Mendel's Principles of Heredity. Camb Univ Press. 1909;
12. Fisher RA. The correlations between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 1918; 399–433.
13. Lindsey AW. A Textbook of Genetics. Macmillan; 1932.
14. Sinnott EW, Dunn LC. Principles of Genetics. 2nd ed. McGraw-Hill, New York; 1932.
15. Snyder LH. The principles of heredity. D.C. Heath and company, Boston; 1935.
16. Avery L, Wasserman S. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. Trends Genet. 1992;8: 312–316. doi:10.1016/0168-9525(92)90263-4
17. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature. 2002;418: 387. doi:10.1038/nature00935
18. Kim D-U, Hayles J, Kim D, Wood V, Park H-O, Won M, et al. Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces pombe*. Nat Biotechnol. 2010;28: 617. doi:10.1038/nbt.1628

19. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2: 2006.0008. doi:10.1038/msb4100050
20. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, et al. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. Nature. 2003;421: 231. doi:10.1038/nature01278
21. Dietzl G, Chen D, Schnorrer F, Su K-C, Barinova Y, Fellner M, et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature. 2007;448: 151. doi:10.1038/nature05954
22. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The Genetic Landscape of a Cell. Science. 2010;327: 425–431. doi:10.1126/science.1180823
23. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016;353: aaf1420. doi:10.1126/science.aaf1420
24. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, et al. Conservation and Rewiring of Functional Modules Revealed by an Epistasis Map in Fission Yeast. Science. 2008;322: 405–410. doi:10.1126/science.1162609
25. Byrne AB, Weirauch MT, Wong V, Koeva M, Dixon SJ, Stuart JM, et al. A global analysis of genetic interactions in Caenorhabditis elegans. J Biol. 2007;6: 8. doi:10.1186/jbiol58
26. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nat Genet. 2006;38: 896. doi:10.1038/ng1844
27. Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, Boutros M. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. Nat Methods. 2011;8: 341. doi:10.1038/nmeth.1581
28. Botstein D, Fink GR. Yeast: an experimental organism for 21st Century biology. Genetics. 2011;189: 695–704. doi:10.1534/genetics.111.130765
29. Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, et al. Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. Science. 2001;294: 2364–2368. doi:10.1126/science.1065810
30. Schuldiner M, Collins SR, Weissman JS, Krogan NJ. Quantitative genetic analysis in Saccharomyces cerevisiae using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. Methods. 2006;40: 344–352. doi:10.1016/j.ymeth.2006.07.034
31. Pan X, Yuan DS, Xiang D, Wang X, Sookhai-Mahadeo S, Bader JS, et al. A Robust Toolkit for Functional Profiling of the Yeast Genome. Mol Cell. 2004;16: 487–496. doi:10.1016/j.molcel.2004.09.035
32. Decourty L, Saveanu C, Zemam K, Hantraye F, Frachon E, Rousselle J-C, et al. Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles. Proc Natl Acad Sci. 2008;105: 5821–5826. doi:10.1073/pnas.0710533105

33. Mani R, St.Onge RP, Hartman JL, Giaever G, Roth FP. Defining genetic interaction. Proc Natl Acad Sci. 2008;105: 3461–3466. doi:10.1073/pnas.0712255105
34. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. Nat Rev Genet. 2007;8: 437. doi:10.1038/nrg2085
35. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, et al. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. Nat Methods. 2010;7: 1017–1024. doi:10.1038/nmeth.1534
36. Jasnos L, Korona R. Epistatic buffering of fitness loss in yeast double deletion strains. Nat Genet. 2007;39: 550. doi:10.1038/ng1986
37. Sanjuán R, Elena SF. Epistasis correlates to genomic complexity. Proc Natl Acad Sci U S A. 2006;103: 14402–14405. doi:10.1073/pnas.0604543103
38. van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, et al. Functional overlap and regulatory links shape genetic interactions between signaling pathways. Cell. 2010;143: 991–1004. doi:10.1016/j.cell.2010.11.021
39. Vizeacoumar FJ, van Dyk N, S.Vizeacoumar F, Cheung V, Li J, Sydorskyy Y, et al. Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. J Cell Biol. 2010;188: 69–81. doi:10.1083/jcb.200909013
40. Zheng J, Benschop JJ, Shales M, Kemmeren P, Greenblatt J, Cagney G, et al. Epistatic relationships reveal the functional organization of yeast transcription factors. Mol Syst Biol. 2010;6: 420. doi:10.1038/msb.2010.77
41. Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, et al. Functional Organization of the S. cerevisiae Phosphorylation Network. Cell. 2009;136: 952–963. doi:10.1016/j.cell.2008.12.039
42. Ohno S. Evolution by Gene Duplication [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 1970. doi:10.1007/978-3-642-86659-3
43. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol. 2003;18: 292–298. doi:10.1016/S0169-5347(03)00033-8
44. Qian W, Zhang J. Genomic evidence for adaptation by gene duplication. Genome Res. 2014;24: 1356–1362. doi:10.1101/gr.172098.114
45. Kellis M, Birren BW, Lander ES. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. Nature. 2004;428: 617. doi:10.1038/nature02424
46. Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, et al. The Ashbya gossypii Genome as a Tool for Mapping the Ancient Saccharomyces cerevisiae Genome. Science. 2004;304: 304–307. doi:10.1126/science.1095781
47. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, et al. Genome evolution in yeasts. Nature. 2004;430: 35. doi:10.1038/nature02579
48. Fares MA, Keane OM, Toft C, Carretero-Paulet L, Jones GW. The Roles of Whole-Genome and Small-Scale Duplications in the Functional Specialization of Saccharomyces cerevisiae Genes. PLoS Genet. 2013;9. doi:10.1371/journal.pgen.1003176

49. Byrne KP, Wolfe KH. The Yeast Gene Order Browser: Combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 2005;15: 1456–1461. doi:10.1101/gr.3672305
50. Ihmels J, Collins SR, Schuldiner M, Krogan NJ, Weissman JS. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. Mol Syst Biol. 2007;3: 86. doi:10.1038/msb4100127
51. Wagner A. Robustness against mutations in genetic networks of yeast. Nat Genet. 2000;24: 355. doi:10.1038/74174
52. Plata G, Vitkup D. Genetic robustness and functional evolution of gene duplicates. Nucleic Acids Res. 2014;42: 2405–2414. doi:10.1093/nar/gkt1200
53. Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, et al. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. Cell. 2014;157: 740–752. doi:10.1016/j.cell.2014.02.054
54. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global Mapping of the Yeast Genetic Interaction Network. Science. 2004;303: 808–813. doi:10.1126/science.1091317
55. Kelley R, Ideker T. Systematic interpretation of genetic interactions using protein networks. Nat Biotechnol. 2005;23: 561. doi:10.1038/nbt1096
56. Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C. Systematic mapping of genetic interaction networks. Annu Rev Genet. 2009;43: 601–625. doi:10.1146/annurev.genet.39.073003.114751
57. Tischler J, Lehner B, Fraser AG. Evolutionary plasticity of genetic interaction networks. Nat Genet. 2008;40: 390–391. doi:10.1038/ng.114
58. Gutin J, Sadeh A, Rahat A, Aharoni A, Friedman N. Condition-specific genetic interaction maps reveal crosstalk between the cAMP/PKA and the HOG MAPK pathways in the activation of the general stress response. Mol Syst Biol. 2015;11. doi:10.15252/msb.20156451
59. Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, et al. Structure and function of a transcriptional network activated by the MAPK Hog1. Nat Genet. 2008;40: 1300. doi:10.1038/ng.235
60. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell. 2016;167: 1853–1866.e17. doi:10.1016/j.cell.2016.11.038
61. Pirkl M, Diekmann M, van der Wees M, Beerenwinkel N, Fröhlich H, Markowetz F. Inferring modulators of genetic interactions with epistatic nested effects models. PLoS Comput Biol. 2017;13: e1005496. doi:10.1371/journal.pcbi.1005496
62. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet. 2002;31: 64–68. doi:10.1038/ng881
63. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, et al. Transcriptional regulatory networks in Saccharomyces cerevisiae. Science. 2002;298: 799–804. doi:10.1126/science.1075090

64. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol. 2008;9: 770. doi:10.1038/nrm2503
65. Bartocci E, Lió P. Computational Modelling, Formal Analysis, and Tools for Systems Biology. PLOS Comput Biol. 2016;12: e1004591. doi:10.1371/journal.pcbi.1004591
66. Fischer HP. Mathematical Modelling of Complex Biological Systems. Alcohol Res Health. 2008;31: 49–59.
67. Glass L, Kauffman SA. The logical analysis of continuous, non-linear biochemical control networks. J Theor Biol. 1973;39: 103–129. doi:10.1016/0022-5193(73)90208-7
68. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol. 1969;22: 437–467. doi:10.1016/0022-5193(69)90015-0
69. Kauffman S, Peterson C, Samuelsson B, Troein C. Random Boolean network models and the yeast transcriptional network. Proc Natl Acad Sci. 2003;100: 14796–14799. doi:10.1073/pnas.2036429100
70. Li F, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci U S A. 2004;101: 4781–4786. doi:10.1073/pnas.0305937101
71. Petri CA. Kommunikation mit Automaten. Bonn: Institut für Instrumentelle Mathematik, Schriften des IIM; 1962.
72. Martiník I. Sequential Object Petri Nets and the Modelling of Multithreading Object-Oriented Programming Systems. 2012; doi:10.5772/48470
73. Zuberek WM. Timed Petri Nets in Performance Exploration of Simultaneous Multithreading. 2012; doi:10.5772/48601
74. Reddy VN, Mavrovouniotis ML, Liebman MN. Petri net representations in metabolic pathways. Proc Int Conf Intell Syst Mol Biol. 1993;1: 328–336.
75. Cordero F, Horváth A, Manini D, Napione L, De Pierro M, Pavan S, et al. Simplification of a complex signal transduction model using invariants and flow equivalent servers. Theor Comput Sci. 2011;412: 6036–6057. doi:10.1016/j.tcs.2011.06.013
76. Koch I, Junker BH, Heiner M. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. Bioinformatics. 2005;21: 1219–1226. doi:10.1093/bioinformatics/bti145
77. Blätke MA, Heiner M, Marwan W. Predicting Phenotype from Genotype through Automatically Composed Petri Nets. Computational Methods in Systems Biology. Springer, Berlin, Heidelberg; 2012. pp. 87–106. doi:10.1007/978-3-642-33636-2_7
78. Bonzanni N, Feenstra KA, Fokkink W, Heringa J. Petri Nets Are a Biologist's Best Friend. Formal Methods in Macro-Biology. Springer, Cham; 2014. pp. 102–116. doi:10.1007/978-3-319-10398-3_8
79. Armitage P, Doll R. The Age Distribution of Cancer and a Multi-stage Theory of Carcinogenesis. Br J Cancer. 1954;8: 1–12.
80. Nordling CO. A New Theory on the Cancer-inducing Mechanism. Br J Cancer. 1953;7: 68–72.
81. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. Nat Genet. 2013;45: 1127. doi:10.1038/ng.2762

82. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502: 333. doi:10.1038/nature12634
83. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet. 2013;45: 1134. doi:10.1038/ng.2760
84. Macintyre G, Ylstra B, Brenton JD. Sequencing Structural Variants in Cancer for Precision Therapeutics. Trends Genet. 2016;32: 530–542. doi:10.1016/j.tig.2016.07.002
85. Land H, Parada LF, Weinberg RA. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature. 1983;304: 596–602.
86. Ruley HE. Adenovirus early region 1A enables viral and cellular transforming genes to transform primary cells in culture. Nature. 1983;304: 602–606.
87. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res. 2012;22: 398–406. doi:10.1101/gr.125567.111
88. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med. 2004;10: 789. doi:10.1038/nm1087
89. Wang X, Fu AQ, McNerney ME, White KP. Widespread genetic epistasis among cancer genes. Nat Commun. 2014;5: 4828. doi:10.1038/ncomms5828
90. O'Neil NJ, Bailey ML, Hieter P. Synthetic lethality and cancer. Nat Rev Genet. 2017;18: 613. doi:10.1038/nrg.2017.47
91. Beijersbergen RL, Wessels LFA, Bernards R. Synthetic Lethality in Cancer Therapeutics. Annu Rev Cancer Biol. 2017;1: 141–161. doi:10.1146/annurev-cancerbio-042016-073434
92. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science. 2017;355: 1152–1158. doi:10.1126/science.aam7344
93. Brunen D, Bernards R. Drug therapy: Exploiting synthetic lethality to improve cancer therapy. Nat Rev Clin Oncol. 2017;14: 331–332. doi:10.1038/nrclinonc.2017.46
94. Park S, Lehner B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. Mol Syst Biol. 2015;11. doi:10.15252/msb.20156102
95. Vandin F, Upfal E, Raphael BJ. De novo discovery of mutated driver pathways in cancer. Genome Res. 2012;22: 375–385. doi:10.1101/gr.120477.111

Growth condition dependency is the major cause of non-responsiveness upon genetic perturbation

Saman Amini[1,2], Frank C.P. Holstege[1], Patrick Kemmeren[1,2]
[1]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands
[2]Center for Molecular Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands

Published in PLOS One. 2017; 12(3):e0173432. [PubMed:28257504]

Abstract

Investigating the role and interplay between individual proteins in biological processes is often performed by assessing the functional consequences of gene inactivation or removal. Depending on the sensitivity of the assay used for determining phenotype, between 66% (growth) and 53% (gene expression) of *Saccharomyces cerevisiae* gene deletion strains show no defect when analyzed under a single condition. Although it is well known that this non-responsive behavior is caused by different types of redundancy mechanisms or by growth condition/cell type dependency, it is not known what the relative contribution of these different causes is. Understanding the underlying causes of and their relative contribution to non-responsive behavior upon genetic perturbation is extremely important for designing efficient strategies aimed at elucidating gene function and unraveling complex cellular systems. Here, we provide a systematic classification of the underlying causes of and their relative contribution to non-responsive behavior upon gene deletion. The overall contribution of redundancy to non-responsive behavior is estimated at 29%, of which approximately 17% is due to homology-based redundancy and 12% is due to pathway-based redundancy. The major determinant of non-responsiveness is condition dependency (71%). For approximately 14% of protein complexes, just-in-time assembly can be put forward as a potential mechanistic explanation for how proteins can be regulated in a condition dependent manner. Taken together, the results underscore the large contribution of growth condition requirement to non-responsive behavior, which needs to be taken into account for strategies aimed at determining gene function. The classification provided here, can also be further harnessed in systematic analyses of complex cellular systems.

Introduction

Understanding the interplay between individual proteins and their role in various biological processes is critical for understanding cellular systems as well as understanding genotype to phenotype relationships. The role of individual proteins is often investigated by assessing the functional consequences of inactivation, often through genetic perturbation. With the advent of efficient gene editing techniques such as CRISPR-Cas9, such reverse genetics approaches are now starting to be even more expansively applied [1]. Pioneering work by the yeast gene deletion consortium using homologous recombination has in the past yielded an exhaustive collection of gene deletion mutants for *Saccharomyces cerevisiae* [2]. Subsequent high-throughput studies have enabled systematic investigation of the functional consequences of deleting individual genes and have been conducted to investigate cellular systems in many different ways [3–6]. However, a large number of deletion strains do not show detectable phenotypic effects and the scale and cause of this non-responsive phenomenon has been the subject of considerable interest [7–14]. In *S. cerevisiae* over 66% of deletion mutants have little or no detectable effect on growth in a single rich medium [2,15]. Similar observations have been made for other organisms [16,17]. Growth is not necessarily the most sensitive assay and as an alternative, genome-wide gene expression can be used, also as a much more detailed phenotypic readout. A recent study investigating the consequences of almost 1,500 gene deletions revealed that 53% of these mutants have a genome-wide expression profile that is essentially the same as wildtype (WT) [3]. This confirms that a large proportion of gene deletions have no phenotype when assayed under a single condition. Lack of a phenotype is due to either redundancy mechanisms or due to growth condition dependency of the deleted gene. An important question addressed insufficiently so far, is the degree to which these different mechanisms contribute to non-responsiveness. Understanding the underlying causes and their relative contribution is important for understanding complex cellular systems and for improving reverse genetic strategies aimed at determining gene function.

There are at least two kinds of redundancy mechanisms that contribute to non-responsive behavior upon genetic perturbation. One factor is the presence of closely related paralogs (homology-based redundancy) [7,10,11,14]. Redundancy can also be achieved through the presence of alternative pathways that can facilitate the same biological process (pathway-based redundancy) [11,18]. Homology-based redundancy is attributed to gene duplicates or proteins with similar functional domains. Even though most close paralog pairs have diverged during evolution [19], there are still pairs that have retained the capability to compensate for each other's loss [10,11,14,20]. Previous estimates indicate that as much as 23% to 29% of non-responsiveness can be ascribed to duplicated genes [7,10]. These estimates are however based on growth-rate as the phenotypic read-out and involved analysis of only a single growth condition. It therefore remains to be seen whether these

estimates are similar when applying a more sensitive phenotypic read-out and taking into account multiple growth conditions. To what degree pathway-based redundancy contributes to non-responsiveness is still unclear. This is mainly due to the fact that the degree to which negative genetic interactions reflect buffering between two genes in parallel pathways has remained unsolved. In addition to negative genetic interactions between parallel pathways, many negative genetic interactions have been detected between seemingly unrelated pathways [21–23]. It has also been suggested that most negative genetic interaction scores are not due to simple redundancy between pathways, but rather reflect network topology [24,25]. Providing clear estimates of the contribution of pathway-based redundancy to non-responsiveness solely based on negative genetic interactions is therefore unsatisfactory. Additional sources of information likely need to be included in order to provide reliable estimates of pathway-based redundancy.

Besides various redundancy mechanisms, condition dependency is another factor that contributes to non-responsiveness upon genetic perturbation. Some genes are almost certainly only required under specific growth conditions. It has been suggested that a large number of seemingly non-responsive mutants have important fitness roles under specific environmental conditions [8,18,26]. These estimates vary between 37% and 97%, using either focused subsets of genes, different assays or a combination of heterozygous and homozygous deletion strains. The degree to which these different approaches affect condition dependency estimates is unclear. A uniform analysis, using a single resource of non-responsive deletion strains is therefore required for reliably estimating the contribution of condition dependency to non-responsiveness, also in combination with redundancy estimates.

In spite of much effort, the causes and mechanisms of non-responsiveness remains elusive [7,11,18,26]. Most studies have either focused on a single explanation or have insufficiently investigated the underlying mechanisms and their relative contributions. Here, using a sensitive phenotypic readout (gene expression) as a starting point, we systematically investigate the underlying causes of and their relative contribution to non-responsiveness upon gene deletion. This reveals that condition dependency is the major determinant of non-responsiveness, explaining approximately 71%. Homology-based and pathway-based redundancy contributes 17% and 12% respectively. For approximately 14% of protein complexes, just-in-time assembly provides a potential mechanistic explanation for how protein complexes can be regulated in a condition dependent manner. The classification of the relative contribution of underlying causes provided here, can be harnessed for other systematic analyses of cellular systems.

Results

A collection of non-responsive mutants

To systematically investigate the underlying causes of and their relative contribution to non-responsiveness observed upon gene deletion, we exploited the availability of a collection of 1,484 deletion mutants expression profiles in the yeast *Saccharomyces cerevisiae* [3]. To obtain the list of responsive and non-responsive mutants, the same definitions were applied as in the original study [3]. In short, a mutant is considered non-responsive (NR) if three or less transcripts are changing significantly (FC > 1.7 & p-value < 0.05) as a consequence of deleting a particular gene. A mutant is classified as responsive (R) if four or more transcripts are changing significantly upon gene deletion. Applying these thresholds, 784 mutants are classified as NR (53%) and 700 mutants are classified as R (47%) as indicated before [3].

To ensure that non-responsiveness is not restricted to a particular class of proteins, the relative contribution of non-responsiveness to different functional classes was investigated [3]. As expected, non-responsiveness is not limited to a specific functional class, but common to a wide range of functional classes (Fig 1). There are differences however in the relative contribution of non-responsiveness to various functional classes. Classes that have a central role in cellular systems such as chromatin factors, Pol II transcription and translation / ribosome biogenesis have a relatively low number of NR mutants. Other functional classes such as cell cycle regulation and meiosis have a high number of NR mutants. This indicates that there is a relationship between the relative importance of different functional classes and degree of non-responsiveness, but also suggests that different causes such as condition dependency (meiosis) and robustness (cell cycle regulation) may contribute to non-responsiveness.

Both redundancy and condition dependency contribute to non-responsiveness

Exclusive requirement of a gene under a specific growth condition is expected to contribute to the number of non-responsive mutations. As a starting point, differences in mRNA transcript and protein levels between responsive and non-responsive mutants were investigated. As also observed before [3], genes within the group of NR mutants have a slightly larger fraction of genes with low transcript [3] and protein [27] levels (Fig 2A and 2B). This also reveals that almost all genes within the group of NR mutants generally exhibit normal expression levels in wildtype (WT), both for transcript as well as protein levels. Their removal however, has little effect on gene expression and raises the question why so many genes are expressed while not being required under the condition investigated.

Another likely cause of non-responsiveness is redundancy. This is confirmed by a strong enrichment of close paralogs in NR mutants (Fig 2C). Interestingly, NR mutants exhibit

markedly less negative genetic interactions [22] (Fig 2D, left panel). Negative genetic interactions do not necessarily reflect redundancy [10,11,24,28,29]. It is also suggested that redundant gene pairs have few negative genetic interactions as they are only expected to show a defect when both genes are absent [10], just as indicated here. NR mutants also have fewer physical interactions [30] and show less growth defects when assayed across many different conditions [8] (Fig 2D, middle and right panel). This is all confirmatory of more specialized, less centrally required roles in cellular biology and that both redundancy and condition dependency are contributing causes of non-responsiveness.

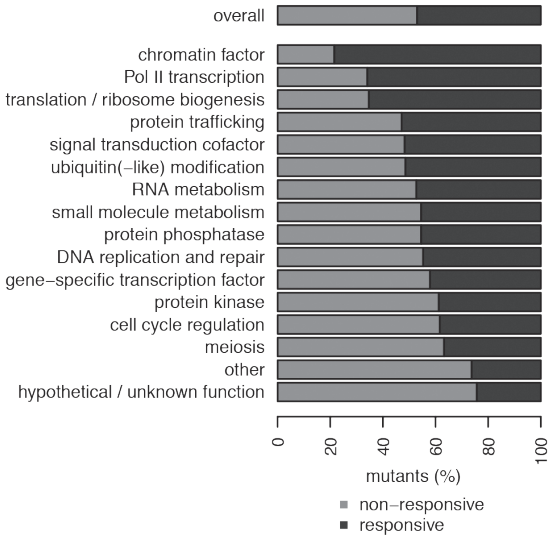


Fig 1. Distribution of non-responsive behavior. Percentage of non-responsive (NR; light grey) and responsive (R; dark grey) mutants within different functional categories, as well as the overall percentage. A mutant is classified as NR if three or less transcripts are changing compared to WT. A mutant is classified as R if four or more transcripts are changing. Functional categories as defined in the original study [3].

Close paralogs as a proxy for homology-based redundancy

To provide a reliable estimate of the relative contribution of complete redundancy to non-responsiveness, close paralog pairs within the non-responsive mutants are used here. Close paralog pairs are among the most recent evolutionary duplications and probably still have a similar function [32,33]. Close paralog pairs that show non-responsive behavior are therefore most likely the best starting point for estimating the relative contribution of redundancy to non-responsiveness. If two genes form a complete redundancy pair, it is also expected that removing either one of these genes does not result in any measurable defect, regardless of the condition investigated. The dataset from Hillenmeyer et al. [8] measures the relative growth defect of individual deletions in many different conditions.

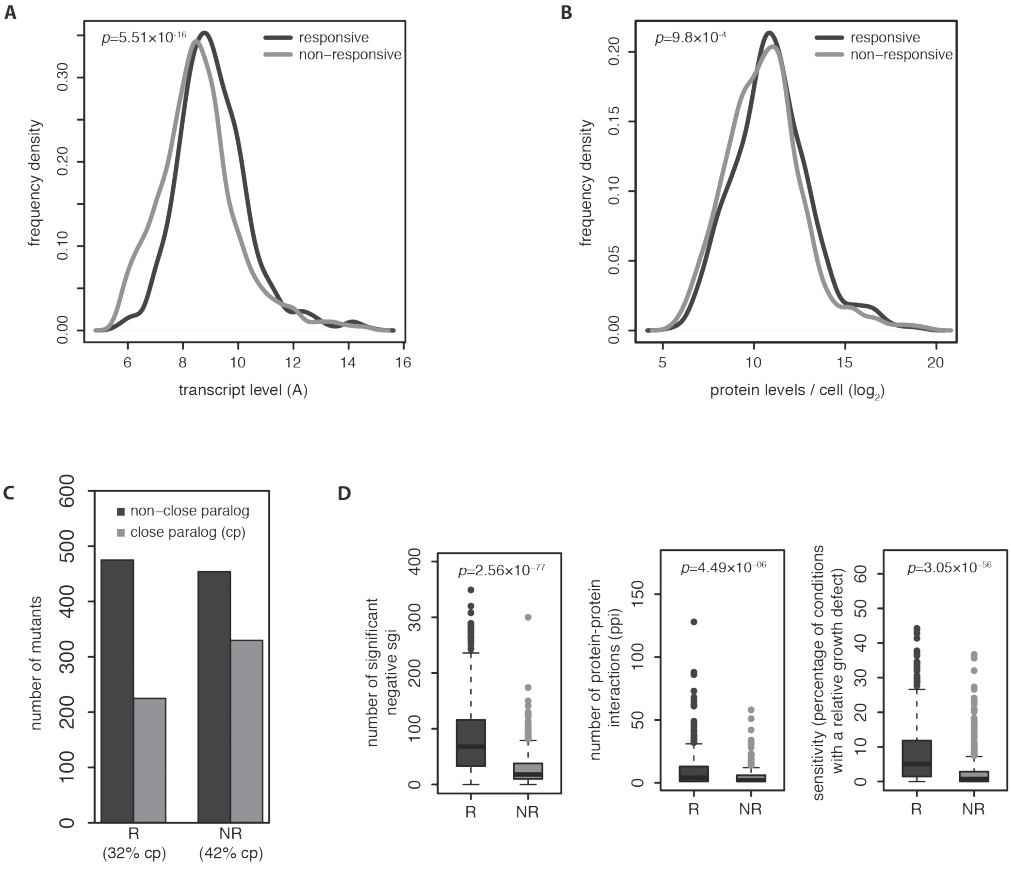


Fig 2. General characteristics of non-responsive and responsive mutants. (A) Frequency density distribution of wildtype (WT) mRNA transcript levels (A) obtained from the average of 200 WT strains, adapted from Kemmeren et al. [3]. *P*-value indicates the difference between transcript levels in the WT strains between NR and R mutants based on a two-sided Mann-Whitney test. **(B)** Frequency density distribution of the number of proteins per WT cell [27] for NR and R mutants, adapted from Kemmeren et al. [3]. Only detectable proteins are depicted. *P*-value indicates the difference between protein levels in NR and R mutants based on a two-sided Mann-Whitney test. **(C)** Number of NR and R mutants with a close paralog (CP) that arose from whole-genome duplication (WGD) [31] or small scale duplications (SSD) [20]. **(D)** Box plots showing the number of significant negative synthetic genetic interaction (SGI) scores [22] ($\varepsilon \leq -0.08$, $p \leq 0.05$; left panel), number of protein-protein interactions [30](PPI, middle panel) and sensitivity to different conditions [8] (percentage of conditions that deletion mutants show a relative growth defect; growth defect > 0, $p \leq 0.05$; right panel) for NR and R mutants. *P*-values are based on a two-sided Mann-Whitney test.

Pairs with a complete redundancy relationship are expected to have low sensitivity to different environmental conditions as measured in this dataset (Materials and Methods). Indeed, close paralog pairs between non-responsive mutants exhibit low sensitivity to different environmental conditions (Fig 3A, left panel). On the other hand, close paralog pairs between non-responsive and responsive mutants or between responsive mutants show a significant increase in sensitivity to different conditions (Fig 3A, middle & right panel; Fig 3B). This effect is stronger when only considering negative genetic interactions within the pairs depicted in Fig 3A (Fig 3C, D), confirming that negative genetic interactions provide useful additional information when investigating redundancy.

Interestingly, a large difference in sensitivity between the various close paralog groups is observed (Fig 3A and 3B), suggesting that there might be a disparity in sequence divergence among the close paralog pairs. The number of non-synonymous substitutions per non-synonymous site (Ka) is often considered an indicator of sequence divergence [14]. Although there seems to be a slight difference in the Ka values for close paralog pairs between non-responsive (NR-NR), between non-responsive and responsive (NR-R) or between responsive mutants (R-R), this is not significant (S1A Fig). When considering percentage sequence identity, a significant difference is only observed when using the domains shared between close paralog pairs (Materials and Methods; S1B and S1C Fig). This effect is lost when using all pairs but close paralog pairs (S1D Fig), suggesting that compensation capability is conserved between pairs with the same ancestral gene.

To ensure that homology-based redundancy estimates are reliable, several sensitivity cutoffs and their corresponding false-discovery rate (FDR) were investigated (Materials and Methods). Selecting close paralog pairs between non-responsive mutants with 5% or less sensitivity leads to the lowest FDR (≈ 10%; S2 Fig). Based on this, the contribution of homology-based redundancy to non-responsiveness is estimated at approximately 17% (Fig 3A, left panel; S1 Table). This encompasses 133 unique genes out of 784 NR mutants that are likely non-responsive due to a complete redundancy relationship with another gene.

Negative genetic interactions indicate pathway-based redundancy

In addition to homology-based redundancy (Fig 3A and 3B), where two genes can directly compensate for each other's loss, two genes can also operate in two redundant pathways. Negative genetic interactions have in the past been associated with redundant pathways [34]. Pairs of non-responsive mutants with negative genetic interactions show less sensitivity to different conditions compared to pairs between non-responsive and responsive or between responsive mutants (Fig 4A and 4B), as also observed before for close paralog pairs (Fig 3). However, taking into account the 5% sensitivity cutoff for estimating the contribution of pathway-based redundancy, the FDR rate is much higher (approximately 70%) compared to homology-based redundancy estimates. This clearly indicates that only using negative

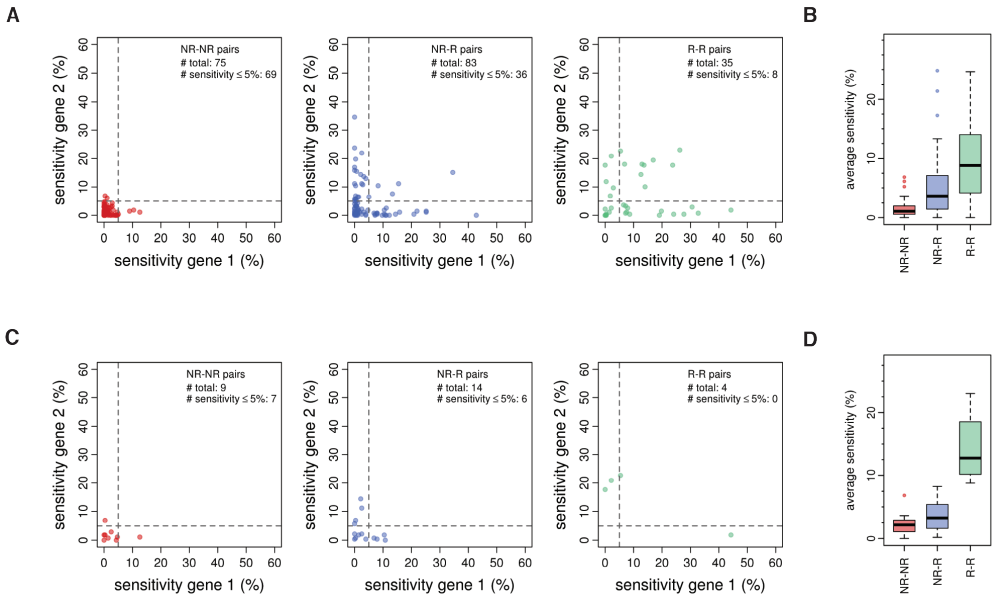


Fig 3. Homolog-based redundancy estimates. (A) Scatter plots depicting the sensitivity of gene 1 and gene 2 for close paralog pairs between non-responsive (NR-NR, left panel), between non-responsive and responsive (NR-R, middle panel) or between responsive (R-R, right panel) mutants. **(B)** Boxplots of the average sensitivity of the same groups as in A. The difference between all three groups is statistically significant (p-value $_{(NR\text{-}NR\ vs.\ NR\text{-}R)} = 3.14\times10^{-9}$; p-value $_{(NR\text{-}NR\ vs.\ R\text{-}R)} = 1.02\times10^{-8}$; p-value $_{(NR\text{-}R\ vs.\ R\text{-}R)} = 1.66\times10^{-3}$) based on a one-sided Mann-Whitney test. **(C)** Same as in A, except that only close paralog pairs with a significant negative SGI score are shown. **(D)** Boxplots of the average sensitivity of the same groups as in C (p-value $_{(NR\text{-}NR\ vs.\ NR\text{-}R)} = 0.128$, p-value $_{(NR\text{-}NR\ vs.\ R\text{-}R)} = 1.4\times10^{-3}$, p-value $_{(NR\text{-}R\ vs.\ R\text{-}R)} = 3.27\times10^{-3}$), statistical test as in B.

genetic interactions as a proxy for pathway redundancy is insufficient for estimating the contribution of pathway-based redundancy to non-responsiveness.

If two genes operate in a completely redundant manner, they should only exhibit a negative genetic interaction score with each other and not with any other gene [10]. Since the removal of a single gene can be completely compensated by the other gene, there should be no measurable phenotypic effect, except when both genes are deleted. The number of negative genetic interactions is indeed significantly lower for pairs between non-responsive mutants compared to responsive mutants (Fig 4C). Using a cutoff on the number of negative genetic interactions (40), a similar degree of reliability (FDR ≤ 10%) as obtained for homology-based redundancy estimates can be obtained (Fig 4D and 4E; Materials and Methods). Pathway-based redundancy estimates were therefore based on pairs between non-responsive mutants that fulfilled four criteria. First, pairs are not close paralogs of each other. Second, pairs have a significant negative genetic interaction. Third, pairs have 40 or less negative genetic interactions with other genes. And last, pairs show,

on average, 5% or less sensitivity in other conditions. In total, 68 pairs, 97 unique genes, can be associated with pathway-based redundancy (Fig 4D, left panel; S2 Table), indicating that approximately 12% (FDR ≈ 8%) of non-responsiveness can be attributed to pathway-based redundancy.

Relative contribution of redundancy and condition dependency to non-responsiveness

By combining the estimates of homology-based redundancy (133 genes, 17%) and pathway-based redundancy (97 genes, 12%), the overall contribution of redundancy to non-responsiveness is approximately 29% (Fig 5A). This number is a rough estimate and the exact percentage also depends on the cutoffs used. Here, stringent cutoffs are used to avoid a high FDR, but therefore potentially interesting pairs might be missed that do not fulfill the stringent criteria. Nevertheless, the analysis indicates that the majority of non-responsiveness, as much as 71%, is attributable to condition dependency (554 genes; Fig 5A; S3 Table). Interestingly, different functional classes do show diverse behavior in the relative contribution of condition dependency and redundancy (Fig 5B). Non-responsiveness for classes such as meiosis, DNA replication and repair have a relatively high contribution of condition dependency. A few classes such as gene-specific transcription factors, protein kinases and protein phosphatases however, show a much higher contribution of homology-based redundancy, indicating that the degree of redundancy differs between different functional classes (Fig 5B).

Mechanistic explanations for homology-based as well as pathway-based redundancy have been proposed before [28,29]. For homology-based redundancy, this is based on two proteins that can completely replace each other's function (Fig 5A). For instance, the redundant kinases Ark1 and Prk1 that both phosphorylate the same consensus target amino acid sequence and are involved in regulating endocytosis and actin skeleton [24,35]. Another example includes Upc2 and Ecm22, two redundant gene-specific transcription factors that bind to the same set of target genes and regulate ergosterol biosynthesis [36,37]. A final example involves Nhp6a and Nhp6b, two highly homologous HMGB proteins involved in modulation of chromatin structure [38]. These examples are all indicative of more direct relationships for homology-based redundancy pairs with proteins having the ability of taking over the exact same function.

For pathway-based redundancy, the concept is that two pathways or proteins can compensate each other's function in a more indirect fashion (Fig 5A). Although the exact relationship between many of these pairs is still unclear, they are often found to be involved in the same process and loss of both leads to significant unexpected defects. One example is the relationship between Ecm1 and Pom34. Ecm1 is a pre-ribosomal factor involved in pre-60S particle export and Pom34 is a subunit of the nuclear pore complex (NPC). In the absence of Pom34, depletion of Ecm1 leads to defects in pre-60S particle export [39].

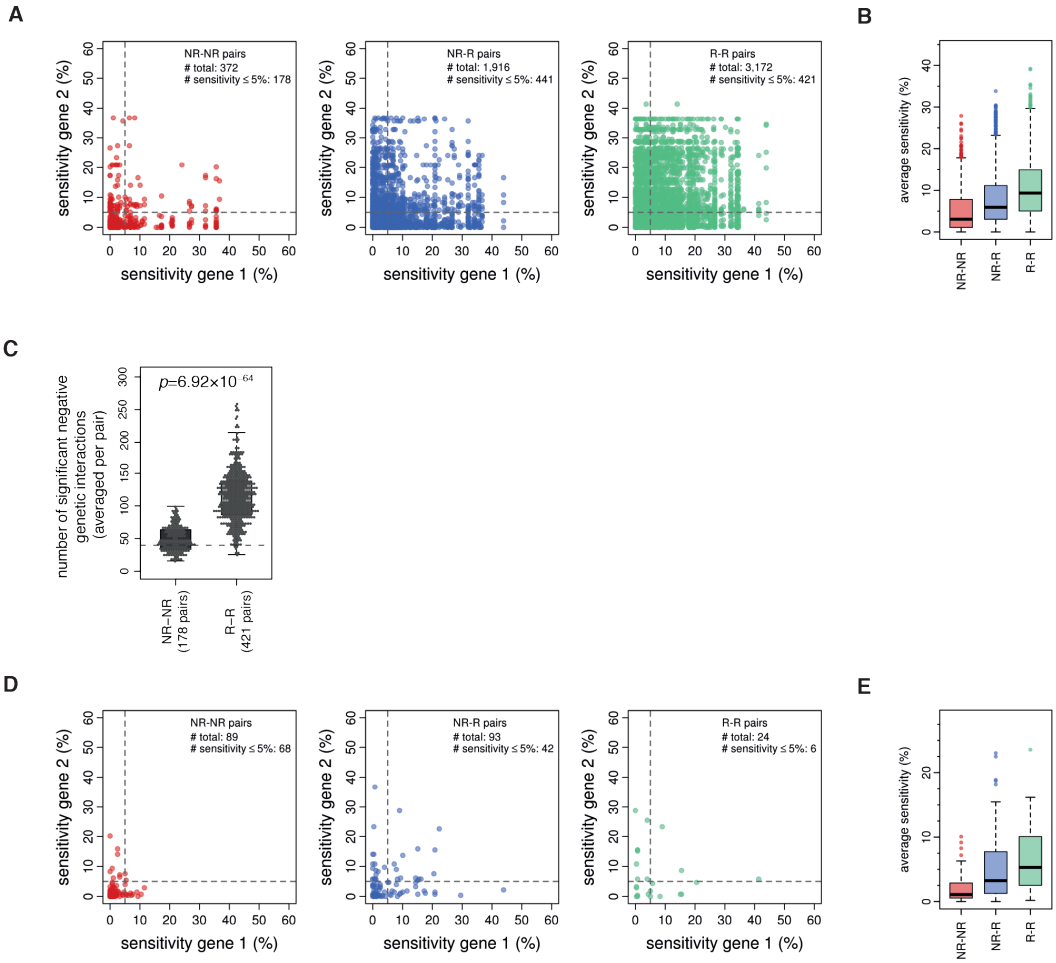


Fig 4. Pathway-based redundancy estimates. (A) Scatter plots depicting the sensitivity of gene 1 and gene 2 for pairs with a significant negative genetic interaction score between non-responsive (NR-NR), between non-responsive and responsive (NR-R) or between responsive (R-R) mutants. Close paralog pairs are excluded. **(B)** Boxplots of the average sensitivity of the same groups as in B, all three are significantly different (p-value < 2.2×10^{-16}). P-values are based on a one-sided Mann-Whitney test. **(C)** Boxplot showing the number of significant negative genetic interactions for NR-NR and R-R pairs with 5% or less sensitivity. Close paralog pairs are excluded. The dashed line indicates the threshold (40) taken to obtain an FDR of 8%. **(D)** Scatter plots shown for the same pairs in A that in addition have, on average, less than 40 negative genetic interactions per pair. **(E)** Boxplots of the average sensitivity of the same groups as in D (p-value $_{(\text{NR-NR vs. NR-R})} = 1.31\times10^{-7}$, p-value $_{(\text{NR-NR vs. R-R})} = 3.64\times10^{-6}$, p-value $_{(\text{NR-R vs. R-R})} = 0.064$), statistical test as in B.

Another example includes Kar5 and Prm3, which are both involved in nuclear membrane fusion during karyogamy and loss of both genes leads to nuclear fusion defects [40]. Note that, as highlighted with these cases, the underlying mechanism for pairs associated with pathway-based redundancy is often unclear.

Potential mechanisms for condition dependent genes

The mechanistic explanations provided for redundancy are useful in understanding the interplay between different proteins and how this relates to cellular processes. For genes/proteins required only under certain conditions, regulation can occur at many different levels, such as regulation of mRNA and protein levels, post-translational modifications (PTMs) that change aspects of protein-protein interactions and cellular localization. Many genes that are non-responsive still have relatively high mRNA and protein expression levels (Fig 2A and 2B), suggesting that these are constitutively expressed despite not being required under the given condition. For cell-cycle related protein complexes, De Lichtenberg et al. [41] have proposed a mechanism termed "just-in-time assembly" to indicate that most members of these protein complexes are constitutively expressed and only a few key subunits need to be regulated in order to control the activity of the entire protein complex. Many proteins are expressed under the condition investigated here, but do not result in a detectable phenotypical defect when removed. This therefore raises the question of whether "just-in-time assembly" may also apply to condition dependency.

To investigate this, all proteins that are part of a protein complex were obtained using a predefined definition of protein complexes [42]. Out of 554 genes attributed to condition dependency 181 are part of 127 unique protein complexes (Fig 5A). To determine the degree of differential regulation within each protein complex, highly regulated subunits were detected based on the degree of variation in expression levels across different deletion mutants (Material and Methods). Out of the 127 protein complexes, 18 complexes contain a highly regulated subunit (≈14%; Fig 5A; S3 Fig). To ensure that these results are not limited to the subset of proteins analyzed here, the same analysis was also performed using all protein coding genes in yeast. This revealed that 69 protein complexes out of 501 protein complexes [42] contain a highly regulated subunit (≈14%), confirming the results obtained when using the non-responsive mutants attributed to condition dependency.

One example of a protein complex with a highly regulated subunit includes the Dcs1-Dcs2 heterodimer that removes capped mRNA fragments left over from mRNA decay [43]. Previous reports indicated that Dcs2 is most sensitive to transcriptional induction and suppresses Dcs1 activity in response to nutrient stress [44,45]. Here, Dcs2 is identified as a highly regulated subunit (Fig 5C), confirming its role as a modulator of the activity of the Dcs1-Dcs2 heterodimer. Another example includes the GID complex, consisting of seven

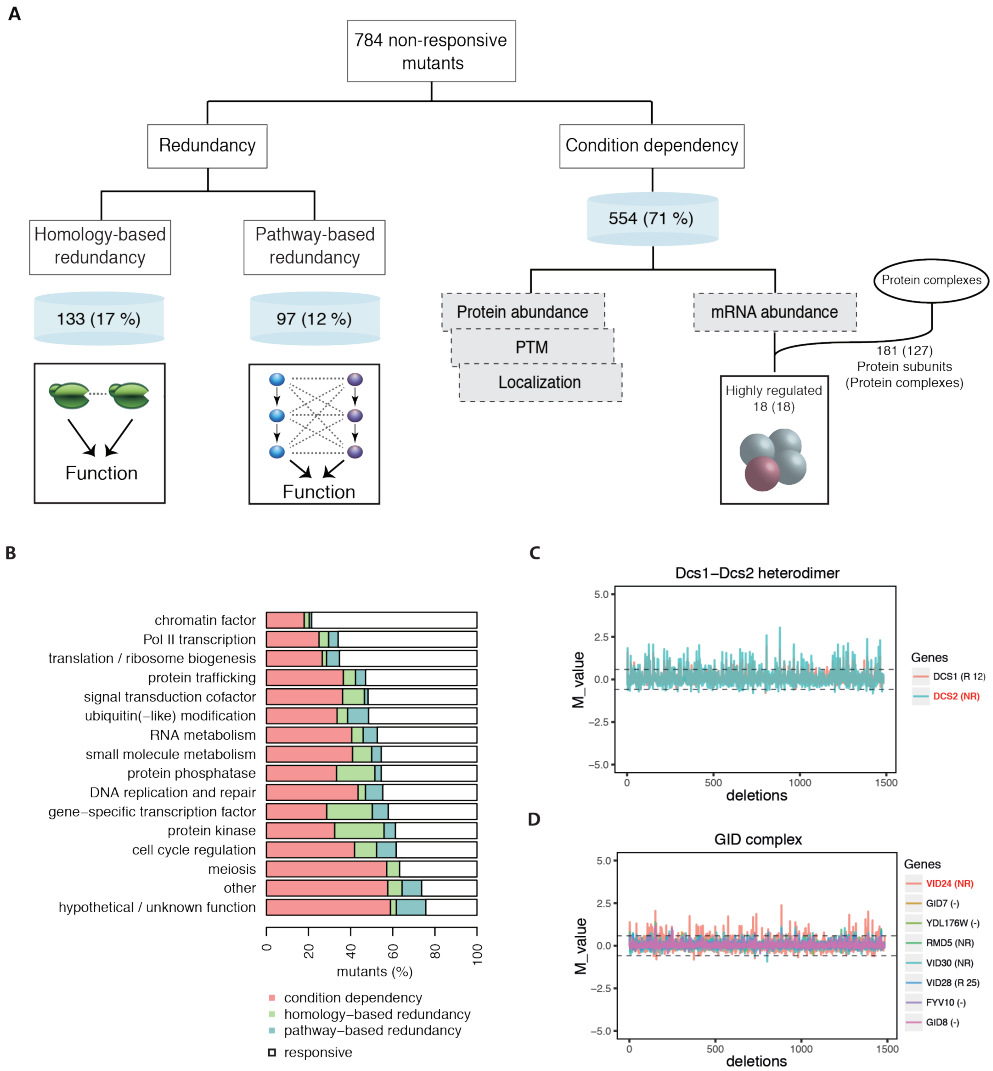


Fig 5. Relative contribution of redundancy and condition dependency and potential mechanisms. (A) Flowchart showing the relative contribution of redundancy and condition dependency to non-responsiveness. Potential mechanisms are indicated for homology-based redundancy, pathway-based redundancy and "just-in-time assembly". **(B)** Relative contribution of redundancy and condition dependency to non-responsiveness within each functional category. **(C)** Line plot showing the mRNA expression changes across 1,484 deletion mutants [3] for the Dcs1-Dcs2 heterodimer. Individual lines indicate the Dcs1 and Dcs2 subunit. The highly regulated subunit is indicated in red. **(D)** Line plot as in C for the GID complex.

subunits [46–48]. The GID complex targets the key enzyme fructose-1,6-biphosphatase (FBPase) for proteosomal degradation during the switch from gluconeogenesis to glycolysis [48]. Vid24, the regulatory subunit of the GID complex, initiates FBPase polyubiquitination and subsequent degradation [48]. Vid24 protein levels are undetectable under gluconeogenic conditions, but rapidly accumulate when glucose is added and decrease together with FBPase [48]. Our analysis also reveals Vid24 as a highly regulated subunit (Fig 5D), consistent with previous reports about the role of Vid24 in regulating the activity of the GID complex [46,48]. Besides these known examples, there are also putative protein complexes that have a highly regulated subunit (S3 Fig) that can be used as a starting point for further unraveling their function and potential regulatory mechanism. Taken together, approximately 14% of protein complexes are regulated similar to the "just-in-time assembly" model, but in a condition-dependent manner.

Discussion

Redundancy estimates

The high proportion of gene deletions that fail to exhibit a detectable defect under a specific condition has muddled efforts aimed at systematic phenotyping. Elucidating gene function requires the ability to elicit a detectable phenotypical readout, such as a growth defect or gene expression changes. Over 66% of deletion mutants in *Saccharomyces cerevisiae* have no detectable growth phenotype [15] and over 53% of non-essential gene deletions hardly show any gene expression changes [3]. Understanding the underlying causes of this non-responsive behavior will aid in elucidating gene function and understanding the complexity of cellular systems. Here, we estimate that approximately 29% of non-responsiveness can be attributed to either homology-based (17%) or pathway-based (12%) redundancy. Previous estimates have indicated that between 23% and 29% of negative genetic interactions can be ascribed to duplicated genes [7,10]. Our estimate of 17% is most likely lower because we require low sensitivity under multiple conditions (Fig 3A) whereas previous estimates are based on growth under a single condition. For complete redundancy pairs, deletion of a single gene should exhibit no detectable phenotype under any given condition. It is therefore expected that the redundancy pairs that we miss, but are included in previous estimates will fail to exhibit a redundancy relationship when investigated under different conditions.

Homology-based redundancy versus pathway-based redundancy

Estimates for the contribution of redundancy to non-responsiveness include both homology-based as well as pathway-based redundancy. The known examples found for pathway-based redundancy indicate a higher degree of uncertainty as they are found to be involved in the same process and loss of both leads to significant defects. The exact

mechanism however, is often unknown. Since there is no direct sequence-based evidence for a redundancy relationship between these pairs, it remains to be seen to what degree pathway-based redundancy relationships can fully compensate the loss of one gene, or whether this is only partial. One can also question whether these should be considered as redundancy relationships, or whether these most likely reflect pathway connectivity or combinatorial effects as already suggested before [24,28]. The most likely candidate pairs providing redundancy are therefore those based on homology (estimated at 17%), whereas most other negative genetic interactions reflect pathway connectivity.

Partial redundancy

Partial redundancy is also expected to contribute to non-responsive behavior. Simplistically, two different types can be differentiated. For the first type, only one of the two genes involved elicits a response under the condition investigated. The other type consists of a pair whereby deletion of either gene does not result in a detectable phenotype under the condition investigated. However, at least one of the genes would elicit a response when removed under a different (environmental) condition. We deliberately did not include these pairs within our redundancy estimates, as the purpose here is to investigate to what degree complete redundancy contributes to non-responsive behavior. Pairs that can fully compensate for each other's loss are expected to do so given any circumstance. It is therefore much more likely that partial redundancy is either an intrinsic property of pathway connectivity, just as for many pathway-based redundancy pairs, or is a different way to achieve condition dependency. The relationship might then indicate a manner to regulate condition dependency or ensure that genes with a similar function, but required under different environmental conditions can behave in a switch-like fashion so that the most efficient protein is available under the right circumstances.

Condition dependency and just-in-time assembly

While condition dependency can be regulated in many different ways, we focused here on a mechanism shown previously for cell cycle related protein complexes [41]. The just-in-time assembly mechanism may provide a mechanistic explanation how protein complexes can be regulated in a condition-dependent manner. Regulation of approximately 14% of all protein complexes can be explained in such a way. This still leaves many protein complexes unexplained and indicates that the just-in-time assembly is not such a widespread mechanism. Interestingly, many of the protein complexes with a highly regulated subunit are involved in metabolism. Almost 50% of these protein complexes are associated with metabolism and this percentage is even higher for heterodimers. Metabolic flux analyses have shown that the activity of metabolic pathways can be regulated in different ways and through several regulatory factors [49]. Although not necessarily the most predominant mode of action, transcriptional regulation of key regulatory factors has been shown as

one way to achieve condition- and tissue-specific control of metabolic pathway activity [50]. The enrichment of heterodimers also fits the idea of enzymatic protein complexes consisting of a regulatory and catalytic subunit. The mechanism found here, might thus be more specific for metabolic pathways and protein complexes involved in metabolism.

The term just-in-time assembly proposed by De Lichtenberg et al. [41] to describe the dynamic formation of protein complexes during the cell cycle might not always be appropriate for the highly regulated subunits found here. The term implies that the entire complex is assembled just before it is needed. This doesn't necessarily need to be the case. For instance, most of the protein complex could already be pre-assembled and only the regulatory subunit joins the partly pre-assembled protein complex when needed. Since it is impossible to assess the exact mechanism by which a protein complex is activated using only gene expression data, we propose to use "just-in-time activation" as a more general term. This also allows other types of just-in-time activation to be included, such as protein phosphorylation or other post-translational modifications. Additional data exploring condition dependent protein expression or post-translational modifications are needed to gain more detailed mechanistic insights.

Materials and Methods

Gene expression data and functional categories

The list of deletion mutants and corresponding gene expression profiles were obtained from Kemmeren et al [3]. Each gene expression profile consists of *p*-values and average mRNA transcript changes (M values; $\log_2$(fold-change)) relative to wildtype (WT). Genes are considered significantly changed when fold-change (FC) > 1.7 and *p*-value < 0.05. Genes that were frequently changing regardless of the deletion mutant (WT variable genes) were excluded as described in the original study [3]. The number of differentially expressed transcripts was used to classify mutants either as responsive (R) or non-responsive (NR) as done previously [3]. Mutants with less than 4 transcript changes relative to WT are considered NR and mutants with 4 or more changes are considered R. Mutants were grouped in different functional categories as before [3]. Both the overall ratio of NR and R mutants as well as for each individual functional category (Fig 1) was calculated.

mRNA transcript levels in WT

A wild-type pool consisting of 200 WTs [3] was used to obtain average mRNA transcript levels (A values; $\log_2$(R*G)) for all genes corresponding to the different deletion mutants. mRNA transcript levels in WT were compared between NR and R mutants. *P*-values are calculated using a Mann-Whitney two-sided test to evaluate differences in mRNA transcript levels between NR and R mutants (Fig 2A).

Protein expression levels, protein-protein interactions and protein complexes

The number of protein molecules per WT cell was obtained from Ghaemmaghami et al. [27]. Proteins with undetectable signals, low signals or experimental problems were excluded. Protein expression levels were compared between NR and R mutants. *P*-values are calculated using a Mann-Whitney two-sided test to evaluate differences in protein levels between NR and R mutants (Fig 2B). Protein-protein interactions were obtained from Collins et al. [30]. The number of protein-protein interactions between NR and R mutants was compared. *P*-values are calculated using a Mann-Whitney two-sided test (Fig 2D, middle panel). Protein complexes (501) were obtained from the curated ''consensus + GO'' set from Benschop et al. [42].

Close paralog genes

For close paralog genes, both small scale duplications (SSD) as well as whole-genome duplications (WGD) are used. A list of SSD pairs was obtained from Guan et al. [20]. WGD (547 pairs) were obtained from Byrne and Wolf [31]. Global protein sequence alignment for each close paralog pair was performed using the Needleman-Wunsch algorithm [51] as implemented in Needle-EMBOSS version 6.6 [52] with default options. Only SSD pairs with 20% or higher identity were kept (751 pairs). Combining both WGD and SSD resulted in a total of 1,298 close paralog pairs. Close paralog pairs for which both genes were profiled [3] were used for further downstream analyses (205 pairs).

Genetic interaction data

Synthetic genetic interaction (SGI) scores were obtained from Costanzo et al. [22]. SGI scores were acquired for all pairwise combinations of NR and R mutants for which a score was calculated (Figs 3 and 4). For determining significant negative genetic interactions, the same cutoff was applied as used originally [22] ($\varepsilon \leq -0.08$, $p \leq 0.05$).

Condition sensitivity

Growth rate of yeast homozygous gene deletions compared against WT under multiple conditions (418 experiments) was obtained from Hillenmeyer et al. [8]. A relative growth rate above zero is an indication of a growth defect of a deletion mutant relative to WT in a particular condition. A two-sided t-test was applied to assign *p*-values. Growth rates from duplicate conditions were averaged, resulting in 278 unique conditions. Only significant growth defects were selected for downstream analysis ($p \leq 0.05$). The sensitivity of each gene is expressed as the percentage of significant relative growth defects within the 278 unique conditions.

Sequence divergence

The number of non-synonymous substitutions per non-synonymous site (Ka) was calculated based on a method introduced by Yang and Nielsen [53]. This method is implemented in the GenomeHistory 2.0 tool, which was used to calculate Ka values [54]. Non-default parameters chosen include: minimum translated ORF: 100; minimum number of aligned residues to accept pair: 100; accepting all BLAST hits with e <= 1e-08 and minimum percentage identity for analysis: 40% (S1A Fig). Protein sequence identity for close paralog pairs was calculated either across the whole protein sequence or only for protein domains using the Needleman-Wunsch algorithm [51] as implemented in Needle-EMBOSS version 6.6 [52] with default options (S1B Fig). InterProscan version 5.10-50 [55] was used to obtain a list of shared domains between genes. Only protein domains identified by Pfam [56] were included in downstream analyses. Sequence identity was calculated between two shared Pfam domains using the same tool as used for complete protein sequences. Gene pairs with at least one shared Pfam domain were included in the downstream analysis. Sequence identity for Pfam domains with the same identifier were averaged before averaging all distinct shared domains across two proteins.

Averages of sequence identity between shared Pfam domains for gene pairs were compared either for close paralog or non-paralog pairs. *P*-values are calculated using a Mann-Whitney two-sided test (S1C and S1D Fig).

Contribution of redundancy and condition dependency to non-responsiveness

Homology-based redundancy pairs had to fulfill the following criteria: the pair is a close paralog pair (either WGD or SSD), both genes are NR and both genes have a condition sensitivity of 5% or less. For calculating the false-discovery rate (FDR), NR-NR close paralog pairs below the sensitivity cutoff were used as true positives and R-R close paralog pairs below the sensitivity cutoff were used as false positives. The lowest false-discovery rate (FDR) was achieved when the applied cutoff was 5% sensitivity (S2 Fig). Pathway-based redundancy pairs had to fulfill the following criteria: the pair has a significant negative genetic interaction score, both genes are NR and both genes have a condition sensitivity of 5% or less. To obtain a similar degree of reliability for pathway-based redundancy estimates, the maximum allowed FDR was set to 10% (as obtained from the homology-based redundancy estimates). In addition to the abovementioned criteria, pathway-based redundancy pairs also have to have less than 40 significant negative genetic interaction scores with other genes in order to obtain a FDR lower than 10%. A pair that isn't classified as homology-based redundancy or pathway-based redundancy is considered condition dependent.

Highly regulated genes

Highly regulated genes were determined within the collection of deletion mutants [3]. For each transcript, the observed standard deviation across different deletion mutants or conditions was compared against 1,000,000 permuted transcript profiles. The permuted transcript profiles were generated by randomly selecting expression values within each column (deletion mutant) a million times. *P*-values were calculated as the fraction of permuted profiles with standard deviations equal or larger than the observed standard deviation. Transcripts with p-values < 0.05 after multiple testing correction (Benjamini & Hochberg) are considered highly regulated. Wildtype variable genes [3] were excluded from the list of highly regulated genes.

Availability of data and materials

All relevant data are available within the manuscript and corresponding Supporting Information.

Funding

Financial support was provided by the Netherlands Organization for Scientific Research (NWO), grant 864.11.010 (SA & PK).

Author Contributions

Conceptualization: PK FCPH SA
Formal analysis: PK FCPH SA
Funding acquisition: PK
Methodology: PK FCPH SA
Software: SA
Supervision: PK FCPH
Visualization: SA
Writing – Original Draft Preparation: PK FCPH SA

Acknowledgements

We would like to thank Philip Lijnzaad, Thanasis Margaritis, Wim de Jonge and Berend Snel for helpful discussions and advice throughout the project.

References

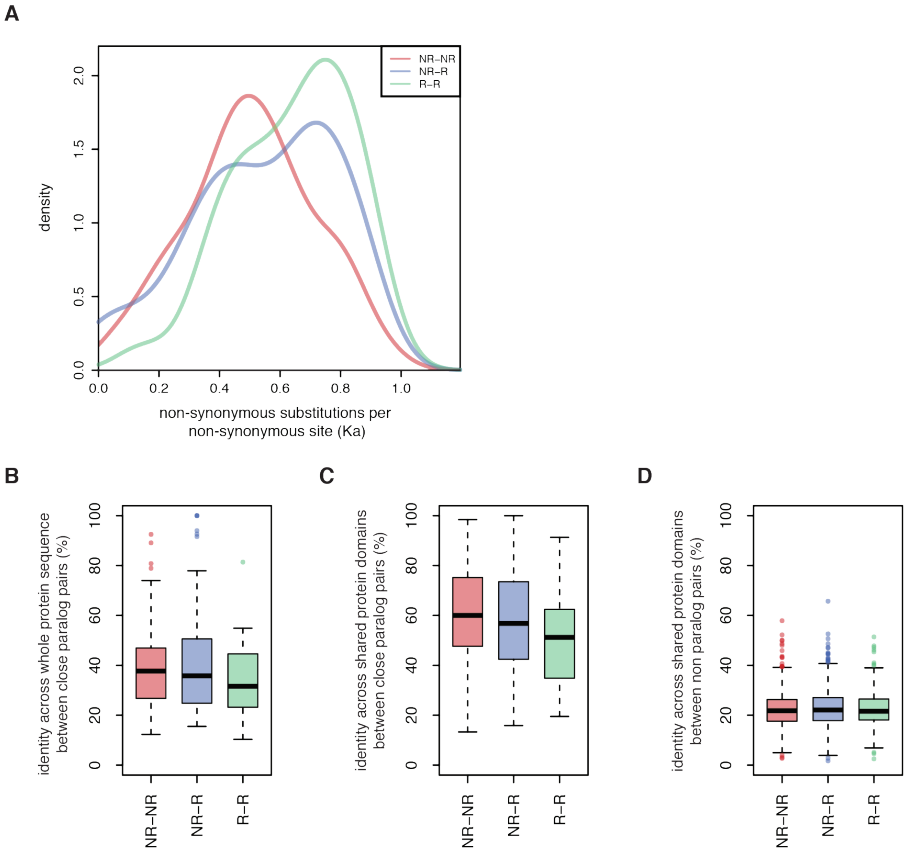
1. Wright AV, Nuñez JK, Doudna JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell. 2016;164: 29–44. doi:10.1016/j.cell.2015.12.035
2. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science. 1999;285: 901–906. doi:10.1126/science.285.5429.901
3. Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, et al. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. Cell. 2014;157: 740–752. doi:10.1016/j.cell.2014.02.054
4. Liberali P, Snijder B, Pelkmans L. Single-cell and multivariate approaches in genetic perturbation screens. Nat Rev Genet. 2015;16: 18–32. doi:10.1038/nrg3768
5. Giaever G, Nislow C. The Yeast Deletion Collection: A Decade of Functional Genomics. Genetics. 2014;197: 451–465. doi:10.1534/genetics.114.161620
6. Usaj MM, Styles EB, Verster AJ, Friesen H, Boone C, Andrews BJ. High-Content Screening for Quantitative Cell Biology. Trends Cell Biol. 2016;26: 598–611. doi:10.1016/j.tcb.2016.03.008
7. Gu Z, Steinmetz LM, Gu X, Scharfe C, Davis RW, Li W-H. Role of duplicate genes in genetic robustness against null mutations. Nature. 2003;421: 63–66. doi:10.1038/nature01198
8. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, et al. The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. Science. 2008;320: 362–365. doi:10.1126/science.1150021
9. Hsiao T-L, Vitkup D. Role of Duplicate Genes in Robustness against Deleterious Human Mutations. PLoS Genet. 2008;4: e1000014. doi:10.1371/journal.pgen.1000014
10. Ihmels J, Collins SR, Schuldiner M, Krogan NJ, Weissman JS. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. Mol Syst Biol. 2007;3: 86. doi:10.1038/msb4100127
11. Wagner A. Robustness against mutations in genetic networks of yeast. Nat Genet. 2000;24: 355–361. doi:10.1038/74174
12. Li J, Yuan Z, Zhang Z. The Cellular Robustness by Genetic Redundancy in Budding Yeast. PLoS Genet. 2010;6: e1001187. doi:10.1371/journal.pgen.1001187
13. Zhang J. Genetic Redundancies and Their Evolutionary Maintenance. In: Soyer OS, editor. Evolutionary Systems Biology. Springer New York; 2012. pp. 279–300. doi:10.1007/978-1-4614-3567-9_13
14. Plata G, Vitkup D. Genetic robustness and functional evolution of gene duplicates. Nucleic Acids Res. 2013; gkt1200. doi:10.1093/nar/gkt1200
15. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002;418: 387–391. doi:10.1038/nature00935

16. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, et al. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature. 2003;421: 231–237. doi:10.1038/nature01278
17. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, et al. Essential Bacillus subtilis genes. Proc Natl Acad Sci U S A. 2003;100: 4678–4683. doi:10.1073/pnas.0730515100
18. Papp B, Pál C, Hurst LD. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature. 2004;429: 661–664. doi:10.1038/nature02636
19. Lynch M, Conery JS. The Evolutionary Fate and Consequences of Duplicate Genes. Science. 2000;290: 1151–1155. doi:10.1126/science.290.5494.1151
20. Guan Y, Dunham MJ, Troyanskaya OG. Functional Analysis of Gene Duplications in Saccharomyces cerevisiae. Genetics. 2007;175: 933–943. doi:10.1534/genetics.106.064329
21. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. Science. 2004;303: 808–813. doi:10.1126/science.1091317
22. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science. 2010;327: 425–431. doi:10.1126/science.1180823
23. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016;353: aaf1420. doi:10.1126/science.aaf1420
24. van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, et al. Functional Overlap and Regulatory Links Shape Genetic Interactions between Signaling Pathways. Cell. 2010;143: 991–1004. doi:16/j.cell.2010.11.021
25. Tischler J, Lehner B, Fraser AG. Evolutionary plasticity of genetic interaction networks. Nat Genet. 2008;40: 390–391. doi:10.1038/ng.114
26. Harrison R, Papp B, Pál C, Oliver SG, Delneri D. Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl Acad Sci. 2007;104: 2307–2312. doi:10.1073/pnas.0607153104
27. Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. Nature. 2003;425: 737–741.
28. Lehner B. Molecular mechanisms of epistasis within and between genes. Trends Genet. 2011;27: 323–331. doi:10.1016/j.tig.2011.05.007
29. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. Nat Rev Genet. 2007;8: 437–449. doi:10.1038/nrg2085
30. Collins SR, Kemmeren P, Zhao X-C, Greenblatt JF, Spencer F, Holstege FCP, et al. Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. Mol Cell Proteomics MCP. 2007;6: 439–50. doi:M600381-MCP200
31. Byrne KP, Wolfe KH. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. Genome Res. 2005;15: 1456–1461. doi:10.1101/gr.3672305
32. Dean EJ, Davis JC, Davis RW, Petrov DA. Pervasive and Persistent Redundancy among Duplicated Genes in Yeast. PLoS Genet. 2008;4: e1000113. doi:10.1371/journal.pgen.1000113

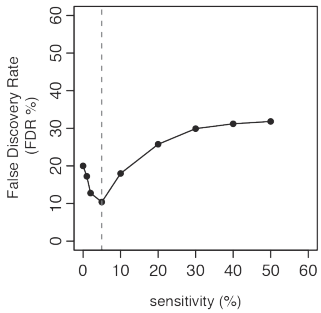
33. Musso G, Costanzo M, Huangfu M, Smith AM, Paw J, Luis B-JS, et al. The extensive and condition-dependent nature of epistasis among whole-genome duplicates in yeast. Genome Res. 2008;18: 1092–1099. doi:10.1101/gr.076174.108
34. Kelley R, Ideker T. Systematic interpretation of genetic interactions using protein networks. Nat Biotechnol. 2005;23: 561–566. doi:10.1038/nbt1096
35. Cope MJ, Yang S, Shang C, Drubin DG. Novel protein kinases Ark1p and Prk1p associate with and regulate the cortical actin cytoskeleton in budding yeast. J Cell Biol. 1999;144: 1203–1218.
36. Vik Å, Rine J. Upc2p and Ecm22p, Dual Regulators of Sterol Biosynthesis in Saccharomyces cerevisiae. Mol Cell Biol. 2001;21: 6395–6405. doi:10.1128/MCB.21.19.6395-6405.2001
37. Sameith K, Amini S, Groot Koerkamp MJA, van Leenen D, Brok M, Brabers N, et al. A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. BMC Biol. 2015;13: 112. doi:10.1186/s12915-015-0222-5
38. Stillman DJ. Nhp6: A small but powerful effector of chromatin structure in Saccharomyces cerevisiae. Biochim Biophys Acta BBA - Gene Regul Mech. 2010;1799: 175–180. doi:10.1016/j.bbagrm.2009.11.010
39. Yao Y, Demoinet E, Saveanu C, Lenormand P, Jacquier A, Fromont-Racine M. Ecm1 is a new pre-ribosomal factor involved in pre-60S particle export. RNA N Y N. 2010;16: 1007–1017. doi:10.1261/rna.2012310
40. Shen S, Tobery CE, Rose MD. Prm3p is a pheromone-induced peripheral nuclear envelope protein required for yeast nuclear fusion. Mol Biol Cell. 2009;20: 2438–2450. doi:10.1091/mbc.E08-10-0987
41. Lichtenberg U de, Jensen LJ, Brunak S, Bork P. Dynamic Complex Formation During the Yeast Cell Cycle. Science. 2005;307: 724–727. doi:10.1126/science.1105103
42. Benschop JJ, Brabers N, van Leenen D, Bakker LV, van Deutekom HWM, van Berkum NL, et al. A consensus of core protein complex compositions for Saccharomyces cerevisiae. Mol Cell. 2010;38: 916–928. doi:10.1016/j.molcel.2010.06.002
43. Liu H. The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases. EMBO J. 2002;21: 4699–4708. doi:10.1093/emboj/cdf448
44. Malys N, Carroll K, Miyan J, Tollervey D, McCarthy JEG. The 'scavenger' m7GpppX pyrophosphatase activity of Dcs1 modulates nutrient-induced responses in yeast. Nucleic Acids Res. 2004;32: 3590–3600. doi:10.1093/nar/gkh687
45. Malys N, McCarthy JEG. Dcs2, a Novel Stress-induced Modulator of m7GpppX Pyrophosphatase Activity that Locates to P Bodies. J Mol Biol. 2006;363: 370–382. doi:10.1016/j.jmb.2006.08.015
46. Regelmann J, Schüle T, Josupeit FS, Horak J, Rose M, Entian K-D, et al. Catabolite Degradation of Fructose-1,6-bisphosphatase in the Yeast Saccharomyces cerevisiae: A Genome-wide Screen Identifies Eight Novel GID Genes and Indicates the Existence of Two Degradation Pathways. Mol Biol Cell. 2003;14: 1652–1663. doi:10.1091/mbc.E02-08-0456

47. Pitre S, Dehne F, Chan A, Cheetham J, Duong A, Emili A, et al. PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs. BMC Bioinformatics. 2006;7: 365. doi:10.1186/1471-2105-7-365
48. Santt O, Pfirrmann T, Braun B, Juretschke J, Kimmig P, Scheel H, et al. The Yeast GID Complex, a Novel Ubiquitin Ligase (E3) Involved in the Regulation of Carbohydrate Metabolism. Mol Biol Cell. 2008;19: 3323–3333. doi:10.1091/mbc.E08-03-0328
49. Metallo CM, Vander Heiden MG. Understanding Metabolic Regulation and Its Influence on Cell Physiology. Mol Cell. 2013;49: 388–398. doi:10.1016/j.molcel.2013.01.018
50. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. Nat Rev Cancer. 2011;11: 85–95. doi:10.1038/nrc2981
51. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 1970;48: 443–453. doi:10.1016/0022-2836(70)90057-4
52. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet TIG. 2000;16: 276–277.
53. Yang Z, Nielsen R. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. Mol Biol Evol. 2000;17: 32–43.
54. Conant GC, Wagner A. GenomeHistory: a software tool and its application to fully sequenced genomes. Nucleic Acids Res. 2002;30: 3378–3386.
55. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinforma Oxf Engl. 2014;30: 1236–1240. doi:10.1093/bioinformatics/btu031
56. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44: D279–D285. doi:10.1093/nar/gkv1344

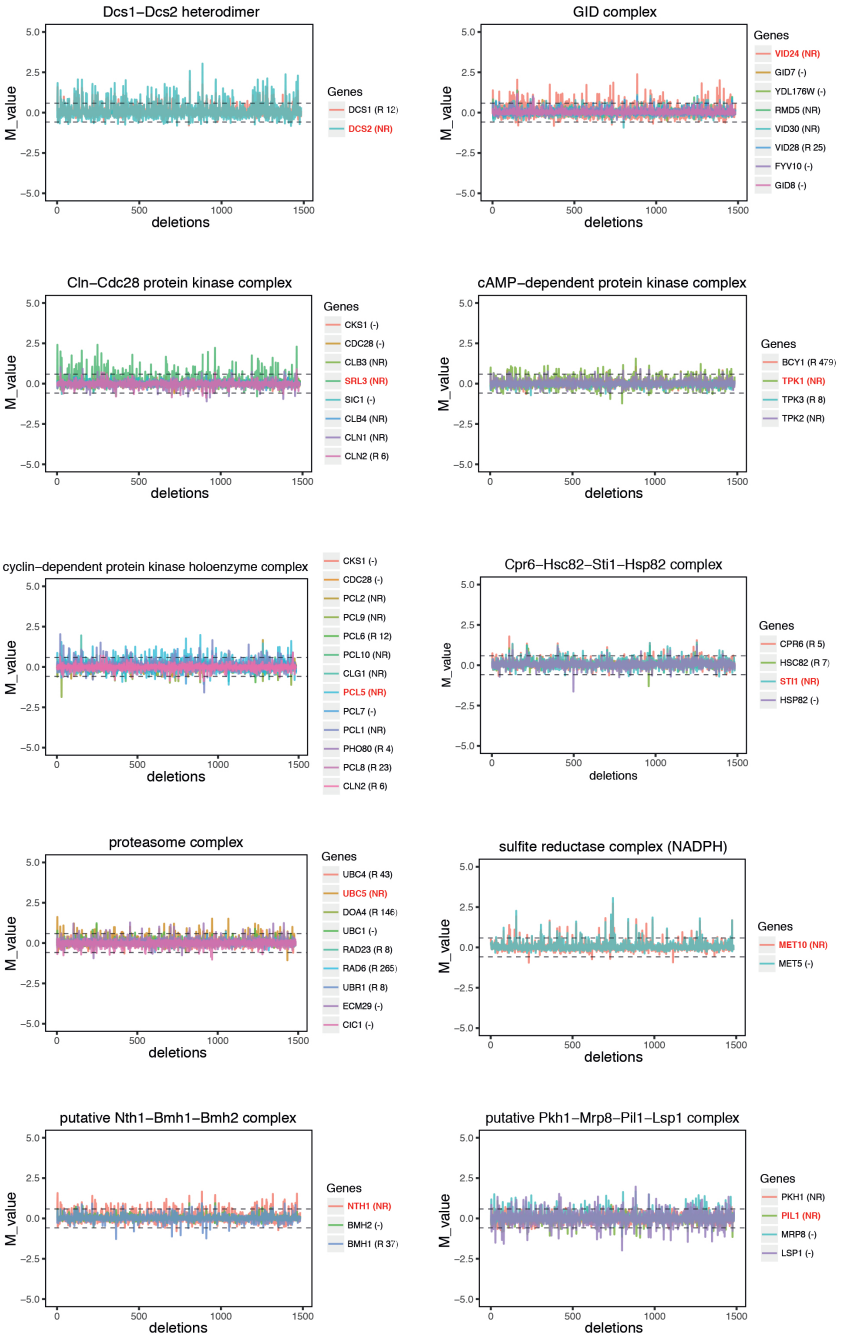
Supporting Information

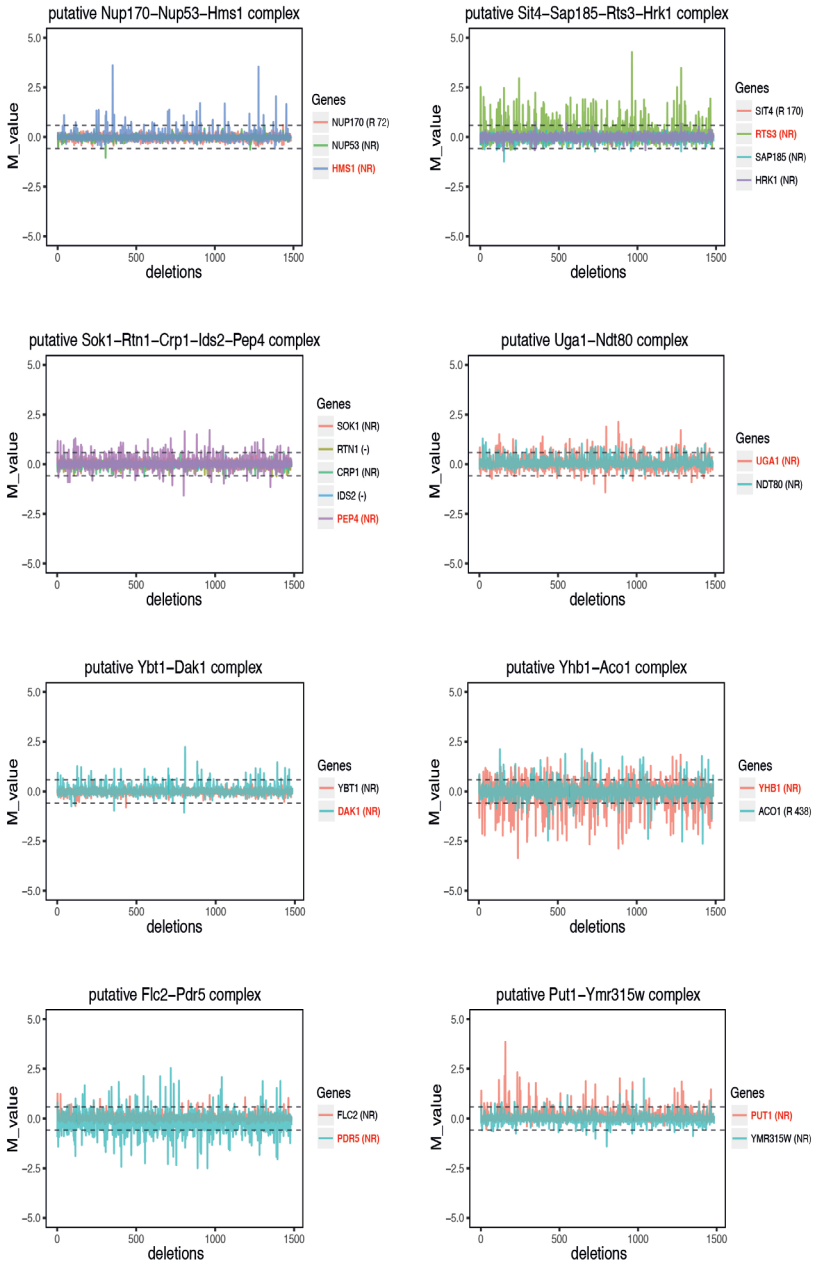


S1 Fig. Sequence divergence in close paralog pairs. (A) Ka values as a measurement of sequence divergence are depicted for NR-NR (75), NR-R (83) and R-R (35) pairs. **(B)** Boxplots showing percentage sequence identity across the entire protein length for the same pairs as in A. **(C)** Boxplots showing percentage sequence identity within shared domains for the same pairs as in A. The difference between NR-NR and R-R pairs is statistically significant (p-value = 0.008). **(D)** Boxplots showing percentage sequence identity within shared Pfam domains (close paralogs excluded). No significant difference is observed between the three groups. The number of investigated pairs is 96,141, 196,672 and 100,128 for NR-NR, NR-R, and R-R mutants, respectively. Of these pairs, 455, 736 and 320 share at least one Pfam domain and are included in this figure.



S2 Fig. Selected sensitivity cutoff for homology-based redundancy. Line plot showing the calculated FDR for different sensitivity cutoffs (Materials and Methods). The dashed line indicates the 5% sensitivity cutoff used to identify the contribution of homology-based redundancy to non-responsiveness (FDR ≈ 10%). The same cutoff was also used to infer the contribution of pathway-based redundancy.


Dcs1−Dcs2 heterodimer
GID complex
Cln−Cdc28 protein kinase complex
cAMP−dependent protein kinase complex
cyclin−dependent protein kinase holoenzyme complex
Cpr6−Hsc82−Sti1−Hsp82 complex
proteasome complex
sulfite reductase complex (NADPH)
putative Nth1−Bmh1−Bmh2 complex
putative Pkh1−Mrp8−Pil1−Lsp1 complex
M_value
deletions
Genes
DCS1 (R 12)
DCS2 (NR)
VID24 (NR)
GID7 (-)
YDL176W (-)
RMD5 (NR)
VID30 (NR)
VID28 (R 25)
FYV10 (-)
GID8 (-)
CKS1 (-)
CDC28 (-)
CLB3 (NR)
SRL3 (NR)
SIC1 (-)
CLB4 (NR)
CLN1 (NR)
CLN2 (R 6)
BCY1 (R 479)
TPK1 (NR)
TPK3 (R 8)
TPK2 (NR)
PCL2 (NR)
PCL9 (NR)
PCL6 (R 12)
PCL10 (NR)
CLG1 (NR)
PCL5 (NR)
PCL7 (-)
PCL1 (NR)
PHO80 (R 4)
PCL8 (R 23)
CPR6 (R 5)
HSC82 (R 7)
STI1 (NR)
HSP82 (-)
UBC4 (R 43)
UBC5 (NR)
DOA4 (R 146)
UBC1 (-)
RAD23 (R 8)
RAD6 (R 265)
UBR1 (R 8)
ECM29 (-)
CIC1 (-)
MET10 (NR)
MET5 (-)
NTH1 (NR)
BMH2 (-)
BMH1 (R 37)
PKH1 (NR)
PIL1 (NR)
MRP8 (-)
LSP1 (-)



S3 Fig. Highly regulated protein complex subunits. Line plots showing the mRNA expression changes across 1,484 deletion mutants for protein complexes with a highly regulated subunit (18 in total). Each line shows the expression changes observed for the individual subunits. Subunits highlighted in red have been identified as highly regulated.

S1 Table.
List of homology-based redundant pairs.

Gene1	Gene2	Duplication scale	Identity percentage	Responsiveness gene1	Responsiveness gene2	Sensitivity gene1	Sensitivity gene2
KCC4	GIN4	WGD	45.6	NR	NR	0	3
PLM2	TOS4	WGD	33.5	NR	NR	0.4	0.7
GLG2	GLG1	WGD	25.3	NR	NR	0	0
UBX7	UBX6	WGD	28.8	NR	NR	0	0.4
PES4	MIP6	WGD	35.7	NR	NR	0	0
RCK1	RCK2	WGD	37.7	NR	NR	0.4	1
IDP2	IDP3	WGD	74	NR	NR	2	1
GBP2	HRB1	WGD	47.2	NR	NR	0	0
PRR2	NPR1	WGD	33.6	NR	NR	0.4	3
YCK1	YCK2	WGD	65.5	NR	NR	5	1
PRK1	ARK1	WGD	38.5	NR	NR	1	0.4
ASM4	NUP53	WGD	29	NR	NR	4	0
PGM1	PGM2	WGD	78.9	NR	NR	1	0.7
STB6	STB2	WGD	41.3	NR	NR	0	1
OSH2	SWH1	WGD	49.7	NR	NR	3	2
NHP6B	NHP6A	WGD	80.8	NR	NR	1	2
GAL83	SIP2	WGD	45.2	NR	NR	0.4	0
MIG3	MIG2	WGD	33	NR	NR	5	0
ALK2	ALK1	WGD	30.6	NR	NR	0	0.4
CNA1	CMP2	WGD	61.6	NR	NR	1	1
FPR4	FPR3	WGD	58.4	NR	NR	0	0.7
RSF2	TDA9	WGD	37.7	NR	NR	0	1
YGR283C	YMR310C	WGD	30.6	NR	NR	3	0.4
DON1	CUE5	WGD	19.7	NR	NR	0.4	0.4
MTH1	STD1	WGD	57.5	NR	NR	0.4	2
MLP2	MLP1	WGD	26.6	NR	NR	3	3
UBP7	UBP11	WGD	32.1	NR	NR	4	0
PDR8	YRR1	WGD	22.4	NR	NR	0.4	3
PYC2	PYC1	WGD	92.5	NR	NR	0.4	0
MF(ALPHA)2	MF(ALPHA)1	WGD	46.7	NR	NR	0.4	0.4
RGM1	USV1	WGD	19.9	NR	NR	1	0.4
OAF1	PIP2	WGD	39.7	NR	NR	0	0.4
PSK1	PSK2	WGD	47.8	NR	NR	0	5
NTG1	NTG2	WGD	39.6	NR	NR	1	1

Gene1	Gene2	Duplication scale	Identity percentage	Responsiveness gene1	Responsiveness gene2	Sensitivity gene1	Sensitivity gene2
PCL9	PCL2	WGD	50.8	NR	NR	1	4
PPH22	PPH21	WGD	89.1	NR	NR	0.7	0.4
CLB3	CLB4	WGD	44.8	NR	NR	0	1
TOD6	DOT6	WGD	28.8	NR	NR	0.4	3
PTC3	PTC2	WGD	59.1	NR	NR	0	3
JSN1	PUF2	WGD	43.2	NR	NR	0.4	0.7
CUP2	HAA1	WGD	12.3	NR	NR	0.7	0.4
SUT1	SUT2	WGD	33	NR	NR	0.4	3
SAK1	TOS3	WGD	20.7	NR	NR	5	0
GRX3	GRX4	WGD	58.6	NR	NR	1	0
KIN1	KIN2	WGD	50.5	NR	NR	2	0
CTH1	TIS11	WGD	40.5	NR	NR	3	0
UPC2	ECM22	WGD	43.9	NR	NR	3	4
VHS1	SKS1	WGD	43.9	NR	NR	0	0
SMP1	RLM1	WGD	22.4	NR	NR	3	3
KIN4	FRK1	WGD	43.8	NR	NR	0	0
ZDS2	ZDS1	WGD	32.4	NR	NR	0	2
NGL3	NGL2	WGD	38.6	NR	NR	1	1
ERT1	GSM1	SSD	22.3	NR	NR	0	0
MAL33	YPR196W	SSD	70.6	NR	NR	3	0
RDS1	YLL054C	SSD	31	NR	NR	0	0
MSH6	MSH3	SSD	20.3	NR	NR	1	0.4
YDR131C	DAS1	SSD	20.5	NR	NR	0.4	5
HPA3	HPA2	SSD	44.1	NR	NR	0.7	0.7
YFL052W	MAL33	SSD	66.4	NR	NR	0.4	3
PKP2	PKP1	SSD	20.4	NR	NR	0	0
MAL13	MAL33	SSD	68.1	NR	NR	0	3
GAT4	GAT3	SSD	32.7	NR	NR	0.4	0
GSM1	RDS2	SSD	20.9	NR	NR	0	4
YKL222C	YRM1	SSD	26.9	NR	NR	0.4	0.7
KNS1	YAK1	SSD	20.4	NR	NR	1	3
RDS2	ERT1	SSD	22	NR	NR	4	0
KIP2	KIP3	SSD	24.1	NR	NR	2	3
CMR3	YPR015C	SSD	29.3	NR	NR	0	0.4

S2 Table.
List of pathway-based redundant pairs.

Gene1	Gene2	Responsiveness gene1	Responsiveness gene2	Sensitivity gene1	Sensitivity gene2	sgi score	*p* value	Negative sgi scores gene1	Negative sgi scores gene2
ECM1	THI2	NR	NR	2	0	-0.1359	0.00769	30	5
ECM1	POM34	NR	NR	2	0.7	-0.3493	3.82e-91	30	50
ECM1	KAR5	NR	NR	2	2	-0.1147	3.53e-21	30	21
AKL1	UGA3	NR	NR	4	0.4	-0.0897	0.00807	36	16
SLX1	CSN9	NR	NR	0	0.7	-0.088	0.00033	48	18
YCR016W	ARO80	NR	NR	0.4	0.4	-0.1279	0.0014	65	9
YCR016W	YGR067C	NR	NR	0.4	0.4	-0.1222	0.0231	65	11
ATG31	INO1	NR	NR	0.4	1	-0.1431	0.0371	26	6
FOB1	DOT1	NR	NR	1	0.7	-0.1699	0.00491	10	47
NUP42	CRN1	NR	NR	0.4	2	-0.2228	0.0135	44	23
RNH202	RNH1	NR	NR	1	0.4	-0.084	1.49e-08	56	6
UBX5	YMR209C	NR	NR	0	0.4	-0.0992	0.00565	15	22
RAD30	GTS1	NR	NR	0.7	0.4	-0.1771	1.87e-06	48	17
RAD30	MGA1	NR	NR	0.7	0	-0.2652	0.000439	48	10
RAD30	CRZ1	NR	NR	0.7	5	-0.0846	0.00876	48	28
SIP1	ACM1	NR	NR	0.4	0	-0.1139	0.000774	48	18
CHZ1	YMR209C	NR	NR	0.4	0.4	-0.1066	0.000785	28	22
TPA1	SRL4	NR	NR	0	0	-0.134	0.0369	39	12
TPA1	YPL247C	NR	NR	0	0.4	-0.1288	1.94e-79	39	10
YGR021W	DNL4	NR	NR	0	0	-0.162	0.0313	65	11
YGR021W	HST2	NR	NR	0	0	-0.1403	0.00929	65	8
PCI8	YLR224W	NR	NR	3	2	-0.1023	0.00257	39	12
SGN1	CUZ1	NR	NR	1	0	-0.129	0.0042	57	22
MND2	MSC3	NR	NR	0.7	0	-0.1461	1.08e-17	34	26
HUL4	PPQ1	NR	NR	0.7	3	-0.082	0.0101	34	31
YUH1	KAR5	NR	NR	0	2	-0.1012	0.000232	46	21
MAE1	RPL38	NR	NR	2	0.4	-0.0872	0.000103	55	24
MAE1	SPG5	NR	NR	2	0.7	-0.2294	1.34e-09	55	5
NDJ1	SFG1	NR	NR	0.4	1	-0.0818	3.06e-05	18	31
ULA1	PPQ1	NR	NR	1	3	-0.094	0.0203	47	31
YBL055C	SMM1	NR	NR	0	0.7	-0.1884	0.0109	24	39
ATG12	UBX5	NR	NR	0	0	-0.0821	0.0435	38	15
EFM2	ATG17	NR	NR	1	3	-0.104	8.27e-06	9	42

Gene1	Gene2	Responsiveness gene1	Responsiveness gene2	Sensitivity gene1	Sensitivity gene2	sgi score	*p* value	Negative sgi scores gene1	Negative sgi scores gene2
RTK1	SMM1	NR	NR	4	0.7	-0.2421	2.78e-10	11	39
GLT1	ULA1	NR	NR	0.7	1	-0.1512	0.0421	23	47
MGT1	TPA1	NR	NR	3	0	-0.2663	0.00105	33	39
MGT1	HUL4	NR	NR	3	0.7	-0.083	0.000724	33	34
TMA64	TMA20	NR	NR	0.7	4	-0.1831	4.12e-06	17	45
CWC15	SIP4	NR	NR	0	1	-0.1559	0.0128	21	58
STB3	NCA2	NR	NR	0	0	-0.0839	0.0268	13	23
CSN9	MAE1	NR	NR	0.7	2	-0.1201	0.0104	18	55
NKP1	ATG16	NR	NR	0.7	0.4	-0.0982	0.00258	18	40
NKP1	RUP1	NR	NR	0.7	1	-0.0821	0.0173	18	42
DOT1	TEX1	NR	NR	0.7	0	-0.2296	0.000835	47	33
CUE3	JEM1	NR	NR	0	3	-0.1064	4.75e-08	39	21
HUL5	ULA1	NR	NR	3	1	-0.0982	0.0169	19	47
FMP48	SGN1	NR	NR	0.4	1	-0.0871	2.29e-06	22	57
FMP48	YPL150W	NR	NR	0.4	0.4	-0.0809	0.0282	22	56
SLI1	BUB2	NR	NR	0.4	2	-0.099	3.37e-09	10	57
YGR250C	RUP1	NR	NR	0	1	-0.0865	1.24e-40	29	42
NDT80	SMM1	NR	NR	0	0.7	-0.3623	2.68e-09	7	39
DSE2	DAL82	NR	NR	2	4	-0.1021	0.0327	15	28
DSE2	RUP1	NR	NR	2	1	-0.0838	0.00192	15	42
YHR202W	BFA1	NR	NR	0	0.7	-0.1323	0.0382	10	50
MET28	ATG16	NR	NR	2	0.4	-0.0834	0.00432	31	40
YJL043W	MCM22	NR	NR	0	2	-0.0884	0.000504	13	54
IME2	YPR078C	NR	NR	0	0	-0.0809	2.16e-08	49	25
YMR1	UBP2	NR	NR	2	1	-0.0875	0.0216	9	48
PUT3	SMM1	NR	NR	0.4	0.7	-0.1959	0.05	11	39
REX2	PCL1	NR	NR	1	0	-0.0882	0.017	22	37
MCP2	MSS11	NR	NR	0	1	-0.1835	8.97e-05	10	27
STE23	YPL150W	NR	NR	0	0.4	-0.1177	8.09e-25	11	56
KAR5	PRM3	NR	NR	2	0.4	-0.3023	0.00108	21	17
YMR209C	RUP1	NR	NR	0.4	1	-0.0829	0.0422	22	42
DSK2	UBP2	NR	NR	2	1	-0.0826	6.36e-23	15	48
CRZ1	UBP2	NR	NR	5	1	-0.1242	0.049	28	48
UBP2	SET6	NR	NR	1	0	-0.0938	3.56e-72	48	32

S3 Table.
List of genes attributed to condition dependency.

Condition Dependancy Genes
FUN30 ,LTE1 ,SAW1 ,CYC3 ,YAL044W-A ,FLC2 ,ACS1 ,YAT1 ,HTB2 ,HTA2 ,YBL010C ,NCL1 ,YBL028C ,SAS3 ,KIP1 ,SEF1 ,ATG8 ,RKM3 ,EDS1 ,HMT1 ,YBR062C ,ECM2 ,NRG2 ,RDH54 ,VID24 ,IML3 ,RAD16 ,BMT2 ,TBS1 ,UBS1 ,PCH2 ,NTC20 ,DER1 ,COS111 ,NGR1 ,TDP1 ,YBR225W ,PBP2 ,SHG1 ,HSM3 ,PPS1 ,SAF1 ,BSD2 ,GLK1 ,KAR4 ,HCM1 ,SSK22 ,YCR087C-A ,STP4 ,LHP1 ,RAD59 ,BDF2 ,MRK1 ,THI3 ,NUR1 ,QRI7 ,PHO2 ,ATG20 ,RDI1 ,ATG9 ,MSH5 ,AIR2 ,SNF3 ,ASF2 ,ACK1 ,RRI1 ,PTP1 ,NTH1 ,MAF1 ,SOK1 ,GAL3 ,NSI1 ,MRH1 ,LYS14 ,YDR034W-B ,ARO3 ,VMS1 ,YOS9 ,UBC5 ,AIM7 ,RTR2 ,BMH2 ,RUB1 ,EKI1 ,CPR1 ,HST4 ,ADR1 ,RAD9 ,CRF1 ,BTT1 , MET32 ,CHL4 ,RMD5 ,RKM4 ,YAP6 ,DIN7 ,GCN2 ,ZIP1 ,HRQ1 ,SSD1 ,SSF2 ,PIB1 ,RAD34 ,GGA1 ,DXO1 ,SIZ1 ,CAD1 ,PPZ2 ,YHP1 ,PKH3 , CWC21 ,PAC11 ,PKH1 ,GRX2 ,YDR514C ,SLF1 ,GRH1 ,SPS1 ,EDC3 ,YEL043W ,HAT2 ,NOP16 ,YPT31 ,FIR1 ,EDC2 ,MEI4 ,ACA1 ,JHD1 , THO1 ,ARG5 ,6 ,UBP9 ,NUP157 ,MAM1 ,COM2 ,PMD1 ,YER137C ,MAG1 ,DDI1 ,UBP5 ,RAD4 ,BCK2 ,RAD24 ,ECM32 ,DMC1 ,YER184C , BLM10 ,HSP12 ,GAT1 ,CAF16 ,HAC1 ,RPO41 ,OTU1 ,SWP82 , GCN20 ,MET10 ,RRT5 ,PHO4 ,IRC5 ,CNN1 ,HXK1 ,RPN14 ,PDR1 ,HOP2 ,PNC1 ,TIF4632 ,YGL082W ,SCY1 ,LCL3 ,MAD1 ,MMS2 ,LIF1 ,PAN2 ,TOS8 ,COQ8 ,SNT2 ,PCL10 ,MPT5 ,ATG1 ,IME4 ,GCN1 ,SKI8 ,CLG1 ,VID30 ,SAP4 ,TAN1 ,KAP114 ,TAD1 ,PDE1 ,PAU11 ,CUL3 ,UGA1 ,KSS1 ,ROM1 ,TWF1 ,PIL1 ,NNF2 ,CLB6 ,SYF2 ,RSR1 ,RTS3 ,RBG2 ,PCT1 ,YCH1 ,MVB12 ,AMA1 ,NAS6 ,PHO81 ,YHB1 ,CPD1 ,RTT102 ,RNH70 ,SHU1 ,YAP3 ,SPO11 ,RIM4 ,NEM1 ,SPO13 ,ERC1 ,PCL5 ,PTC7 ,RTC3 ,YHR127W ,CRP1 ,SPO12 ,LIN1 ,REC104 ,ATG7 ,LNP1 ,EGD2 ,NVJ1 ,DOT5 ,SNL1 ,HOP1 ,SPO22 ,AIR1 ,XBP1 ,PFK26 ,SDP1 ,PRM5 ,RPI1 ,SIM1 ,RRT14 ,ASG1 ,CSM2 ,MPH1 ,MSL1 ,FLO11 ,MAD3 ,MAD2 ,YJL049W ,IKS1 ,BIT61 ,TAX4 ,SRS2 ,SAP185 ,FAR1 ,TPK1 ,HAL5 ,ATG27 ,SWE1 ,UBP12 ,REC107 ,YJR030C ,GEA1 ,ISY1 ,RAD7 ,TOR1 ,HAM1 ,YJR084W ,IME1 ,YJR096W ,HIR3 ,BYE1 ,HCS1 ,PAN3 ,TCD2 ,TUL1 ,PHD1 ,MSN4 ,YNK1 ,NUP100 ,SRX1 ,CUE2 ,YKL091C ,HSL1 ,YKL107W ,APN1 ,PRR1 ,SHE2 ,NNK1 ,CNB1 ,PTK1 ,YRA2 ,TOF2 ,HEL1 ,ALY1 ,DAL80 ,CAF4 ,GAP1 ,NAP1 ,DYN1 ,TIF1 ,OAF3 ,GPT2 ,SRL3 ,PTR2 ,SDC25 ,HIF1 ,YLL032C ,UBI4 ,VPS13 ,RNP1 ,YBT1 ,MHT1 ,CMS1 ,PML1 ,PSR2 ,MLH2 ,TRX1 ,GIS3 ,HRT3 ,REX3 ,YPS1 ,DCN1 ,CKI1 ,SLX4 ,PUT1 ,DPH6 ,RNH203 ,HRD3 ,RSA3 ,BNA5 ,IRC20 ,RED1 ,NEJ1 ,YLR271W ,PIG1 ,MEC3 ,YLR296W ,UBC12 ,IMH1 ,EST2 ,BUD6 ,REC102 ,NUP2 ,DCR2 ,NMD4 ,GRX8 ,PSY3 ,FBP1 ,CTF3 ,SWC7 ,REH1 ,CST9 ,SKI2 ,DUS3 ,YLR419W ,TUS1 ,MAG2 ,GMC2 ,RIF2 ,YLR455W ,RAD33 ,YOX1 ,SRC1 ,SML1 ,HUG1 ,DAK1 ,TCB3 ,HMG1 ,WAR1 ,YML082W ,UFO1 ,RAD10 ,CUE4 ,BUL2 ,NAB6 ,YML119W ,PHO84 ,AIM34 ,CLU1 ,CSI1 ,RSF1 ,UBX4 ,NAT4 ,YMR074C ,YPK2 ,ASI1 ,POM152 ,REC114 ,PSO2 ,CIN4 ,NDE1 ,YIM1 ,TPP1 ,MLH1 ,HOT1 ,ECM5 ,MMT1 ,CTL1 ,CLN1 ,RAD14 ,ESC1 ,TRI1 ,FAA4 ,RKR1 ,GFD1 ,ROY1 ,TPS3 ,CAT8 ,TDA1 ,YME2 ,FKS3 ,YMR315W ,DIA1 ,ASI3 ,PUB1 ,FAP1 ,HHF2 ,HHT2 ,SFB2 ,ARP5 ,FKH2 ,RNH201 ,DMA2 ,TEP1 ,CPT1 ,FYV6 ,MEP2 ,NOP13 ,SPS18 ,SSB2 ,MER1 ,MGS1 ,ATG4 ,SQS1 ,LAP3 ,ATG2 ,VPS75 ,GIS2 ,SIP3 ,TOF1 ,CAF120 ,CUS2 ,TRF5 ,SKP2 ,ATG3 ,NRM1 ,MPP6 ,HUB1 ,SOL1 ,YNR063W ,AIF1 ,TOP1 ,HRD1 ,NGL1 ,AIM39 ,PSH1 ,CRT10 ,REX4 ,ATG19 ,MHF1 ,DUF1 ,MDY2 ,SKM1 ,MSN1 ,RRI2 ,YGK3 ,YOL159C ,SGT2 ,AHC1 ,HST3 ,STI1 ,HMS1 ,ETT1 ,TMC1 ,YNG1 ,SKI7 ,PTC5 ,TMA46 ,RAS1 ,RGS2 ,CEX1 ,PDR5 ,LCB4 ,DCS2 ,GSP2 ,IES4 ,SLK19 ,MCA1 ,PTP2 ,WTM2 ,WTM1 ,MKK1 ,TUM1 ,HRK1 ,PAC1 ,MBF1 ,SPS4 ,MIP1 ,TEA1 ,UBC11 ,TYE7 ,REV1 ,MEK1 ,MSC6 ,RAD17 ,RDR1 ,PHR1 ,HAT1 ,ECM23 ,EGD1 ,MET31 ,SGF11 ,LEE1 ,CWC27 ,RGL1 ,ERI1 ,YPL109C ,DBP1 ,VPS30 ,HHO1 ,ATG5 ,PEP4 ,MLH3 ,ATG29 ,TRE1 ,UIP4 ,DDC1 ,AFT2 ,TPK2 ,YPL216W ,ENV7 ,CIN2 ,GAL4 ,VIK1 ,YPL260W ,REC8 ,YPR022C ,CSR2 ,NTO1 ,MCM16 ,SMK1 ,HOS1 ,LTP1 ,OPY2 ,ASR1 ,ISR1 ,DBF20 ,BSP1 ,YPR174C ,GDB1 ,ATG13 ,ARR1

A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions

Katrin Sameith[1], Saman Amini[1], Marian J.A. Groot Koerkamp[1], Dik van Leenen[1], Mariel Brok[1], Nathalie Brabers[1], Philip Lijnzaad[1], Sander R. van Hooff[1], Joris J. Benschop[1], Tineke L. Lenstra[1], Eva Apweiler[1], Sake van Wageningen[1], Berend Snel[2], Frank C.P. Holstege[1], Patrick Kemmeren[1]

[1]Molecular Cancer Research, University Medical Centre Utrecht, Universiteitsweg 100, Utrecht, the Netherlands

[2]Theoretical Biology and Bioinformatics, Department of Biology, Utrecht University, Padualaan 8, Utrecht, the Netherlands

Published in BMC Biol. 2015:13: 112. [PubMed: 26700642]

Abstract

Genetic interactions, or non-additive effects between genes, play a crucial role in many cellular processes and disease. Which mechanisms underlie these genetic interactions has hardly been characterized. Understanding the molecular basis of genetic interactions is crucial in deciphering pathway organization and understanding the relationship between genotype, phenotype and disease. To investigate the nature of genetic interactions between gene-specific transcription factors (GSTFs) in *Saccharomyces cerevisiae*, we systematically analyzed 72 GSTF pairs by gene expression profiling double and single deletion mutants. These pairs were selected through previously published growth-based genetic interactions as well as through similarity in DNA binding properties. The result is a high-resolution atlas of gene expression-based genetic interactions that provides systems-level insight into GSTF epistasis. The atlas confirms known genetic interactions and exposes new ones. Importantly, the data can be used to investigate mechanisms that underlie individual genetic interactions. Two molecular mechanisms are proposed, "Buffering by induced dependency" and "Alleviation by derepression". These mechanisms indicate how negative genetic interactions can occur between seemingly unrelated parallel pathways and how positive genetic interactions can indirectly expose parallel rather than same-pathway relationships. The focus on GSTFs is important for understanding the transcription regulatory network of yeast as it uncovers details behind many redundancy relationships, some of which are completely new. In addition, the study provides general insight into the complex nature of epistasis and proposes mechanistic models for genetic interactions, the majority of which do not fall into easily recognizable within- or between-pathway relationships.

Background

Predicting the phenotype of an individual organism based on its genotype is a major challenge. Such relationships can be complex, also because individual alleles can produce unexpected phenotypes in combination. The term epistasis has been used in distinct ways by classical and population geneticists [1,2]. Epistasis was first introduced by Bateson to refer to the masking of one mutation by another [1]. Later, the term epistasis was generalized by Fisher to any non-additive genetic interaction whereby the combination of two mutations yields a phenotype that is unexpected based on the effect of the respective individual mutations [2]. Throughout the article, epistasis refers to the general definition by Fisher and is used synonymously with genetic interaction to refer to any unanticipated combinatorial effect.

Cell growth has frequently been used to study genetic interactions on a large scale [3–12]. Genetic interactions are measured by the extent to which growth defects of double deletion mutants deviate from their expected value. A widely applied model assumes that the expected double mutant fitness should be equivalent to the product of the two single mutants [13]. Genetic interactions scored by the difference between observed and expected fitness can broadly be classified into two groups: negative and positive genetic interactions. A genetic interaction is negative if the fitness observed for a double mutant is worse than expected based on the fitness of the respective single mutants. Conversely, an interaction is positive if the observed fitness is better than expected. The largest study available to date investigated the existence of genetic interactions for 5.4 million gene pairs in the model eukaryote *Saccharomyces cerevisiae*, identifying approximately 170,000 interactions [12]. The extent and pervasiveness of genetic interactions is evident from this study. Understanding genetic interactions and the mechanisms underlying them are therefore of obvious importance for understanding genotype-phenotype relationships.

Several mechanisms have been proposed for genetic interactions (reviewed in [14,15]). The most intuitive explanation for a negative genetic interaction is redundancy, where two genes can substitute for one another by their ability to take over the exact same function [16]. Only simultaneous deletion of both genes has an effect on that function. A second explanation for a negative genetic interaction extends the concept of redundancy from genes to molecular pathways that function in parallel [14,15]. Positive genetic interactions have been suggested to occur more often between genes functioning in the same pathway or complex [4,17,18]. Deletion of one gene causes dysfunction of the entire pathway or complex such that deletion of a second gene in the same pathway or complex has no further consequence. Although the interpretation of negative and positive genetic interactions as relationships between and within pathways is appealing, it leaves large parts of the epistatic landscape completely unexplained. First, duplicated, redundant genes only explain a small subset of negative genetic interactions [19,20]. Second, many negative genetic interactions

are detected between seemingly unrelated rather than parallel pathways [5,6,12]. And third, the vast majority of positive genetic interactions occur between genes encoding proteins in different pathways or complexes rather than the same [5,12]. A theoretical model, "induced essentiality", has been proposed in the past [21] that provides an explanation how negative genetic interactions can occur between unrelated pathways. In this particular model, inactivation of one process results in an alternative condition that requires activation of another process. It does, however, leave non-essential genetic interactions unexplained and also lacks experimental data. Taken together, the molecular mechanisms underlying most genetic interactions are poorly characterized and further investigation is needed to provide a better mechanistic understanding of genetic interactions.

Transcription plays a major role in the relationship between genotype and phenotype. Depending on the state or fate of a cell, different genes are expressed at different levels. This is in part mediated by gene-specific transcription factors (GSTFs). Understanding the basis of genetic interactions between GSTFs is therefore likely to be important for understanding the transcription regulatory network. To study genetic interactions between GSTFs, genome-wide gene expression was monitored for 72 GSTF double deletion mutants and their corresponding single mutants. Gene expression has previously proven useful to study genetic interactions in more detail than is possible through growth defects [22–25]. The high-resolution expression atlas generated here, provides a systems-level overview of the epistatic landscape between GSTFs and reveals underlying mechanistic details. Besides revealing new redundancy relationships, this study also proposes two molecular mechanisms. These mechanisms, which we term "buffering by induced dependency" and "alleviation by derepression", provide explanations for negative and positive genetic interactions that were previously not understood.

Results

Growth-based genetic interaction scores

A genetic interaction between two genes can be studied by different phenotypes, of which cell growth is most frequently used. Here, growth is used in combination with genome-wide gene expression to investigate genetic interactions between GSTFs. *S. cerevisiae* has an estimated 215 GSTFs (Additional file 1; Methods). Selection of GSTF pairs likely having a genetic interaction is based on two distinct criteria. Pairs with a significant growth-based genetic interaction score as determined from a previous large-scale study [11] (47 pairs) and/or with similarity in their DNA binding properties [26–28] (50 pairs; Methods) were selected, resulting in a total of 90 pairs. Fitness of each deletion mutant is defined by its growth rate during exponential growth relative to WT. Replicate relative growth rates (RGRs) are highly reproducible and were averaged for subsequent analyses (Additional file 2A; single mutants: $R = 0.96$, $p < 2.23 \times 10^{-308}$; double mutants: $R = 0.98$, $p < 2.23 \times 10^{-308}$).

To score the genetic interaction $e_{growth,XY}$ between two GSTFs *X* and *Y*, fitness observed for the respective double mutant $W_{xDyΔ}$ is compared to the fitness that is expected based on both single mutants $W_{xΔ}*W_{yΔ}$ ($e_{growth,XY} = W_{xDyΔ} - W_{xΔ}*W_{yΔ}$) [13]. The resulting genetic interaction scores largely agree with the initial scores used for selecting GSTF pairs [27] (Additional file 2B; R = 0.63, $p = 2.64x10^{-5}$), taking into account differences in the growth procedures (liquid culture versus agar plates) and media used (synthetic complete (SC) versus yeast extract peptone dextrose (YPD)).

Gene expression profiles of GSTF single and double deletion mutants

To investigate mechanisms of genetic interactions between GSTFs, gene expression profiles were generated for 154 single and double GSTF deletion mutants. Wildtype (WT) cultures were grown and profiled alongside deletion mutants on each day to control for biological and technical variation. With the exception of single mutants that behave like WT, each mutant was grown and profiled four times from two independent cultures. Statistical modelling of the data results in an average expression profile for each mutant consisting of *P* values and fold changes (FC) for each gene relative to the average expression in a collection of WT cultures (Methods). Gene expression profiles for 72 GSTF pairs and their corresponding single deletion mutants successfully passed all quality controls and were used for further analysis (Additional file 1). These profiles provide the basis for a high-resolution atlas of genetic interactions between GSTFs.

A gene expression atlas of epistasis

A well appreciated mechanism of genetic interactions is complete redundancy [16], where two proteins can substitute for one another. For example, the two GSTFs Ecm22 and Upc2 redundantly activate sterol biosynthesis genes [29]. Deletion of either *ECM22* or *UPC2* does not affect growth and neither induces many expression changes ($RGR_{ecm22Δ} = 1$ and $RGR_{upc2Δ} = 1.03$; Fig 1A, left and middle panel). Loss of one GSTF can almost completely be compensated by the presence of the second GSTF. Simultaneous deletion of both GSTFs however results in slow growth and many gene expression changes ($RGR_{ecm22Δ\ upc2Δ} = 0.5$; Fig 1A, right panel). In addition to redundancy, other types of genetic interactions are exposed, for example between the two GSTFs Gzf3 and Gln3 involved in nitrogen regulation [30,31]. Whereas deletion of *GZF3* has no effect on growth and expression ($RGR_{gzf3Δ} = 1.02$; Fig 1B, left panel), deletion of *GLN3* results in a growth defect and many expression changes ($RGR_{gln3Δ} = 0.8$; Fig 1B, middle panel). Intriguingly, these effects are suppressed in the double mutant ($RGR_{gln3Δ\ gzf3Δ} = 0.99$; Fig 1B, right panel). These two examples show that both positive and negative genetic interactions can be studied using gene expression and that the observed expression changes may also be indicative of the genetic interaction type.

A potential advantage of investigating genetic interactions using gene expression is that changes can be compared at the level of individual genes. The effect of a genetic interaction between two GSTFs *X* and *Y* on a gene *i* can be measured as the deviation between the expression change observed in the double mutant M_{xDyDi} and the expected expression change, given each single mutant $M_{xDi}+M_{yDi}$ ($e_{txpn,xyi} = |M_{xDyDi} - (M_{xDi}+M_{yDi})|$). If gene *i* is unaffected, observed and expected expression changes will be highly similar such that $e_{txpn,XYi}$ is close to zero. In other words, the stronger gene *i* is affected, the more $e_{txpn,XYi}$ deviates from zero. The genetic interaction between the GSTFs *X* and *Y* is then scored by counting the total number of genes for which an unexpected expression change can be observed in the respective double mutant. A FC of 1.5 is chosen as a threshold ($e_{txpn,XY} = \sum_{all\ genes\ i} f(i)$, with $f(i)=1$, if $e_{txpn,XYi} > \log_2(1.5)$; 0, else). Based on this threshold, the genetic interaction between Ecm22 and Upc2 involves 801 genes (Fig 1C, left panel), the genetic interaction between Gln3 and Gzf3 involves 110 genes (Fig 1C, right panel). With a few exceptions, genetic interaction scores derived by growth or expression are highly related (Fig 1D; R = 0.75, $P = 5.9\text{x}10^{-8}$, using absolute values for growth-based scores). GSTF double mutants that grow unexpectedly slow or fast, often also show unexpected expression changes. However, as is exemplified below, expression-based genetic interaction scores provide a higher level of detail compared to growth-based genetic interaction scores and can be further used to investigate the underlying mechanisms of genetic interactions.

Different expression patterns define the epistatic landscape

The expression of individual genes can be affected by a genetic interaction between two GSTFs in several ways. In response to deletion of a single GSTF, expression of a gene can be decreased ($P \leq 0.01$, FC < 1), unchanged ($P > 0.01$) or increased ($P \leq 0.01$, FC > 1), relative to WT. When comparing expression changes in two GSTF single mutants and their corresponding double mutant, for example *emc22*Δ and *upc2*Δ, and accounting for quantitative effects as well, 20 different expression patterns are observed that can be divided into six different types (Fig 2A). The most intuitive expression pattern is buffering, where individual deletion of either GSTF does not affect expression, but simultaneous deletion results in many changes, including increased and/or decreased expression levels. Suppression is observed when expression changes elicited by one single mutant are suppressed by deletion of a second GSTF. Quantitative buffering is defined by expression changes induced by one single mutant that are amplified by deletion of a second GSTF. In contrast, quantitative suppression is observed if expression changes elicited by one GSTF single mutant are dampened by additional deletion of a second GSTF. Masking takes place if expression of a gene is increased (or decreased) in one single mutant, but this expression change is masked by decreased (or increased) expression in response to deletion of a second GSTF. Last, inversion is observed if expression of a gene is increased (or decreased) by individual deletion of either GSTF single mutant, but decreased (or increased) upon

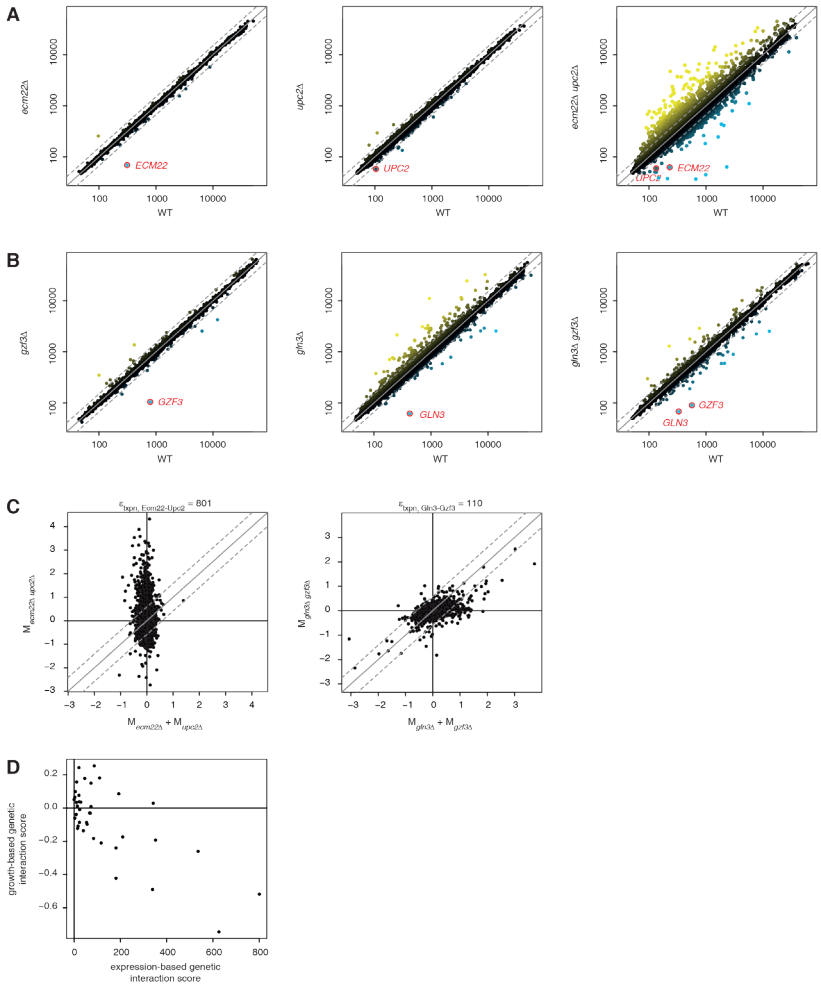


Fig 1. Genetic interactions measured by gene expression. (**A**) Genome-wide expression levels in the GSTF deletion mutants *emc22*Δ, *upc2*Δ and *ecm22*Δ *upc2*Δ (vertical, from left to right) versus reference WT (horizontal). Individual genes are depicted by solid circles, deleted genes are highlighted in red. Color range from yellow for increased expression levels relative to WT, black for unchanged expression, blue for decreased expression. A FC of 1.5 is depicted by a dashed grey line. (**B**) Genome-wide expression levels in the GSTF deletion mutants *gzf3*Δ, *gln3*Δ and *gln3*Δ *gzf3*Δ versus reference WT. Representation as in A. (**C**) Expected expression changes (M) in the double mutants *ecm22*Δ *upc2*Δ (horizontal, left panel) and *gln3*Δ *gzf3*Δ (right panel) versus observed expression changes (vertical). Individual genes are depicted by solid circles. Expected expression changes are calculated as the sum of expression changes in the respective single mutants. A FC of 1.5 is depicted by a dashed grey line. The number of genes outside the FC threshold is stated above each scatterplot. (**D**) Growth-based genetic interaction scores (vertical; liquid culture growth) are plotted versus expression-based genetic interaction scores (horizontal). Individual GSTF pairs are depicted by solid circles.

deletion of both GSTFs. Based on the different epistatic effects observed, an epistasis profile can be derived for all GSTF pairs.

With this approach, an atlas can be built consisting of the epistatic effects on expression levels by any two GSTFs under investigation (Fig 2B). The number of individual genes showing expression-based genetic interactions varies throughout the GSTF pairs. It is immediately noticeable that buffering is predominant in epistasis profiles of many GSTF pairs (dark-red color in Fig 2B), as expected based on the fact that genetic interactions are investigated between a single functional class of proteins (GSTFs). The abundance of buffering effects holds particularly for GSTF pairs with strong genetic interactions, affecting expression of many genes (Fig 2B, top). Classification of expression patterns in GSTF single and double mutants facilitates an abstract view on the epistatic landscape that can be further harnessed to reveal different types of genetic interactions.

Expression patterns expose the nature of genetic interactions

Hierarchical clustering was applied to group GSTF pairs with similar genetic interactions (Fig 3A). Here, the identity of individual genes was disregarded. Instead, GSTF pairs with similar patterns of epistatic effects are clustered together. GSTF pairs separate into several distinct groups, indicating that they have different types of genetic interactions. GSTF pairs clustered on the left are generally characterized by suppressive effects (Fig 3A, green branch), whereas GSTF pairs clustered on the right are generally characterized by buffering effects (Fig 3A, blue branch). From this analysis it is clear that buffering effects appear to be a good predictor for slow growth and vice versa (Fig 3B and 3C). If expression is changed only upon deletion of two GSTFs, the respective double mutant often grows slower than expected (negative interaction, e.g. Ecm22-Upc2). In turn, if expression changes induced by deletion of one GSTF are suppressed by additional deletion of the second GSTF, the respective double mutant sometimes grows faster than expected (positive genetic interaction, e.g. Gln3-Gzf3).

To investigate the difference between selecting GSTF pairs either on growth-based genetic interactions or similarity in DNA binding, hierarchical clustering was applied to both groups individually (Additional file 3). Several interesting differences emerge from this comparison. First, of all GSTF pairs selected based on growth (Additional file 3A) approximately 87% (33 out of 38 pairs) show an expression-based genetic interaction. Only 17 of the forty (43%) GSTF pairs selected based on similarity in DNA binding (Additional file 3B) have an expression-based genetic interaction. The fact that individual as well as simultaneous deletion of the remaining 23 pairs hardly affects expression may in part be due to condition specificity. Two of the three selection criteria used for similarity in DNA binding are based on *in vitro* (promoter affinity scores) and *in silico* (similarity of DNA binding domain) criteria, with no evidence for functional relevance under the condition investigated. For example, Msn2 and Msn4 are two redundant GSTFs involved in stress

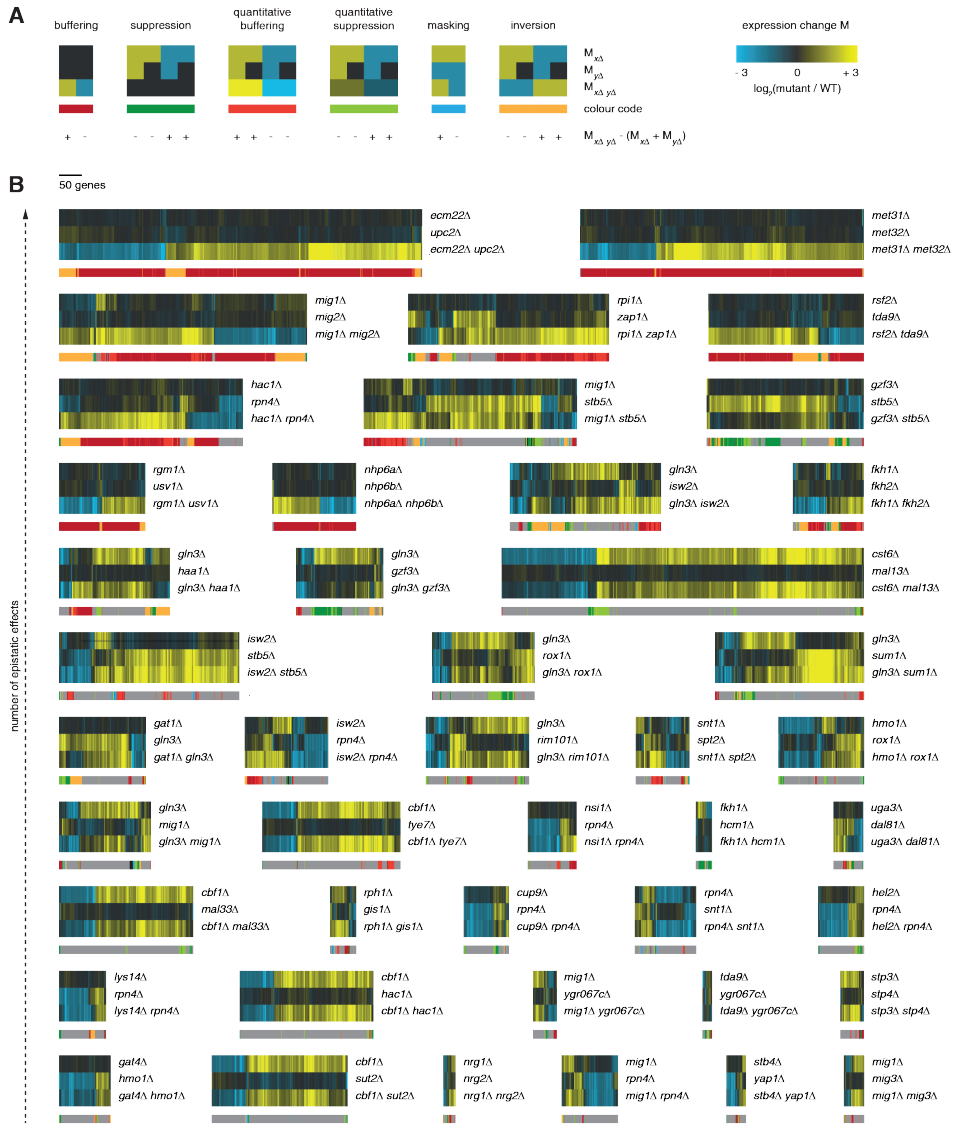


Fig 2. Gene expression atlas of GSTF pairs. (**A**) Cartoon of expression changes (horizontal) in GSTF single and double mutants (vertical). Color range from yellow for increased expression ($p \leq 0.01$; FC > 0), black for unchanged expression ($p > 0.01$), and blue for decreased expression ($p \leq 0.01$; FC < 0), as depicted in the top-right corner. Types of epistatic effects are color-coded, shown below the cartoon expression data. At the bottom, it is stated for each expression pattern, whether the observed expression change ($M_{x\Delta\,x\Delta}$) is more positive (+) or more negative (-) than expected ($M_{x\Delta} + M_{y\Delta}$). (**B**) Expression changes (horizontal) in GSTF single and double mutants (vertical). Color range as in A. Types of epistatic effects on individual genes are depicted below the expression changes (colors as in A; grey depicts non-epistatic expression changes). GSTF pairs are sorted according to the amount of epistatic effects (increasing from bottom to top). GSTF pairs with less than 10 epistatic expression changes are not shown.

response [32] and are therefore likely inactive in non-stress conditions as used in this study. Second, pairs selected based on similarity in DNA binding mostly show buffering effects (Additional file 3B, red branch) and only few pairs are characterized by suppressive relationships (Additional file 3B, green branch). Pairs selected based on growth on the other hand show much more suppressive effects (Additional file 3A, green branch). These pairs also show both positive genetic interactions (Additional file 3A; 4 pairs) as well as negative genetic interactions (8 pairs), whereas pairs selected based on similarity in DNA binding are mostly showing negative genetic interactions (Additional file 3B; 4 pairs) and little positive genetic interactions (1 pair). This all indicates that pairs selected based on similarity in DNA binding show a strong bias towards selecting redundant relationships, whereas pairs selected based on growth show a broader spectrum of genetic interaction types.

One group of six GSTF pairs is particularly distinctive (Fig 3A, red branch). First, GSTF pairs in this group are strongly epistatic, with buffering effects on many genes (Fig 3B, red bars). Second, six out of the seven GSTF pairs show very few expression changes that are not epistatic (Fig 3B). Third, most pairs contain a DNA binding domain of the same type (Fig 3A, diamonds). Together, these three characteristics are indicative of redundancy relationships. Indeed, redundancy relationships have previously been described for Ecm22-Upc2 [29], Met31-Met32 [33], Nhp6a-Nhp6b [34] and Mig1-Mig2 [35].

Rgm1 and Usv1 redundantly activate genes involved in respiratory ATP synthesis

Another interesting GSTF pair that clusters tightly with the redundant GSTF pairs Met31-Met32, Nhp6a-Nhp6b and Ecm22-Upc2 is Rgm1-Usv1 (Fig 3C, red branch). Besides having a binding site of the same type, Rgm1 and Usv1 have thus far not been reported to genetically interact. Individual deletion of either RGM1 or USV1 has no consequences for growth and almost no consequences for expression ($RGR_{rgm1\Delta} = 0.99$; $RGR_{usv1\Delta} = 1$; Fig 4A). Simultaneous deletion, on the other hand, results in slower growth and many expression changes ($RGR_{rgm1\Delta\ usv1\Delta} = 0.89$; Fig 4A). Rgm1 and Usv1 contain a DNA binding domain of the same type (Zinc finger pair [36]), and *in vitro*-derived promoter affinity scores of Rgm1 and Usv1 are highly correlated [27,28] (R = 0.94; Methods). Binding affinities of both GSTFs are significantly increased for genes that are unaffected in the single mutants *rgm1*Δ and usv1Δ, but show decreased expression in the double mutant *rgm1*Δ *usv1*Δ (Fig 4A, gene set 4). Rgm1 and Usv1 therefore likely activate transcription of these genes redundantly (Fig 4B). The top functional category enriched among Rgm1- and Usv1-activated genes is "ATP synthesis coupled electron transport" ($p = 1.69 \times 10^{-23}$), a respiration related process. Yeast cells preferentially produce energy through fermentation, but switch to respiration when fermentable carbon sources such as glucose are depleted [37,38]. Rgm1 and Usv1 are probably also active at basal levels when glucose is available,

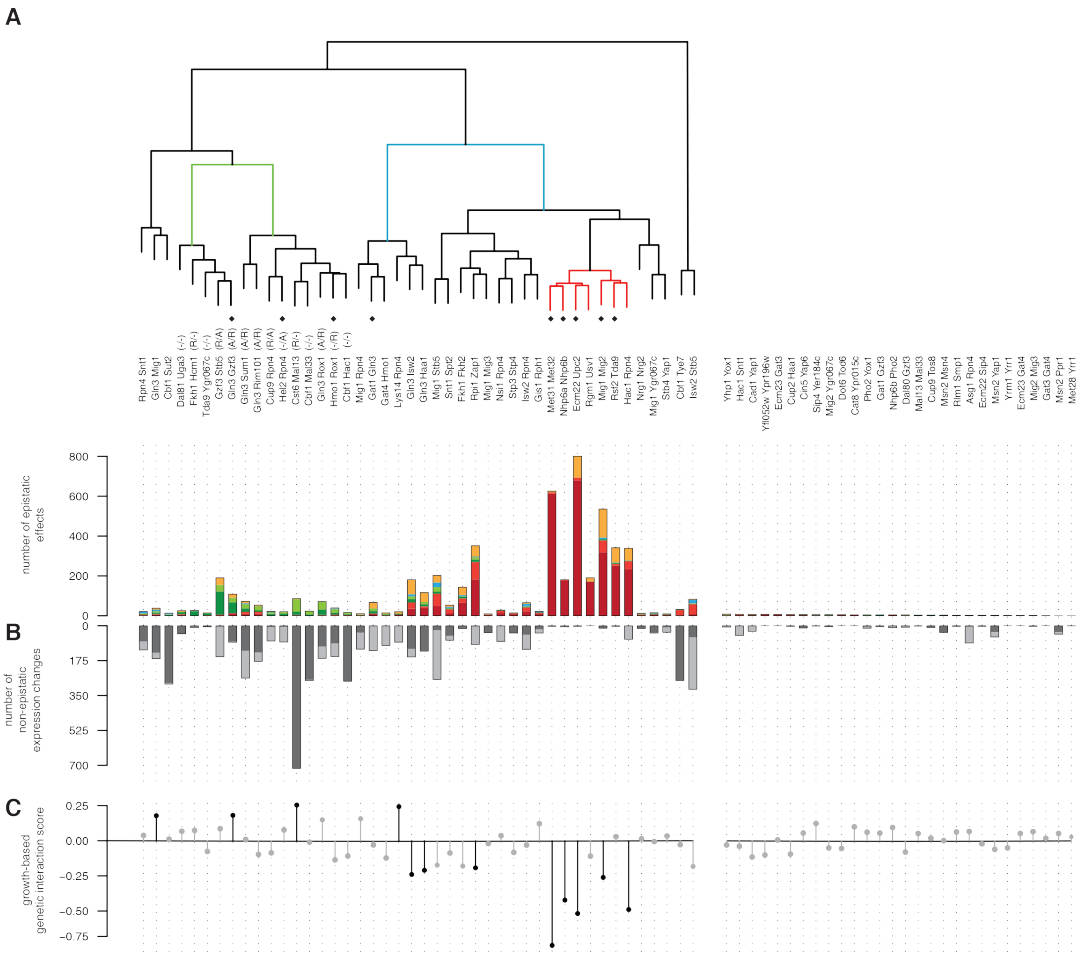


Fig 3. GSTFs have different genetic interactions. (**A**) Hierarchical clustering of GSTF pairs based on their epistatic effects. Average linkage clustering was applied to group GSTF pairs with similar epistatic effects. The identity of genes was disregarded. Instead, the number of total occurrences for each of the six different epistatic effects (Fig 2A) were used. Similarities between GSTF pairs were calculated based on cosine correlation. Colored branches depict example groups described in the text. GSTFs marked with a diamond have DNA binding domains of the same type. The number of epistatic effects underlying the clustering are shown as bar-plots below the dendrogram (colors as in Fig 2A). GSTF pairs with epistatic effects on less than ten genes are not included in the clustering, but are shown on the right. (**B**) Number of non-epistatic expression changes. Dark grey for the first named GSTF, light grey for the second. Counted are all genes with significantly changed expression ($p \leq 0.01$) above a FC of 1.5. Left-to-right ordering as in A. (**C**) Growth-based genetic interaction scores depicted by solid circles. Significant scores are visualized in black, grey otherwise. Vertical lines for visual purpose only. Left-to-right ordering as in A.

since expression changes are measured during exponential fermentative growth. The expression levels of Usv1 and Rgm1 as well as their target genes are increased during growth phases that require respiration (Fig 4C, lag phase, diauxic shift to stationary phase). Interestingly, the expression level of Rgm1 and Usv1 differs during shift from fermentation to respiration, indicating that Rgm1 and Usv1 may not be completely redundant under these conditions (Fig 4C, diauxic shift to stationary phase). This hypothesis is further supported by growth assays of *rgm1*Δ, usv1Δ and *rgm1*Δ usv1Δ on different carbon sources (Fig 4D). If Rgm1 and Usv1 are completely redundant under any given growth condition, deletion of one of the two factors is not expected to affect growth. Indeed, no growth defect is visible for any mutant during growth on the fermentable carbon sources glucose and raffinose. On the other hand, during growth on galactose, a less preferred fermentable carbon source, as well as the non-fermentable source glycerol, the single mutants *rgm1*Δ and *usv1*Δ grow markedly slower than WT and this growth defect is amplified in the double mutant *rgm1*Δ *usv1*Δ. Taken together, these results provide evidence that Rgm1 and Usv1 act redundantly, at least under exponential growth on glucose, to activate genes involved in respiratory ATP synthesis.

Buffering by induced dependency as a potential mechanism underlying negative genetic interactions

Negative genetic interactions have often been associated with redundant genes and this is reflected by GSTF pairs such as Ecm22-Upc2 [29], Met31-Met32 [33] and Mig1-Mig2 [35]. Buffering can also occur between genes in parallel pathways that can compensate for each other's loss [15]. A related mechanism termed "induced essentiality" has been proposed [21]. So far this has remained a theoretical model with no examples reported. Here, at least one GSTF pair potentially exhibits a closely related mechanism. In contrast to "induced essentiality", the potential mechanism observed here, does not lead to lethality, but instead results in a stress response. We therefore term this as "buffering by induced dependency". Hac1 and Rpn4 activate transcription of genes involved in two different pathways that are linked to the processing of inappropriately folded proteins, the unfolded protein response [39] (UPR, Hac1) and the endoplasmic reticulum associated degradation (ERAD) by the proteasome (Rpn4) [40]. Deletion of HAC1 neither affects growth or expression ($RGR_{hac1\Delta}$ = 1.05; Fig 5A), indicating that the UPR is inactive in WT (Fig 5D, WT panel). Deletion of RPN4, on the other hand, induces a mild growth defect and results in decreased expression of its proteasomal target genes ($RGR_{rpn4\Delta}$ = 0.9; Fig 5A, gene set 3; Fig 5B). It further results in increased expression of Hac1 target genes including KAR2 [26,41] (Fig 5C). This agrees with a previous observation that disruption of the ERAD pathway leads to activation of the UPR due to accumulation of misfolded and unfolded proteins in the endoplasmic reticulum [42] (Fig 5D, *rpn4*Δ panel).

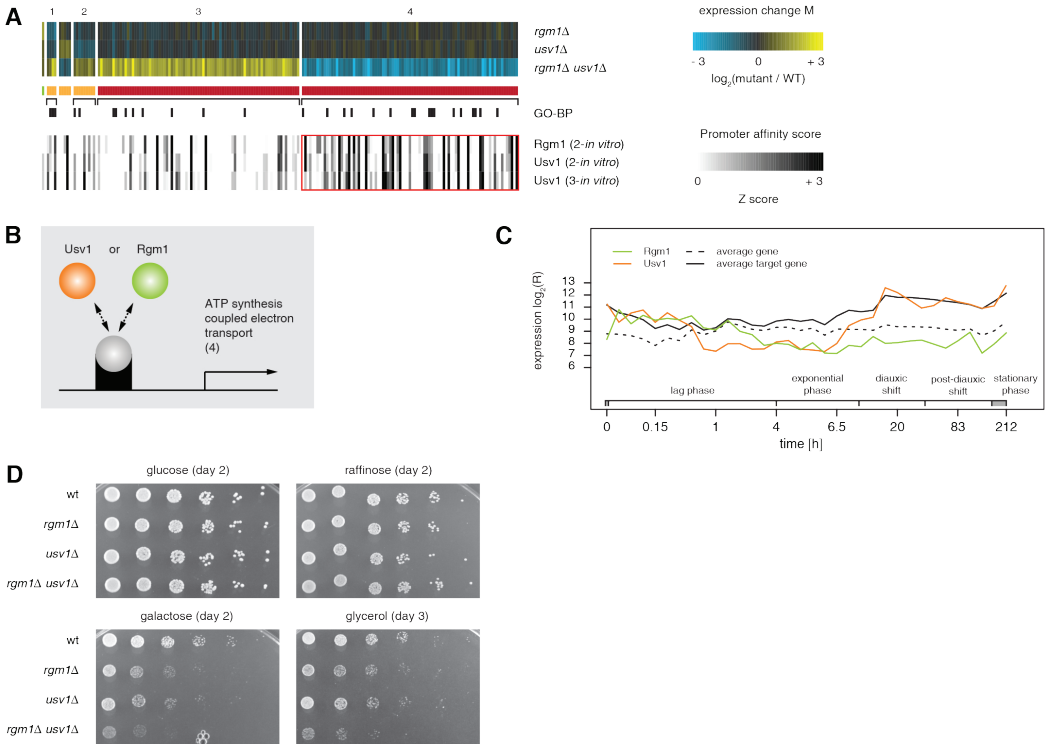


Fig 4. Rgm1 and Usv1 activate genes involved in respiratory ATP synthesis. (**A**) Genetic interaction between Rgm1 and Usv1. Co-expressed genes are separated into different sets based on increased, unchanged or decreased expression levels in the GSTF single and double mutants (top). Color scale as in Fig 2A. Types of epistatic effects on individual genes are depicted below the expression changes (colors as in Fig 2A). Gene sets are sorted by size. Annotations of the top GO-BP terms enriched in individual gene sets are presented below the expression data; gene set 1: monocarboxylic acid metabolic process ($p = 3.64 \times 10^{-4}$); gene set 2: agglutination involved in conjugation with cellular fusion ($p = 3.3 \times 10^{-4}$); gene set 3: iron ion transport ($p = 2.03 \times 10^{-8}$); gene set 4: ATP synthesis coupled electron transport ($p = 1.69 \times 10^{-23}$). Available DNA binding data showing significant overlap with at least one gene set are presented at the bottom. These are *in vitro* derived promotor affinity scores ("2-*in vitro*", "3-*in vitro*" as calculated from [28] and [27] respectively; Methods). Promotor affinity scores range from zero (white) to three (black) as depicted. Significant correspondence with expression data is depicted by red boxes. (**B**) Cartoon depicting the proposed genetic interaction between Rgm1 and Usv1. (**C**) Log-transformed expression levels of Rgm1 (green solid), Usv1 (orange solid), and average expression levels of their activated genes (black solid; gene set 4 in A) as well as all genes (black dashed) throughout different growth phases [58]. (**D**) Spot-assays showing growth on different carbon sources.

Expression changes elicited by the double mutant *hac1*Δ *rpn4*Δ indicate that disruption of both the ERAD and UPR induces severe stress (Fig 5D, *hac1*Δ *rpn4*Δ panel). Simultaneous repression of transcripts involved in translation (Fig 5A, gene set 8) and induction of transcripts involved in respiration (Fig 5A, gene set 9) are hallmarks of a stress response [43]. Moreover, selective and non-selective autophagy may be activated [44] (suggested by increased expression of the autophagy-related gene ATG8) coupled to vacuolar degradation [45] (suggested by increased expression of the peptidase PRC1 and proteinase PRB1). The activation of these processes may help the cells to survive, but are not sufficient to compensate the disruption of both ERAD and UPR as is reflected by a strong growth defect and many expression changes in the double mutant *hac1*Δ *rpn4*Δ ($RGR_{hac1\Delta\ rpn4\Delta}$ = 0.45). Additional follow-up experimentation will have to be performed in order to further substantiate the proposed model. Taken together, the genetic interaction between Hac1 and Rpn4 suggests how two pathways can buffer each other in a non-redundant, non-essential manner, whereby one pathway is only required because the other pathway has been inactivated: buffering by induced dependency (Fig 5E and 5F).

Alleviation by derepression as a potential mechanism underlying positive genetic interactions

Many GSTF pairs are characterized by suppressive effects in the double mutant (Fig 2B, 3). For two pairs, these expression changes are also reflected in a positive growth-based genetic interaction (Fig 3C). Positive genetic interactions have been suggested to occur more often between genes functioning in the same pathway or complex [4,17,18]. This leaves the majority of positive genetic interactions unexplained [15]. Similarly, same-pathway or same-complex relationships are not obvious or reported for the GSTF pairs with suppressive effects (Fig 3B, green branch). Instead, all GSTFs within this branch, for which systematic classification into either activator or repressor was available [46], form activator-repressor pairs. We therefore focused on one of these pairs to propose a likely mechanism. Gln3 and Gzf3 are two of four closely related GATA GSTFs involved in different aspects of nitrogen catabolite repression, a process that prevents the production of enzymes and permeases for the utilization of non-preferred nitrogen sources when a preferred nitrogen source is available [47]. Transcription activation through Gln3 and Gat1 is counteracted by repression through Gzf3 and Dal80 [47]. Deletion of GLN3 alone results in a growth defect and changed expression of many genes ($RGR_{gln3\Delta}$ = 0.8; Fig 6A), including decreased expression of the well-known target gene GLN1 [48] (Fig 6B). Gln1 is an enzyme that synthesizes glutamine from glutamate and ammonium [49]. Arginine can serve as a source for glutamine synthesis and glutamine is an input into amino acid and nucleotide biosynthesis [47]. Increased activity of enzymes involved in these processes may help to compensate for lower amounts of available glutamine. Indeed, as a secondary response to the limited activity of Gln1, expression of genes involved in arginine biosynthesis as well as de-novo nucleotide

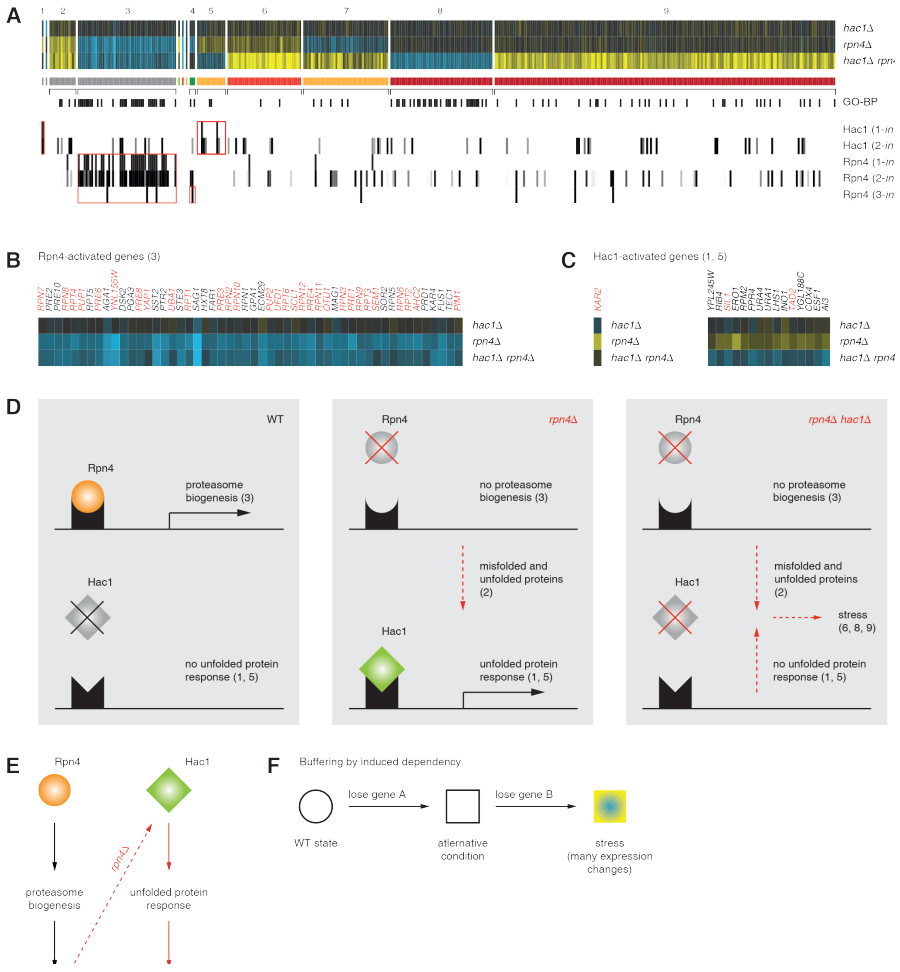


Fig 5. Buffering by induced dependency. (**A**) Genetic interaction between Hac1 and Rpn4. Representation as described for Fig 4A. Grey depicts non-epistatic expression changes. Available DNA binding data showing significant overlap with at least one gene set are presented below the expression data. These are *in vivo* binding targets ("1-*in vivo*" as in [26]) as well as *in vitro* derived promoter affinity scores ("2-*in vitro*","3-*in vitro*" as calculated from [28] and [27] respectively; Methods). Promotor affinity scores range from zero (white) to three (black) as depicted. Significant correspondence with expression data is depicted by red boxes. Top GO-BP terms are: gene set 2: protein refolding ($P = 2.85 \times 10^{-6}$); gene set 3: modification-dependent catabolic process ($P = 1.29 \times 10^{-32}$); gene set 4: transmembrane transport ($P = 0.04$); gene set 5: 'de novo' pyrimidine nucleobase biosynthetic process ($P = 9.79 \times 10^{-3}$); gene set 6: response to water deprivation ($P = 0.03$); gene set 7: glutamine family amino acid biosynthetic process ($P = 2.7 \times 10^{-4}$); gene set 8: ribosome biogenesis ($P = 1.55 \times 10^{-21}$); gene set 9: oxidation-reduction process ($P = 1.1 \times 10^{-6}$). (**B**) Rpn4-activated genes (zoom-up of gene set 3 in A). Red labels for annotated target genes of Rpn4 *in vivo* [26]. (**C**) Hac1-activated genes (zoom-up of gene sets 1 and 5 in A). Red labels for annotated target genes of Hac1 *in vivo* [26]. (**D**) Cartoon depicting the proposed genetic interaction between Hac1 and Rpn4. Consequences of individual deletions are indicated in red. (**E**) Summarized model to describe the proposed genetic interaction between Hac1 and Rpn4. (**F**) Generalized model for "buffering by induced dependency".

biosynthesis is increased (Fig 6A, gene sets 10,11; Fig 6D, *gln3*Δ panel). In contrast to *gln3*Δ, deletion of GZF3 has no effect on growth and results in very few expression changes ($RGR_{gzf3\Delta}$ = 1.02; Fig 6A). Deletion of GZF3 does however result in Increased expression for a second activator Gat1 [50], as well as two other GATA regulated genes MEP2 and DAL3 [47,51] (Fig 6C; Fig 6D, *gzf3*Δ panel). Deletion of GZF3 in the genetic background of *gln3*Δ alleviates the growth defect and suppresses many of the expression changes observed upon deletion of GLN3 alone ($RGR_{gln3\Delta\ gzf3\Delta}$ = 0.99; Fig 6A, gene sets 9-11 and gene sets 6, 12 respectively). A likely explanation is that derepression of Gat1, Mep2 and/or Dal3 in *gzf3*Δ can compensate for the loss of GLN3 (Fig 6D, *gln3*Δ *gzf3*Δ panel). Although additional evidence is needed to further validate the results, the example of Gln3 and Gzf3 suggests a molecular mechanism underlying positive genetic interactions. Within this mechanistic model, the effects of deleting one gene are alleviated by deleting a second gene, through derepression of a third gene (Fig 6E and 6F, "alleviation by derepression"). Since many of the other pairs with suppression effects consist of an activator and a repressor, alleviation by derepression may also hold for other pathways, providing a mechanistic explanation for how positive genetic interactions can nevertheless occur between different pathways.

Discussion

Positive genetic interactions between parallel, non-redundant pathways

We systematically investigated the epistatic landscape between yeast GSTFs by monitoring genome-wide gene expression changes of single and double deletion mutants. As a resource, the generated expression atlas can be harnessed in several ways. In addition to providing insights into the epistatic network between yeast GSTFs, the atlas exposes mechanistic details for negative and positive genetic interactions.

A number of molecular mechanisms have been proposed to underlie positive genetic interactions. For example, it has been suggested that positive genetic interactions could occur more often between genes encoding proteins of the same pathway or complex [4,17,18]. The reasoning behind this expectation is that deletion of any individual gene will cause dysfunction of the entire pathway or complex, and thus, deletion of a second gene will have no further consequence. Moreover, positive genetic interactions are suggested to occur in signaling pathways where two proteins have opposing influences on pathway activity. This (Batesonian) epistasis has been successfully applied to order genes in signaling pathways [22,23]. However, it has been shown that positive genetic interactions occur more frequently between different complexes or pathways rather than the same [5,12]. A molecular mechanism for such a positive genetic interaction is proposed in this study ("alleviation by derepression"), based on the epistatic effects observed for the two GSTFs Gln3 and Gzf3. The effects of inactivation of one gene/pathway may be buffered

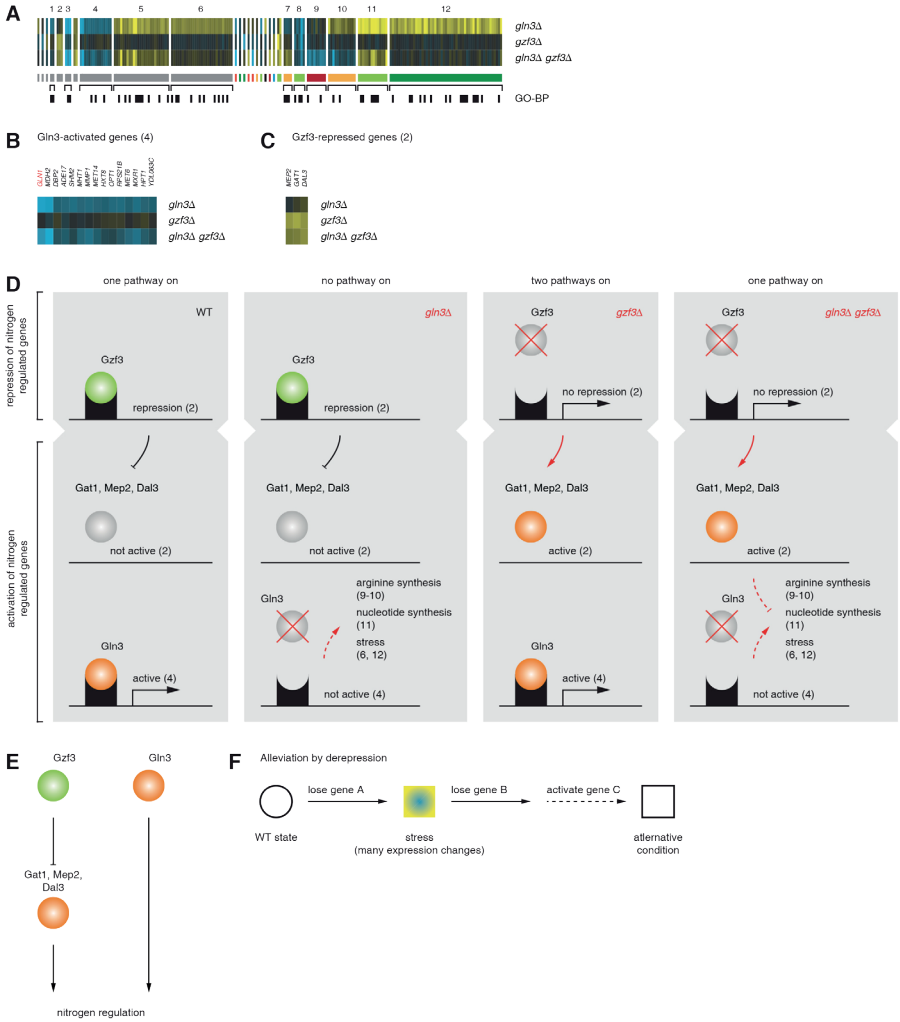


Fig 6. Alleviation by derepression. (**A**) Genetic interaction between Gln3 and Gzf3. Representation as described for Fig 5A. Top GO-BP terms are: gene set 1: generation of precursor metabolites and energy (P = 5.86x10^{-3}); 3: 'de novo' pyrimidine nucleobase biosynthetic process (P = 9.58x10^{-5}); 4: methionine biosynthetic process (P = 4.78x10^{-3}); 5: oxidation reduction process (P = 4.32x10^{-4}); 6: generation of precursor metabolites and energy (P = 1.09x10^{-6}); 7: oxidation reduction process (P = 0.01); 8: response to pheromone involved in conjugation with cellular fusion (P = 3.12x10^{-4}); 9: arginine biosynthetic process (P = 4.44x10^{-3}); 10: arginine biosynthetic process (P = 5.45x10^{-3}); 11: nucleotide biosynthetic process (P = 1.11x10^{-6}); 12: oxidation reduction process (P = 5.28x10^{-6}). (**B**) Gln3-activated genes (zoom-up of gene set 4 in A). Red label to highlight the known target gene GLN1 [48]. (**C**) Gzf3-repressed genes (zoom-up of gene set 2 in A). (**D**) Cartoon depicting the proposed relationship between Gln3 and Gzf3. Consequences of individual deletions are indicated in red. (**E**) Summarized model to describe the proposed genetic interaction between Gln3 and Gzf3. (**F**) Generalized model for "alleviation by derepression".

by the activation of another gene/pathway. However, that gene/pathway is repressed. Only upon deletion of the repressor, can buffering take place and the effects of the single mutant are suppressed. In addition to Gln3-Gzf3, many other GSTF pairs with suppressive effects consist of an activator and a repressor. The molecular mechanism proposed for Gln3-Gzf3 is therefore likely not an exception and may be applicable to these pairs as well.

Negative interactions between parallel, non-redundant pathways

The perhaps simplest cause of negative genetic interactions is genetic redundancy, where two genes can completely substitute for one another. The first systematic genetic interaction surveys, however, quickly revealed that genetic redundancy accounts for only a small subset of all negative genetic interactions [19,20]. Extending the concept of redundancy to molecular pathways does not explain either why many negative interactions are observed between seemingly unrelated pathways [5,12]. In this study, a molecular mechanism is proposed that may provide this missing link. The GSTF pair Hac1-Rpn4 alludes how a negative genetic interaction can be caused by the regulatory responses of a cell to deletion of a gene rather than by simple redundancy ("buffering by induced dependency"). The observed relationship is akin to a theoretical model ("induced essentiality") that has previously been suggested [21] to explain many synthetic lethal interactions. The model introduced here, is closely related. In both models, loss of one gene leads to a rearrangement into an alternative condition, where a second gene has become important for a particular cellular process. The key difference is that whereas in the "induced essentiality" model, the second gene has become essential, in our model this is not the case. It does, however, lead to (severe) stress when also removed as the cell is unable to appropriately cope with the resulting loss of function. The extent to which these mechanisms apply to other negative genetic interactions remains to be determined, but "buffering by induced dependency" proposes a mechanistic explanation for negative genetic interactions that cannot be explained by simple redundancy.

Genome-wide expression as a tool to understand the nature of genetic interactions

The development of synthetic genetic arrays [52] alongside computational methods for data analysis [17] facilitated the study of genetic interactions in a high-throughput manner. On the basis of growth as a fitness measure, the epistatic landscape, both static and dynamic, has been quantified in several large-scale studies [3–12]. In addition to growth, genome-wide expression has been used to study the nature of genetic interactions. In a study of kinase/phosphatase genetic interactions [25] this increased resolution revealed "mixed epistasis", whereby two yeast kinases and/or phosphatases exert different epistatic effects on different genes. Mixed epistasis is also observed between GSTFs, but to a lesser degree when compared to kinase/phosphatase pairs. This may be due to the fact that

expression changes occur as direct consequences of the inactivity of a GSTF. Inactivity of a kinase or phosphatase first has to be communicated through a signaling pathway, offering additional possibilities for interconnectivity. As large-scale fitness studies continue to reveal the full spectrum of the epistatic network, the additional use of high-resolution phenotypes such as expression is beneficial to increase our understanding of the underlying molecular mechanisms. In turn, increased mechanistic understanding facilitates a better interpretation of present and future large-scale genetic interaction studies and elucidate complex genotype-to-phenotype relationships.

Conclusions

Here, we have investigated the nature of genetic interactions between gene-specific transcription factors in baker's yeast. Systematic analysis of 72 GSTF pairs results in a high-resolution atlas of gene expression-based genetic interactions. The atlas exposes both novel genetic interactions as well as confirms known ones. More importantly, the data is used to investigate mechanisms underlying genetic interactions. Two mechanisms, "buffering by induced dependency" and "alleviation by derepression", are proposed. These mechanisms indicate how negative genetic interactions can occur between seemingly unrelated parallel pathways as well as how positive genetic interactions can indirectly expose parallel rather than same-pathway relationships. The study provides general insights into the complex nature of genetic interactions and proposes mechanistic models for genetic interactions that help us understand the full spectrum of genetic interactions and their contribution to cellular processes and pathway organization.

Methods

Selection of GSTF pairs

A list of 215 putative GSTFs was compiled based on 1) the presence of a DNA binding domain and 2) evidence for specific DNA binding (Additional file 1). For 12 putative GSTFs, the latter criterion was not fulfilled. These were included nevertheless, because they contain a domain that previously has been associated with specific DNA binding for another GSTF. Putative genetically interacting GSTF pairs were selected based on two distinct criteria. First, GSTF pairs were selected that exhibit significant genetic interactions based on growth [11]. Significance of a genetic interaction was estimated by z-transformation of the genetic interaction scores. A single genetic interaction score was compared to all other scores in the entire dataset (28 pairs), as well as to all other scores of one of the two GSTFs of interest (37 pairs; 47 pairs in total). Second, GSTF pairs were selected based on evidence for common DNA binding. These are GSTF pairs with similar DNA binding domains (19 pairs), GSTF pairs with common *in vivo* target genes [26] (ten pairs), and GSTF pairs with

similar promoter affinity profiles calculated from *in vitro* data [27,28] (30 pairs, 50 pairs in total for the second criterion). Altogether, 90 GSTF pairs were selected (Additional file 1).

Yeast strains

All strains are isogenic to S288c. Single mutants (Additional file 1) were taken from the Saccharomyces Genome Deletion library and obtained from Euroscarf (Frankfurt, Germany) or Open Biosystems (Huntsville, AL, USA). Double mutants were generated in duplicate by haploid transformation, random spore analysis, or tetrad dissection, in an identical genetic background as the single mutants (Additional file 1). All single mutants and most double mutants carry the mating type matα and are in the genetic background of BY4742. Few double mutants carry the mating type matA and are in the genetic background of BY4741. In six strains from the collection, gene expression profiles revealed different defects. Note that such defects may be common to all copies of the collection but could also have arisen due to our handling of these strains. All these strains were remade. Of the selected 90 double mutants, 18 failed quality control criteria even after remake. They were therefore excluded from further analyses. In total, 154 deletion mutants passed our quality control criteria (Additional file 1).

Gene expression profiling

Full details of all gene expression profiling procedures have been described before [46]. In summary, deletion mutants were grown in rich medium (SC, supplemented with 2% glucose) and harvested in early mid-log phase. WT cultures were grown alongside and processed in parallel. Dual-channel 70-mer oligonucleotide arrays were employed with WT RNA as common reference. All steps after RNA isolation were automated using robotic liquid handlers. These procedures were first optimized for accuracy (correct FC) and precision (reproducible result), using spiked-in RNA calibration [53]. After quality control, normalization and dye-bias correction [54], statistical analysis was performed for each mutant versus a collection of WT cultures [25,46,55]. Single mutants differing from WT as well as all double mutants were profiled another two times from an independently inoculated culture. The reported FC is then the average of four replicate mutant expression profiles versus the average of all WTs. Genes that show variable expression changes in the WT collection were excluded from further analyses (57 WT variable genes in total), as well as YDL196W.

Growth-based genetic interaction scores

Strains were grown in a Tecan Infinite F200 microplate reader, an automated system to incubate and measure optical densities in microplates. Growth measurements were taken every ten minutes, until cells were harvested at an OD_{600} of 0.6. Measurements that fall into exponential growth phase were selected to calculate growth rates. These are generally

between an OD_{600} of 0.3 and the last measurement at an OD_{600} of 0.6. Growth rates were calculated as the slope of the selected log-transformed measurements.

The fitness *W* of a deletion mutant was determined as the fraction between the average growth rate of WT and the growth rate of the mutant. Mutants growing slower than WT hence result in a fitness smaller than one. The genetic interaction $e_{growth,XY}$ between two GSTFs X and Y was scored by comparing the fitness of the respective double mutant $W_{xDy\Delta}$ with the fitness expected based on both single mutants $W_{x\Delta} * W_{y\Delta}$ ($e_{growth,XY}=W_{xDy\Delta}-W_{x\Delta} * W_{y\Delta}$) [13]. Significance of genetic interaction scores is derived by z-transformation. Background genetic interaction scores (100,000 in total) were calculated by randomly selecting triplets of WTs and applying the same calculation as applied to evaluate genetic interactions between GSTFs. Resulting *p* values were corrected for multiple testing using Benjamini-Hochberg, adjusted *P* values lower than 0.05 were considered significant. Fitness values of all single and double mutants, as well as calculated genetic interaction scores can be found in Additional file 4.

Expression-based genetic interaction scores

For a given GSTF pair, only genes with a statistically significant expression change in at least one of the two respective single mutants or in the double mutant, were considered ($P \leq 0.01$). The epistatic effect of two GSTFs X and Y on the expression of a gene *i* was measured as the deviation between the expression change observed in the double mutant M_{xDyDi} and the expression change expected given the single mutants $M_{xDi}+M_{yDi}$ ($e_{txpn,XYi}=|M_{xDyDi} - (M_{xDi}+M_{yDi})|$). The overall genetic interaction between the GSTFs X and Y was then scored by counting the total number of genes for which an unexpected expression change can be observed in the respective double mutant, whereby an FC of 1.5 was chosen as a threshold ($e_{txpn,XY}=\sum_{all\ genes\ i} f(i)$, with $f(i)=1$, if $e_{txpn,XYi}>\log_2(1.5)$; 0, else).

For each GSTF pair, genes with epistatic expression changes (all genes *i* where $e_{txpn,XYi}>\log_2(1.5)$) were further divided into different sets based on the observed expression patterns. Depending on whether expression levels are increased relative to WT ($P \leq 0.01$, FC > 0), unchanged ($P > 0.01$), or decreased ($P \leq 0.01$, FC < 0) in either single and/or in the double mutant, 20 common expression patterns were observed and divided into the six different types: buffering, suppression, quantitative buffering, quantitative suppression, masking and inversion. Less frequent epistatic patterns were categorized as miscellaneous and excluded from downstream analysis.

Functional enrichment analyses

For functional enrichment analyses, a hypergeometric testing procedure was performed using Gene Ontology (GO) Biological Process (BP) annotations [56] as obtained from the Saccharomyces Cerevisiae Database [57]. The background population was set to 6,359 (the number of genes annotated in GO) and *P* values were corrected for multiple testing using Bonferroni.

GSTF promoter affinity scores calculated from in vitro data

GSTF promoter affinity scores were calculated from [27,28]. First, signal intensities and enrichment scores for GSTFs that have been measured in vitro on multiple protein binding microarrays were averaged within each of the two datasets. The affinity by which a GSTF binds to the promoter of a potential target gene was then estimated by adding up signal intensities for each DNA 8-mer sequence, with an enrichment score greater than or equal to 0.45, within 600 base pairs upstream of the translation start site. Last, the resulting promoter affinity profile of one GSTF to all possible promoters was z-transformed to correct for experimental variation.

Overlap between expression changes and DNA binding

For each of the exemplified GSTF pairs, genes with expression levels that changed dependently (all genes *i* where $e_{txpn,XYi} > \log_2(1.5)$) or independently (all genes *i* where $e_{txpn,XYi} < \log_2(1.5)$) of the respective genetic interaction were first sorted into different sets depending on the observed expression patterns. Overlap between genes in a given set and genes whose promoter is known to be bound by the respective GSTFs *in vivo* [26] (selected parameters are $P = 0.005$, no conservation restriction) was evaluated by Fisher's exact test. Furthermore, a Mann-Whitney test was applied to test whether genes in a given set exhibit higher GSTF promoter affinity scores in vitro (calculated from [27,28] as described above) than all other genes. Resulting *P* values were corrected for multiple testing using Benjamini-Hochberg, adjusted *P* values below 0.05 were considered significant.

Availability of supporting data

The dataset supporting the results of this article is available in the ArrayExpress repository, E-MTAB-1385, http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1385/, as well as in the GEO repository, GSE42536, http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58144. The data is also available as flat-file or in TreeView format from http://www.holstegelab.nl/publications/GSTF_geneticinteractions.

Abbreviations

GSTF: gene-specific transcription factor
RGR: relative growth rate
WT: wildtype
FC: fold change
ERAD: endoplasmic reticulum associated degradation
SC: synthetic complete
YPD: yeast extract peptone dextrose

Competing interests

The authors declare that they have no competing interests

Author's contributions

Arranged funding: PK, FCPH. Carried out and analyzed experiments: MJAGK, DvL, MB, NB, JJB, TLL, EA, SvW. Bioinformatics analyses and interpretation: FCPH, PK, KS, SRH, PL, SA, BS. Figures: KS, SA. Contributed to manuscript text: FCPH, PK, KS, SA.

Acknowledgements

We would like to thank Cheuk Ko for technical assistance; Sebastiaan van Heesch, Thanasis Margaritis and Tony Miles for contributing data. Furthermore, we would like to thank Thanasis Margaritis, Eoghan O'Duibhir and Loes van de Pasch for helpful discussions throughout the project. Financial support provided by the Netherlands Bioinformatics Centre (NBIC) and the Netherlands Organization of Scientific Research (NWO), grants 016108607 (FH), 81702015 (FH), 05071057 (FH), 91106009 (FH), 021002035 (TLL), 863.07.007 (PK), 864.11.010 (PK), 70057407 (JJB).

References

1. Bateson W. FACTS LIMITING THE THEORY OF HEREDITY. Science. 1907;26: 649–660. doi:10.1126/science.26.672.649
2. Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 1918;52: 399–433.
3. Jasnos L, Korona R. Epistatic buffering of fitness loss in yeast double deletion strains. Nat Genet. 2007;39: 550–554. doi:10.1038/ng1986
4. St Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, et al. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet. 2007;39: 199–206. doi:10.1038/ng1948
5. Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat Genet. 2011;43: 656–662. doi:10.1038/ng.846
6. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. Science. 2004;303: 808–813. doi:10.1126/science.1091317
7. Davierwala AP, Haynes J, Li Z, Brost RL, Robinson MD, Yu L, et al. The synthetic genetic interaction spectrum of essential genes. Nat Genet. 2005;37: 1147–1152. doi:10.1038/ng1640
8. Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD. A DNA integrity network in the yeast Saccharomyces cerevisiae. Cell. 2006;124: 1069–1081. doi:10.1016/j.cell.2005.12.036
9. Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, et al. Functional organization of the S. cerevisiae phosphorylation network. Cell. 2009;136: 952–963. doi:10.1016/j.cell.2008.12.039
10. Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ, et al. Rewiring of genetic networks in response to DNA damage. Science. 2010;330: 1385–1389. doi:10.1126/science.1195618
11. Zheng J, Benschop JJ, Shales M, Kemmeren P, Greenblatt J, Cagney G, et al. Epistatic relationships reveal the functional organization of yeast transcription factors. Mol Syst Biol. 2010;6: 420. doi:10.1038/msb.2010.77
12. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The genetic landscape of a cell. Science. 2010;327: 425–431. doi:10.1126/science.1180823
13. Mani R, St.Onge RP, Hartman JL, Giaever G, Roth FP. Defining genetic interaction. Proc Natl Acad Sci. 2008;105: 3461–3466. doi:10.1073/pnas.0712255105
14. Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C. Systematic mapping of genetic interaction networks. Annu Rev Genet. 2009;43: 601–625. doi:10.1146/annurev.genet.39.073003.114751
15. Lehner B. Molecular mechanisms of epistasis within and between genes. Trends Genet TIG. 2011;27: 323–331. doi:10.1016/j.tig.2011.05.007
16. Hartman JL, Garvik B, Hartwell L. Principles for the Buffering of Genetic Variation. Science. 2001;291: 1001–1004. doi:10.1126/science.1056072

17. Collins SR, Schuldiner M, Krogan NJ, Weissman JS. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. Genome Biol. 2006;7: R63. doi:10.1186/gb-2006-7-7-r63
18. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature. 2007;446: 806–810. doi:10.1038/nature05649
19. Ihmels J, Collins SR, Schuldiner M, Krogan NJ, Weissman JS. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. Mol Syst Biol. 2007;3: 86. doi:10.1038/msb4100127
20. VanderSluis B, Bellay J, Musso G, Costanzo M, Papp B, Vizeacoumar FJ, et al. Genetic interactions reveal the evolutionary trajectories of duplicate genes. Mol Syst Biol. 2010;6: 429. doi:10.1038/msb.2010.82
21. Tischler J, Lehner B, Fraser AG. Evolutionary plasticity of genetic interaction networks. Nat Genet. 2008;40: 390–391. doi:10.1038/ng.114
22. van de Peppel J, Kettelarij N, van Bakel H, Kockelkorn TTJP, van Leenen D, Holstege FCP. Mediator Expression Profiling Epistasis Reveals a Signal Transduction Pathway with Antagonistic Submodules and Highly Specific Downstream Targets. Mol Cell. 2005;19: 511–522. doi:10.1016/j.molcel.2005.06.033
23. Van Driessche N, Demsar J, Booth EO, Hill P, Juvan P, Zupan B, et al. Epistasis analysis with global transcriptional phenotypes. Nat Genet. 2005;37: 471–477. doi:10.1038/ng1545
24. Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, et al. Structure and function of a transcriptional network activated by the MAPK Hog1. Nat Genet. 2008;40: 1300–1306. doi:10.1038/ng.235
25. van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, et al. Functional Overlap and Regulatory Links Shape Genetic Interactions between Signaling Pathways. Cell. 2010;143: 991–1004. doi:16/j.cell.2010.11.021
26. MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E. An improved map of conserved regulatory sites for Saccharomyces cerevisiae. BMC Bioinformatics. 2006;7: 113. doi:10.1186/1471-2105-7-113
27. Zhu C, Byers KJRP, McCord RP, Shi Z, Berger MF, Newburger DE, et al. High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. 2009;19: 556–566. doi:10.1101/gr.090233.108
28. Badis G, Chan ET, van Bakel H, Pena-Castillo L, Tillo D, Tsui K, et al. A library of yeast transcription factor motifs reveals a widespread function for Rsc3 in targeting nucleosome exclusion at promoters. Mol Cell. 2008;32: 878–887. doi:10.1016/j.molcel.2008.11.020
29. Vik A null, Rine J. Upc2p and Ecm22p, dual regulators of sterol biosynthesis in Saccharomyces cerevisiae. Mol Cell Biol. 2001;21: 6395–6405.

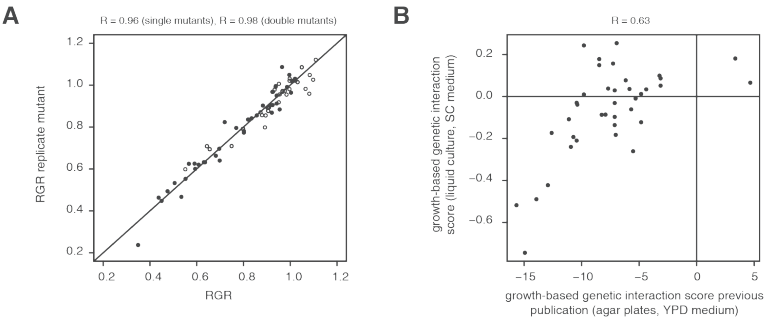
30. Soussi-Boudekou S, Vissers S, Urrestarazu A, Jauniaux JC, André B. Gzf3p, a fourth GATA factor involved in nitrogen-regulated transcription in Saccharomyces cerevisiae. Mol Microbiol. 1997;23: 1157–1168.
31. Blinder D, Magasanik B. Recognition of nitrogen-responsive upstream activation sequences of Saccharomyces cerevisiae by the product of the GLN3 gene. J Bacteriol. 1995;177: 4190–4193.
32. Martínez-Pastor MT, Marchler G, Schüller C, Marchler-Bauer A, Ruis H, Estruch F. The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). EMBO J. 1996;15: 2227–2235.
33. Blaiseau PL, Isnard AD, Surdin-Kerjan Y, Thomas D. Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. Mol Cell Biol. 1997;17: 3640–3648.
34. Costigan C, Kolodrubetz D, Snyder M. NHP6A and NHP6B, which encode HMG1-like proteins, are candidates for downstream components of the yeast SLT2 mitogen-activated protein kinase pathway. Mol Cell Biol. 1994;14: 2391–2403.
35. Westholm JO, Nordberg N, Murén E, Ameur A, Komorowski J, Ronne H. Combinatorial control of gene expression by the three yeast repressors Mig1, Mig2 and Mig3. BMC Genomics. 2008;9: 601. doi:10.1186/1471-2164-9-601
36. Böhm S, Frishman D, Mewes HW. Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. Nucleic Acids Res. 1997;25: 2464–2469.
37. Zaman S, Lippman SI, Zhao X, Broach JR. How Saccharomyces responds to nutrients. Annu Rev Genet. 2008;42: 27–81. doi:10.1146/annurev.genet.41.110306.130206
38. Broach JR. Nutritional control of growth and development in yeast. Genetics. 2012;192: 73–105. doi:10.1534/genetics.111.135731
39. Mori K, Kawahara T, Yoshida H, Yanagi H, Yura T. Signalling from endoplasmic reticulum to nucleus: transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway. Genes Cells Devoted Mol Cell Mech. 1996;1: 803–817.
40. Mannhaupt G, Schnall R, Karpov V, Vetter I, Feldmann H. Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. FEBS Lett. 1999;450: 27–34.
41. Cox JS, Walter P. A Novel Mechanism for Regulating Activity of a Transcription Factor That Controls the Unfolded Protein Response. Cell. 1996;87: 391–404. doi:10.1016/S0092-8674(00)81360-4
42. Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell. 2000;101: 249–258.
43. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, et al. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell. 2000;11: 4241–4257.
44. Yorimitsu T, Klionsky D. Autophagy: molecular machinery for self-eating. Cell Death Differ. 2005;12: 1542–1552. doi:10.1038/sj.cdd.4401765

45. van den Hazel HB, Kielland-Brandt MC, Winther JR. Review: Biosynthesis and function of yeast vacuolar proteases. Yeast. 1996;12: 1–16. doi:10.1002/(SICI)1097-0061(199601)12:1<1::AID-YEA902>3.0.CO;2-N
46. Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, et al. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. Cell. 2014;157: 740–752. doi:10.1016/j.cell.2014.02.054
47. Magasanik B, Kaiser CA. Nitrogen regulation in Saccharomyces cerevisiae. Gene. 2002;290: 1–18.
48. Mitchell AP, Magasanik B. Regulation of glutamine-repressible gene products by the GLN3 function in Saccharomyces cerevisiae. Mol Cell Biol. 1984;4: 2758–2766.
49. Mitchell AP. The GLN1 locus of Saccharomyces cerevisiae encodes glutamine synthetase. Genetics. 1985;111: 243–258.
50. Georis I, Feller A, Vierendeels F, Dubois E. The yeast GATA factor Gat1 occupies a central position in nitrogen catabolite repression-sensitive gene activation. Mol Cell Biol. 2009;29: 3803–3815. doi:10.1128/MCB.00399-09
51. Marini AM, Soussi-Boudekou S, Vissers S, Andre B. A family of ammonium transporters in Saccharomyces cerevisiae. Mol Cell Biol. 1997;17: 4282–4293.
52. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science. 2001;294: 2364–2368. doi:10.1126/science.1065810
53. van Bakel H, Holstege FCP. In control: systematic assessment of microarray performance. EMBO Rep. 2004;5: 964–969. doi:10.1038/sj.embor.7400253
54. Margaritis T, Lijnzaad P, van Leenen D, Bouwmeester D, Kemmeren P, van Hooff SR, et al. Adaptable gene-specific dye bias correction for two-channel DNA microarrays. Mol Syst Biol. 2009;5: 266. doi:10.1038/msb.2009.21
55. Lenstra TL, Benschop JJ, Kim T, Schulze JM, Brabers NACH, Margaritis T, et al. The Specificity and Topology of Chromatin Interaction Pathways in Yeast. Mol Cell. 2011;42: 536–549. doi:16/j.molcel.2011.03.026
56. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25: 25–29. doi:10.1038/75556
57. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. 2012;40: D700-705. doi:10.1093/nar/gkr1029
58. Radonjic M, Andrau J-C, Lijnzaad P, Kemmeren P, Kockelkorn TTJP, van Leenen D, et al. Genome-Wide Analyses Reveal RNA Polymerase II Located Upstream of Genes Poised for Rapid Response upon S. cerevisiae Stationary Phase Exit. Mol Cell. 2005;18: 171–183. doi:10.1016/j.molcel.2005.03.010

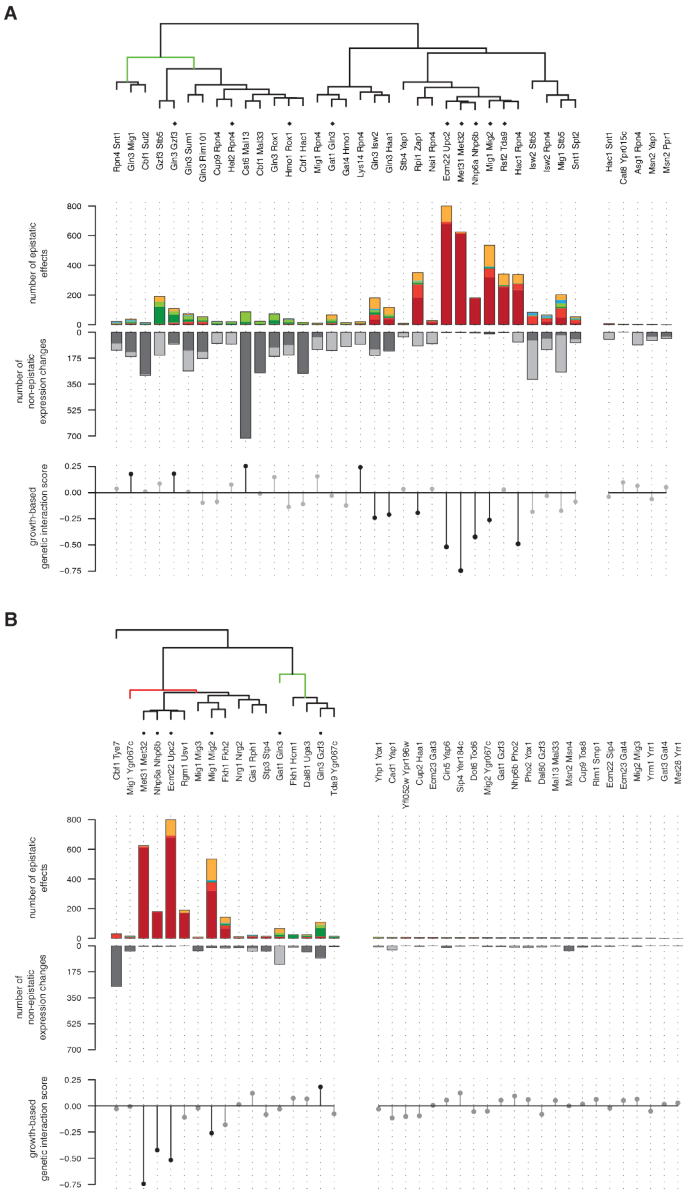
Additional file 1.
List of putative GSTFs and strains used.

systematic name	gene symbol	systematic name	gene symbol	systematic name	gene symbol	systematic name	gene symbol
Ykl112W	Abf1	Ygl237C	Hap2	Yfr034C	Pho4	Yml081W	Tda9
Ymr072W	Abf2	Ybl021C	Hap3	Yor363C	Pip2	Yor337W	Tea1
Yer045C	Aca1	Yor358W	Hap5	Ylr014C	Ppr1	Ybr083W	Tec1
Ylr131C	Ace2	Ycr065W	Hcm1	Ykl015W	Put3	Ydr362C	Tfc6
Ydr448W	Ada2	Ycl066W	Hmlalpha1	Ypr186C	Pzf1	Ybr240C	Thi2
Ydr216W	Adr1	Ycl067C	Hmlalpha2	Ynl216W	Rap1	Ybl054W	Tod6
Ygl071W	Aft1	Ydr174W	Hmo1	Yor380W	Rdr1	Ygl096W	Tos8
Ypl202C	Aft2	Ycr097W	Hmra1	Ycr106W	Rds1	Yor344C	Tye7
Ymr042W	Arg80	Ycr096C	Hmra2	Ypl133C	Rds2	Ydl170W	Uga3
Yml099C	Arg81	Yor032C	Hms1	Ybr049C	Reb1	Ydr207C	Ume6
Ydr421W	Aro80	Yjr147W	Hms2	Ybr267W	Rei1	Ydr213W	Upc2
Ypr199C	Arr1	Ygl073W	Hsf1	Ylr176C	Rfx1	Ydr520C	Urc2
Yil130W	Asg1	Ygl253W	Hxk2	Ymr182C	Rgm1	Ypl230W	Usv1
Ykl185W	Ash1	Ydr123C	Ino2	Ykl038W	Rgt1	Ydr049W	Vms1
Yor113W	Azf1	Yol108C	Ino4	Yhl027W	Rim101	Yml076C	War1
Ykr099W	Bas1	Yor304W	Isw2	Ypl089C	Rlm1	Yil101C	Xbp1
Ynl039W	Bdp1	Ykl032C	Ixr1	Ygr044C	Rme1	Yml007W	Yap1
Ydr423C	Cad1	Ypl054W	Lee1	Ypr065W	Rox1	Yhl009C	Yap3
Ymr280C	Cat8	Ylr451W	Leu3	Yer169W	Rph1	Yir018W	Yap5
Yjr060W	Cbf1	Ydr034C	Lys14	Yil119C	Rpi1	Ydr259C	Yap6
Ymr213W	Cef1	Ymr021C	Mac1	Ydl020C	Rpn4	Yol028C	Yap7
Ymr168C	Cep3	Ygr288W	Mal13	Ydr303C	Rsc3	Ydr026C	Ydr026C
Ylr098C	Cha4	Ybr297W	Mal33	Yhr056C	Rsc30	Ydr266C	Ydr266C
Yor028C	Cin5	Ycr040W	Matalpha1	Yfr037C	Rsc8	Yer130C	Yer130C
Ynl027W	Crz1	Ycr039C	Matalpha2	Yjr127C	Rsf2	Yer184C	Yer184C
Yil036W	Cst6	Yor298C-A	Mbf1	Yol067C	Rtg1	Yfl052W	Yfl052W
Ygl166W	Cup2	Ydl056W	Mbp1	Ybl103C	Rtg3	Ygr067C	Ygr067C
Ypl177C	Cup9	Ymr043W	Mcm1	Yor077W	Rts2	Ygr071C	Ygr071C
Ykr034W	Dal80	Yir017C	Met28	Ybl066C	Sef1	Ydr451C	Yhp1
Yir023W	Dal81	Ypl038W	Met31	Yor140W	Sfl1	Yjl206C	Yjl206C
Ynl314W	Dal82	Ydr253C	Met32	Ylr403W	Sfp1	Ykl222C	Ykl222C
Yml113W	Dat1	Ynl103W	Met4	Yjl089W	Sip4	Yll054C	Yll054C
Yer088C	Dot6	Ygr249W	Mga1	Ydr409W	Siz1	Ylr278C	Ylr278C
Ylr228C	Ecm22	Ygl035C	Mig1	Yhr206W	Skn7	Ynr063W	Ynr063W

systematic name	gene symbol	systematic name	gene symbol	systematic name	gene symbol	systematic name	gene symbol
Ypl021W	Ecm23	Ygl209W	Mig2	Ynl167C	Sko1	Yml027W	Yox1
Ybr239C	Ert1	Yer068W	Mot2	Ycr033W	Snt1	Ypr015C	Ypr015C
Ypr104C	Fhl1	Ymr070W	Mot3	Ygl131C	Snt2	Ypr022C	Ypr022C
Yil131C	Fkh1	Ymr037C	Msn2	Ymr016C	Sok2	Ypr196W	Ypr196W
Ynl068C	Fkh2	Ykl062W	Msn4	Yjl127C	Spt10	Yor172W	Yrm1
Yer109C	Flo8	Ymr164C	Mss11	Yer148W	Spt15	Yor162C	Yrr1
Ygl254W	Fzf1	Yhr124W	Ndt80	Yer161C	Spt2	Yjl056C	Zap1
Ypl248C	Gal4	Ydl002C	Nhp10	Ycr018C	Srd1	Ybr150C	Tbs1
Yfl021W	Gat1	Ypr052C	Nhp6A	Ydr169C	Stb3	Yfl031W	Hac1
Ymr136W	Gat2	Ybr089C-A	Nhp6B	Ymr019W	Stb4	Yol089C	Hal9
Ylr013W	Gat3	Ydr043C	Nrg1	Yhr178W	Stb5	Ylr256W	Hap1
Yir013C	Gat4	Ybr066C	Nrg2	Yhr084W	Ste12	Ylr266C	Pdr8
Yel009C	Gcn4	Yal051W	Oaf1	Ydr463W	Stp1	Ydr323C	Pep7
Ypl075W	Gcr1	Ykr064W	Oaf3	Yhr006W	Stp2	Ykl043W	Phd1
Ydr096W	Gis1	Ybr060C	Orc2	Ylr375W	Stp3	Yer111C	Swi4
Yer040W	Gln3	Yfl044C	Otu1	Ydl048C	Stp4	Ydr146C	Swi5
Yjl103C	Gsm1	Ydr081C	Pdc2	Ydr310C	Sum1	Ypl128C	Tbf1
Yjl110C	Gzf3	Ygl013C	Pdr1	Ygl162W	Sut1		
Ypr008W	Haa1	Ybl005W	Pdr3	Ypr009W	Sut2		



Additional file 2. Replicate growth-based genetic interactions. (**A**) Replicate fitness values derived from growth in liquid culture. Values are expressed as growth rates, relative to WT (RGR). GSTF single mutants are depicted as open circles, double mutants as solid circles. (**B**) Genetic interaction scores derived from growth on agar plates [11] (YPD medium, horizontal) versus genetic interaction scores derived from growth in liquid culture (SC medium, vertical, this study).



Additional file 3. Contribution of selection criteria to genetic interaction types. (**A**) Hierarchical clustering of GSTF pairs selected on growth-based genetic interaction scores, represented as in Fig 3. Clustering was performed on the epistatic effects. GSTF pairs marked with a solid circle where also selected based on similarity in DNA binding. Colored branches depict example groups described in the text. (**B**) Hierarchical clustering of GSTF pairs selected based on similarity in DNA binding, represented as in A. GSTF pairs marked with a solid circle also exhibit a genetic interaction as derived by growth on agar plates [11].

Additional file 4.
Genetic interaction scores between GSTF pairs.

GSTF pair	RGR (observed)	RGR (expected)	RGR (single mutants)	Genetic interaction score	adjusted *p*
Met31 / Met32	0,293	1,038	1.02 / 1.017	-0,745	8,96E-22
Ecm22 / Upc2	0,501	1,018	0.995 / 1.024	-0,518	1,12E-10
Hac1 / Rpn4	0,449	0,938	1.045 / 0.898	-0,489	1,05E-09
Nhp6a / Nhp6b	0,45	0,873	0.913 / 0.955	-0,422	2,18E-07
Mig1 / Mig2	0,631	0,892	0.902 / 0.989	-0,261	0,003659
Gln3 / Isw2	0,608	0,848	0.795 / 1.067	-0,24	0,008669467
Gln3 / Haa1	0,634	0,844	0.795 / 1.062	-0,21	0,028196387
Rpi1 / Zap1	0,614	0,807	0.955 / 0.845	-0,193	0,041482932
Isw2 / Stb5	0,596	0,778	1.067 / 0.729	-0,183	0,053242898
Fkh1 / Fkh2	0,889	1,07	0.986 / 1.085	-0,181	0,053242898
Mig1 / Stb5	0,484	0,658	0.902 / 0.729	-0,173	0,061346529
Hmo1 / Rox1	0,519	0,655	0.676 / 0.969	-0,136	0,163288516
Gat4 / Hmo1	0,552	0,675	0.998 / 0.676	-0,123	0,217384944
Cad1 / Yap1	1,009	1,124	1.099 / 1.023	-0,115	0,237120351
Rgm1 / Usv1	0,887	0,996	0.989 / 1.007	-0,109	0,251883659
Cbf1 / Hac1	0,597	0,705	0.675 / 1.045	-0,108	0,251883659
Yfl052w / Ypr196w	0,961	1,062	1.018 / 1.043	-0,101	0,284262605
Gln3 / Rim101	0,669	0,766	0.795 / 0.964	-0,097	0,28745307
Cup2 / Haa1	1,026	1,121	1.056 / 1.062	-0,095	0,28745307
Snt1 / Spt2	0,782	0,869	0.891 / 0.975	-0,087	0,299202556
Cup9 / Rpn4	0,915	1,001	1.115 / 0.898	-0,086	0,299202556
Stp3 / Stp4	0,971	1,054	0.904 / 1.166	-0,083	0,299202556
Dal80 / Gzf3	0,998	1,079	1.061 / 1.017	-0,081	0,299286698
Tda9 / Ygr067c	0,99	1,066	0.967 / 1.101	-0,076	0,299286698
Msn2 / Yap1	0,984	1,045	1.022 / 1.023	-0,061	0,368219797
Dot6 / Tod6	0,976	1,031	1.023 / 1.008	-0,055	0,381473185
Mig2 / Ygr067c	1,038	1,089	0.989 / 1.101	-0,051	0,381473185
Yrm1 / Yrr1	0,96	1,011	0.976 / 1.036	-0,051	0,381473185
Hac1 / Snt1	0,893	0,931	1.045 / 0.891	-0,039	0,433866156
Isw2 / Rpn4	0,927	0,958	1.067 / 0.898	-0,031	0,441089691
Gat1 / Gln3	0,772	0,802	1.009 / 0.795	-0,03	0,441089691
Yhp1 / Yox1	0,958	0,989	1.046 / 0.946	-0,03	0,441089691
Cbf1 / Tye7	0,601	0,629	0.675 / 0.932	-0,028	0,441089691
Ecm22 / Sip4	1	1,022	0.995 / 1.027	-0,022	0,463784686

GSTF pair	RGR (observed)	RGR (expected)	RGR (single mutants)	Genetic interaction score	adjusted *p*
Cbf1 / Mal33	0,672	0,682	0.675 / 1.011	-0,01	0,489638433
Mig1 / Ygr067c	0,988	0,993	0.902 / 1.101	-0,006	0,476218503
Msn2 / Msn4	0,991	0,99	1.022 / 0.969	0,001	0,459823418
Ecm23 / Gat3	0,947	0,942	0.937 / 1.005	0,006	0,441089691
Gln3 / Sum1	0,787	0,778	0.795 / 0.979	0,01	0,441089691
Cbf1 / Sut2	0,696	0,685	0.675 / 1.014	0,012	0,441089691
Nrg1 / Nrg2	0,993	0,978	0.992 / 0.986	0,014	0,43682998
Gat3 / Gat4	1,021	1,004	1.005 / 0.998	0,017	0,431701826
Cup9 / Tos8	1,038	1,02	1.115 / 0.914	0,018	0,431701826
Met28 / Yrr1	1,033	1,005	0.97 / 1.036	0,028	0,381473185
Rsf2 / Tda9	0,946	0,917	0.947 / 0.967	0,029	0,381473185
Stb4 / Yap1	1,025	0,991	0.969 / 1.023	0,034	0,379227196
Nsi1 / Rpn4	0,91	0,874	0.974 / 0.898	0,036	0,374365876
Rpn4 / Snt1	0,838	0,8	0.898 / 0.891	0,038	0,368219797
Ecm23 / Gat4	0,987	0,935	0.937 / 0.998	0,052	0,299286698
Mal13 / Mal33	0,996	0,944	0.934 / 1.011	0,052	0,299286698
Msn2 / Ppr1	0,946	0,894	1.022 / 0.875	0,052	0,299286698
Cin5 / Yap6	1,11	1,055	1.035 / 1.019	0,055	0,299286698
Gat1 / Gzf3	1,08	1,026	1.009 / 1.017	0,055	0,299286698
Pho2 / Yox1	0,971	0,91	0.962 / 0.946	0,061	0,299202556
Rlm1 / Smp1	1,057	0,995	0.934 / 1.066	0,062	0,299202556
Mig2 / Mig3	1,054	0,989	0.989 / 1	0,065	0,299202556
Asg1 / Rpn4	1,022	0,957	1.066 / 0.898	0,066	0,299202556
Dal81 / Uga3	0,987	0,92	0.965 / 0.953	0,067	0,299202556
Fkh1 / Hcm1	1,02	0,947	0.986 / 0.96	0,073	0,28745307
Hel2 / Rpn4	0,856	0,779	0.868 / 0.898	0,077	0,284262605
Gzf3 / Stb5	0,828	0,741	1.017 / 0.729	0,086	0,251883659
Nhp6b / Pho2	1,014	0,919	0.955 / 0.962	0,094	0,236609521
Cat8 / Ypr015c	1,013	0,914	0.98 / 0.932	0,099	0,217384944
Gis1 / Rph1	0,984	0,862	0.913 / 0.944	0,122	0,133288411
Sip4 / Yer184c	1,065	0,942	1.027 / 0.917	0,123	0,133288411
Gln3 / Rox1	0,919	0,77	0.795 / 0.969	0,149	0,061346529
Mig1 / Rpn4	0,967	0,81	0.902 / 0.898	0,157	0,053242898
Gln3 / Mig1	0,895	0,717	0.795 / 0.902	0,178	0,032725425
Gln3 / Gzf3	0,989	0,808	0.795 / 1.017	0,181	0,032324779
Lys14 / Rpn4	1,026	0,782	0.872 / 0.898	0,243	0,003264142
Cst6 / Mal13	0,791	0,537	0.575 / 0.934	0,254	0,002252662

The ability of transcription factors to fine-tune gene expression is a crucial component of the mechanism underlying inversion, a frequently observed genetic interaction pattern

Saman Amini[1,2,¶], Annika Jacobsen[3,¶], Olga Ivanova[3], Philip Lijnzaad[1], Jaap Heringa[3], Frank C. P. Holstege[1], K. Anton Feenstra[3], Patrick Kemmeren[1,2]

[1] Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands

[2] Center for Molecular Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands

[3] Centre for Integrative Bioinformatics (IBIVU), Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

¶ These authors contributed equally to this work.

Submitted to PLOS Genetics

Abstract

Genetic interactions, a phenomenon whereby combinations of mutations lead to unexpected effects, reflect how cellular processes are wired and play an important role in complex genetic diseases. Understanding the molecular basis of genetic interactions is crucial for deciphering pathway organization as well as understanding the relationship between genetic variation and disease. Several putative molecular mechanisms have been linked to different genetic interaction types. However, differences in genetic interaction patterns and their underlying mechanisms have not yet been compared systematically between different functional gene classes. Here, differences in the occurrence and types of genetic interactions are compared for two classes, gene-specific transcription factors (GSTFs) and signaling genes (kinases and phosphatases). Genome-wide gene expression data for 63 single and double deletion mutants in baker's yeast reveals that the two most common genetic interaction patterns are buffering and inversion. Buffering is typically associated with redundancy and is well understood. In inversion, genes show opposite behavior in the double mutant compared to the corresponding single mutants. The underlying mechanism is poorly understood. Although both classes show buffering and inversion patterns, the prevalence of inversion is much stronger in GSTFs. To decipher potential mechanisms, a modelling approach was employed, where genes are represented as nodes and relationships between genes as edges. This allowed over 9 million possible three and four node models to be exhaustively enumerated. The results show that a quantitative difference in interaction strength is a strict requirement for obtaining inversion. In addition, this difference is frequently accompanied with a second gene that shows buffering. Taken together, these results provide a mechanistic explanation for inversion whereby different efficacies between two redundant processes and a third compensatory process can be fine-tuned. Furthermore, the ability of transcription factors to fine-tune expression of their targets provides a logical explanation why inversion is more prevalent for GSTFs.

Introduction

Understanding the relationship between genotype and phenotype of an organism is a major challenge [1,2]. One of the difficulties for unravelling genotype-phenotype relationship has been genetic interactions, when combinations of mutations lead to phenotypic effects that are unexpected based on the phenotypes of the individual mutations [3–5]. Large-scale analyses of single and double deletion mutants have revealed that genetic interactions are pervasive in many model organisms [6–11]. Recently, efforts have been initiated to investigate genetic interactions in human cell lines too, using large-scale RNA interference and Crispr-Cas9 knock downs [12–15]. Our understanding of the molecular mechanisms that underlie genetic interactions lags behind our ability to detect genetic interactions. Understanding the molecular basis of genetic interactions and their interplay with cellular processes is important for unraveling how different processes are connected [16–18], to what degree genetic interactions shape pathway architecture [6], as well as for understanding the role genetic interactions play in human disease [5,19].

One of the phenotypes that is frequently used to investigate genetic interactions is cell growth [6,20–28]. Based on this phenotype, genetic interactions can be broadly subdivided in two types, negative genetic interactions where the double mutant is growing slower than expected given the growth rate of the single deletion mutants, and positive genetic interactions where the double mutant is growing faster than expected [3]. Negative genetic interactions have frequently been associated with a redundancy relationship between two functionally related genes [29]. The redundancy mechanisms by which two genes can compensate for each other's loss has been linked with close paralog genes or redundant pathways [30,31]. Positive genetic interactions have been associated with genes participating in the same protein complex or pathway [32]. There are however many exceptions to these rules and it also has become clear that there are many different potential mechanisms underlying these genetic interactions [3,18].

Another phenotype that has been less frequently used for investigating genetic interactions is gene expression [16,17,33–36]. Expression-based genetic interaction profiling provides detailed information at the molecular level which is beneficial for unraveling mechanisms of genetic interactions [16,17,33–36]. Unlike growth-based profiling, which gives a subdivision into either positive or negative interactions, expression-based genetic interaction profiling provides further subdivision into more specific genetic interaction patterns including buffering, quantitative buffering, suppression, quantitative suppression, masking and inversion. A more detailed sub classification that includes information on expression of downstream genes, can also contribute to understanding the mechanisms by which two genes interact [16,17,37].

To provide mechanistic insights into biological networks, Boolean modelling has been used successfully [38,39]. It has also been applied to unravel regulatory networks

underlying genetic interaction patterns between kinases and phosphatases [16]. Due to their intrinsically simple nature, such Boolean network models allow exhaustive enumeration of network topologies. The outcomes of these models can then be easily compared to the patterns observed in experimental data. Boolean operators however, are limited to on and off values and cannot easily accommodate quantitative measurements, which limits the types of genetic interaction patterns that can be investigated using this approach. Unravelling the regulatory network underlying genetic interaction patterns would potentially benefit from application of modelling approaches that allow some degree of quantitativeness to be introduced while still being computationally feasible to exhaustively explore all potential models. In this way, Petri nets may be considered an extension of Boolean modelling that provides more flexibility, in particular by choosing different network edge strengths, without the need to incorporate detailed prior quantitative knowledge [40–44]. Petri net modelling would therefore allow investigation of all possible genetic interaction patterns in an exhaustive and semi-quantitative manner.

It is evident that genetic interactions are widespread in *Saccharomyces cerevisiae* [6] as well as other organisms [7,8]. Nevertheless, extensive characterization of the molecular mechanisms underlying genetic interactions, as well as a comparison of the molecular mechanisms underlying genetic interactions between different functional classes have, as yet, not been performed. Here, two functional classes, gene specific transcription factors (GSTFs) and signaling related genes (kinases and phosphatases) have been compared with regard to negative genetic interaction patterns and the possible underlying molecular mechanisms. This revealed that the two most common genetic interaction patterns are buffering and inversion. The prevalence of inversion however, is much stronger in GSTFs. The underlying mechanism of Inversion, where genes show opposite behavior in the double mutant compared to the corresponding single mutants, is poorly understood. Exhaustive enumeration of network topologies using Petri net modelling reveals that the minimum requirement for observing inversion is having a quantitative difference in interaction strength (edge weight). In addition, this quantitative edge difference is frequently accompanied by an intermediate node, that displays a buffering pattern. The proposed model provides a mechanistic explanation when observing inversion, thereby further aiding a better understanding of genetic interactions. GSTFs, more so than kinases/phosphatases, can modulate or fine-tune the activation levels of their target genes, which suggests quantitative differences in regulating downstream target genes are important for the functioning of GSTFs. This is consistent with the fact that inversion occurs much more often between GSTFs than between signaling genes, as well as our observation that quantitative edge differences are required for inversion to occur and provides a logical explanation why inversion is more prevalent for transcription factors.

Results

A single dataset to compare mechanisms of genetic interactions between gene-specific transcription factors and kinases/phosphatases

To investigate potential differences in mechanisms of genetic interactions between different groups of genes with a different function, data from two previously published datasets were combined [16,17]. The first dataset includes genome-wide gene expression measurements of 154 single and double gene-specific transcription factor (GSTF) deletion mutants [17]. The second dataset contains genome-wide gene expression measurements of 54 single and double kinase/phosphatase (K/P) deletion mutants [16]. These studies applied different criteria to select for interacting pairs. To avoid potential biases, here an identical selection criteria was applied to both datasets to select for pairs that are considered to have a redundancy relationship. Selection was based on pairs having a significant growth-based negative genetic interaction score ($p < 0.05$, Methods). In addition, for a given double mutant, at least one of the corresponding single mutants has an expression profile similar to wildtype (WT) (eight or more transcripts changing significantly ($p < 0.05$, fold-change > 1.7)). These selection criteria yield a uniform dataset consisting of 11 GSTF double mutants and 15 kinase/phosphatase double mutants as well as their respective single mutants (63 single and double mutants in total; S1 Table).

Genetic interaction profiles indicate a large degree of buffering

Genetic interactions can be investigated in different ways. Here, both growth as well as genome-wide gene expression is used to compare genetic interactions between GSTFs and kinases/phosphatases, as described before [17]. In short, a growth-based genetic interaction score $\varepsilon_{growth,XY}$ between two genes X and Y is obtained by comparing the observed fitness for double mutant $W_{x\Delta y\Delta}$ to the fitness that is expected based on both single mutants $W_{x\Delta} \times W_{y\Delta}$ ($\varepsilon_{growth,XY} = W_{x\Delta y\Delta} - W_{x\Delta} \times W_{y\Delta}$) [45]. A gene expression-based genetic interaction score between two genes X and Y is calculated in two consecutive steps [17]. First, the effect of a genetic interaction between two genes X and Y on any downstream gene *i* is calculated as the deviation between the expression change observed in the double mutant $M_{i,x\Delta y\Delta}$ and the expected expression change based on the corresponding single mutants $M_{i,x\Delta} + M_{i,y\Delta}$ ($\varepsilon_{txpn\_i,XY} = |M_{i,x\Delta y\Delta} - (M_{i,x\Delta} + M_{i,y\Delta})|$). The overall genetic interaction score between gene X and Y is then obtained by counting the total number of genes for which $\varepsilon_{txpn\_i,XY}$ is greater than 1.5 [17]. Gene expression changes from single and double mutants were subsequently grouped into the six genetic interaction patterns, buffering, suppression, quantitative buffering, quantitative suppression, masking and inversion, as previously described (Fig 1A) [17]. When comparing the genetic interaction profiles between GSTFs (Fig 1B) and

kinases/phosphatases (Fig 1C), it is immediately clear that buffering is prevalent in most genetic interaction profiles regardless of the functional class.

Removal of a slow growth associated expression signature for improved identification of direct effects

Hierarchical clustering was applied to group pairs with similar genetic interaction patterns (S1 Fig), thereby disregarding the identity of individual downstream genes. From this clustering, it is clear that there is no distinct separation between pairs consisting of GSTFs and kinases/phosphatases. Instead, most pairs are characterized by large buffering effects, grouped together in a single large cluster (S1A Fig, red branch labeled as 1). This is not surprising, since all pairs are selected for having a significant growth-based negative genetic interaction score. This in turn is based on double mutants growing slower than expected based on the single mutants. Slow growing strains are known to display a common gene expression signature [46,47]. This slow growth gene expression signature is caused by a change in the distribution of cells over different cell cycle phases [48]. To facilitate investigating mechanisms of genetic interactions, such effects are better disregarded. As described previously [48], the dataset was transformed by removing the slow growth signature (Methods). Removing the slow growth signature and thereby reducing effects due to a cell cycle population shift, in general improves identification of direct target genes of GSTF pairs (S2 Fig) as shown before for individual GSTFs [48].

Discerning potential mechanisms with slow growth corrected genetic interaction profiles

Hierarchical clustering of the slow growth corrected genetic interaction profiles was then applied to unravel potential differences in observed genetic interactions patterns between GSTFs and K/P (Fig 2A-C). Three striking differences emerge when comparing this clustering with the clustering of the original, untransformed data (S1 Fig). First, pairs are grouped into four distinct clusters, whereas previously, most were grouped into a single large cluster. Second, a cluster of predominantly kinase/phosphatase pairs emerges (Fig 2A, green branch, labeled as 1). These contain mixtures of different genetic interaction patterns, corresponding with what has previously been termed ‘mixed epistasis’ [16]. Third, a smaller cluster dominated by buffering appears (Fig 2A, red branch, labeled as 2). This cluster also has strong growth-based negative genetic interaction scores (Fig 2C), which is known to be associated with redundancy.

The ‘buffering’ cluster, with its strong growth-based negative interactions, mostly consists of pairs with a high sequence identity (average 43.7%) compared to the others (average 21%). These include Nhp6a-Nhp6b, Met31-Met32, Ecm22-Upc2 and Ark1-Prk1, for all of which redundancy relationships have been described previously [49–52]. The high sequence identity here indicates a homology-based redundancy, in which both genes

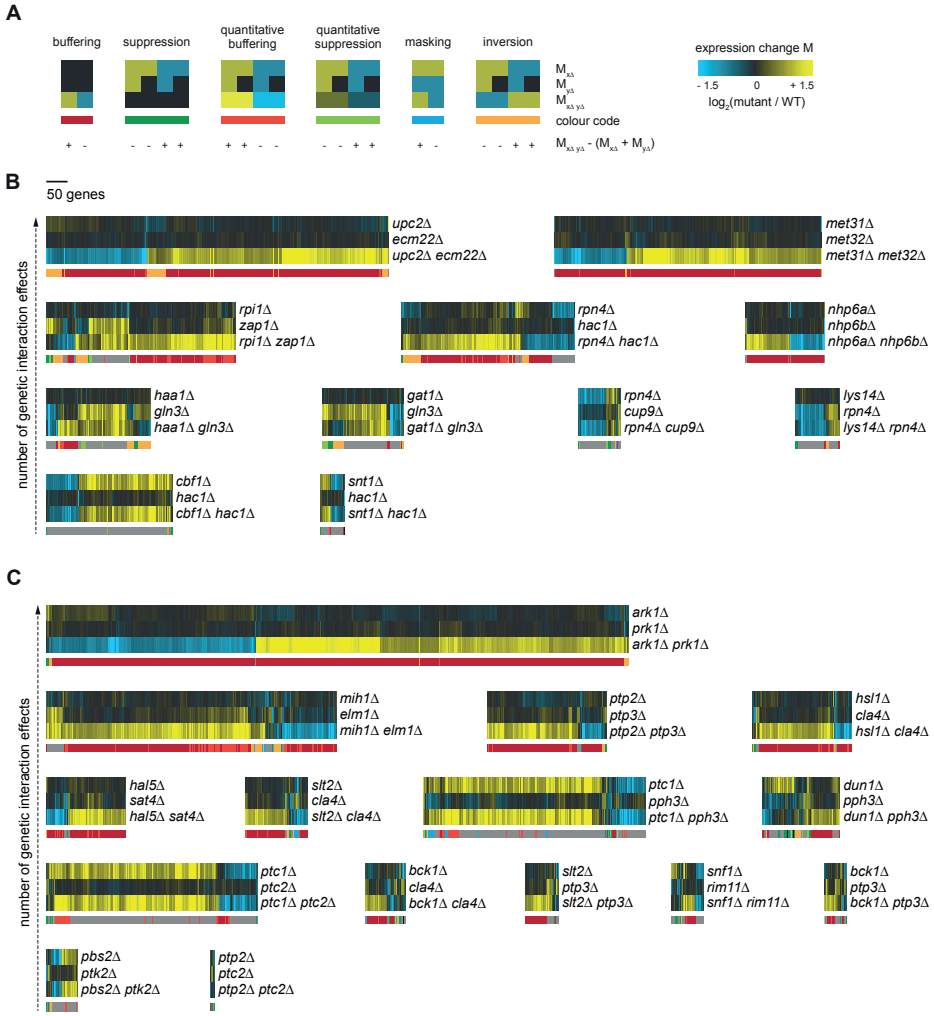


Fig 1. Genetic interaction profiles of GSTF and kinase/phosphatase pairs. (A) Cartoon depicting expression changes in single and double mutants with different genetic interaction patterns color coded underneath. At the bottom, the direction of expression differences between the observed expression change ($M_{x\Delta y\Delta}$) and expected ($M_{x\Delta}+M_{y\Delta}$) is stated. Color scale from yellow for an increase in expression levels compared to WT ($p \leq 0.01$, $\log_2(FC) > 0$), black for unchanged expression ($p > 0.01$) and blue for a decrease in expression levels compared to WT ($p \leq 0.01$, $\log_2(FC) < 0$). **(B)** Expression changes compared to WT (horizontal) in GSTF single and double mutants (vertical). Different colors underneath the gene expression profiles represent different genetic interaction patterns as indicated in A. Pairs are sorted based on the number of genetic interaction effects, increasing from bottom to top. **(C)** Expression changes compared to WT (horizontal) in kinase and phosphatase single and double mutants (vertical). Layout and ordering as in B.

can perform the same function [30,31,53,54]. The only exception here, is the kinase/phosphatase pair Elm1-Mih1. This pair may be explained through pathway-based redundancy where two parallel pathways can compensate for each other's function [55]. Elm1 is a serine/threonine kinase, and Mih1 a tyrosine phosphatase, which are both involved in cell cycle control (S3 Fig, left panel) [56,57]. Mih1 directly regulates the cyclin-dependent kinase Cdc28, a master regulator of the G2/M transition [57]. Elm1, on the other hand, indirectly regulates Cdc28 activity by promoting Swe1 degradation through the recruitment of Hsl1 [58,59]. The timing of entry into mitosis is controlled by balancing the opposing activities of Swe1 and Mih1 on Cdc28, and both Swe1 and Mih1 are key in the checkpoint mediated G2 arrest [60,61]. Deletion of Elm1 does not result in many gene expression changes (Fig 1C) which can be explained through compensatory activity of Mih1 (S3 Fig, middle panel). Downregulation of Mih1 activity has also been suggested before as an effective mechanism to counter stabilization of Swe1, as neither stabilization of Swe1 or elimination of Mih1 in itself is sufficient to promote G2 delay, but simultaneous stabilization of Swe1 and elimination of Mih1 does cause G2 arrest [59]. Simultaneous deletion of Elm1 and Mih1 leads to higher levels of inactive Cdc28 causing a G2 delay and stress (S3 Fig, right panel) [59] . All pairs within this cluster can therefore be associated with a redundancy mechanism.

Taken together, these results suggest that the clustering of the slow growth corrected genetic interaction profiles is able to discern potential differences in mechanisms. Even though most pairs in the four clusters (Fig 2A) show negative genetic interactions (Fig 2C), different mechanisms are likely underlying each individual cluster.

Inversion is associated with a specific subset of GSTFs

Within the slow growth corrected genetic interaction profiles another interesting cluster stands out: the orange branch where five out of six pairs involve GSTFs which predominantly show the inversion pattern (Fig 2A, orange branch, labeled as 3). This suggests that inversion may be strongly associated with GSTFs, whereas this does not seem to be the case for kinases and phosphatases. Indeed, when investigating the GSTF pairs within this cluster, it is clear that these display a much higher percentage of inversion compared to kinases and phosphatases (Fig 2D; p=0.00026). Moreover, GSTF pairs within this cluster also have a higher percentage of genes showing inversion compared to other GSTF pairs (Fig 2D; p= 0.0043). This indicates that not only is inversion more frequently associated with GSTFs in general, but one particular subset of GSTFs is also predominantly defined by inversion.

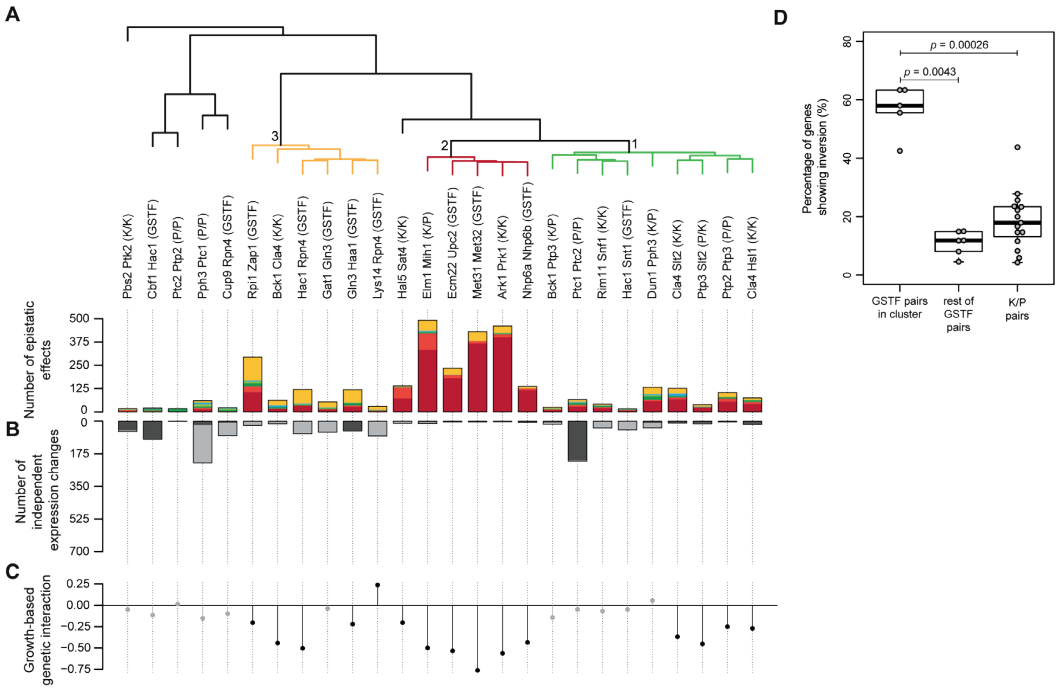


Fig 2. Hierarchical clustering of slow growth corrected genetic interaction profiles is better suited to discern underlying mechanisms. **(A)** Hierarchical clustering of all pairs according to their genetic interaction effects after slow growth correction. Average linkage clustering was applied to group pairs with similar genetic interaction patterns. The number of occurrences for each genetic interaction pattern (Fig 1A) was used and the identity of individual genes was disregarded. Similarity between pairs was calculated using cosine correlation. Branch depicted in red, label 2, indicates pairs that are dominated by buffering. Branch depicted in orange, label 3, indicates pairs dominated by inversion. Branch depicted in green, label 1, indicates pairs explained by mixed epistasis. The number of genetic interaction effects underlying the clustering are shown as bar plots below the dendrogram (colors as in Fig 1A). **(B)** Number of genes showing no genetic interaction pattern but significantly changing in one of the mutants compared to WT ($p \leq 0.01$, FC > 1.5). Dark gray for the first named gene, light gray for the second named gene. **(C)** Growth-based genetic interaction scores depicted by solid circles. Significant genetic interaction scores are shown in black, gray otherwise. Ordering of pairs is the same as in A and B. **(D)** Boxplot highlighting the difference between the percentage of genes showing inversion for GSTF pairs within the orange branch (Fig 2A), GSTF pairs outside this cluster and K/P pairs. *P*-values are based on a two-sided Mann-Whitney test.

An exhaustive modelling approach to explore potential mechanisms underlying inversion

Unlike buffering, where redundancy is a likely mechanistic explanation, the underlying mechanism of inversion is still unknown [17]. To investigate potential mechanisms of inversion, an exhaustive exploration was initiated. Previously, Boolean modelling has been applied to exhaustively explore all mechanisms underlying two genetic interaction patterns for the Fus3-Kss1 kinase phosphatase pair [16]. However, to explore all potential mechanisms underlying inversion, a Boolean approach is not sufficient as more subtle, quantitative effects, may be needed to obtain inversion. At the same time, it is required that the modelling remains computationally feasible. For this purpose, a modelling approach based on Petri nets was devised to exhaustively evaluate all possible three and four node models (Fig 3, Methods). Interactions between nodes (edges) can be activating (positive) or inhibiting (negative). In order to incorporate quantitative differences, both strong and weak edges were used (Methods). Counting all possible combinations of different edges results in 152,587,890,625 possible edge weight matrices. To reduce the number of models, three conditions were imposed, as used previously [16]. In short, nodes contain no self-edges, the number of incoming edges on any node is limited to two and the model includes at least two edges from one of the regulators (R1, R2) to the downstream genes (G1, G2). Applying these requirements and filtering for mirror edge weight matrices results in 2,323,936 matrices. By including AND/OR logics the final number of models to be evaluated was 9,172,034 (Methods). Petri net simulations were then run and genetic interaction patterns determined for G1 and G2, analogous to what was done for the real data (Methods) (Fig 1A). Depending on the topology, Petri net models can be stochastic, in other words, they do not show the same behavior when simulated multiple times and therefore result in unstable models. Only 2.3% of the models were found to be unstable, i.e. showed inconsistent genetic interaction patterns for G1 and G2 across five times simulation runs. Thus, stochasticity hardly influences the observation of genetic interaction patterns in our simulations (Fig 3). Nevertheless, unstable models were excluded from further analysis. In total, 168,987 models (1.8%) show inversion in either G1, G2, or both downstream nodes.

A quantitative difference in interaction strength is a strict requirement when observing inversion

To investigate which potential regulatory patterns underlie the 168,987 models showing inversion, low complexity models with few edges were analyzed first. Two interesting observations can be made. First, although there are many high complexity models involving four nodes and many edges (up to eight), three nodes and three edges are sufficient to explain inversion (Fig 4A). Second, only two three-node models exist that show inversion (Fig 4A). These two models only differ in the strength of the inhibiting edge from R1 to R2. Both models involve inhibition of R2 through R1 and weak activation of G1 by R1 in

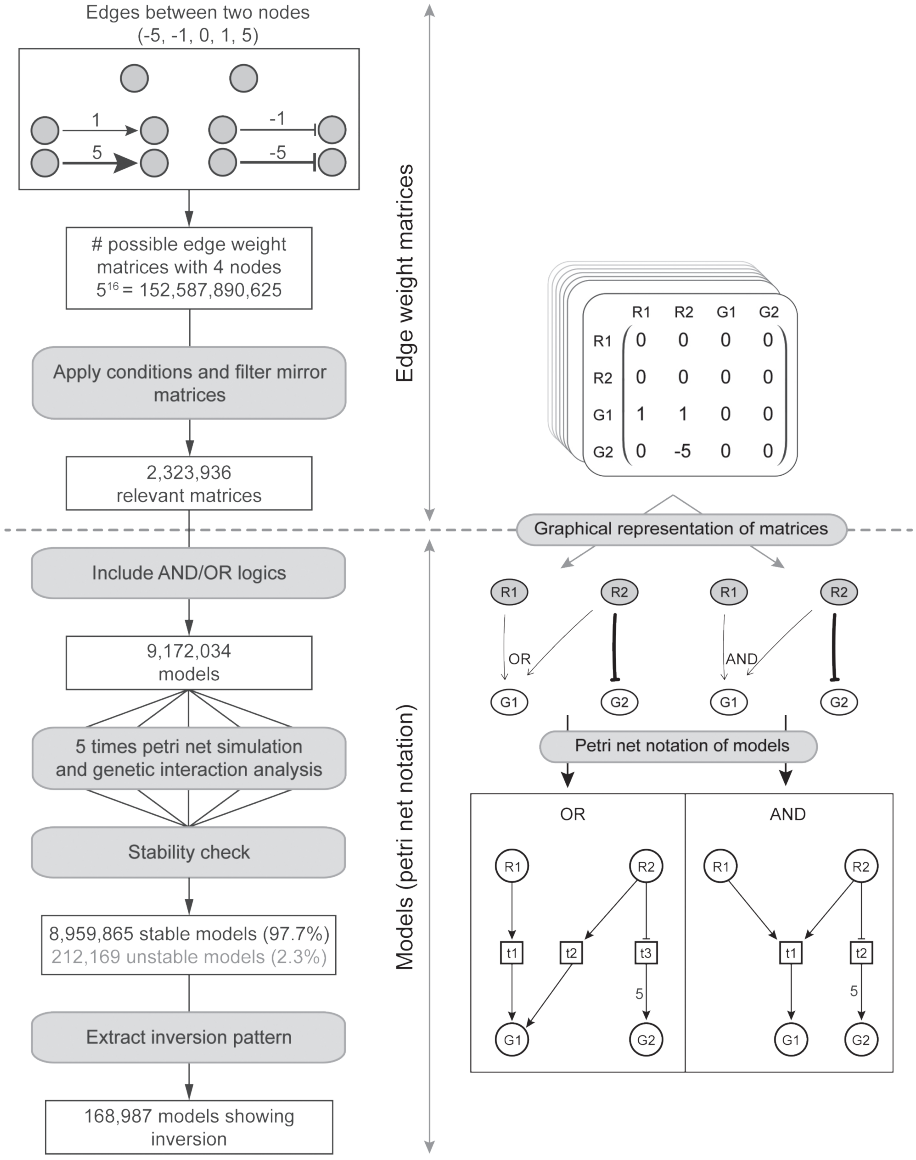


Fig 3. Schematic overview of Petri net simulation pipeline. Schematic overview of the pipeline implemented for performing Petri net simulations. Area above the dashed line indicates a series of steps where edge weight matrices are used. Area below the dashed line indicates steps where models or Petri net notation are used.

combination with a strong activation of G1 by R2, i.e. a quantitative edge difference between the incoming edges of G1. Deletion of R1 in these two models results in activation of R2, and therefore upregulation of G1 due to a strong activating edge. Deletion of R2 however, will not result in any changes compared to WT as it is normally inhibited by R1. Deletion of both R1 and R2 will lead to downregulation of G1 as the weak activating edge from R1 to G1 is lost. Taken together, the analysis of the low complexity models indicates that a quantitative difference in interaction strength is required to explain inversion.

To investigate whether this requirement also holds for higher complexity models, all models containing two to eight edges were further analyzed. Inversion models were grouped by the number of edges (complexity) and then analyzed for their relative frequency of having a quantitative edge difference (Fig 4B, top left panel, note that the number of possible models grows exponentially with the number of edges). Almost all of these models show a quantitative edge difference, with only a very small fraction (1.3% overall) of models not having a quantitative edge difference. The other genetic interaction patterns show different behavior, indicating that the relative ratio of quantitative versus non-quantitative edges is not an inherent network property. Based on both the low complexity models as well as the high complexity models showing inversion, it is evident that a quantitative difference in interaction strength of two genes or pathways acting on a downstream gene is required to explain inversion.

A quantitative difference in interaction strength is frequently accompanied by an intermediate buffering node

With the exception of the two models discussed above, all other inversion models consist of four nodes with two regulator nodes and two downstream effector nodes. To better understand the interplay between all four nodes, besides the node displaying inversion (G1), the second downstream gene (G2) was also analyzed for the occurrence of different genetic interaction patterns (Fig 5A). Most G2 nodes tend to have no genetic interaction pattern (27%). The most common genetic interaction patterns are buffering (23%) and quantitative buffering (18%). These both are very alike in their genetic interaction pattern (Fig 1A) and only show slight differences in their quantitative behavior. They may therefore be considered as part of the same superclass of “buffering”. The buffering node is frequently positioned upstream of the inversion node, and always downstream of R1/R2 (Fig 5B). The combination of inversion and buffering is also significantly overrepresented within inversion models when compared to all models (Table 1, $p < 0.005$). Taken together this shows that a quantitative difference in interaction strength of two genes or pathways acting on a downstream gene is frequently accompanied by an intermediate gene or pathway that displays buffering.

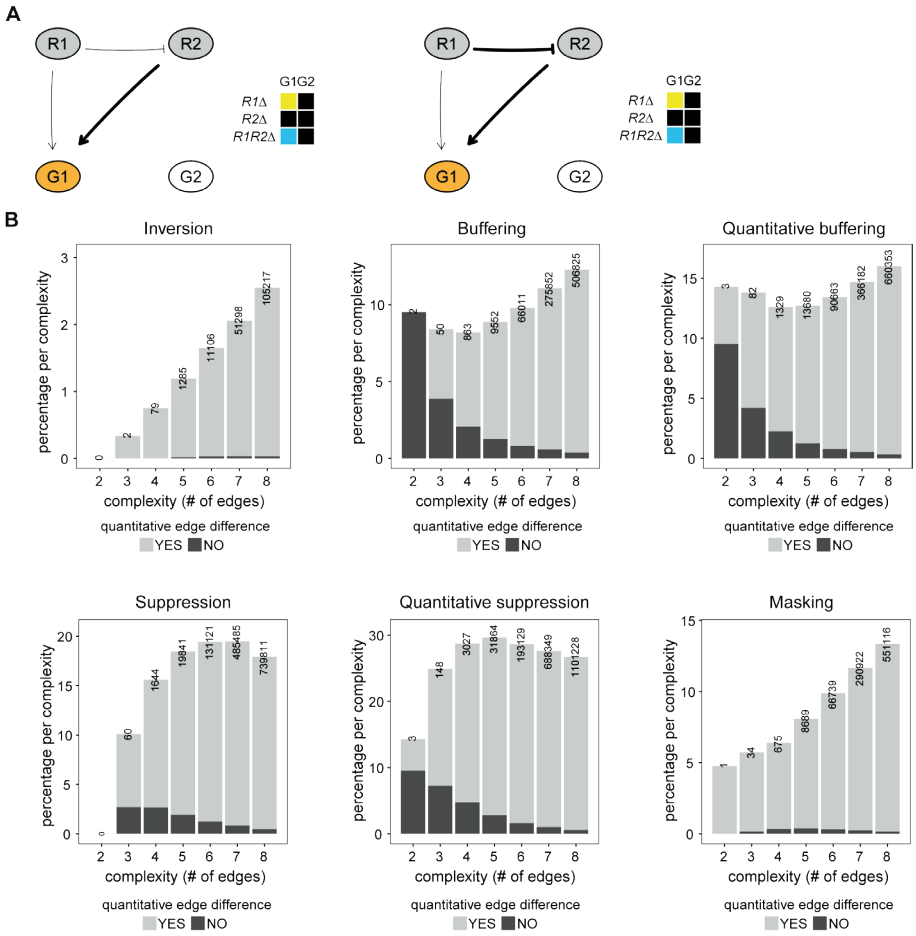


Fig 4. A quantitative edge difference is the minimum requirement for observing inversion. (A) Petri net simulation results for the only two models with three nodes that result in inversion (indicated in orange) for the G1 node. Heat maps indicate the $\log_2$(FC) of the number of tokens in simulated deletion mutants (single and double mutant) relative to the WT situation. Thicker lines indicate edges with a strong effect. **(B)** For each genetic interaction pattern (inversion, buffering, quantitative buffering, suppression, quantitative suppression and masking), the percentage of models showing that particular genetic interaction pattern is shown, split up per complexity (number of edges). The percentage per complexity is calculated as the number of models showing a particular genetic interaction pattern for a certain complexity, divided by the total number of models for that complexity. Bar plots are subdivided into two types of models, models that have quantitative differences between edge weights (bright gray) and models that have no quantitative differences between edge weights (dark gray). The number of models showing the particular genetic interaction pattern per complexity is shown on top of each bar plot.

Table 1.
Models with a quantitative edge difference and intermediate buffering node

		Quantitative edge difference	
		YES	NO
Buffering or quantitative buffering	YES	69,333 (41.03%) * 1,754,000 (23.66%) #	1,098 (0.65%) * 65,618 (0.89%) #
	NO	97,418 (57.65%) * 5,412,614 (73.01%) #	1,138 (0.67%) * 180,808 (2.44%) #

Inversion models are indicated with *. All models are indicated with #. The combination of a quantitative difference in edge strength and buffering is enriched for the inversion models (41% *vs.* 24%, $p < 0.005$).

Pdr3 likely acts as the intermediate buffering gene in mediating the inversion pattern observed for Hac1-Rpn4

One interesting pair of genes within the GSTF cluster dominated by the inversion pattern (Fig 2A, yellow branch) is Hac1-Rpn4. This pair displays a substantial amount of both inversion as well as buffering (Fig 6A). Hac1 and Rpn4 are both involved in the processing of inappropriately folding proteins, either by activating genes of the unfolded protein response [62] (UPR, Hac1) or via the endoplasmic reticulum-associated degradation [63] (ERAD, Rpn4). Two genes that display inversion, Pdr5 and Pdr15, show stronger expression changes compared to the other genes in the same gene set (Fig 6A, gene set 1). Both Pdr5 and Pdr15 are multidrug transporters involved in the pleiotropic drug response [64]. Expression of these two genes is tightly regulated by Pdr1 and Pdr3 [65,66]. Pdr5 is also positively regulated by expression of Yap1, a basic leucine zipper transcription factor that is required for oxidative stress tolerance [67]. Deletion of *PDR3* or *YAP1* also leads to significant downregulation of *PDR5* and *PDR15* (Fig 6B, left panel), whereas deletion of *PDR1* results in a slight upregulation of only *PDR15*. Based on these results, as well as different reports in literature [65–67], it is unclear which transcription factor is mediating the inversion effect observed for Hac1-Rpn4. However, when investigating the degree of mRNA expression for *PDR1*, *PDR3* and *YAP1* in the *rpn4*Δ, *hac1*Δ and *hac1*Δ *rpn4*Δ mutants, only *PDR3* shows a clear upregulation in the *hac1*Δ *rpn4*Δ double mutant and hardly any change in the respective single deletion mutants (Fig 6B, right panel). This is consistent with the role of the intermediate buffering gene as derived from our Petri net modelling results. This provides a possible mechanistic explanation where Pdr3 acts as the intermediate buffering gene in regulating Pdr5 and Pdr15 (Fig 6C).

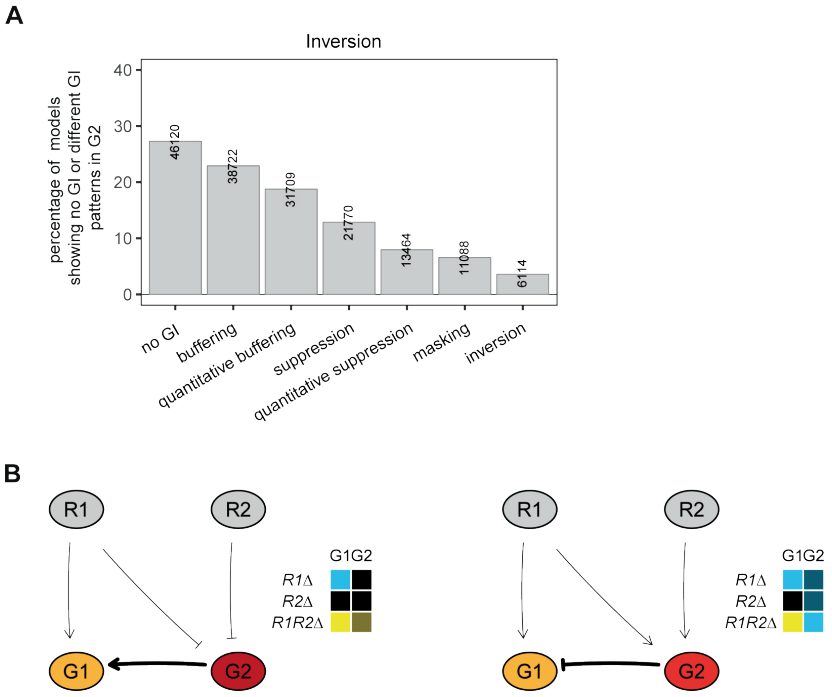
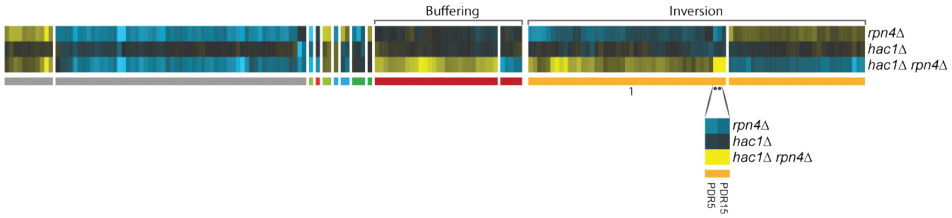


Fig 5. Inversion is frequently accompanied by buffering. (A) Bar plots showing the percentage of models that either have no genetic interaction or a different genetic interaction pattern in node G2 when node G1 is displaying inversion. The number of models per category is shown on top of each bar plot. **(B)** Petri net simulation results for two models with four nodes with node G1 always displaying inversion and node G2 displaying either buffering (left) or quantitative buffering (right). Heat maps as in Fig 4A.

A



B

	pdr3Δ	yap1Δ	pdr1Δ	rpn4Δ	hac1Δ	hac1Δ rpn4Δ
Pdr3	-0.9276	-0.3081	0.3508	-0.09	0.045	0.5431
Yap1	-0.0931	-3.4225	-0.0958	-0.6903	0.0004	-0.4687
Pdr1	-0.0462	0.0051	-2.6044	-0.3960	-0.0383	-0.3338
Pdr5	-1.4097	-1.0937	0.1322	-0.7909	-0.1017	2.0443
Pdr15	-1.279	-0.8107	0.3307	-0.6101	-0.184	1.7136

C

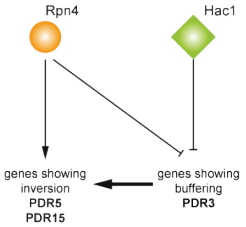


Fig 6. Pdr3 acts as an intermediate gene for observing inversion in PDR5 and PDR15. (A) Expression changes compared to WT (horizontal) in *rpn4Δ*, *hac1Δ*, and *hac1Δ rpn4Δ* mutants (vertical). Different colors underneath the gene expression profiles represent different genetic interaction patterns as indicated in Fig 1A. **(B)** Expression changes of Pdr5, Pdr15 and their known regulators compared to WT in the regulators single mutants (middle) and *rpn4Δ*, *hac1Δ* and *hac1Δ rpn4Δ* mutants (right). Yellow indicates upregulation compared to WT, blue indicates downregulation compared to WT **(C)** Proposed model to explain inversion pattern for PDR5 and PDR15 based on the results of Petri net simulation in fig 5A.

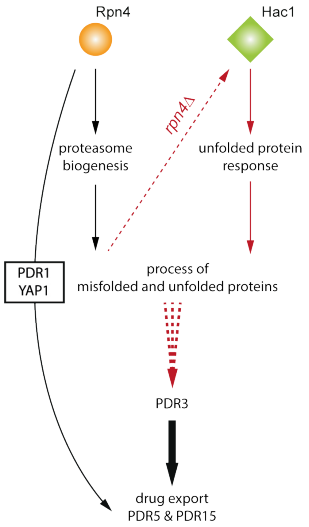


Fig 7. Combination of buffering by induced dependency and proposed model for inversion. Carton depiction of proposed model for genetic interaction between Rpn4 and Hac1. Red arrows indicate the consequence of disrupted genes and pathways.

Discussion

Genome-wide gene expression measurements to investigate the genetic interaction landscape

To investigate genetic interactions in a high-throughput manner, growth-based assays have frequently been deployed, resulting in the identification of an overwhelming number of both negative and positive genetic interactions [6,20–28]. Based on these surveys, several theoretical mechanisms have been proposed to explain genetic interactions [3,18,68,69]. More efforts, also using different types of assays, are however still needed to systematically and thoroughly investigate the underlying mechanisms. Alongside growth-based genetic interactions, genome-wide gene expression measurements have been applied to elucidate potential molecular mechanisms underlying genetic interactions [16,17,33–36]. Expression-based genetic interactions provide a more in-depth characterization of the genetic interaction landscape. This study shows that buffering is the most frequently occurring pattern underlying most negative genetic interactions. These are however to a large degree related to slow growing strains, hindering the investigation of the underlying mechanisms. By applying a slow growth transformation that removes a cell cycle associated gene expression signature, many of such effects can be filtered out [48]. The transformation results in distinct clusters that can be more easily aligned with potential underlying mechanisms. Recent advances using Crispr-Cas9 single and double knock-down screens, followed by single cell RNA sequencing have also shown that results are greatly influenced by the cell-cycle phase in which different cells are found [35,70]. It is therefore essential for future studies on genetic interactions to incorporate methods that decompose such large confounding effects, as they greatly influence the ability to deduce mechanism.

Systematic modelling to understand mechanisms of genetic interactions

To infer underlying mechanisms from the genetic interaction landscape as obtained from genome-wide gene expression measurements, systematic modelling approaches are needed [3,18]. Various modelling techniques have been instrumental in understanding various aspects of experimental data (reviewed in [71]). Different modelling methods have different applications, depending on the question asked and available data types. To infer the underlying mechanisms for many genetic interactions, an approach is needed that is able to exhaustively explore the complete genetic interaction landscape while at the same time incorporating (semi-) quantitative values. Here, using Petri net modelling, we have been able to exhaustively explore more than nine million models in a semi-quantitative manner. Inversion, a pattern strongly associated with a group of GSTF pairs was investigated in more detail, resulting in the striking conclusion that a quantitative difference in interaction strength is needed to explain inversion. The approach taken here,

by combining slow growth corrected genome-wide gene expression measurements with the exhaustive semi-quantitative Petri-net modelling thus highlights the benefits of using such an approach to understand mechanisms of genetic interactions. Applying this approach to other types of genetic interactions or across many more genetic interaction pairs can help us in further characterizing mechanisms of genetic interactions and relating these to pathway organization and cellular states.

Inversion as a way to fine-tune different efficacies between two redundant processes and a third, compensatory process

Previously, a mechanism termed "buffering by induced dependency" was proposed to explain parts of the genetic interaction patterns observed between Rpn4 and Hac1 and links the endoplasmic reticulum-associated degradation (ERAD) with the unfolded protein response (UPR), two distinct processes dealing with misfolded and unfolded proteins [17]. By combining this previously proposed mechanism with the proposed model for inversion here, most genetic interaction patterns observed for Rpn4 and Hac1 can be well explained (Fig 7). The combined model introduces a third, compensatory process, the pleiotropic drug response. Even though the exact relationship between ERAD, UPR and pleiotropic drug response is as yet unclear, the interplay between UPR and drug export has been shown in mammalian cells [72]. In yeast, Pdr5 and Pdr15 have been implicated in cellular detoxification [66,73] and may also be required for cellular detoxification under normal growth conditions [73]. Both Pdr5 and Pdr15 have been reported to be regulated through Pdr1 and Yap1 [67,74]. Rpn4 in turn, has been shown to also regulate expression of Pdr1 and Yap1 [75,76], as also confirmed by downregulation of both genes in *rpn4*Δ (Fig 6B, right panel) and downregulation of their downstream target genes Pdr5 and Pdr15 (Fig 6B, right panel). It is therefore likely that in the wildtype situation when Rpn4 is active, both ERAD and the pleiotropic drug response are functioning (Fig 7). Deletion of RPN4 leads to deactivation of the ERAD and pleiotropic drug response pathways and activation of the UPR through Hac1 (Fig 7). Deletion of both RPN4 and HAC1 results in a major growth defect and accumulation of misfolded and unfolded proteins, most likely leading to a stronger activation of the pleiotropic drug response through Pdr3 compared to the wildtype situation (Fig 6B, left panel; Fig 7) [65,66]. Taken together, this model thus provides a potential regulatory mechanism in which two redundant processes, each with slightly different efficacies, can be fine-tuned through a third, compensatory process. The requirement to fine-tune slightly different efficacies of different cellular processes then also provides a logical explanation why inversion is observed more frequently for gene-specific transcription factors since these allow for more fine-grained control than protein kinases and phosphatases.

Materials and Methods

Selection of GSTF and kinase/phosphatase pairs

Two selection criteria were applied to select genetically interacting GSTF and kinase/phosphatase pairs. First, one of the mutants of each individual pair should show genome-wide gene expression measurements similar to wildtype (WT). DNA microarray data from Kemmeren et al [77] was used to determine whether a single deletion mutant is similar to WT. A deletion mutant is considered similar to WT when fewer than eight genes are changing significantly ($p < 0.05$, FC > 1.7) in the deletion mutant gene expression profile, as previously described [16]. Second, selected pairs should show a significant growth-based negative genetic interaction score. Growth-based genetic interaction scores for GSTF [28] and kinase/phosphate [26] pairs were converted to Z-scores. A negative Z-score significance of $p < 0.05$ after multiple testing correction was used as the significance threshold. Applying these selection criteria resulted in 11 GSTF pairs and 15 kinase/phosphatase pairs (S1 Table).

Genome-wide gene expression measurements

Genome-wide gene expression measurements of single and double mutant GSTF pairs were obtained from Sameith et al [17]. Genome-wide gene expression measurements of single and double mutant kinase/ phosphatase pairs were obtained from van Wageningen et al [16]. Statistical analysis of these gene expression profiles was performed as previously described [77]. In summary, mutants were grown in Synthetic Complete (SC) medium with 2% glucose and harvested during exponential growth. WT cultures were grown alongside mutants in parallel to monitor for day to day effects. For each mutant statistical analysis was performed versus a collection of WT [16,77]. Reported FC for each transcript in both single and double mutants is the average of four replicate expression profiles over a WT pools consisted of 200 WT strains.

Growth-based genetic interaction scores

Growth measurements for single and double mutant GSTF and kinase/phosphatase pairs were obtained from Sameith et al [17] and van Wageningen et al [16] respectively. Growth-based genetic interaction scores were calculated for both GSTF and kinase/phosphatase pairs as performed before [17]. In summary, the fitness W of single and double mutants was determined as the ratio between the WT growth rate and the mutant growth rate. The growth-based genetic interaction score $\varepsilon_{growth,XY}$ was calculated as the deviation of the observed fitness in a double mutant from the expected fitness based on the respective single mutants ($\varepsilon_{growth,XY} = W_{x\Delta y\Delta} - W_{x\Delta} \cdot W_{y\Delta}$). P values were assigned to genetic interaction scores based on the mean and standard deviation of a generated background distribution [17]. P values were corrected for multiple testing using Benjamini-Hochberg. Adjusted p values

lower than 0.05 were considered significant. Fitness values of all single and double mutants, as well as calculated genetic interaction scores can be found in S1 Table.

Expression-based genetic interaction scores

Expression-based genetic interaction scores were calculated for both GSTF and kinase/phosphatase pairs as described before [17]. In summary, the effect of a genetic interaction between two genes X and Y on gene *i* is calculated as the deviation between the observed expression change in the double mutant and the expected expression change based on the corresponding single mutants ($\varepsilon_{txpn\_i,XY} = |M_{i,x\Delta y\Delta} - (M_{i,x\Delta} + M_{i,y\Delta})|$). The overall genetic interaction score between X and Y is calculated as the sum all genes *i* for which $\varepsilon_{txpn\_i,XY} > \log2(1.5)$. All genetic interaction scores consisting of at least 10 genes were kept for further downstream analyses. Genes with similar gene expression changes were divided into the 6 different patterns (buffering, quantitative buffering, suppression, quantitative suppression, masking, inversion), as previously described [17] (Fig 1A).

Clustering of expression-based genetic interaction scores

Genetic interaction profiles for both classes of proteins were grouped together based on the number of occurrences of the six different patterns using hierarchical clustering. Average linkage was applied for the clustering. Identity of genes in each genetic interaction profile was disregarded.

Slow growth transformation

Slow growth signature transformation of the gene expression profiles was performed as previously described [48]. In short, for each mutant, the correlation of its expression profile with the first principal component of 1,484 deletion strains [77] was removed, thus minimizing correlation with the relative growth rate. The transformation reduces correlation with the relative growth rate from 0.29 to 0.10 on average [48].

Model generation

Exhaustive modelling of possible network topologies underlying the genetic interaction patterns was carried out by creating Petri net models consisting of four nodes, representing two regulator genes (R1 and R2) and two downstream genes (G1 and G2). With four nodes and directed edges, there are 4^2=16 possible edges, and 2^{16}=65536 possible edge weight matrices, which is a tractable number. However, each interaction can in addition be positive or negative, and weak or strong (and absent), leading to $5^{16}=1.5\cdot10^{11}$ possible interaction graphs (edge weight matrices), which becomes intractable. Many of these models, however, will be irrelevant for the understanding the biological behavior of genetic interaction patterns of two genes. To exclude these types of models, the following conditions were applied: 1) No self-edges are allowed. 2) The number of incoming edges on any node

must be limited to two. 3) At least two incoming edges from at least one of the regulators (upstream nodes) to the genes (downstream nodes). Applying these conditions reduces the number of relevant edge weight matrices to 9,287,616. Furthermore, most generated matrices have mirror counterparts, therefore only one of the matrices was included in downstream analyses. Applying this filtering step results in 2,323,936 matrices. Fig 3 gives an overview of the various filtering steps, and shows which representation of the models was relevant in different stages of the filtering. Edge weight matrices were generated in R, version 3.2.2 (the function expand.grid was used to generate all combinations of edges per row in a given matrix).

Petri net simulations

Regulatory effects of two potentially interacting genes (R1 and R2) on two downstream genes (G1 and G1) were simulated using a Petri net approach [42,44,78,79] to recapitulate genetic interaction patterns observed in the gene expression data.

In the Petri net notation, nodes in a given model are represented by places (denoted as circles). Interactions between nodes always go via a transition (denoted as squares), connected via directed arcs (drawn as arrows). An incoming arc to a transition can be either activating or inhibiting. The weight on arcs going to a transition is always fixed to 1. The weight on arcs going from a transition to a place depends on the edge weight between two nodes, 1 for weak and 5 for strong (Fig 3).

For nodes with two incoming edges, one has to decide how these two inputs should be combined: does the transition require both inputs to be activated (AND logic), or can one or the other activate it (OR logic). To incorporate this, for each pair of incoming edges with the same weight, two Petri net models were generated: one using the AND logic, and one using the OR logic (Fig 3, bottom right panel). For two incoming edges with different weights only the Petri net model using the OR logic was generated. For cases with two incoming edges to a node with two different directions, activation and inhibition, inhibition dominates.

To simulate the regulatory effects of two upstream genes (R1 and R2), 200 tokens were provided to represent the mRNA resources for each regulator, except when one of the regulators has an incoming edge from the other regulator as shown in (S4A Fig). Each step in the simulation process comprises of firing all enabled transitions (maximal parallel execution) [80,81]. A transition is enabled to fire when resources (tokens) in the input place(s) match or exceed the weight(s) on the respective incoming arc(s) to the transition (S4B Fig). In total 50 consecutive transition firing steps were performed.

To incorporate deletion mutants in the simulation process, tokens were removed from corresponding regulators. To prevent accumulation of tokens in deleted regulators, each outgoing arc from a transition to the corresponding deleted places were also removed in simulated deletion strains. The number of tokens in G1 and G2 after 50 steps of firing

transitions in single and double mutants were compared with that in the WT situation where both R1 and R2 are active. To avoid division by zero one token was added to the total number of tokens in G1 and G1. These fold changes were then log2 transformed (M values).

Simulation-based genetic interaction scores for G1 and G2 were calculated based on the deviation between observed M values in the double mutant and the expected M value based on the single mutants, as follows: $\varepsilon_{sim,R1R2i} = |M_{R1\Delta R2\Delta i} - (M_{R1\Delta i} + M_{R2\Delta i})|$, where *i* can be either G1 or G2. Each node with $\varepsilon_{sim,R1R2i} > \log_2(1.7)$ was further divided into genetic interaction patterns, as defined before based on gene expression data [17]. Simulated expression levels for single and double mutants are considered to be increased relative to WT when $M > \log_2(1.7)$ and decreased when $M < -\log_2(1.7)$.

Functional enrichment test

Functional enrichment analysis was performed using a hypergeometric testing procedure on Gene Ontology (GO) biological process (BP) annotations [67] obtained from the *Saccharomyces Cerevisiae* Database [68]. The background population of genes was set to 6,359 and *p* values were corrected for multiple testing using Bonferroni.

Visualization of models

Models were visualized in R, version 3.2.2, using diagram package (version 1.6.3). Weak and strong activation/inhibition edges are represented as thin and thick lines, respectively.

Acknowledgements

We would like to thank Wim de Jonge, Thanasis Margaritis and all the other lab members for helpful discussions and advices throughout the project. Financial support was provided by the Netherlands Organisation for Scientific Research (NWO); grant number 864.11.010 (PK, SA).

Author Contributions

Conceptualization: Saman Amini, Frank C. P. Holstege, K. Anton Feenstra, Patrick Kemmeren

Formal Analysis: Saman Amini, Annika Jacobsen, Olga Ivanova, Philip Lijnzaad

Funding Acquisition: Patrick Kemmeren

Investigation: Saman Amini, Annika Jacobsen, Olga Ivanova, Philip Lijnzaad, Jaap Heringa, Frank C. P. Holstege, K. Anton Feenstra, Patrick Kemmeren

Methodology: Saman Amini, Annika Jacobsen, Frank C. P. Holstege, K. Anton Feenstra, Patrick Kemmeren

Resources: Frank C. P. Holstege, Patrick Kemmeren

Software: Saman Amini, Annika Jacobsen, Olga Ivanova

Supervision: Jaap Heringa, Frank C. P. Holstege, K. Anton Feenstra, Patrick Kemmeren
Visualization: Saman Amini, Annika Jacobsen, Olga Ivanova
Writing – Original Draft preparation: Saman Amini, Patrick Kemmeren
Writing – Review & Editing: Annika Jacobsen, Olga Ivanova, Philip Lijnzaad, Jaap Heringa, Frank C. P. Holstege, K. Anton Feenstra, Patrick Kemmeren

References

1. Badano JL, Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. Nat Rev Genet. 2002;3: 779–789. doi:10.1038/nrg910
2. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Hum Genet. 2013;132: 1077–1130. doi:10.1007/s00439-013-1331-2
3. 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: 111–133. doi:10.1146/annurev-genom-082509-141730
4. Phillips PC. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9: 855–867. doi:10.1038/nrg2452
5. Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nat Rev Genet. 2014;15: 722–733. doi:10.1038/nrg3747
6. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016;353. doi:10.1126/science.aaf1420
7. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in Caenorhabditis elegans identifies common modifiers of diverse signaling pathways. Nat Genet. 2006;38: 896–903. doi:10.1038/ng1844
8. Babu M, Arnold R, Bundalovic-Torma C, Gagarinova A, Wong KS, Kumar A, et al. Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in Escherichia coli. PLoS Genet. 2014;10: e1004120. doi:10.1371/journal.pgen.1004120
9. Bakal C, Linding R, Llense F, Heffern E, Martin-Blanco E, Pawson T, et al. Phosphorylation Networks Regulating JNK Activity in Diverse Genetic Backgrounds. Science. 2008;322: 453–456. doi:10.1126/science.1158739
10. Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, Boutros M. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. Nat Methods. 2011;8: 341–346. doi:10.1038/nmeth.1581
11. Roguev A, Talbot D, Negri GL, Shales M, Cagney G, Bandyopadhyay S, et al. Quantitative genetic-interaction mapping in mammalian cells. Nat Methods. 2013;10: 432–437. doi:10.1038/nmeth.2398
12. Vizeacoumar FJ, Arnold R, Vizeacoumar FS, Chandrashekhar M, Buzina A, Young JTF, et al. A negative genetic interaction map in isogenic cancer cell lines reveals cancer cell vulnerabilities. Mol Syst Biol. 2013;9: 696. doi:10.1038/msb.2013.54
13. Billmann M, Horn T, Fischer B, Sandmann T, Huber W, Boutros M. A genetic interaction map of cell cycle regulators. Mol Biol Cell. 2016;27: 1397–1407. doi:10.1091/mbc.E15-07-0467
14. Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat Biotechnol. 2017;35: 463–474. doi:10.1038/nbt.3834

15. Shen JP, Zhao D, Sasik R, Luebeck J, Birmingham A, Bojorquez-Gomez A, et al. Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. Nat Methods. 2017;14: 573–576. doi:10.1038/nmeth.4225
16. van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, et al. Functional overlap and regulatory links shape genetic interactions between signaling pathways. Cell. 2010;143: 991–1004. doi:10.1016/j.cell.2010.11.021
17. Sameith K, Amini S, Groot Koerkamp MJA, van Leenen D, Brok M, Brabers N, et al. A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions. BMC Biol. 2015;13: 112. doi:10.1186/s12915-015-0222-5
18. Lehner B. Molecular mechanisms of epistasis within and between genes. Trends Genet TIG. 2011;27: 323–331. doi:10.1016/j.tig.2011.05.007
19. Moore JH, Williams SM. Epistasis and its implications for personal genetics. Am J Hum Genet. 2009;85: 309–320. doi:10.1016/j.ajhg.2009.08.006
20. Jasnos L, Korona R. Epistatic buffering of fitness loss in yeast double deletion strains. Nat Genet. 2007;39: 550–554. doi:10.1038/ng1986
21. St Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, et al. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet. 2007;39: 199–206. doi:10.1038/ng1948
22. Szappanos B, Kovács K, Szamecz B, Honti F, Costanzo M, Baryshnikova A, et al. An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat Genet. 2011;43: 656–662. doi:10.1038/ng.846
23. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. Science. 2004;303: 808–813. doi:10.1126/science.1091317
24. Davierwala AP, Haynes J, Li Z, Brost RL, Robinson MD, Yu L, et al. The synthetic genetic interaction spectrum of essential genes. Nat Genet. 2005;37: 1147–1152. doi:10.1038/ng1640
25. Pan X, Ye P, Yuan DS, Wang X, Bader JS, Boeke JD. A DNA integrity network in the yeast Saccharomyces cerevisiae. Cell. 2006;124: 1069–1081. doi:10.1016/j.cell.2005.12.036
26. Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, et al. Functional Organization of the S. cerevisiae Phosphorylation Network. Cell. 2009;136: 952–963. doi:10.1016/j.cell.2008.12.039
27. Bandyopadhyay S, Mehta M, Kuo D, Sung M-K, Chuang R, Jaehnig EJ, et al. Rewiring of genetic networks in response to DNA damage. Science. 2010;330: 1385–1389. doi:10.1126/science.1195618
28. Zheng J, Benschop JJ, Shales M, Kemmeren P, Greenblatt J, Cagney G, et al. Epistatic relationships reveal the functional organization of yeast transcription factors. Mol Syst Biol. 2010;6: 420. doi:10.1038/msb.2010.77
29. Hartman JL, Garvik B, Hartwell L. Principles for the buffering of genetic variation. Science. 2001;291: 1001–1004.

30. Amini S, Holstege FCP, Kemmeren P. Growth condition dependency is the major cause of non-responsiveness upon genetic perturbation. PloS One. 2017;12: e0173432. doi:10.1371/journal.pone.0173432
31. Ihmels J, Collins SR, Schuldiner M, Krogan NJ, Weissman JS. Backup without redundancy: genetic interactions reveal the cost of duplicate gene loss. Mol Syst Biol. 2007;3: 86. doi:10.1038/msb4100127
32. Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature. 2007;446: 806–810. doi:10.1038/nature05649
33. Gutin J, Sadeh A, Rahat A, Aharoni A, Friedman N. Condition-specific genetic interaction maps reveal crosstalk between the cAMP/PKA and the HOG MAPK pathways in the activation of the general stress response. Mol Syst Biol. 2015;11: 829. doi:10.15252/msb.20156451
34. Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, et al. Structure and function of a transcriptional network activated by the MAPK Hog1. Nat Genet. 2008;40: 1300–1306. doi:10.1038/ng.235
35. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell. 2016;167: 1853–1866.e17. doi:10.1016/j.cell.2016.11.038
36. Pirkl M, Diekmann M, van der Wees M, Beerenwinkel N, Fröhlich H, Markowetz F. Inferring modulators of genetic interactions with epistatic nested effects models. PLoS Comput Biol. 2017;13: e1005496. doi:10.1371/journal.pcbi.1005496
37. Wong ASL, Choi GCG, Lu TK. Deciphering Combinatorial Genetics. Annu Rev Genet. 2016;50: 515–538. doi:10.1146/annurev-genet-120215-034902
38. Kauffman S, Peterson C, Samuelsson B, Troein C. Random Boolean network models and the yeast transcriptional network. Proc Natl Acad Sci U S A. 2003;100: 14796–14799. doi:10.1073/pnas.2036429100
39. Li F, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. Proc Natl Acad Sci U S A. 2004;101: 4781–4786. doi:10.1073/pnas.0305937101
40. Bonzanni N, Feenstra KA, Fokkink W, Heringa J. Petri Nets Are a Biologist's Best Friend. Formal Methods in Macro-Biology. Springer, Cham; 2014. pp. 102–116. doi:10.1007/978-3-319-10398-3_8
41. Bonzanni N, Garg A, Feenstra KA, Schütte J, Kinston S, Miranda-Saavedra D, et al. Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. Bioinforma Oxf Engl. 2013;29: i80-88. doi:10.1093/bioinformatics/btt243
42. Bonzanni N, Krepska E, Feenstra KA, Fokkink W, Kielmann T, Bal H, et al. Executing multicellular differentiation: quantitative predictive modelling of C.elegans vulval development. Bioinformatics. 2009;25: 2049–2056. doi:10.1093/bioinformatics/btp355
43. Chaouiya C, Remy E, Thieffry D. Qualitative Petri Net Modelling of Genetic Networks. Transactions on Computational Systems Biology VI. Springer, Berlin, Heidelberg; 2006. pp. 95–112. doi:10.1007/11880646_5

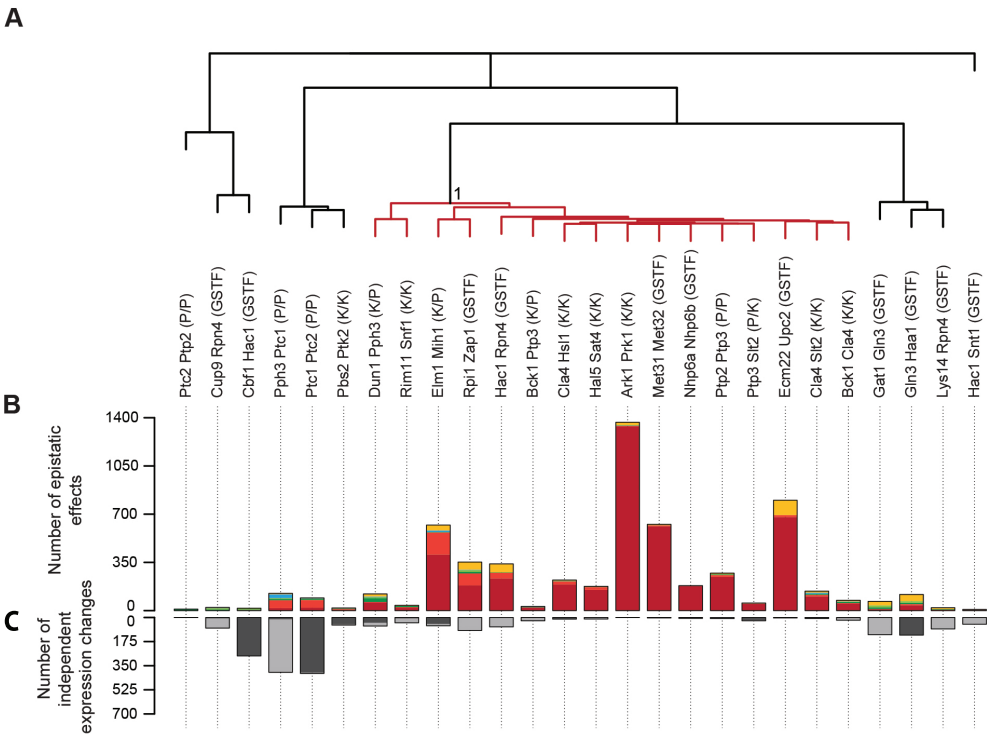
44. Jacobsen A, Heijmans N, Verkaar F, Smit MJ, Heringa J, Amerongen R van, et al. Construction and Experimental Validation of a Petri Net Model of Wnt/β-Catenin Signaling. PLOS ONE. 2016;11: e0155743. doi:10.1371/journal.pone.0155743
45. Mani R, St Onge RP, Hartman JL, Giaever G, Roth FP. Defining genetic interaction. Proc Natl Acad Sci U S A. 2008;105: 3461–3466. doi:10.1073/pnas.0712255105
46. Regenberg B, Grotkjaer T, Winther O, Fausbøll A, Akesson M, Bro C, et al. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. Genome Biol. 2006;7: R107. doi:10.1186/gb-2006-7-11-r107
47. Keren L, Zackay O, Lotan-Pompan M, Barenholz U, Dekel E, Sasson V, et al. Promoters maintain their relative activity levels under different growth conditions. Mol Syst Biol. 2013;9: 701. doi:10.1038/msb.2013.59
48. O'Duibhir E, Lijnzaad P, Benschop JJ, Lenstra TL, van Leenen D, Groot Koerkamp MJA, et al. Cell cycle population effects in perturbation studies. Mol Syst Biol. 2014;10: 732.
49. Costigan C, Kolodrubetz D, Snyder M. NHP6A and NHP6B, which encode HMG1-like proteins, are candidates for downstream components of the yeast SLT2 mitogen-activated protein kinase pathway. Mol Cell Biol. 1994;14: 2391–2403.
50. Blaiseau PL, Isnard AD, Surdin-Kerjan Y, Thomas D. Met31p and Met32p, two related zinc finger proteins, are involved in transcriptional regulation of yeast sulfur amino acid metabolism. Mol Cell Biol. 1997;17: 3640–3648.
51. Vik A null, Rine J. Upc2p and Ecm22p, dual regulators of sterol biosynthesis in Saccharomyces cerevisiae. Mol Cell Biol. 2001;21: 6395–6405.
52. Cope MJ, Yang S, Shang C, Drubin DG. Novel protein kinases Ark1p and Prk1p associate with and regulate the cortical actin cytoskeleton in budding yeast. J Cell Biol. 1999;144: 1203–1218.
53. Keane OM, Toft C, Carretero-Paulet L, Jones GW, Fares MA. Preservation of genetic and regulatory robustness in ancient gene duplicates of Saccharomyces cerevisiae. Genome Res. 2014;24: 1830–1841. doi:10.1101/gr.176792.114
54. Plata G, Vitkup D. Genetic robustness and functional evolution of gene duplicates. Nucleic Acids Res. 2014;42: 2405–2414. doi:10.1093/nar/gkt1200
55. Boone C, Bussey H, Andrews BJ. Exploring genetic interactions and networks with yeast. Nat Rev Genet. 2007;8: 437–449. doi:10.1038/nrg2085
56. Bouquin N, Barral Y, Courbeyrette R, Blondel M, Snyder M, Mann C. Regulation of cytokinesis by the Elm1 protein kinase in Saccharomyces cerevisiae. J Cell Sci. 2000;113 (Pt 8): 1435–1445.
57. Russell P, Moreno S, Reed SI. Conservation of mitotic controls in fission and budding yeasts. Cell. 1989;57: 295–303.
58. Thomas CL, Blacketer MJ, Edgington NP, Myers AM. Assembly interdependence among the S. cerevisiae bud neck ring proteins Elm1p, Hsl1p and Cdc12p. Yeast Chichester Engl. 2003;20: 813–826. doi:10.1002/yea.1003

59. McMillan JN, Longtine MS, Sia RA, Theesfeld CL, Bardes ES, Pringle JR, et al. The morphogenesis checkpoint in Saccharomyces cerevisiae: cell cycle control of Swe1p degradation by Hsl1p and Hsl7p. Mol Cell Biol. 1999;19: 6929–6939.
60. Morgan DO. Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol. 1997;13: 261–291. doi:10.1146/annurev.cellbio.13.1.261
61. Russell P. Checkpoints on the road to mitosis. Trends Biochem Sci. 1998;23: 399–402.
62. Mori K, Kawahara T, Yoshida H, Yanagi H, Yura T. Signalling from endoplasmic reticulum to nucleus: transcription factor with a basic-leucine zipper motif is required for the unfolded protein-response pathway. Genes Cells Devoted Mol Cell Mech. 1996;1: 803–817.
63. Mannhaupt G, Schnall R, Karpov V, Vetter I, Feldmann H. Rpn4p acts as a transcription factor by binding to PACE, a nonamer box found upstream of 26S proteasomal and other genes in yeast. FEBS Lett. 1999;450: 27–34.
64. Golin J, Ambudkar SV, May L. The yeast Pdr5p multidrug transporter: how does it recognize so many substrates? Biochem Biophys Res Commun. 2007;356: 1–5. doi:10.1016/j.bbrc.2007.02.011
65. Katzmann DJ, Burnett PE, Golin J, Mahé Y, Moye-Rowley WS. Transcriptional control of the yeast PDR5 gene by the PDR3 gene product. Mol Cell Biol. 1994;14: 4653–4661.
66. Wolfger H, Mahé Y, Parle-McDermott A, Delahodde A, Kuchler K. The yeast ATP binding cassette (ABC) protein genes PDR10 and PDR15 are novel targets for the Pdr1 and Pdr3 transcriptional regulators. FEBS Lett. 1997;418: 269–274.
67. Miyahara K, Hirata D, Miyakawa T. yAP-1- and yAP-2-mediated, heat shock-induced transcriptional activation of the multidrug resistance ABC transporter genes in Saccharomyces cerevisiae. Curr Genet. 1996;29: 103–105.
68. Boucher B, Jenna S. Genetic interaction networks: better understand to better predict. Front Genet. 2013;4: 290. doi:10.3389/fgene.2013.00290
69. Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C. Systematic mapping of genetic interaction networks. Annu Rev Genet. 2009;43: 601–625. doi:10.1146/annurev.genet.39.073003.114751
70. Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell. 2016;167: 1867–1882.e21. doi:10.1016/j.cell.2016.11.048
71. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. Nat Rev Mol Cell Biol. 2008;9: 770–780. doi:10.1038/nrm2503
72. YAN M-M, NI J-D, SONG D, DING M, HUANG J. Interplay between unfolded protein response and autophagy promotes tumor drug resistance. Oncol Lett. 2015;10: 1959–1969. doi:10.3892/ol.2015.3508
73. Mamnun YM, Schüller C, Kuchler K. Expression regulation of the yeast PDR5 ATP-binding cassette (ABC) transporter suggests a role in cellular detoxification during the exponential growth phase. FEBS Lett. 2004;559: 111–117. doi:10.1016/S0014-5793(04)00046-8

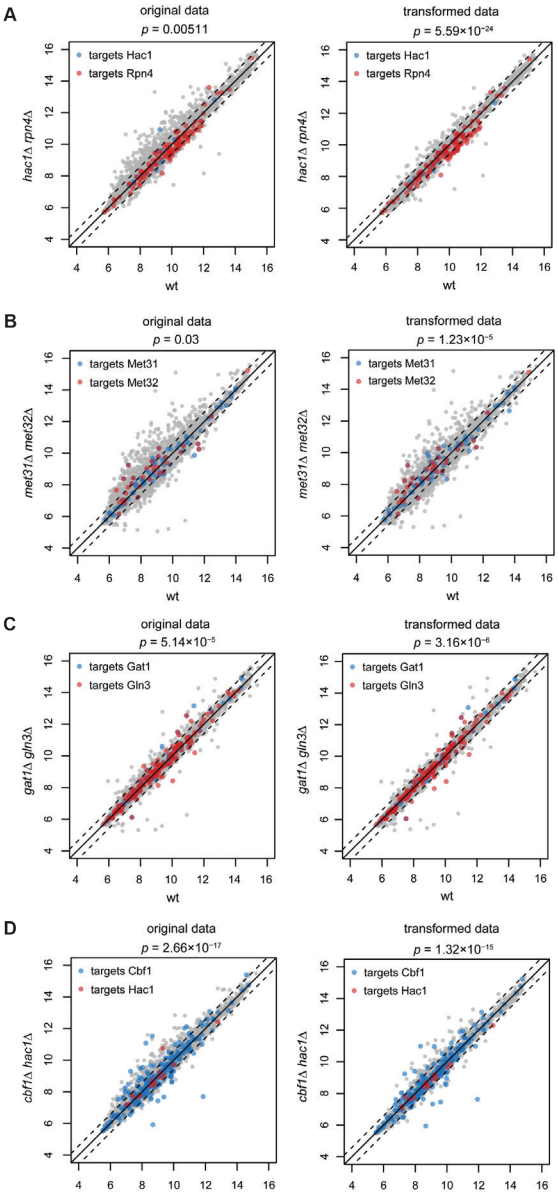
74. DeRisi J, van den Hazel B, Marc P, Balzi E, Brown P, Jacq C, et al. Genome microarray analysis of transcriptional activation in multidrug resistance yeast mutants. FEBS Lett. 2000;470: 156–160.
75. Salin H, Fardeau V, Piccini E, Lelandais G, Tanty V, Lemoine S, et al. Structure and properties of transcriptional networks driving selenite stress response in yeasts. BMC Genomics. 2008;9: 333. doi:10.1186/1471-2164-9-333
76. Spasskaya DS, Karpov DS, Mironov AS, Karpov VL. Transcription factor Rpn4 promotes a complex antistress response in Saccharomyces cerevisiae cells exposed to methyl methanesulfonate. Mol Biol. 2014;48: 141–149. doi:10.1134/S0026893314010130
77. Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, et al. Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. Cell. 2014;157: 740–752. doi:10.1016/j.cell.2014.02.054
78. Krepska E, Bonzanni N, Feenstra A, Fokkink W, Kielmann T, Bal H, et al. Design Issues for Qualitative Modelling of Biological Cells with Petri Nets. Formal Methods in Systems Biology. Springer, Berlin, Heidelberg; 2008. pp. 48–62. doi:10.1007/978-3-540-68413-8_4
79. C. A. Petri. Kommunikation mit Automaten. Bonn: Institut für Instrumentelle Mathematik, Schriften des IIM; 1962.
80. Burkhard H-D. On Priorities of Parallelism. Logic of Programs and Their Applications, Proceedings. London, UK, UK: Springer-Verlag; 1983. pp. 86–97. Available: http://dl.acm.org/citation.cfm?id=648062.747425
81. Bonzanni N, Feenstra KA, Fokkink W, Krepska E. What Can Formal Methods Bring to Systems Biology? FM 2009: Formal Methods. Springer, Berlin, Heidelberg; 2009. pp. 16–22. doi:10.1007/978-3-642-05089-3_2

S1 Table.
Single and double mutant GSTF and kinase/phosphatase pairs.

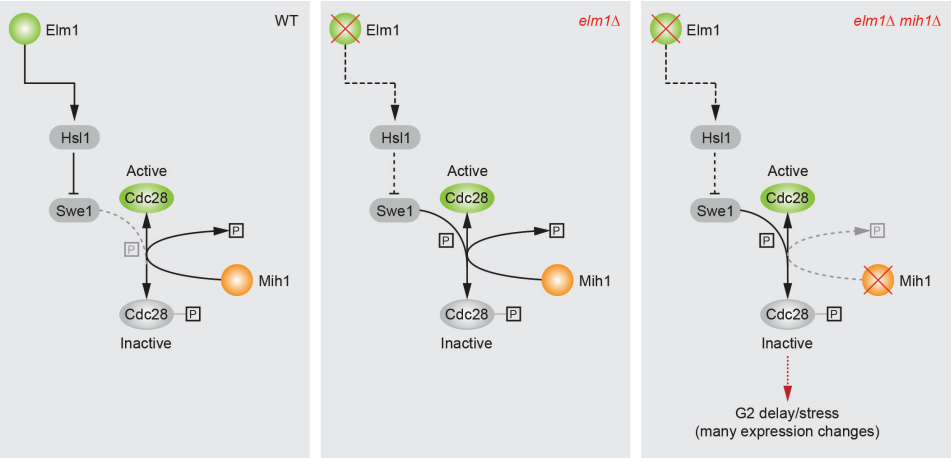
Gene pair	GSTF or K/P	RGR (observed)	RGR (expected)	RGR (single mutants)	Genetic interaction score	Adjusted *p*
Met31 / Met32	GSTF	0,296	1,059	1.031 / 1.028	-0,763	2,05E-21
Ecm22 / Upc2	GSTF	0,506	1,040	1.005 / 1.034	-0,534	8,17E-11
Hac1 / Rpn4	GSTF	0,453	0,958	1.056 / 0.907	-0,504	8,76E-10
Nhp6a / Nhp6b	GSTF	0,455	0,891	0.923 / 0.965	-0,436	1,12E-7
Gln3 / Haa1	GSTF	0,640	0,861	0.803 / 1.073	-0,221	1,34E-2
Rpi1 / Zap1	GSTF	0,621	0,824	0.965 / 0.854	-0,203	2,43E-2
Cbf1 / Hac1	GSTF	0,603	0,720	0.682 / 1.056	-0,117	1,91E-1
Cup9 / Rpn4	GSTF	0,924	1,022	1.127 / 0.907	-0,098	2,52E-1
Hac1 / Snt1	GSTF	0,902	0,951	1.056 / 0.901	-0,049	3,88E-1
Gat1 / Gln3	GSTF	0,780	0,818	1.019 / 0.803	-0,039	4,12E-1
Lys14 / Rpn4	GSTF	1,036	0,799	0.881 / 0.907	0,238	3,97E-3
Pbs2 / Ptk2	K / K	0,840	0,889	0.949 / 0.937	-0,049	3,88E-1
Ptc2 / Ptp2	P / P	1,003	0,989	1.006 / 0.983	0,014	4,35E-1
Pph3 / Ptc1	P / P	0,412	0,564	0.967 / 0.583	-0,152	9,56E-2
Bck1 / Cla4	K / K	0,431	0,874	0.973 / 0.898	-0,443	7,61E-8
Hal5 / Sat4	K / K	0,718	0,921	0.980 / 0.940	-0,203	2,43E-2
Elm1 / Mih1	K / P	0,316	0,816	0.849 / 0.961	-0,500	9,78E-10
Ark1 / Prk1	K / K	0,412	0,976	0.989 / 0.986	-0,563	7,64E-12
Bck1 / Ptp3	K / P	0,810	0,954	0.973 / 0.980	-0,143	1,14E-1
Ptc1 / Ptc2	P / P	0,538	0,587	0.583 / 1.006	-0,048	3,88E-1
Rim11 / Snf1	K / K	0,874	0,942	0.977 / 0.965	-0,068	3,39E-1
Dun1 / Pph3	K / P	0,785	0,731	0.755 / 0.967	0,054	3,39E-1
Cla4 / Slt2	K / K	0,496	0,865	0.898 / 0.964	-0,369	1,07E-5
Ptp3 / Slt2	P / K	0,492	0,944	0.980 / 0.964	-0,453	4,11E-8
Ptp2 / Ptp3	P / P	0,712	0,963	0.983 / 0.980	-0,251	3,97E-3
Cla4 / Hsl1	K / K	0,612	0,882	0.898 / 0.983	-0,270	2,30E-3



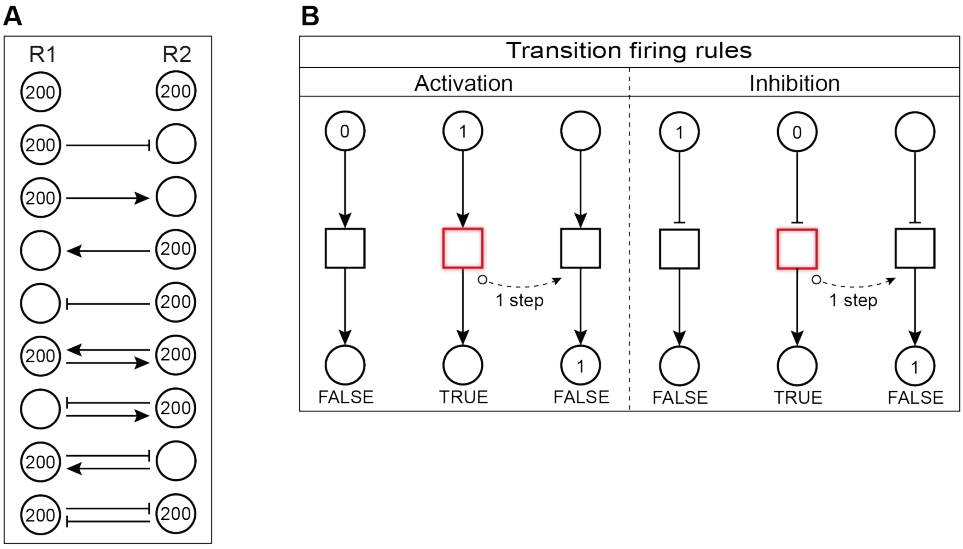
S1 Fig. Buffering dominates genetic interaction profiles. (A) Hierarchical clustering of all pairs according to their genetic interaction effects. Average linkage clustering was applied to group pairs with similar genetic interaction patterns. The number of occurrences for each genetic interaction pattern was used and the identity of individual genes was disregarded. Similarity between pairs was calculated using the cosine correlation. Most pairs are grouped together in a single branch (indicated in red), which is dominated by buffering. **(B)** The number of genetic interaction effects underlying the clustering are shown as bar plots below the dendrogram (colors as in Fig 1A). **(C)** Number of genes showing no genetic interaction pattern but significantly changing in one of the mutants compared to WT ($p \leq 0.01$, FC > 1.5). Dark gray for the first named gene, light gray for the second named gene.



S2 Fig. Slow growth correction improves identification of GSTF targets. Scatter plots showing gene expression levels in the GSTF double mutant pairs *hac1*Δ *rpn4*Δ (**A**), *met31*Δ *met32*Δ (**B**), *gat1*Δ *gln3*Δ (**C**) and *cbf1*Δ *hac1*Δ (**D**) versus WT before (left) or after (right) slow growth correction. Individual transcripts are represented as dots. The dashed line indicates a FC of 1.7. Dots depicted in blue and red correspond to targets of the first and second gene in a named GSTF pair. *P*-values are calculated using a hypergeometric testing procedure to test the enrichment of GSTF targets among genes that change more than 1.7 fold before (left) or after (right) slow growth correction.



S3 Fig. The genetic interaction between Elm1 and Mih1 can be explained through pathway redundancy. Cartoon depicting the proposed genetic interaction between Elm1 and Mih1. **(left panel)** WT situation where the activity of Cdc28 is not disrupted by Swe1 phosphorylation. **(Middle panel)** Deletion of Elm1 leads to derepression of Swe1 activity. The increase of Swe1 activity can be compensated by Mih1. **(Right panel)** Deletion of both Elm1 and Mih1 will cause an increase of phosphorylated Cdc28 (inactive form), which in turn can lead to G2 delay/stress and therefore many gene expression changes.



S4 Fig. Provided tokens to places in WT condition and transition firing rules. **(A)** Provided tokens to regulators depending on edges between them. **(B)** Transition firing rules for activation and inhibition edges depending on the presence of tokens in upstream places.

An initial map of genetic interactions in pediatric cancer

Saman Amini[1,2,*], Josephine T. Daub[1,*], Frank C. P. Holstege[1], Patrick Kemmeren[1,2]
[1] Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands Abstract
[2] Center for Molecular Medicine, University Medical Centre Utrecht, Utrecht, The Netherlands
[*] Equal contribution

Abstract

Cancers are thought to arise from the combined effects of alterations in individual genes and such combinatorial effects can be viewed as a type of genetic interaction. Genetic interactions occur when the outcome of combining two or more genetic alterations cannot be predicted by the individual effect of single alterations. One of the methods to detect genetic interactions in cancer is by statistical analysis of large numbers of tumors. Applying such a strategy in adult cancers has led to identification of co-occurring and mutually exclusive pairs of DNA alterations. Little is known about genetic interactions in pediatric cancer. The detection of DNA alterations in large numbers of pediatric tumors has only recently begun. This study describes a framework for detecting genetic interactions in pediatric cancer. Two methods are applied to detect co-occurring and mutually exclusive pairs of genes in 22 different pediatric cancer types. The two genetic interaction analyses are performed for both individual cancer types and across all data. A map of genetic interactions is presented by combining the results obtained from the four approaches. The results show a limited number of genetic interactions, in part probably due to the as yet low number of samples. The results nevertheless also contain mutually exclusive mutations that have previously been reported, confirming the utility of the approach. Our framework therefore provides a basis for future detection of genetic interactions between different types of alterations and in larger datasets.

Introduction

A primary focus of cancer biology is to obtain insight into cancer initiation and progression. Many driver genes have been identified that differentiate tumor cells from normal cells in both adult and pediatric cancer types [1–4]. Despite the discovery of such driver genes we still lack the ability to understand how mutations in genes interact together to convey a complex phenotype [5]. It is thought that cancers do not arise from the effect of individual genes but are rather the result of widespread genetic interactions between them [6]. Genetic interactions occur when the effect of combining two or more alterations in the genome cannot be predicted by combining the individual effects of the separate alterations. Genetic interactions are known to be pervasive in model organisms such as yeast [7]. Efforts have been also initiated to map genetic interactions in adult cancer cells [8–11]. One of the primary goals of such studies is to detect synthetic lethality between two alterations. Synthetic lethality is a type of genetic interaction in which simultaneous disruption of two genes results in cell death [12]. For example, cancer cells that harbor mutations in *BRCA1* or *BRCA2* are highly dependent on the function of *PARP1* [13]. Targeting *PARP1* which is involved in single-stranded DNA break repair, results in cell death in cancer cells that have mutations in *BRCA1* or *BRCA2* [14]. Exploiting these types of genetic interactions for therapeutic purposes can therefore be beneficial for devising more precise cancer treatments.

Multiple strategies have been used to detect genetic interactions in cancer cells [8–10,15]. One of these strategies is to use statistical analysis to detect pairs of genes that are either frequently or rarely co-mutated in large collections of tumor genomes [16,17]. Applying this method to adult cancers has led to the identification of co-occurring and mutually exclusive pairs of alterations [15,17–19]. A pair is co-occurring when two alterations co-occur more often than expected by chance, while a pair is considered mutually exclusive when two alterations co-occur less often than expected by chance. One of the interpretations for mutual exclusivity is synthetic lethality [19]. A commonly used strategy to find genetic interactions in cancer is to start with a gene-sample mutation matrix. This matrix contains information about the presence or absence of mutated genes for each individual. The number of co-occurrences for each gene pair can be obtained from such a matrix. Then, co-occurring and mutually exclusive pairs can be inferred by comparing the number of co-occurrences in the real dataset with large numbers of permuted matrices.

Several tools have been developed to identify co-occurring and mutually exclusive interactions between frequently mutated genes in adult cancers [15,17–22]. One of the important requirements is to have large numbers of samples to have enough statistical power. The number of samples for many adult cancers is not a limiting factor [2]. However, this can be an issue in rare diseases such as pediatric cancers. Furthermore, pediatric cancers are fundamentally different form adult cancers. For example, they usually have a different

cell of origin compared to adult tumors [23]. In addition they are thought to require a lower number of driver mutations for tumorigenesis and usually exhibit a much lower number of passenger mutations [23]. Little is known about genetic interactions in pediatric cancers. The exhaustive characterization of genetic alterations has only recently begun in pediatric cancers using whole-genome and exome sequencing approaches [24]. Larger numbers of sequenced tumors for pediatric cancers are now available which makes it feasible to detect co-occurring and mutually exclusive altered genes [3]

This work describes a framework for detecting genetic interactions in pediatric cancer. Two existing methods have been applied to detect co-occurring and mutually exclusive pairs in 22 different pediatric cancers. These methods are denoted as permutation and Weighted Sampling based Mutual Exclusivity (WeSME) and have previously been used to detect genetic interactions in adult cancers [15,17]. The main difference between these methods is in the randomization procedure that accounts for the mutation frequencies of samples. Most genetic interaction studies have only focused on detecting interactions between significantly mutated genes (SMGs) [3,15,17]. SMGs show significantly higher mutation rates compared to background [25]. Although a logical choice, restriction to SMGs potentially excludes finding all important interactions. The result of our approach is an initial map of genetic interactions for pediatric cancer, without restriction to SMGs. Our framework can be used in future to detect genetic interactions between other types of alterations including copy number variations (CNVs), structural variants (SV) and DNA methylation events.

Results

A pediatric cancer data set for detecting genetic interactions

To systematically detect genetic interactions in pediatric cancers, a collection of tumor samples comprising 961 tumors from 22 distinct cancer types was used as a starting point [3]. The cohort covers major childhood cancer types with an emphasis on central nervous system tumors including atypical teratoid/rhabdoid tumors (**ATRT**), embryonal tumors with multilayered rosettes (**ETMR**), the four medulloblastoma groups WNT (**MB-WNT**), SHH (**MB-SHH**), Group3 (**MB-GR3**) and Group4 (**MB-GR4**), pilocytic astrocytoma (**PA**), high-grade glioma with and without histone 3 K27M mutations (**HGG-K27M**, **HGG-other**), and infratentorial and supratentorial ependymoma (**EPD-IT**, **EPD-ST**). The remaining cancer types cover acute myeloid leukemia (**AML**), acute lymphoblastic leukemia (**ALL**), Burkitt's lymphoma (**BL**), neuroblastoma (**NB**), Wilms' tumor (**WT**), osteosarcoma (**OS**), Ewing's sarcoma (**EWS**), hepatic hepatoblastoma (**HB**), adrenocortical carcinoma (**ACT**), rhabdomyosarcoma (**RMS**), and retinoblastoma (**RB**). The dataset consists of paired and single end Illumina-based sequencing including whole-genome sequences (WGS) and whole-exome sequences (WES). Both primary and relapse tumors

were included. Combining primary and relapse tumors from the same individual in genetic interaction analysis may introduce a confounding effect due to having two related samples. Thus, only primary tumors were used in the detection of genetic interactions to avoid any biases. Furthermore, only samples sequenced with paired-end reads were included to avoid potential biases due to using different sequencing technologies. After removing relapse tumors and single-end sequenced samples, a total of 869 tumors remained in the final dataset used for investigating genetic interactions. The number of samples per cancer type varies from 8 for ACC to 105 for PA with most cancer types containing at least 20 samples (Fig 1A).

Selection of genetic alterations

Genetic alterations that were considered for the detection of genetic interactions included somatic single nucleotide variants (SNVs) as well as small insertions and deletions (indels). Only SNVs and indels occurring in exons that are expected to have a functional consequence were used for further analysis (Methods). We identified all genes that harbor one or more mutations, not limiting the analysis to just the SMGs. The number of mutated genes per individual varies greatly across different pediatric cancer types (Fig 1B). PA and RB cancer types have the lowest number of mutated genes (median = 3). On the other side of the spectrum, BL has the highest number of mutated genes (median = 28). Several individual tumor samples have no detected mutations of the types considered here. Moreover, to avoid false positives, three samples in cancer type HGG-other with an extremely high number of mutated genes (denoted as hypermutators) were excluded from genetic interaction analysis (Fig 1B).

Systematic identification of genetic interactions

For detecting genetic interactions, an ensemble approach was applied, combining the results from two slightly different methods. First, a permutation approach consisting of several steps was applied to identify co-occurring and mutually exclusive pairs of mutated genes (Fig 2, Methods). As a first step, mutation matrices listing which genes are mutated in each individual were constructed per cancer type and the number of co-occurrences was obtained for all gene pair combinations. For the second step, a randomization procedure was employed that also accounts for the heterogeneous distribution of mutated genes and samples to create a null distribution of co-occurrence counts. For the third step, *P*-values were computed for all the gene pairs within each cancer type. Finally, a null distribution of *P*-values was obtained by testing for genetic interactions for 100 randomly selected permuted matrices. This null distribution was used to infer empirical false discovery rates (FDR, Methods). Gene pairs with FDR < 0.2 were considered as significant.

A second approach was also applied to detect genetic interactions. This second approach uses WeSME and is based on the same principle as the matrix permutation approach,

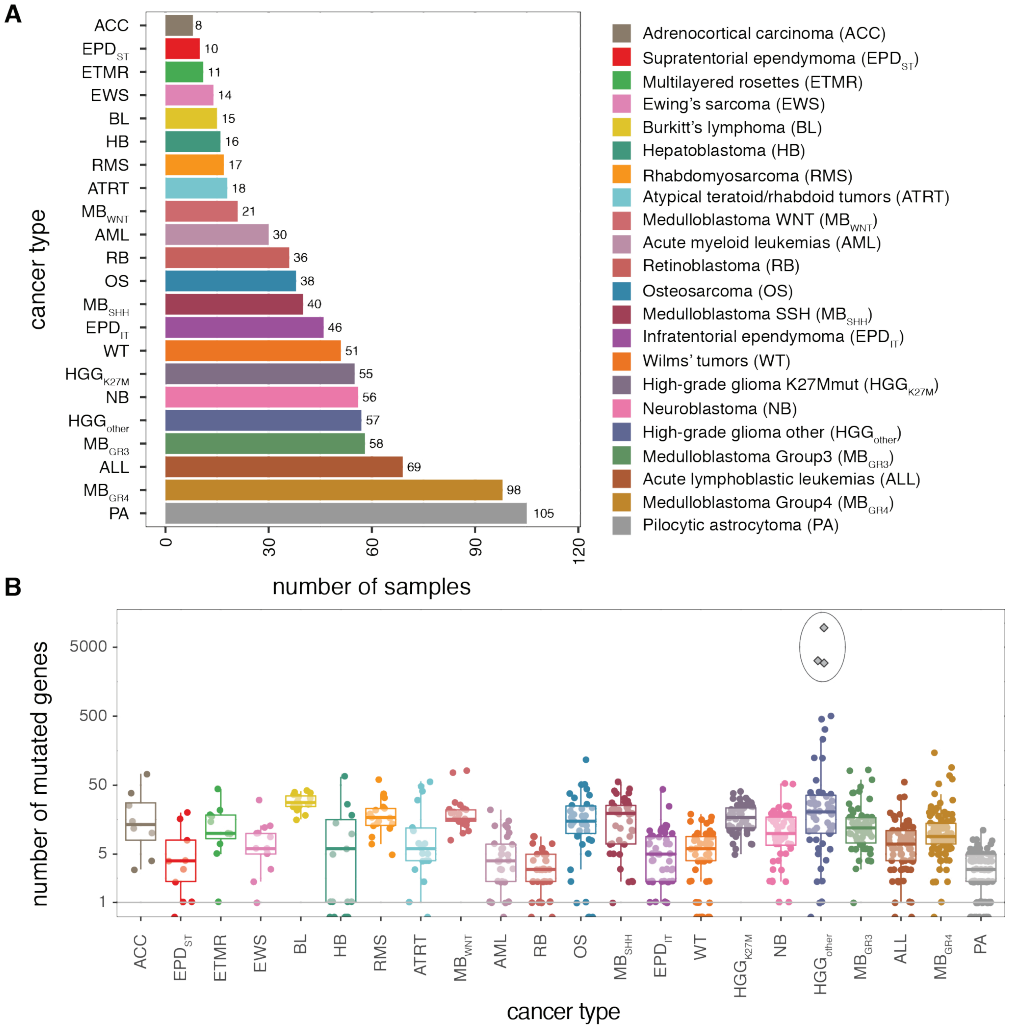


Fig 1. Number of samples and mutated genes per cancer type. (A) The number of samples per cancer type in the cohort that was used in this study. Only primary and paired-end sequenced samples are included. **(B)** Number of mutated genes identified in each individual per cancer type. Points on the horizontal axis represent samples without any mutated genes. Three samples within HGGother depicted by diamond shape symbols are hypermutators which are excluded from the genetic interaction analysis. Mutated genes are defined as genes that harbor at least one exonic mutation with a likely functional consequence. Only SNVs and small indels were considered.

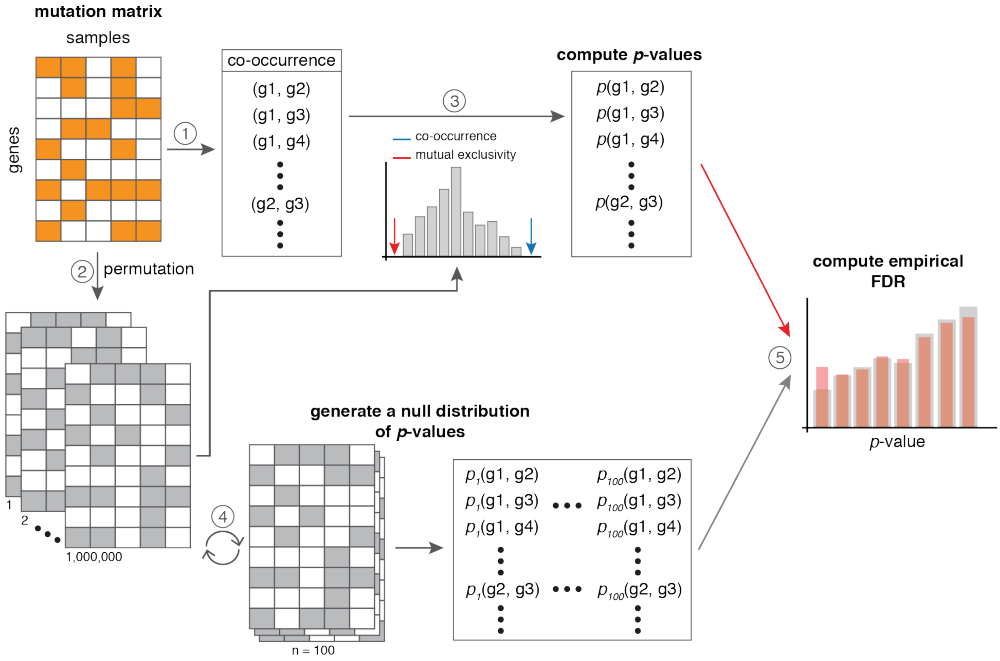


Fig 2. Permutation approach to identify co-occurrence and mutual exclusivity. The permutation pipeline in which genetic interactions are detected consist of 5 steps. A detailed explanation of each step is described in the Methods section.

starting with a mutation matrix and counting the number of mutually exclusive samples in gene pairs, where low counts are indicative of co-occurrence [17]. However, in order to save computation time, WeSME does not perform matrix permutations to infer significance but uses a weighted sampling scheme based on the mutation frequencies of samples. Gene pairs with an FDR < 0.2 and a *P*-value < 0.1 were considered as significant candidates. Genetic interactions detected in both approaches where combined and considered strong candidates if the genetic interaction occurs in both tests.

Detected genetic interactions in individual cancer types

The investigation of genetic interactions was first performed per cancer type. Using the combined results from the per cancer type analyses of the two different methods 14 genetic interactions were identified. Significant genetic interactions were detected for cancer types HGG-other, HGG-K27M, MB-SHH, and MB-WNT (Fig 3). As is discussed below, reasons for detecting genetic interactions in only four cancer types out of 22 include having insufficient samples (Fig 1). The majority of detected interactions consist of mutually exclusive pairs. Half of the detected genetic interactions have an FDR < 0.1 in both tests (Supplemental table 1). The PIK3CA-H3F3A pair is only significant in the WeSME test. Six

genetic interactions (five pairs in MB-SHH and one pair in MB-WNT) were only significant in the permutation test (Fig 3, Supplemental table 1). Several genes are part of significant pairs in more than one cancer type. For example, TP53 is mutually exclusive with ACVR1 and PTCH1 in HGG-K27M and MB-SHH cancer types, respectively. Interestingly, one of the identified pairs (KMT2D-DDX3X) was identified as mutually exclusive in the MB-WNT cancer type but co-occurring in MB-SHH.

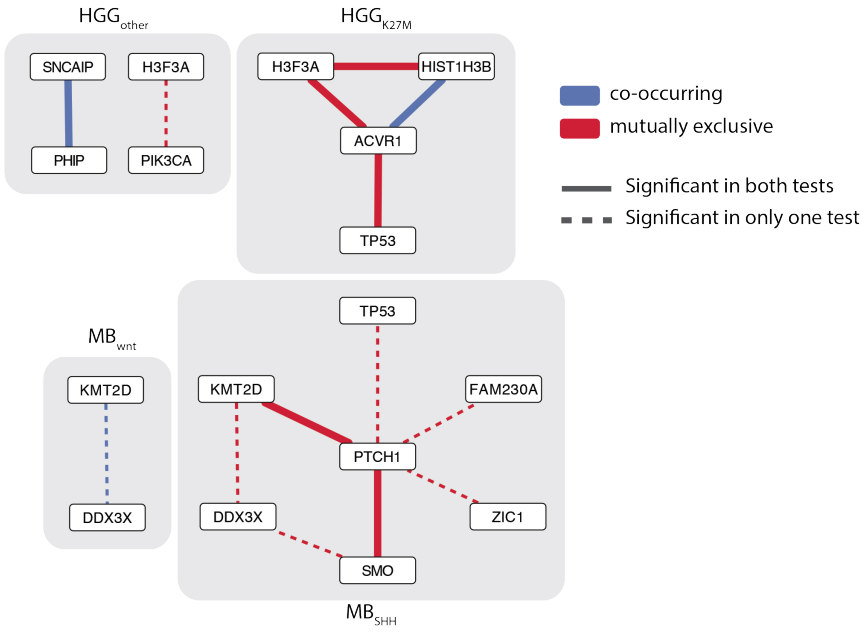
A mutually exclusive interaction was identified between PTCH1 and SMO in the MB-SHH cancer type. Both of these genes are members of the sonic hedgehog signalling (SHH) pathway [26]. While normally the SHH pathway is tightly regulated, hyperactivation of this pathway is found in many solid tumors both in adult and pediatric cancers [27–30]. Inactivation of the SHH pathway is achieved through the inhibition activity of the transmembrane protein Patched1 (PTCH1) on the transmembrane protein Smoothened (SMO) [31,32]. Mutations in *PTCH1* can result in removing the inhibition on SMO which leads to nuclear accumulation of GL1 and activation of target genes that promote several oncogenic properties of tumor cells [32]. Activating mutations in *SMO* are also associated with the same response [33]. The mutual exclusivity observed here between *PTCH1* and *SMO* can therefore be explained based on the different types of mutations and their function in a linear pathway.

In the HGG-K27M subtype, one co-occurring and three mutually exclusive interactions were identified by both tests (Fig 3). A mutually exclusive interaction was detected between *H3F3A* and *HIST1H3B*. *H3F3A* encodes for the H3.3 histone and *HIST1H3B* for the H3.1 histone. It has been reported that individual mutations in these genes occur in pediatric diffuse intrinsic pontine glioma (DIPG), a subtype of HGG [34,35]. As described previously, mutations in *H3F3A* and *HIST1H3B* define two subgroups of DIPG [34]. This is confirmed by the mutually exclusive interaction in our analysis. Mutations in *ACVR1* (encoding for Activin A Receptor Type 1 protein) were also mutually exclusive with *H3F3A*. Co-occurring relationship between *ACVR1* and *HIST1H3B* also fits previous findings, as *ACVR1* mutations are exclusively found in H3.1-mutated tumors [34]. Taken together, the detected genetic interactions between *H3F3A*, *HIST1H3B* and *ACVR1* depicted by the triangle shape (Fig 3) reflects what is known from previously reported subtypes of HGG-KM27, confirming that our analysis is able to detect meaningful genetic interactions.

Genetic interactions in the pan-cancer dataset

The previous analysis was performed on each cancer type individually. To detect genetic interactions across all pediatric cancers, all tumors from different cancer types were combined in a single analysis. Similar to the analysis of individual cancer types, the heterogeneous distribution of mutations across mutated genes, samples and cancer types was taken into account. The analysis was again carried out by both permutation and

WeSME tests. The result of the pan-cancer analysis revealed seven genetic interactions including five mutually exclusive and two co-occurring. Four interactions were detected in both tests and three were only detected in the WeSME test. Five of the detected interactions were also found in the individual cancer type analysis. The genes involved in these interactions primarily occur in HGG cancer subtypes (Fig 4, supplemental table 2). Two genetic interactions were found in the pan-cancer analysis only (COL5A2-FOLH1 and HLA-DRB1-TP53). The mutation rates of these genes were very low in individual cancers (Supplemental table 2). Therefore, they only become significant when the co-occurrence or



mutual exclusivity numbers are aggregated.

Fig 3. Genetic interaction network for four cancer types. Mutually exclusive and co-occurring interactions between mutated genes were detected in four cancer types. Nodes represent mutated genes and edges represent the type of detected genetic interactions. Co-occurrence and mutual exclusivity are depicted in blue and red colours, respectively. Solid edges depict genetic interactions that are significant in both tests. Dashes edges depict genetic interactions that are significant in only one of the tests.

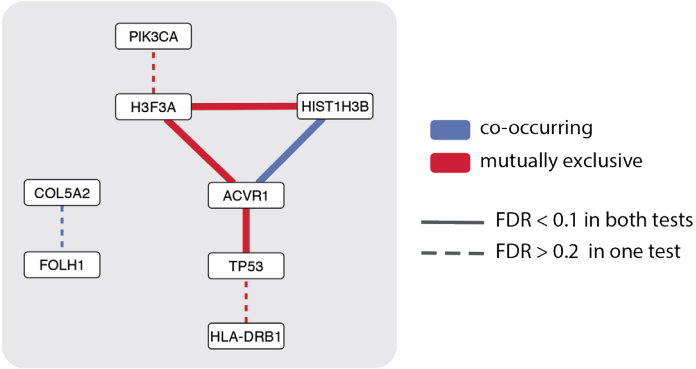


Fig 4. Pan-cancer genetic interaction network. Mutually exclusive and co-occurrence interactions between genes in the pan-cancer analysis. Nodes represent gene names and edges represent genetic interactions. Co-occurrence and mutual exclusivity are depicted in blue and red colours, respectively. Solid edges depict genetic interactions that are significant in both methods (permutation and WeSME). Dashes edges depict genetic interactions that are significant in only one of the tests.

Discussion

This study presents the first extensive search for genetic interactions in pediatric cancer, without restriction to significantly mutated genes only. Two different methods were applied to detect genetic interactions. The result is an initial map of genetic interactions between genes that harbor somatic mutations. Genetic interactions that are detected by both permutation and WeSME methods are considered as strong candidates. The genetic interaction analysis was performed for individual cancer types as well as pan-cancer. Two genetic interactions were found in the pan-cancer analysis that were not detected in the cancer type specific analysis. Only combining samples in a pan-cancer analysis led to the detection of these genetic interactions. Therefore, pan-cancer analysis is helpful for detecting genetic interactions between less frequently mutated genes.

Genetic interactions were only detected in four cancer types which do not have the highest number of samples (Fig 1A). One of the reasons that we were not able to detect any genetic interactions for some of the cancer types could be due to a small sample size. Including more samples from different cohorts would likely be extremely useful for detecting more genetic interactions. Despite the high number of samples for cancer type PA, no genetic interactions were detected. One possibility is that PA does not harbor any genetic interactions and requires for example only a single driver alteration. Including other types of alterations such as CNVs and DNA methylation events may also be useful to exhaustively detect all interesting interactions [15]. Alterations such as structural

variants frequently occur in many cancer types and are specifically abundant in pediatric cancers [36]. Annotating genes that are affected by structural variants is a challenging task. However, including them in the detection of genetic interactions will most likely lead to novel candidate gene pairs.

In this study two different methods were applied to detect genetic interactions between somatic mutations. In the cancer type specific analysis, 50% of the genetic interactions were detected in both permutation and WeSME methods. Six genetic interactions (five in MB-SHH and one in MB-WNT) were only detected using the permutation method showing the sensitivity of this approach (Fig 3). One genetic interaction in HGG-other was only detected in the WeSME method. On the other hand, two genetic interactions were detected only by the WeSME method in the pan-cancer analysis. The reason that these two pairs were not detected in the permutation method was because the genes involved in those genetic interactions were excluded in the permutation approach to save computational time (Methods). Overall, there is a strong agreement between the two investigated methods as demonstrated by the similarity between obtained *P*-values (Supplementary table1 and table 2). A disadvantage of using the permutation test is the creation of large numbers of permuted matrices which is time consuming. More computation time is needed with increasing numbers of samples and alterations. Indeed, the analysis with WeSME is notably faster [17]. Therefore, using WeSME for mutation matrices containing more samples and alterations might be a better choice, particularly when germline mutations are included.

Recently, there has been considerable interest in identifying synthetic lethality between genes in cancer for therapeutic interventions [37]. One of the reasons for observing mutual exclusivity between two mutated genes is synthetic lethality. To determine whether the mutually exclusive interaction between *PTCH1* and *SMO* is due to synthetic lethality, further validations are required. However, it is likely that mutual exclusivity interaction detected between these two genes is because they function in the same linear pathway and it is not due to synthetic lethality. A mutation in either one of these genes is probably sufficient for activation of the SHH pathway. Thus, a second mutation in the other gene is not necessary for tumor progression.

Taken together, this study presents an initial effort to detect genetic interactions between somatic mutations in pediatric cancer. Our approach provides a framework which can be applied to detect genetic interactions between different types of alterations and larger cohorts. Future investigation is required to detect genetic interactions and unravel the molecular mechanisms underlying them to better understand pediatric cancer and improve treatments.

Methods

Pediatric cancer genome data

A comprehensive cohort including 22 different pediatric cancer types compiled by collaborators from the German Cancer Research Center (DKFZ) was used in this study [3]. The dataset consists of 961 tumors. In order to avoid false positives, 82 relapse samples were excluded, as well as three hypermutator samples from HGG-other cancer type, and 10 single end sequenced samples. The final dataset used to identify genetic interactions includes 866 samples, comprising 538 samples sequenced using whole genome sequencing (WGS) and 328 samples sequenced using whole exome sequencing (WES).

Mutated genes

Mutation data are available and can be downloaded from publicly available data portals such as http://pedpancan.com, but are limited to frequently mutated genes. Through communication with collaborators we were able to gain access to the high confident, but unfiltered results of their variant calling procedure. In brief, in this procedure raw FASTQ files were processed by the standardized alignment and variant calling pipeline developed by and applied in the ICGC Pan-Cancer project (https://github.com/ICGC-TCGA-PanCancer). The human genome assembly hs37d5 (ncbi.nlm.nih.gov/assembly/2758) was used as a reference genome and GENCODE19 (gencodegenes.org/releases/19.html) for gene annotation. Germline SNVs and indels were determined based on their presence in the matched control tissue.

Only somatic variants were used in the current study. All exonic mutations that might have a functional consequence (frameshift, non-frameshift, non-synonymous, stopgain, stoploss) according to ANNOVAR annotation were included. In total, 46,524 SNVs and 2,169 indels remained for downstream analysis. Genes with at least one of the abovementioned mutations were considered as ‘mutated genes’ and were obtained for each sample.

In cases where mutations were annotated to multiple genes, one of these genes was selected using a voting system based on annotation fields derived from Gencode version 19 (v19, used by DKFZ in their variant calling pipeline) and Gencode version 27 (v27, currently the most recent version). In this voting system, the first ranking gene was chosen after sorting the genes on the following properties: Gencode v19 status (KNOWN, NOVEL, PUTATIVE), Gencode v19 type (protein-coding, other), Gencode v27 type (protein-coding, other), total number of exonic alterations for this gene (higher numbers ranking higher), total number of exonic alterations in single genes (so alterations not overlapping other genes) and gene name not containing ‘-’ (which usually indicates a read-through gene).

Permutation approach

To determine co-occurring and mutually exclusive pairs of mutations, a permutation approach was carried out for each mutation matrix. Mutation matrices for each cancer type were constructed that contain the presence or absence of mutated genes for each sample (Fig 2). Genes with only one mutation in a mutation matrix were excluded to reduce the computational time. Permatswap function in vegan R library was used to generate 1,000,000 permuted matrices. In each permutation the margins of the permuted matrix were the same as the real mutation matrix. In other words, the total number of mutated genes per gene and per sample were maintained the same. Permutation was carried out for each cancer type separately to control for any biases in mutation frequencies. Two empirical *P*-values for co-occurrence (P_{co}) and mutual exclusivity (P_{me}) were calculated for each pair of mutated genes. The proportion of permutations in which the observed co-occurrence was equal to or higher than in the real data was taken as *P*-value for co-occurrence (P_{co}). The proportion of permutations in which the observed co-occurrence was equal to or lower than the real data was taken as *P*-value for mutual exclusivity (P_{me}).

To correct for multiple hypothesis testing, standard FDR calculations could not be applied due to a bias toward high *P*-values. Therefore, we used an empirical approach for FDR calculations. We randomly selected 100 matrices from the permuted matrices and performed genetic interaction analysis to generate a null distribution of *P*-values. Next, the generated *P*-values were used as a null distribution for computing an empirical FDR (Fig 2). For each computed *P*-value P^*, the FDR was estimated as follows (if all hypotheses with a *P*-value < P^* would be rejected):

Where is the estimated number of false positives and is the total number of rejected null hypotheses. was estimated from the proportion of gene pairs in the permuted matrices with *P*-value . Finally, a Q value was determined by taking the lowest estimated FDR among all observed *P*-values . Gene pairs were considered as significant if they scored a Q value < 0.2.

WeSME approach

As the matrix permutation approach is rather time consuming, we also applied a faster genetic interaction test and compared the results of both tests to infer their overlap. This test, called WeSME, starts similar to the permutation approach with a gene-sample mutation matrix from which for each gene pair the number of mutual exclusive samples is counted [17]. Instead of permuting this matrix many times to compute a *P*-value, WeSME uses a weighted sampling approach based on the mutation rate of the samples. The method also reduces computation time by restricting the number of resamplings as it starts with a small null distribution and only increases the number of resamplings (to a maximum of N=10,000) for candidate gene pairs, namely those that have a low *P*-value estimated from the initial null distribution.

To infer the false discover rate, a similar empirical FDR approach is used as with the matrix permutation method, by permuting the mutation matrix 300 times and applying the test on the permuted matrices to create a *P*-value null distribution. WeSME places genes in either of two mutation rate bins (with 2% of the samples being mutated as threshold) and compares the observed *P*-values with a *P*-value distribution of gene pairs from similar bins. Since the null *P*-values are split into three distributions representing the combinations of mutation rate bins ([low,low], [low, high] and [high,high]), 300 permutations were performed instead of 100 as was done in the permutation test. As WeSME can use three different *P*-value distributions to infer the FDR, 300 permutations were performed instead of 100. The WeSME python script was downloaded from https://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/index.cgi#wesme and modified to make it suitable for the current study, in particular to run a pan cancer analysis. In contrast to the original script, the *P*-values of all gene-pairs were used for FDR calculations instead of only keeping gene-pairs with *P*-values below 0.1. Furthermore, only gene pairs with an FDR < 0.2 and a *P*-value < 0.1 were considered as significant candidates for further analysis.

Author Contributions

Conceptualization: PK FCPH JD SA
Formal analysis: PK FCPH JD SA
Funding acquisition: PK
Methodology: PK FCPH JD SA
Software: JD SA
Supervision: PK FCPH
Visualization: JD SA
Writing – Original Draft Preparation: JD SA
Writing – review & editing: PK FCPH JD SA

References:

1. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502: 333. doi:10.1038/nature12634
2. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499: 214–218. doi:10.1038/nature12213
3. Gröbner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, et al. The landscape of genomic alterations across childhood cancers. Nature. 2018; doi:10.1038/nature25480
4. Ma X, Liu Y, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. Nature. 2018; doi:10.1038/nature25795
5. Sahni N, Yi S, Zhong Q, Jailkhani N, Charloteaux B, Cusick ME, et al. Edgotype: the link between genotype and phenotype. Curr Opin Genet Dev. 2013;23: 649–657. doi:10.1016/j.gde.2013.11.002
6. Ashworth A, Lord CJ, Reis-Filho JS. Genetic Interactions in Cancer Progression and Treatment. Cell. 2011;145: 30–38. doi:10.1016/j.cell.2011.03.020
7. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016;353: aaf1420. doi:10.1126/science.aaf1420
8. Srivas R, Shen JP, Yang CC, Sun SM, Li J, Gross AM, et al. A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. Mol Cell. 2016;63: 514–525. doi:10.1016/j.molcel.2016.06.022
9. Laufer C, Fischer B, Billmann M, Huber W, Boutros M. Mapping genetic interactions in human cancer cells with RNAi and multiparametric phenotyping. Nat Methods. 2013;10: 427–431. doi:10.1038/nmeth.2436
10. Najm FJ, Strand C, Donovan KF, Hegde M, Sanson KR, Vaimberg EW, et al. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol. 2017;36: 179. doi:10.1038/nbt.4048
11. Boettcher M, Tian R, Blau JA, Markegard E, Wagner RT, Wu D, et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. Nat Biotechnol. 2018;36: 170. doi:10.1038/nbt.4062
12. Dobzhansky T. Genetics of Natural Populations. Xiii. Recombination and Variability in Populations of Drosophila Pseudoobscura. Genetics. 1946;31: 269–290.
13. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. N Engl J Med. 2009;361: 123–134. doi:10.1056/NEJMoa0900212
14. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science. 2017;355: 1152–1158. doi:10.1126/science.aam7344

15. Park S, Lehner B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. Mol Syst Biol. 2015;11. doi:10.15252/msb.20156102
16. Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell. 2014;158: 1199–1209. doi:10.1016/j.cell.2014.07.027
17. Kim Y-A, Madan S, Przytycka TM. WeSME: uncovering mutual exclusivity of cancer drivers and beyond. Bioinformatics. 2017;33: 814–821. doi:10.1093/bioinformatics/btw242
18. Babur Ö, Gönen M, Aksoy BA, Schultz N, Ciriello G, Sander C, et al. Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. Genome Biol. 2015;16: 45. doi:10.1186/s13059-015-0612-6
19. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res. 2012;22: 398–406. doi:10.1101/gr.125567.111
20. Vandin F, Upfal E, Raphael BJ. De novo discovery of mutated driver pathways in cancer. Genome Res. 2012;22: 375–385. doi:10.1101/gr.120477.111
21. Constantinescu S, Szczurek E, Mohammadi P, Rahnenführer J, Beerenwinkel N. TiMEx: a waiting time model for mutually exclusive cancer alterations. Bioinformatics. 2016;32: 968–975. doi:10.1093/bioinformatics/btv400
22. Szczurek E, Beerenwinkel N. Modeling Mutual Exclusivity of Cancer Mutations. PLOS Comput Biol. 2014;10: e1003503. doi:10.1371/journal.pcbi.1003503
23. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. Science. 2013;339: 1546–1558. doi:10.1126/science.1235122
24. Downing JR, Wilson RK, Zhang J, Mardis ER, Pui C-H, Ding L, et al. The Pediatric Cancer Genome Project. Nat Genet. 2012;44: 619–622. doi:10.1038/ng.2287
25. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC, et al. MuSiC: Identifying mutational significance in cancer genomes. Genome Res. 2012;22: 1589–1598. doi:10.1101/gr.134635.111
26. Simpson F, Kerr MC, Wicking C. Trafficking, development and hedgehog. Mech Dev. 2009;126: 279–288. doi:10.1016/j.mod.2009.01.007
27. Reifenberger J, Wolter M, Weber RG, Megahed M, Ruzicka T, Lichter P, et al. Missense Mutations in SMOH in Sporadic Basal Cell Carcinomas of the Skin and Primitive Neuroectodermal Tumors of the Central Nervous System. Cancer Res. 1998;58: 1798–1803.
28. Wolter M, Reifenberger J, Sommer C, Ruzicka T, Reifenberger G. Mutations in the Human Homologue of the Drosophila Segment Polarity Gene patched (PTCH) in Sporadic Basal Cell Carcinomas of the Skin and Primitive Neuroectodermal Tumors of the Central Nervous System. Cancer Res. 1997;57: 2581–2585.
29. Xie J, Johnson RL, Zhang X, Bare JW, Waldman FM, Cogen PH, et al. Mutations of the PATCHED Gene in Several Types of Sporadic Extracutaneous Tumors. Cancer Res. 1997;57: 2369–2372.

30. Oue T, Yoneda A, Uehara S, Yamanaka H, Fukuzawa M. Increased expression of the hedgehog signaling pathway in pediatric solid malignancies. J Pediatr Surg. 2010;45: 387–392. doi:10.1016/j.jpedsurg.2009.10.081
31. Jiang J, Hui C. Hedgehog Signaling in Development and Cancer. Dev Cell. 2008;15: 801–812. doi:10.1016/j.devcel.2008.11.010
32. Hui C, Angers S. Gli Proteins in Development and Disease. Annu Rev Cell Dev Biol. 2011;27: 513–537. doi:10.1146/annurev-cellbio-092910-154048
33. Wicking C, Smyth I, Bale A. The hedgehog signalling pathway in tumorigenesis and development. Oncogene. 1999;18: 7844–7851. doi:10.1038/sj.onc.1203282
34. Castel D, Philippe C, Calmon R, Le Dret L, Truffaux N, Boddaert N, et al. Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes. Acta Neuropathol (Berl). 2015;130: 815–827. doi:10.1007/s00401-015-1478-0
35. Project SJCRH-WUPCG, Wu G, Broniscer A, McEachron TA, Lu C, Paugh BS, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet. 2012;44: 251–253. doi:10.1038/ng.1102
36. Dupain C, Harttrampf AC, Urbinati G, Geoerger B, Massaad-Massade L. Relevance of Fusion Genes in Pediatric Cancers: Toward Precision Medicine. Mol Ther Nucleic Acids. 2017;6: 315–326. doi:10.1016/j.omtn.2017.01.005
37. Cruz FF de la, Gapp BV, Nijman SMB. Synthetic Lethal Vulnerabilities of Cancer. Annu Rev Pharmacol Toxicol. 2015;55: 513–531. doi:10.1146/annurev-pharmtox-010814-124511

S1 Table.
Genetic interactions in cancer type analysis.

Cancer type	Gene 1	Gene 2	co/me	Mutation rate		# Co-occurrence	# Mutual exclusivity		# Samples	P value (WeSME P value)	Q value (WeSME Q value)
				Gene 1	Gene2		Gene 1	Gene 2			
HGG-other	SNCAIP	PHIP	co	3	3	3	0	0	48 (53)	2.77E-03 (5.02E-03)	**0.07 (0.087)**
	PIK3CA	H3F3A	me	7	8	0	7	8		2.42E-02 (3.40E-03)	0.42 **(0.107)**
HGG-K27M	ACVR1	HIST1H3B	co	12	11	9	3	2	54 (55)	4.00E-06 (2.00E-05)	**0.004 (0)**
	ACVR1	TP53	me	12	27	2	10	25		1.07E-02 (3.16E-03)	**0.013 (0.004)**
	H3F3A	HIST1H3B	me	41	11	0	41	11		1.00E-06 (3.16E-03)	**0.001 (0.004)**
	ACVR1	H3F3A	me	12	41	2	10	39		3.00E-06 (0)	**0.002 (0)**
MB-SHH	PTCH1	SMO	me	14	4	0	14	4	36 (40)	7.23E-02 (6.05E-02)	**0.1 (0.065)**
	KMT2D	PTCH1	me	4	14	0	4	14		7.27E-02 (6.05E-02)	**0.1 (0.065)**
	SMO	DDX3X	me	4	11	0	4	11		1.35E-01 (1.18E-01)	**0.113** (0.130)
	KMT2D	DDX3X	me	4	11	0	4	11		1.35E-01 (1.18E-01)	**0.1325** (0.130)
	ZIC1	PTCH1	me	3	14	0	3	14		1.42E-01 (1.37E-01)	**0.136** (0.197)
	PTCH1	TP53	me	14	3	0	14	3		1.42E-01 (1.37E-01)	**0.143** (0.197)
	PTCH1	FAM230A	me	14	3	0	14	3		1.42E-01 (1.37E-01)	**0.143** (0.197)
MB-WNT	KMT2D	DDX3X	co	4	7	3	1	4	21	9.94E-02 (1.86E-01)	**0.198** (>0.5)

gene1: mutated gene 1
gene2: mutated gene 2
co/me: type of genetic interaction, co-occurring or mutually exclusive
mutation rate: the number of samples with mutations in gene1 or gene2
#co-occurrence: the number of samples with mutations in both genes
#mutual exclusivity gene1: the number of samples with mutations in gene1 but not in gene2
#mutual exclusivity gene2: the number of samples with mutations in gene2 but not in gene1
#samples: the total number of samples per cancer type
P value (WeSME P value): the computed P value obtained from permutation method (the computed P value from WeSME method if it does not match with P values from the Permutation method)
Q value (WeSME Q value): the computed Q value obtained from permutation method (the computed Q value from WeSME method if it does not match with Q values from the Permutation method)
Bold Q values represent significant calls

Supplemental table 2.

Genetic interactions in pan-cancer analysis.

Gene 1	Gene 2	co/ me	P value (WeSME P value)	Q value (WeSME Q value)	Cancer type	Sign in cancer type	# Co-occurrence (WeSME)	Mutation rate		# Mutual exclusivity (WeSME)		# samples (WeSME)
								Gene1	Gene2	Gene 1	Gene 2	
ACVR1	HIST1H3B	co	4.00E-06 (3.60E-05)	0 **(0.003)**	HGGother		- (0)	2	1	- (2)	- (1)	48 (53)
					HGG-K27M	x	9	12	11	3	2	54 (55)
COL5A2	FOLH1	co	- (5.00E-04)	- **(0.115)**	HGG-Other		1	1	1	0	0	48 (53)
					MB-Group3		1	1	1	0	0	43 (58)
					MB-Group4		0	1	2	1	2	90 (97)
					NB		1	1	1	0	0	56
ACVR1	H3F3A	me	1.00E-06 (0)	**0 (0)**	HGG-Other		0	2	8	2	8	48 (53)
					HGG-K27M	x	2	12	41	10	39	54 (55)
H3F3A	HIST1H3B	me	1.00E-06 (0)	**0 (0)**	HGG-Other		- (0)	8	1	- (8)	- (1)	48 (53)
					HGG-K27M	x	0	41	11	41	11	54 (55)
ACVR1	TP53	me	1.20E-03 (2.00E-04)	**0.007 (0.001)**	HGG-Other		0	2	22	2	22	48 (53)
					HGG-K27M	x	2	12	27	10	25	54 (55)
H3F3A	PIK3CA	me	2.63E-02 (3.40E-03)	0.305 **(0.103)**	HGG-Other	x	0	8	7	8	7	48 (53)
					HGG-K27M		3	41	5	38	2	54 (55)
HLA-DRB1	TP53	me	4.74E-02 (1.62E-02)	0.377 **(0.189)**	ALL		0	3	9	3	6	48 (67)
					EWS		- (0)	1	1	- (1)	- (1)	14
					HGG-Other		0	2	22	2	22	48 (53)
					MB-SHH		- (0)	1	3	- (1)	- (3)	36 (40)
					MB-Group3		- (0)	3	1	- (3)	- (1)	43 (58)
					OS		- (0)	1	3	- (1)	- (3)	29 (35)

gene1: mutated gene 1
gene2: mutated gene 2
co/me: type of genetic interaction, co-occurring or mutually exclusive

mutation rate: the number of samples with mutations in gene1 or gene2
#co-occurrence: the number of samples with mutations in both genes
#mutual exclusivity gene1: the number of samples with mutations in gene1 but not in gene2
#mutual exclusivity gene2: the number of samples with mutations in gene2 but not in gene1
#samples: the total number of samples per cancer type
P value (WeSME P value): the computed P value obtained from permutation method (the computed P value from WeSME method if it does not match with P values from the Permutation method)
Q value (WeSME Q value): the computed Q value obtained from permutation method (the computed Q value from WeSME method if it does not match with Q values from the Permutation method)
Sign in cancer type: x depicts whether the detected genetic interaction was also significant in cancer type specific analysis

General discussion

The work described in this thesis focuses mainly on investigating the mechanisms underlying genetic interactions in the yeast *Saccharomyces cerevisiae*. Large-scale studies have revealed that genetic interactions are pervasive, but often the molecular mechanisms underlying them are not well known. In **chapter 2**, the relative contribution of redundancy and condition dependency to non-responsiveness upon genetic perturbations has been investigated. The analyses show that only a small portion of non-responsiveness is due to redundancy. **Chapter 3** investigates the nature of genetic interactions between gene-specific transcription factors (GSTFs) in yeast. The result is a high-resolution atlas of genetic interactions for 72 GSTF pairs. Systematic analysis of double and single mutants confirms known genetic interactions and exposes new ones. In **chapter 4**, a comparison of genetic interactions between two different functional classes of proteins (GSTFs and kinases/phosphatases) was conducted. Computational modelling using Petri nets was applied to investigate the differences. Exhaustive enumeration of more than 9 million models led to a potential regulatory mechanism for inversion, a genetic interaction pattern that is more prevalent in GSTFs compared to kinases/phosphatases. Additionally, in **chapter 5**, a preliminary effort to detect genetic interactions by applying statistical analysis has been presented. Two different methods were applied to detect genetic interactions using DNA sequencing data from pediatric tumors. The result is an initial map of genetic interactions between somatic alterations in pediatric cancers. Detailed implications of these findings are discussed in those chapters. The remainder of this chapter focuses on the future outlook and directions toward the completion of our knowledge about genetic interactions.

Genetic interactions in yeast

Genetic interactions occur when a combination of mutations give rise to an unexpected phenotype based on the respective phenotypes of single mutants [1]. Negative genetic interactions occur when the combined effect of two mutations is more severe than expected. Positive genetic interactions occur when the combined effect of two mutations is less severe than expected. Systematic investigation of genetic interactions in model organisms such as budding and fission yeast, *Caenorhabditis elegans* and *Drosophila melanogaster* cells have revealed that genetic interactions are widespread [2–7]. To date, a global map of all pair-wise genetic interactions has been completed under a single laboratory condition in budding yeast [3]. Such a comprehensive network of genetic interactions is beneficial for investigating the functional organization of a cell and is useful for predicting gene and pathway function. Despite the tremendous success in detecting genetic interactions in yeast, multiple steps can still be envisioned for the future. The next steps should make use of available data and further develop experimental and computational approaches to expand the search space for detecting genetic interactions as well as understanding them at the molecular level. This can be achieved by three different strategies: screening for genetic interactions in multiple laboratory conditions, screening for higher order genetic interactions and using more sensitive/detailed phenotypes.

Genetic interaction assays under multiple conditions

One of the strategies to understand the relation between genotype and phenotype is to investigate the effect of gene deletions. This effect can be measured using growth defect or global gene expression changes. However, the majority of individual gene deletions in large-scale assays do not show a fitness defect or any gene expression changes [8–10]. As investigated in chapter 2, redundancy and condition dependency are the main explanations for this non-responsive behaviour. Negative genetic interactions are frequently associated with redundancy [11,12]. However, many negative genetic interactions are not due to a simple redundancy relationship [13]. Furthermore, it is anticipated that genetic interactions are highly variable across different conditions. For one case, this was confirmed by doubling the number of detected genetic interactions when yeast cells were treated with DNA damaging agent [14]. Two redundant genes that have a high sequence similarity are expected to compensate for each other's loss regardless of the condition they are investigated in. Most large-scale genetic interaction assays have been conducted under only a single laboratory condition [2,3]. Therefore, drawing conclusions about the redundancy contribution to non-responsiveness based on negative genetic interactions in one condition may be unrealistic. To fully estimate the contribution of redundancy to non-responsiveness, large-scale genetic interaction studies should also be conducted under multiple laboratory conditions. This will help to identify gene pairs that show consistent redundancy across many laboratory conditions. It will also expose the degree of plasticity of genetic interaction networks across different conditions [15].

The scale of condition-specific genetic interactions may be enormous, therefore an efficient strategy in combination with current techniques for detecting genetic interactions is required. Due to the feasibility of large-scale genetic screens in yeast, screening for condition-specific genetic interactions is not out of reach. Single mutations have been already exposed to more than 400 conditions in yeast [16]. The combination of such technologies and genetic interaction screens can make it feasible to detect genetic interactions under multiple conditions. A smaller scale strategy would be to investigate only single mutants that show condition specific fitness defects.

Higher order mutations

Interestingly, the non-responsiveness phenomenon is prevalent in double mutant studies as well. The majority of yeast double mutants do not show any fitness defect [2,3]. The same reasoning of redundancy and condition dependency could apply here as well. As mentioned above, one of the strategies to determine to what extent redundancy plays a role in non-responsiveness is to investigate genetic interactions under multiple conditions. However, redundancy could occur on a much higher scale for example, at the level of three or more genes. In this situation, non-responsiveness would not be uncovered using traditional double mutant analyses. A method known as triple mutant analysis (TMA) has

been developed for systematic generation and quantification of triple mutants [17]. Such a method can be applied for detecting genetic interactions between triplets on a much larger scale.

Detecting higher order genetic interactions on a large scale will be a challenging task due to the vast numbers of possible mutations. One of the requirements for double mutant assays is the availability of a library of single deletion mutants [18]. A first step in generating triplet mutants would be the availability of a library of double mutants which can be screened against single mutants. Maintaining a large library consisting of all possible pair-wise mutants is likely an enormous challenge. However, the size of such a library can be reduced to a scale that is more convenient. The analysis of pair-wise genetic interactions in yeast shows that duplicated genes tend to show fewer genetic interactions with the rest of the genome compared to non-duplicated genes [2,19]. Therefore, generating a library of double mutants for only duplicated genes might be a good start.

More sensitive and detailed readouts

One of the most frequently used phenotypes in genetic interaction studies is growth. This phenotype is great for generating and quantifying fitness of double and single mutants on a large scale. Besides growth, there are several other phenotypic readouts that can be used to detect genetic interactions. Employing these phenotypes can provide more sensitive or detailed data. A more sensitive phenotype than growth is morphological changes in the cell. For example, a study of spindle morphogenesis revealed four times more genetic interactions than the equivalent growth-based study [20]. Similar results were obtained from studying genetic interactions using multiple phenotypes obtained from automated imaging in cultured *Drosophila* cells [21]. More sensitive phenotypes can detect interesting genetic interactions which have been missed out by growth-based assays.

Another more detailed phenotype to study genetic interaction is genome-wide gene expression changes. This phenotype proves to be extremely useful in determining how different processes are connected [22–24]. The advantage of using this phenotype is that it is highly sensitive as well as detailed. In this thesis genome-wide gene expression changes were applied to understand genetic interactions at the molecular level. In chapter 3 and 4, six frequently occurring genetic interactions patterns where identified including buffering, quantitative buffering, suppression, quantitative suppression, masking and inversion. This highlights the amount of details that genome-wide gene expression changes can offer for every gene in the genome. Subsequently, the obtained detailed information can be used for understanding the molecular mechanisms of genetic interactions. However, the selection criterion for choosing pairs to study using gene expression changes was based on having a significant growth-based genetic interaction score. This probably means that some interesting pairs were not investigated as the approach relied on a less sensitive phenotype for selecting the pairs.

Conducting large-scale studies using more sensitive phenotypes is becoming feasible as the time and costs of high-throughput microscopy and sequencing are declining. However, smaller scale genetic interaction screens using more sensitive phenotypes such as gene expression changes can be also conducted to understand certain biological processes. For example, protein folding in the endoplasmic reticulum is a complex process which can be compromised by both external stress and mutations [25]. To fully understand how different components of such a complex biological process work together, a highly sensitive/detailed phenotype and condition-specific genetic interaction screen can be beneficial.

Modelling approaches to understand genetic interactions

Applying more sensitive and detailed phenotypes to study genetic interactions will result in high-resolution data. As mentioned above, using gene expression changes six different commonly occurring genetic interaction patterns were identified. In chapter 4, Petri net modelling was applied to enumerate more than 9 million models. The main focus in that chapter was to determine the mechanisms underlying inversion, a genetic interaction pattern more dominant in GSTFs compared to kinases/phosphatases. However, the devised modelling technique can also give more insight into other genetic interaction patterns. For example, the suppression pattern which is more associated with positive genetic interaction scores (chapter 3). Investigating the topology and characteristics of models showing suppression could reveal new mechanisms.

The investigated models in chapter 4 only consisted of three and four nodes in which two of them were deleted to simulate single and double mutants. Therefore, only two downstream genes were investigated regarding genetic interaction patterns. On the other hand, some of the investigated pairs showed all six genetic interactions patterns simultaneously. To run model simulations that can explain the entire genetic interaction profile of such a pair, at least 6 different downstream nodes are needed. Exploring all possible models in larger than four node networks by computational modelling will be extremely challenging due to a daunting number of possibilities. However, by integrating known interactions of signalling components described in literature and databases the number of models to be enumerated can be significantly decreased [26].

Genetic interactions in cancer

In the last few decades many genetic changes have been discovered that can cause cancer to develop, grow, and spread [27]. Many genomes can now be sequenced in a single day which may trigger a significant shift in cancer treatment. The access to large numbers of tumor sequences has made it possible to detect co-occurring and mutually exclusive alterations [28–30]. As described in chapter 5, co-occurrence is defined by two mutations occurring together in the same tumor more often than expected. And mutual exclusivity is defined by two mutations occurring together less often than expected. One of the

reasons for observing mutual exclusivity between two alterations is synthetic lethality. Synthetic lethality is a type of negative genetic interactions in which the combination of two perturbations results in cell death, but individual perturbation of single genes does not affect cellular fitness [31]. Synthetic lethality shows great promise to find new therapeutic strategies to target cancer cells that are harboring loss-of-function mutations [32,33]. A number of synthetic lethal interactions have been detected (reviewed by Beijersbergen et al. [34]), the most clinically successful of which is the interaction between *BRCA* and *PARP* [32]. Cells that harbor loss-of-function mutations in *BRCA1* or *BRCA2* lead to a partial disruption of repairing double-stranded DNA breaks which makes them highly dependent on the function of PARP involved in DNA repair [35]. Inhibition of PARP1 by olaparib, a drug approved by the Food and Drug Administration (FDA), causes cell death in cancer cells that harbor mutations in *BRCA1* and *BRCA2* [36]. This therapeutic strategy is particularly interesting because it only targets cancer cells, unlike many other therapies for cancer [37]. More drugs that are thought to induce synthetic lethality are in clinical trial and the hunt for more continues [38].

There is therefore considerable interest in identifying synthetic lethal relationships, with the intention of finding new cancer drugs [34,37,39]. However, cooperative interactions which are reflected in the number of co-occurrence of two alterations may also be important in tumorigenesis. Cooperative interaction between *MYC* and *RAS* genes, has been long recognized [40]. The combination of mutations in *MYC* and *RAS* are able to form tumors in mice, whereas individually mutated genes are not sufficient for tumor initiation. Large numbers of co-occurrence between cancer driver alterations have been reported across many adult cancers [30]. To what extent co-occurring alterations have an effect on promoting tumor initiation and progression still remains to be discovered. Meanwhile, multiple complementary approaches have been initiated to systematically map genetic interactions in cancer cells.

Generating a map of genetic interactions in cancer cells

Efforts to detect genetic interactions between genes in different human cancer cells are currently underway. These efforts can be divided into three complementary approaches. First, exploiting genetic interactions conserved from yeast to human. It has been shown that genetic interactions can be conserved in different yeast species over long evolutionary time periods [4,41,42]. A comparison of genetic interaction networks between budding and fission yeast revealed a high level of conservation within protein complexes, but variable between them [4]. Plasticity of genetic interaction networks might be a widespread phenomenon which can explain condition-specific genetic interactions as well. This plasticity is also revealed in cancer-type dependent genetic interactions between driver alterations in different cell types [30].

Multiple examples of conserved genetic interactions from yeast to human cells have been reported [43–45]. Interestingly, most of these interactions are within fundamental cellular processes such as DNA synthesis and repair, chromosome segregation and cell cycle regulation [46–48]. Such interactions are also more likely to be the same between different cell types which can be used to find more generally applicable therapeutic strategies. The overall degree and scale of network conservation remains largely uncharted. Therefore, a map of genetic interactions can be obtained by integrating potentially conserved genetic interactions from model organisms to human cells. This will not lead to a complete picture of genetic interactions, but provides an initial starting point and can be complemented with other approaches.

A second strategy for detecting genetic interactions is analysing co-occurring and mutually exclusive relationships between genetic alterations detected through DNA sequencing of individual tumors. In chapter 5, we have provided a framework to detect genetic interactions between somatic alterations in pediatric cancers using a large cohort of tumors. Similar approaches have been applied in adult cancers to detect co-occurring and mutually exclusive alterations [28,30,49]. In adult cancers more interactions are detected for cancer types in which the number of samples is higher [30]. Since the occurrence of pediatric cancers is much lower than that of adult cancers, interaction detection in pediatric cancer suffers from a lower number of samples, in particular for those pediatric cancer types that are very rare. As demonstrated in chapter 5, the number of detected genetic interactions was limited in this initial study. One of the reasons could be the low number of samples to detect any meaningful genetic interactions. Therefore, an international collaboration to collect more samples will be constructive to boost the statistical power.

So far, most studies have focused on using single nucleotide variations, small insertions and deletions that result in loss of function. Including more types of genetic alterations could provide more information and therefore the possibility of detecting more genetic interactions. Such alterations could involve copy number variations (CNV), structural variants (SV), and DNA methylations events. Additionally, many germline mutations in cancer-predisposing genes have been detected in pediatric cancer [50]. Including these germline mutations in genetic interaction analysis can lead to detecting interesting interactions as well [51]. However, including too many alterations with too few samples may not be beneficial for detecting interactions due to the high-dimensionality curse. Aggregating information at a higher level such as pathways will potentially help in dealing with the high-dimensionality problem [52]. This will result in far fewer tests and detection of functional interactions between pathways.

A third approach for mapping genetic interactions is to use combinatorial disruption of genes in cancer cells. The complete genetic interaction network in yeast has been built by evaluating the growth defect of all possible pair-wise combinations of genes [3]. However, in human cells this can be quite challenging. The number of possible pair-wise combinations

of double mutants is 200,000,000 considering approximately 20,000 human protein-coding genes. Taking into account different cell types will make these types of screens in human cells even more challenging. The current methods to target gene pairs in mammalian cells suffer from low accuracy and efficiency. RNA interference (RNAi) based large-scale screens have been applied to identify synthetic lethal interactions by combinatorial disruptions in mammalian cells [53]. RNAi has also been used in combination with drug treatments (reviewed by Chan & Giaccia [33]). However, RNAi-based technologies have some drawbacks. First, the gene inhibition is never complete. Second, off-target effects due to binding to other non-intended mRNAs introduce noise that is difficult to distinguish from the on-target signal [38]. Nevertheless, astonishing advances in CRISPR technologies can also simultaneously perturb two genes, thus representing a promising outlook in screening for genetic interactions in human cells. For example, by knocking out two genes in the same cell Han et al. succeeded to screen all possible pair-wise combinations of 207 genes in K562 leukemia cells [54]. When using two guide RNAs (gRNAs) in one cell, one problem is recombination of two homologous sequences present in vectors carrying gRNAs which results in reduced editing efficiency [55]. Recently, two independent studies have found a solution to solve this problem by using orthogonal Cas9 nucleases from *S. pyogenes* (spCas9) and *S. aureus* (saCas9) [56,57]. Continued improvements in new genome editing technologies will accelerate the discovery of genetic interactions in cancer cells. With the current rate of technological advances, screening all possible pair-wise genetic interactions in human cells may not be such a wild idea.

From yeast to cancer

For more than 5000 years humans have used the power of yeast to brew beer and leaven bread. In the last few decades, yeast has also become a prime eukaryotic model organism to study cell biology. Highly collaborative large-scale studies in yeast have led to entirely new fields of 'functional genomics' and 'systems biology' [58]. These fields try to find how genes and proteins interact and work together to deliver a certain phenotype. It is now possible to look beyond the function of single genes and proteins. To comprehend how a cell works as a system all the components should be first understood. One of these components is genetic interaction networks which have been studied extensively in yeast. In 2001, Hartman et al. suggested that genetic interactions may play a crucial role in our ability to understand the relationship between genotype and phenotype [59]. Since then, yeast has been at the forefront of studying genetic interactions. A global genetic interaction map of yeast is now available [3]. However, as mentioned above we still lack comprehensive knowledge about genetic redundancy, plasticity of genetic interaction networks in different conditions and higher order genetic interactions. More importantly, we still lack the ability to understand genetic interactions at the molecular level. In chapter 3 and 4 we have shown that studying genetic interactions using a more sensitive and detailed phenotype such as genome-wide

gene expression changes combined with computational modelling can be beneficial for unravelling the molecular mechanisms of genetic interactions.

Pioneered by yeast research, efforts have also been initiated to screen for pairwise genetic interactions in human cancer cells [37,53,54,57]. Applying experimental and computational advances in yeast to the cancer field will be instrumental in understanding cancer cells. As concluded from the work described in this thesis, more thorough investigations using sensitive and detailed phenotypes are required to understand genetic interactions in human cells. Recent genome editing approaches such as CRISPR coupled with single cell RNA sequencing can provide such a rich genomic readout [60]. This high-resolution readout combined with computational modelling approaches can provide more insight into the molecular mechanisms of genetic interactions and eventually a better understanding of genotype-phenotype relationships.

References

1. Phillips PC. Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat Rev Genet. 2008;9: nrg2452. doi:10.1038/nrg2452
2. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, et al. The Genetic Landscape of a Cell. Science. 2010;327: 425–431. doi:10.1126/science.1180823
3. Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016;353: aaf1420. doi:10.1126/science.aaf1420
4. Roguev A, Bandyopadhyay S, Zofall M, Zhang K, Fischer T, Collins SR, et al. Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science. 2008;322: 405–410. doi:10.1126/science.1162609
5. Byrne AB, Weirauch MT, Wong V, Koeva M, Dixon SJ, Stuart JM, et al. A global analysis of genetic interactions in Caenorhabditis elegans. J Biol. 2007;6: 8. doi:10.1186/jbiol58
6. Lehner B, Crombie C, Tischler J, Fortunato A, Fraser AG. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. Nat Genet. 2006;38: 896. doi:10.1038/ng1844
7. Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, Boutros M. Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. Nat Methods. 2011;8: 341. doi:10.1038/nmeth.1581
8. Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, et al. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. Science. 1999;285: 901–906.
9. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, et al. Functional profiling of the Saccharomyces cerevisiae genome. Nature. 2002;418: 387–391. doi:10.1038/nature00935
10. Kemmeren P, Sameith K, van de Pasch LAL, Benschop JJ, Lenstra TL, Margaritis T, et al. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. Cell. 2014;157: 740–752. doi:10.1016/j.cell.2014.02.054
11. Plata G, Vitkup D. Genetic robustness and functional evolution of gene duplicates. Nucleic Acids Res. 2013; gkt1200. doi:10.1093/nar/gkt1200
12. Wagner A. Robustness against mutations in genetic networks of yeast. Nat Genet. 2000;24: 355–361. doi:10.1038/74174
13. Tischler J, Lehner B, Fraser AG. Evolutionary plasticity of genetic interaction networks. Nat Genet. 2008;40: 390–391. doi:10.1038/ng.114
14. St Onge RP, Mani R, Oh J, Proctor M, Fung E, Davis RW, et al. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. Nat Genet. 2007;39: 199–206. doi:10.1038/ng1948
15. Ideker T, Krogan NJ. Differential network biology. Mol Syst Biol. 2012;8: 565. doi:10.1038/msb.2011.99

16. Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, et al. The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. Science. 2008;320: 362–365. doi:10.1126/science.1150021
17. Haber JE, Braberg H, Wu Q, Alexander R, Haase J, Ryan C, et al. Systematic Triple Mutant Analysis Uncovers Functional Connectivity Between Pathways Involved in Chromosome Regulation. Cell Rep. 2013;3: 2168–2178. doi:10.1016/j.celrep.2013.05.007
18. Giaever G, Nislow C. The Yeast Deletion Collection: A Decade of Functional Genomics. Genetics. 2014;197: 451–465. doi:10.1534/genetics.114.161620
19. VanderSluis B, Bellay J, Musso G, Costanzo M, Papp B, Vizeacoumar FJ, et al. Genetic interactions reveal the evolutionary trajectories of duplicate genes. Mol Syst Biol. 2010;6: 429. doi:10.1038/msb.2010.82
20. Vizeacoumar FJ, van Dyk N, S.Vizeacoumar F, Cheung V, Li J, Sydorskyy Y, et al. Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. J Cell Biol. 2010;188: 69–81. doi:10.1083/jcb.200909013
21. Fischer B, Sandmann T, Horn T, Billmann M, Chaudhary V, Huber W, et al. A map of directional genetic interactions in a metazoan cell. eLife. 2015;4: e05464. doi:10.7554/eLife.05464
22. van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, de Castro IJ, et al. Functional overlap and regulatory links shape genetic interactions between signaling pathways. Cell. 2010;143: 991–1004. doi:10.1016/j.cell.2010.11.021
23. Gutin J, Sadeh A, Rahat A, Aharoni A, Friedman N. Condition-specific genetic interaction maps reveal crosstalk between the cAMP/PKA and the HOG MAPK pathways in the activation of the general stress response. Mol Syst Biol. 2015;11. doi:10.15252/msb.20156451
24. Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, et al. Structure and function of a transcriptional network activated by the MAPK Hog1. Nat Genet. 2008;40: 1300. doi:10.1038/ng.235
25. Lin JH, Walter P, Yen TSB. Endoplasmic reticulum stress in disease pathogenesis. Annu Rev Pathol. 2008;3: 399–425. doi:10.1146/annurev.pathmechdis.3.121806.151434
26. Jacobsen A, Heijmans N, Verkaar F, Smit MJ, Heringa J, Amerongen R van, et al. Construction and Experimental Validation of a Petri Net Model of Wnt/β-Catenin Signaling. PLOS ONE. 2016;11: e0155743. doi:10.1371/journal.pone.0155743
27. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502: 333. doi:10.1038/nature12634
28. Kim Y-A, Madan S, Przytycka TM. WeSME: uncovering mutual exclusivity of cancer drivers and beyond. Bioinformatics. 2017;33: 814–821. doi:10.1093/bioinformatics/btw242
29. Leiserson MDM, Reyna MA, Raphael BJ. A weighted exact test for mutually exclusive mutations in cancer. Bioinformatics. 2016;32: i736–i745. doi:10.1093/bioinformatics/btw462
30. Park S, Lehner B. Cancer type-dependent genetic interactions between cancer driver alterations indicate plasticity of epistasis across cell types. Mol Syst Biol. 2015;11. doi:10.15252/msb.20156102

31. Dobzhansky T. Genetics of Natural Populations. Xiii. Recombination and Variability in Populations of Drosophila Pseudoobscura. Genetics. 1946;31: 269–290.
32. Brunen D, Bernards R. Drug therapy: Exploiting synthetic lethality to improve cancer therapy. Nat Rev Clin Oncol. 2017;14: 331–332. doi:10.1038/nrclinonc.2017.46
33. Chan DA, Giaccia AJ. Harnessing synthetic lethal interactions in anticancer drug discovery. Nat Rev Drug Discov. 2011;10: 351. doi:10.1038/nrd3374
34. Beijersbergen RL, Wessels LFA, Bernards R. Synthetic Lethality in Cancer Therapeutics. Annu Rev Cancer Biol. 2017;1: 141–161. doi:10.1146/annurev-cancerbio-042016-073434
35. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers. N Engl J Med. 2009;361: 123–134. doi:10.1056/NEJMoa0900212
36. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science. 2017;355: 1152–1158. doi:10.1126/science.aam7344
37. Nijman SMB. Synthetic lethality: general principles, utility and detection using genetic screens in human cells. FEBS Lett. 2011;585: 1–6. doi:10.1016/j.febslet.2010.11.024
38. Cruz FF de la, Gapp BV, Nijman SMB. Synthetic Lethal Vulnerabilities of Cancer. Annu Rev Pharmacol Toxicol. 2015;55: 513–531. doi:10.1146/annurev-pharmtox-010814-124511
39. O'Neil NJ, Bailey ML, Hieter P. Synthetic lethality and cancer. Nat Rev Genet. 2017;18: 613. doi:10.1038/nrg.2017.47
40. Land H, Parada LF, Weinberg RA. Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature. 1983;304: 596–602.
41. Dixon SJ, Fedyshyn Y, Koh JLY, Prasad TSK, Chahwan C, Chua G, et al. Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes. Proc Natl Acad Sci U S A. 2008;105: 16653–16658. doi:10.1073/pnas.0806261105
42. Frost A, Elgort MG, Brandman O, Ives C, Collins SR, Miller-Vedam L, et al. Functional Repurposing Revealed by Comparing S. pombe and S. cerevisiae Genetic Interactions. Cell. 2012;149: 1339–1352. doi:10.1016/j.cell.2012.04.028
43. Deshpande R, Asiedu MK, Klebig M, Sutor S, Kuzmin E, Nelson J, et al. A comparative genomic approach for identifying synthetic lethal interactions in human cancer. Cancer Res. 2013;73: 6128–6136. doi:10.1158/0008-5472.CAN-12-3956
44. McManus KJ, Barrett IJ, Nouhi Y, Hieter P. Specific synthetic lethal killing of RAD54B-deficient human colorectal cancer cells by FEN1 silencing. Proc Natl Acad Sci U S A. 2009;106: 3276–3281. doi:10.1073/pnas.0813414106
45. Srivas R, Shen JP, Yang CC, Sun SM, Li J, Gross AM, et al. A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. Mol Cell. 2016;63: 514–525. doi:10.1016/j.molcel.2016.06.022
46. Hartwell LH, Szankasi P, Roberts CJ, Murray AW, Friend SH. Integrating Genetic Approaches into the Discovery of Anticancer Drugs. Science. 1997;278: 1064–1068. doi:10.1126/science.278.5340.1064

47. Pel DM van, Barrett IJ, Shimizu Y, Sajesh BV, Guppy BJ, Pfeifer T, et al. An Evolutionarily Conserved Synthetic Lethal Interaction Network Identifies FEN1 as a Broad-Spectrum Target for Anticancer Therapeutic Development. PLOS Genet. 2013;9: e1003254. doi:10.1371/journal.pgen.1003254
48. Reid RJD, Du X, Sunjevaric I, Rayannavar V, Dittmar J, Bryant E, et al. A Synthetic Dosage Lethal Genetic Interaction Between CKS1B and PLK1 Is Conserved in Yeast and Human Cancer Cells. Genetics. 2016;204: 807–819. doi:10.1534/genetics.116.190231
49. Thomas RK, Baker AC, DeBiasi RM, Winckler W, LaFramboise T, Lin WM, et al. High-throughput oncogene mutation profiling in human cancer. Nat Genet. 2007;39: 347. doi:10.1038/ng1975
50. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. N Engl J Med. 2015;373: 2336–2346. doi:10.1056/NEJMoa1508054
51. Carter H, Marty R, Hofree M, Gross AM, Jensen J, Fisch KM, et al. Interaction Landscape of Inherited Polymorphisms with Somatic Events in Cancer. Cancer Discov. 2017;7: 410–423. doi:10.1158/2159-8290.CD-16-1045
52. Fang G, Wang W, Paunic V, Heydari H, Costanzo M, Liu X, et al. Discovering genetic interactions bridging pathways in genome-wide association studies. bioRxiv. 2017; 182741. doi:10.1101/182741
53. Roguev A, Talbot D, Negri GL, Shales M, Cagney G, Bandyopadhyay S, et al. Quantitative genetic-interaction mapping in mammalian cells. Nat Methods. 2013;10: 432. doi:10.1038/nmeth.2398
54. Han K, Jeng EE, Hess GT, Morgens DW, Li A, Bassik MC. Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. Nat Biotechnol. 2017;35: 463. doi:10.1038/nbt.3834
55. Zhu S, Li W, Liu J, Chen C-H, Liao Q, Xu P, et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. Nat Biotechnol. 2016;34: 1279. doi:10.1038/nbt.3715
56. Najm FJ, Strand C, Donovan KF, Hegde M, Sanson KR, Vaimberg EW, et al. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. Nat Biotechnol. 2017;36: 179. doi:10.1038/nbt.4048
57. Boettcher M, Tian R, Blau JA, Markegard E, Wagner RT, Wu D, et al. Dual gene activation and knockout screen reveals directional dependencies in genetic networks. Nat Biotechnol. 2018;36: 170. doi:10.1038/nbt.4062
58. Botstein D, Fink GR. Yeast: An Experimental Organism for 21st Century Biology. Genetics. 2011;189: 695–704. doi:10.1534/genetics.111.130765
59. Hartman JL, Garvik B, Hartwell L. Principles for the Buffering of Genetic Variation. Science. 2001;291: 1001–1004. doi:10.1126/science.1056072
60. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell. 2016;167: 1853-1866.e17. doi:10.1016/j.cell.2016.11.038

Appendices

Summary in English

All living cells store their genetic information in the same form of double-stranded molecules known as DNA. The individual blocks that make up DNA are called nucleotides. These nucleotides have nicknames indicated by A, T, C, G. Through a process called transcription, information in the DNA is converted into small RNA messages. These RNA molecules travel from the nucleus to the cytoplasm where they are translated to specific proteins by ribosomes. A gene is a sequence of DNA which codes for a protein with a specific function in the cell. Changes in the sequences of genes have profound influence on health. Many diseases, including cancer, have been linked to mutations in specific genes. The outcome of mutations in genes can be linked to certain diseases. Usually these diseases are caused by mutations in only one or two genes. On the other hand, most complex diseases such as many cancers are caused by mutations in many genes. Predicting the outcome of these mutated genes is a challenging task in biology. One of the hurdles to predict the phenotype of an individual based on their genotype is the presence of genetic interactions, also known as epistasis. A genetic interaction is a phenomenon in which the effect of the mutation of one gene depends on the presence of other mutations. The general goal of the work described in this thesis is aimed at understanding the mechanisms underlying genetic interactions, answering the question how do genetic interactions arise?

The simplest example of a genetic interaction is a redundancy relationship between two genes. Redundancy occurs when one gene can take over the function of another gene, for example if they both code for highly similar proteins. In the case of a complete redundancy relationship, inactivation of either gene on its own will result in no detectable defect. Only simultaneous deletion of both genes will result in obvious defects. Examples of man-made redundant components are all around us from bridges and skyscrapers to computer systems, power plants and aircrafts. In engineering, designing a reliable system often requires redundant components. The purpose of redundancy in man-made structures is to prevent dangerous situations. The result of million years of evolution in biology is a complex genetic interaction network in which redundancy happens to be a byproduct it. The complexity of genetic interaction networks comes with other unintuitive genetic interaction types which are described extensively in **chapter 1**.

The role of a gene is often investigated by assessing functional consequences of their removal from the cell. Depending on the sensitivity of the assay used for assessing the effect of gene removal, between 66% and 53% of yeast gene deletions show no detectable defect when analyzed under a single condition. It is known that this non-responsive behavior is caused by redundancy or condition dependency. Understanding the underlying causes of redundancy and their relative contribution to non-responsive behavior upon genetic perturbation is extremely important for designing efficient strategies aimed at elucidating gene function and unraveling complex cellular systems. In **chapter 2**, we provide a

systematic classification of the underlying causes of and their relative contribution to non-responsive behavior upon gene deletion. The overall contribution of redundancy to non-responsive behavior is estimated to be 29%. The major determinant of non-responsiveness is condition dependency (71%).

Cell growth has frequently been used to study genetic interactions on a large scale. Genetic interactions are measured by the extent to which growth defects of double deletion mutants deviate from their expected value based on the growth defects of the single mutants. Genetic interactions scored by the difference between observed and expected fitness can broadly be classified into two groups: negative and positive genetic interactions. A genetic interaction is negative if the fitness observed for a double mutant is worse than expected based on the fitness of the respective single mutants. An interaction is positive if the observed fitness is better than expected. Depending on the state or fate of a cell, different genes are expressed at different levels. This is in part mediated by gene-specific transcription factors (GSTFs). Understanding the basis of genetic interactions between GSTFs is therefore important for understanding the transcription regulatory network. **Chapter 3,** investigates genetic interactions between GSTFs. The expression levels of all genes in the yeast genome was monitored for 72 GSTF double deletion mutants and their corresponding single mutants. This high-resolution expression atlas provides a systems-level overview of the genetic interaction landscape between GSTFs and reveals underlying mechanistic details. Besides revealing new redundancy relationships, this study also proposes two new molecular mechanisms of genetic interactions. These mechanisms, which we term "buffering by induced dependency" and "alleviation by derepression", provide explanations for negative and positive genetic interactions that were previously not understood.

Several molecular mechanisms have been proposed for different types of genetic interactions. However, differences in occurrence and underlying molecular mechanisms of genetic interactions have not yet been compared between different functional gene classes. In **chapter 4**, we compared genetic interaction patterns identified using gene expression profiling for two classes of genes: gene specific transcription factors and signalling related genes. We employed exhaustive modelling to unravel putative molecular differences underlying different genetic interaction patterns. Our study proposes a new mechanistic explanation for a certain genetic interaction pattern that is more associated with gene specific transcription factors compared to kinases/phosphatases. Overall, our findings and the methodologies implemented can be valuable for understanding the molecular mechanisms underlying genetic interactions.

Cancers are thought to arise from the combined effects of alterations in individual genes and such combinatorial effects can be viewed as a type of genetic interaction. Applying statistical analyses in adult cancers led to identification of co-occurring and mutually exclusive pairs of mutated genes. However, little is known about genetic interactions in

pediatric cancer. The detection of DNA alterations in large numbers of pediatric tumors has only recently begun. **Chapter 5** describes a framework for detecting genetic interactions in pediatric cancer. Two methods are applied to detect co-occurring and mutually exclusive pairs of genes in 22 different pediatric cancer types. The two genetic interaction analyses are performed both for individual cancer types and across all data. A map of genetic interactions is presented by combining the results obtained from the four approaches. The results show a limited number of genetic interactions, in part probably due to the low number of samples. The results nevertheless also contain mutually exclusive mutations that have previously been reported, confirming the utility of the approach. Our framework therefore provides a basis for future detection of genetic interactions between different types of mutations and in larger datasets.

Finally, in **chapter 6** I present an in-depth discussion and future perspective of the genetic interaction field in both yeast and cancer.

Summary in Dutch

Alle levende cellen slaan hun genetische informatie op dezelfde manier op, in de vorm van dubbelstrengs moleculen die bekend staan als DNA. De individuele bouwstenen waaruit DNA bestaat worden nucleotiden genoemd. Deze nucleotiden hebben bijnamen die worden aangeduid met A, T, C, G. Via een proces dat 'transcriptie' wordt genoemd, wordt informatie in het DNA omgezet in kleine RNA-berichten. Deze RNA moleculen reizen van de kern naar het cytoplasma waar ze door ribosomen worden omgezet naar specifieke eiwitten. Een gen is een DNA-sequentie die codeert voor een eiwit met een specifieke functie in de cel. Veranderingen in de sequenties van genen hebben een diepgaande invloed op de gezondheid. Veel ziekten, waaronder kanker, zijn in verband gebracht met mutaties in specifieke genen. Het resultaat van mutaties in genen kan worden gekoppeld aan bepaalde ziekten. Gewoonlijk worden deze ziekten veroorzaakt door mutaties in slechts één of twee genen. Aan de andere kant worden de meeste complexe ziektes zoals de meeste kankers veroorzaakt door mutaties in meerdere genen. Het voorspellen van de uitkomst van deze gemuteerde genen is een uitdagende taak in de biologie. Een van de hindernissen om het fenotype van een individu op basis van het genotype te voorspellen, is genetische interactie, ook wel epistasie genoemd. Genetische interactie is een verschijnsel waarbij het effect van de mutatie van één gen afhangt van de aanwezigheid van andere gemuteerde genen. Het algemene doel van het werk beschreven in dit proefschrift is gericht op het begrijpen van de onderliggende mechanismen van genetische interacties.

Het eenvoudigste voorbeeld van genetische interacties is een redundantie-relatie tussen twee genen. Redundantie treedt op wanneer een gen de functie van een ander gen kan overnemen, bijvoorbeeld als ze beide coderen voor zeer vergelijkbare eiwitten. In het geval van een volledige redundantie-relatie resulteert inactivering van elk van beide genen op zichzelf in geen detecteerbaar defect. Maar gelijktijdige deletie van beide genen zal wel resulteren in duidelijke gebreken. Voorbeelden van door de mens gemaakte redundante componenten zijn overal om ons heen, van bruggen en wolkenkrabbers tot computersystemen, krachtcentrales en vliegtuigen. In werktuigbouwkunde vereist het ontwerpen van een betrouwbaar systeem vaak overtollige componenten. Redundantie wordt in door de mens gemaakte structuren gebruikt om gevaarlijke situaties te voorkomen. Het resultaat van miljoenen jaren evolutie in de biologie is een complex netwerk van genetische interacties waarbij redundantie een bijproduct is. De complexiteit van netwerken met genetische interacties omvat ook andere niet intuïtieve typen genetische interacties die uitgebreid worden beschreven in **hoofdstuk 1**.

De rol van een gen wordt vaak onderzocht door functionele gevolgen van hun verwijdering uit de cel te beoordelen. Afhankelijk van de gevoeligheid van de test die wordt gebruikt voor het beoordelen van de verwijdering van het gen, vertoont 66% tot 53% van gendeleties in gist geen detecteerbaar defect wanneer ze in een enkele conditie

worden geanalyseerd. Het is bekend dat dit niet-responsieve gedrag wordt veroorzaakt door redundantie of afhankelijkheid van de conditie. Het begrijpen van de onderliggende oorzaken van redundantie en hun relatieve bijdrage aan niet-responsief gedrag bij genetische verstoring is uitermate belangrijk voor het ontwerpen van efficiënte strategieën gericht op het ophelderen van genfuncties en het ontrafelen van complexe cellulaire systemen. In **hoofdstuk 2** bieden we een systematische classificatie van de onderliggende oorzaken van en hun relatieve bijdrage aan niet-responsief gedrag na gendeletie. De totale bijdrage van redundantie aan niet-responsief gedrag wordt geschat op 29%. De belangrijkste oorzaak van niet-responsiviteit is afhankelijkheid van de conditie (71%).

Celgroei wordt vaak gebruikt om genetische interacties op grote schaal te bestuderen. Genetische interacties worden gemeten door de mate waarin groeidefecten van dubbele deletiemutanten afwijken van hun verwachte waarde op basis van de groeidefecten van de enkele deletie mutanten. Genetische interacties die worden gescoord door het verschil tussen waargenomen en verwachte groei, kunnen globaal worden ingedeeld in twee groepen: negatieve en positieve genetische interacties. Een genetische interactie is negatief als de groei waargenomen voor een dubbele mutant slechter is dan verwacht op basis van de groei van de respectievelijke afzonderlijke mutanten. Een interactie is positief als de waargenomen groei beter is dan verwacht. Afhankelijk van de toestand van een cel, worden verschillende genen op verschillende niveaus tot expressie gebracht. Dit wordt gedeeltelijk gemedieerd door gen-specifieke transcriptiefactoren (GSTF's). Het begrijpen van de basis van genetische interacties tussen GSTF's is daarom belangrijk voor het begrijpen van het transcriptie regulatie netwerk. **Hoofdstuk 3** onderzoekt genetische interacties tussen GSTF's. De expressieniveaus van alle genen in het gistgenoom werden gemeten in 72 GSTF dubbele deletiemutanten en hun overeenkomstige enkele mutanten. Deze hoge-resolutie expressieatlas biedt een systeem-niveau overzicht van het genetische interactielandschap tussen GSTF's en onthult onderliggende mechanistische details. Naast het onthullen van nieuwe redundantierelaties, stelt deze studie ook twee nieuwe moleculaire mechanismen van genetische interacties voor. Deze mechanismen, die we 'buffering door geïnduceerde afhankelijkheid' en 'verlichting door derepressie' noemen, bieden verklaringen voor negatieve en positieve genetische interacties die voorheen niet werden begrepen.

Verschillende moleculaire mechanismen zijn voorgesteld voor verschillende soorten genetische interacties. Verschillen in het voorkomen en de onderliggende moleculaire mechanismen van genetische interacties zijn echter nog niet vergeleken tussen verschillende functionele genklassen. In **hoofdstuk 4** vergeleken we genetische interactiepatronen geïdentificeerd met behulp van genexpressieprofilering voor twee klassen van genen: gen-specifieke transcriptiefactoren en genen betrokken bij signaal transductie. We gebruikten uitputtend modelleren om mogelijke moleculaire verschillen te ontrafelen die ten grondslag liggen aan verschillende genetische interactiepatronen. Onze studie stelt een nieuwe mechanistische verklaring voor voor een bepaald genetisch interactiepatroon dat meer

geassocieerd is met genspecifieke transcriptiefactoren in vergelijking met eiwit-kinasen en-fosfatasen. Al met al kunnen onze bevindingen en de geïmplementeerde methodologieën waardevol zijn voor het begrijpen van de moleculaire mechanismen die ten grondslag liggen aan genetische interacties.

Kankers worden verondersteld te ontstaan uit de gecombineerde effecten van veranderingen in individuele genen en dergelijke combinatorische effecten kunnen worden beschouwd als een type genetische interactie. Het toepassen van statistische analyse bij kankers in volwassenen leidde tot de identificatie van samen voorkomende en elkaar uitsluitende paren gemuteerde genen. Er is echter weinig bekend over genetische interacties bij pediatrische kankers. De detectie van DNA-veranderingen in grote aantallen pediatrische tumoren is pas sinds kort begonnen. **Hoofdstuk 5** beschrijft een analysemethode voor het detecteren van genetische interacties bij kinderen met kanker. Twee methoden worden toegepast om samenvoorkomende en wederzijds exclusieve genenparen in 22 verschillende soorten kanker bij kinderen te detecteren. De twee genetische interactie-analyses worden uitgevoerd voor zowel individuele kankertypes als voor alle kankers samen. Een kaart van genetische interacties wordt gepresenteerd door de resultaten van de vier benaderingen te combineren. De resultaten laten een beperkt aantal genetische interacties zien, deels vanwege het lage aantal geanalyseerde tumoren. De resultaten bevatten echter ook wederzijds uitsluitende mutaties die al eerder zijn gerapporteerd, wat het nut van de aanpak bevestigt. Onze analysemethode biedt daarom een basis voor de toekomstige detectie van genetische interacties tussen verschillende soorten mutaties in grotere datasets.

Ten slotte presenteer ik in **hoofdstuk 6** een discussie en een toekomstperspectief van het onderzoek naar genetische interacties in zowel gist als kanker.

Acknowledgements

It would be impossible for me to thank everyone who has supported me during my PhD. I have been extremely lucky to be surrounded by helpful colleagues and wonderful friends. For the completion of this thesis I have relied on many people. I would therefore like to express my sincere gratitude to those who helped me during this journey. Even if your name isn't here, I still appreciate everything you have done for me, no matter how small.

First, I would like to thank my co-promotor **Dr. Kemmeren**. Dear Patrick, there is no way that I could thank you enough in such a small space. Thank you for guiding and supporting me over the years. Your guidance helped me throughout my years of research, and during the writing of this thesis.

I would also like to express my sincere gratitude to my promotor **Professor dr. Holstege**. Dear Frank, thank you for your continuous support, patience, motivation, and for sharing your immense knowledge. You have set an example of excellence as a scientist, mentor and instructor.

My gratitude goes out to all the thesis assessment committee members and defense ceremony opponents. I would like to offer my special thanks to **Anton**, **Annika** and **Olga** for your scientific input. Thanks for spending hours discussing how to implement the Petri net modelling, that was great fun!

Dear lab members, I would like to thank you for creating a wonderful atmosphere. It was certainly a pleasure to interact with most of you on a daily basis. I also had amazing times with you outside of the lab, including all the lab retreats, dinners, borrels and game evenings.

I would like to thank my amazing paranymphs. **Wim**, thank you for the incredible friendship and scientific input during these years. I appreciate your determination to convince me to join every single social event. Sorry, that I couldn't join all of them. **Ajit**, thanks for being a great friend. We became very close and I'm sure this friendship will continue for many years.

Marian, **Mariel**, **Cheuk** and **Dik,** I appreciate your hard work over the many years in generating high quality data that I could use in my research. **Josephine**, I am very glad that we had the opportunity to work together for the last chapter. I also enjoyed our conversations about food, sports and holidays. I wish you all the best!

Philip, thank you for all your thoughtful input over these years. I always felt comfortable coming to you to ask any question about statistics or programming. I was never disappointed, you always had the proper answer or directed me to the right resources. **Hinri**, you were a wonderful roomie. Your skills in programming and problem solving always amazed me. You also reflect these skills in gardening and other home improvement projects. Thanks for sharing some of your knowledge with me. **Thanasis**, thank you for your scientific input throughout my PhD. It was always fun to hang out with you and your lovely family. I very

much look forward to seeing Sophia. **Jeff**, it's inspiring to see your enthusiasm for your projects. Keep up the good work and good luck with your PhD.

I owe special thanks to all the people I've shared a lab with, including **Jayne, Ellen, Eduard, Tito, Willemijn, Marit, Chris van Run, Chris Tomuta, Sander Van Hooff, Sander Lambo, Jan Smout, Jurrian and Christina**.

I would like to thank all the members of the Molecular Cancer Research department at the UMC Utrecht. Thank you for organizing the great events within the CSnD PhD program including the PhD retreats, masterclasses and most importantly the borrels. I am also grateful to all the members of the HPC team for allowing me to run heavy jobs on the cluster during my PhD.

Special thanks to all my great colleagues and friends at Princess Máxima Center including **Rijndert, Waleed, Nune, Juli, Fabienne, Zeljko, Margit, Jiangyan, Irene, Nil, Evelyn, Lianne, Thomas, Jordy, Priscilla, Lars, Axel, Lindy, Camilla and so many other people** who I couldn't find a way to include here without making a huge list of names. You are all amazing and I admire your determination in building such a great institute!

Dear football lovers, **Xavi, Mauro, Koen, Tim, Wouter, Lucas, Abel, Enric, Nico, Geert, Eelco, Axel, Alex,** thank you all for joining the weekly games which allowed me to let my frustrations out. **Simona**, thank you for introducing me to the famous pub quiz and allowing me to join the pub quiz team. I would like to thank all the members, especially the team captain **Bas,** for striving to be first every week. Special thanks to my friends for spending quality times with me which was also very important in the development of this thesis: **Marta & Shephard, Giorgio, Markus & Marielle, Welu, Hemn, Ruben, Juli, Ana, Charlotte, Audrey and Henok.**

Dear **Kim**, thank you for your love, constant support, and most importantly for keeping me sane over the past few years. Thank you for being my best friend. Lastly and most importantly, I would like to thank all my family members, especially my parents. **Mom and dad**, thank you for your encouraging words and inspiring me to pursue my dreams. I am especially grateful for your emotional and financial support throughout these years. This journey would not have been possible if not for you, and I dedicate this achievement to you.

List of publications

Amini S, Holstege FC, Kemmeren P. "Growth condition dependency is the major cause of non-responsiveness upon genetic perturbation." PLoS One. 2017; 12(3):e0173432. [PubMed: 28257504]

Sameith K, **Amini S**, Groot Koerkamp MJA, van Leenen D, Brok M, Brabers N, Lijnzaad P, van Hooff SR, Benschop JJ, Lenstra TL, Apweiler W, van Wageningen S, Snel B, Holstege FC, Kemmeren P. "A high-resolution gene expression atlas of epistasis between gene-specific transcription factors exposes potential mechanisms for genetic interactions." BMC Biol. 2015;13: 112. [PubMed: 26700642]

Kim H, Zheng S, **Amini SS**, Virk SM, Mikkelsen T, Brat DJ, Grimsby J, Sougnez C, Muller F, Hu J, Sloan AE, Cohen ML, Van Meir EG, Scarpace L, Laird PW, Weinstein JN, Lander ES, Gabriel S, Getz G, Meyerson M, Chin L, Barnholtz-Sloan JS, Verhaak RG. "Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Res. 2015; 25(3):316–327. [PubMed: 25650244]

Manuscripts in preparation

Amini S*, Jacobsen A*, Ivanova O, Lijnzaad P, Heringa J, Holstege FC, Feenstra KA, Kemmeren P. "The ability of transcription factors to fine-tune gene expression is a crucial component of the mechanism underlying inversion, a frequently observed genetic interaction pattern." Under revision in PLOS Genetics.

Jacobsen A*, Ivanova O*, **Amini S***, Kemmeren P, Heringa J, Feenstra KA. "A framework for exhaustive simulation of epistatic patterns using Petri net models"

Amini S*, Daub JT*, Holstege FC, Kemmeren P. "An initial map of genetic interactions in pediatric cancer."

*** Equal contribution**

Curriculum Vitae

Saman Amini was born on August 8th 1985 in Baneh, Iran. In 2007, Saman received his bachelor's degree in Animal Sciences from University of Tabriz in Iran. In the same year, he moved to Tehran to continue his study in Genetics and Animal Breeding. In 2010, he completed his master's degree from the University of Tehran. Saman moved to the Netherlands in 2011 to pursue his passion for both biology and computers. He enrolled in the Bioinformatics master's program at Wageningen University. From there he moved to the US for a six months internship in the lab of Dr. Roel Verhaak at the University of Texas MD Anderson Cancer Center, Houston. In 2013 he received his second master's degree, in Bioinformatics. In the same year, he moved back to the Netherlands to begin his PhD under the supervision of Dr. Patrick Kemmeren and Professor dr. Frank Holstege at the University Medical Center Utrecht. During the last two years of his PhD he was also affiliated with Princess Máxima Center for pediatric oncology. The results of his PhD research are presented in this thesis. As from April 2018, Saman joined the international financial services company PricewaterhouseCoopers as a consultant.