

For the complementation system to work, the yeast cells used to identify a gene must contain two particular genetic markers. First, it must contain the mutation of interest (the one we have isolated in our screen). Because the correct clone will be detected by complementation, the mutation must be recessive. The second important marker is one used to select transformants, such as a *ura3* mutation. Cells lacking a functional *ura3* gene cannot synthesize uracil and depend on uracil in the medium to survive (they are uracil auxotrophs).

The mutant yeast cells identified in the screen are then transformed with the library DNA (that contains all 10^8 plasmids). On average, yeast cells take up one plasmid during transformation. The transformants (those cells that took up a plasmid of the library) are identified by the selectable marker, in our example, *URA3*. Therefore, the transformed cells are selected on plates lacking uracil in the media. Then, those cells that survive in medium lacking uracil (the transformants) are screened for those that complement the phenotype conferred by the mutation of interest. For example, if the mutants were not able to grow on galactose, the transformants will be screened for those that can use galactose as a carbon source. This is done by replica plating to plates that have galactose. These strains are strong candidates to contain a plasmid clone that carries the wild-type *GAL1* gene (see figure 6-2).

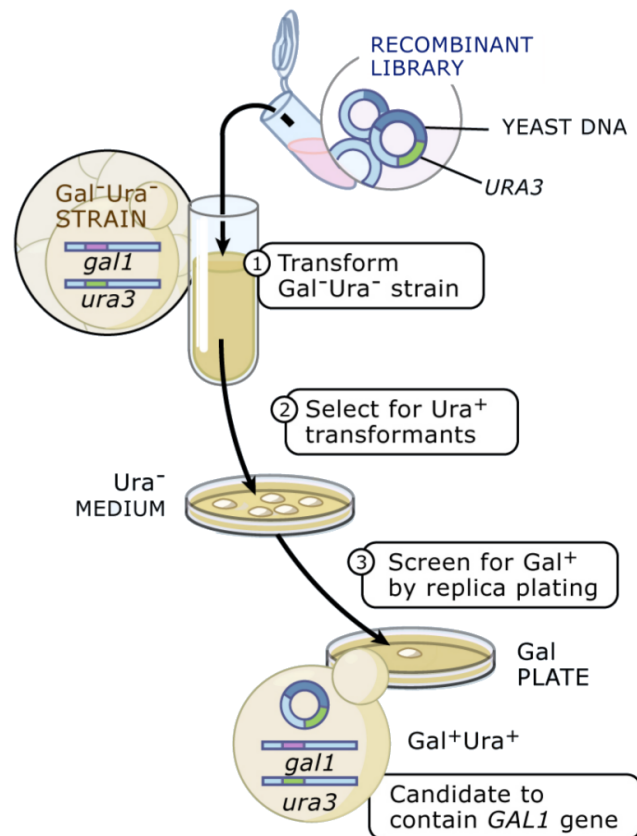


Figure 6-2 Candidate clones are identified after transformation into the appropriate yeast mutant. (adapted from ergito.com)

After identification which one of the plasmids present in the genomic library complements our gene of interest, we know which genomic region is able to complement our phenotype, i.e., contains the gene that is causing the mutant phenotype. However, we still do not know exactly which gene it is, because the average piece of genomic *S. cerevisiae* DNA in each plasmid will be large enough to potentially contain two to three genes. The first step in pinpointing the exact gene of interest on a

clone is to determine the precise genomic segment in the recombinant plasmid. We can do this by sequencing the DNA at the ends of the complementing plasmid. Then, by searching a database of the entire yeast genome sequence, we can identify the exact genomic fragment in the plasmid. In this way, we can reveal the genes present in the insert that was able to complement a certain mutation.

To establish which gene on the plasmid is the one that complements the mutation, we can use two different methods. First, we can make new plasmids, each containing a single gene from those on the original plasmid. Only one of these genes should complement the mutation and identify the gene. The second approach is to perform classical complementation tests between the mutant of interest and mutations in each of the candidate genes. In yeast, we can make use of the synthetic genetic arrays (SGA), collections of mutants where each ORF has been systematically disrupted and replaced by a marker. This array is a useful tool to study gene function in yeast.

Genome editing

Now that we know which genes are affected in our mutations isolated from the screen, we can go on to analyze the gene functions in more detail. Targeted mutagenesis or genome editing allows to change the genome in a desired way, e.g., to replace a mutant allele with the corresponding wild-type allele, to delete a gene or to introduce directed point mutations at a chosen position on a chromosome.

Often, we chose to study a gene based on a particularly interesting mutant phenotype. One example are the mutants that arrest growth at particular points in the cell cycle, the *cdc* mutants. As discussed earlier, such a phenotype could be caused by very different types of mutations. Therefore, an early step after identifying the gene is to construct a deletion of it that results in a null mutant. Comparison of the phenotype of a null mutant with that of the original mutant will determine if the original mutant phenotype results from a loss of function. The analysis will also show whether or not the gene is essential for viability.

Remember that yeast is especially suited to make such deletions because of their efficient homologous recombination system. A precise deletion of any yeast gene can be made using a DNA fragment synthesized by PCR. To do so, a PCR reaction is set up to amplify a selectable marker (e.g., an antibiotic resistance, here against the compound G418) using primers that contain a short region of homology to the DNA sequences flanking the open reading frame of the gene to be deleted on the chromosome. This PCR product is then transformed into yeast cells where the homologous regions present on the PCR product enable recombination at the correct site such that the gene of interest is replaced by the PCR product (see figure 6-3). A diploid yeast strain is used in this experiment to ensure survival of the cells in case the deletion would cause lethality in a haploid situation.

Due to the selectable marker present on the PCR product (in this case, the resistance to the antibiotic G418), yeast cells that integrated the PCR product can be selected for. The resulting spores from this diploid are then dissected and selected using the antibiotic to isolate deletion mutants for a specific gene.

Alternatively to antibiotic resistance, yeast geneticists frequently use auxotrophy markers. For example, in a strain lacking the *HIS3* gene (genotype *his3*), which is required for the synthesis of histidine, the wild type *HIS3* gene can be used as a selection marker based on the fact that once reintroduced in these cells, the transformed cells acquire histidine prototrophy, i.e., the ability to grow in a medium lacking histidine. In addition to *HIS3*, classical auxotrophy markers are *LEU2* (leucine biosynthesis gene), *URA3* (uracil biosynthesis) or *TRP1* (tryptophan biosynthesis).

Since the *S. cerevisiae* genome is known, a deletion mutation of every gene could be made using this method. This collection of mutants where each open reading frame (ORF) is systematically disrupted and replaced by a marker is called the synthetic genetic array (SGA). It consists of about 4800 different yeast strains. These 4800 strains correspond to the 4800 non-essential genes in yeast. About 1800 additional genes are essential and therefore cannot be deleted in haploid cells, hence, their absence

from the collection. Thus, once a gene is identified in a screen, it is straightforward to determine its null phenotype: if it is not in the collection, this means that the null mutation is lethal; if it is in the collection, the null mutant can be recovered from that collection and its phenotype studied in detail.

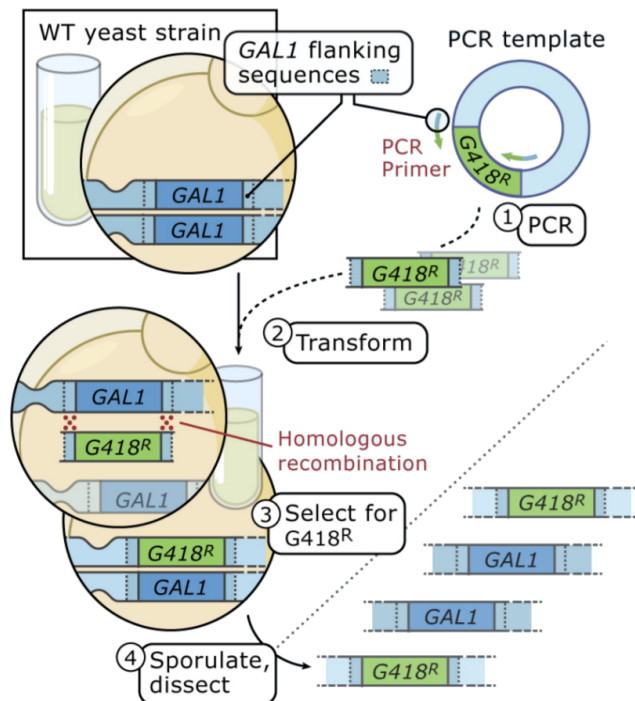


Figure 6-3 A precise deletion of any yeast gene can be made using a DNA fragment synthesized by PCR. The primers used for the PCR contain a short region that is homologous to sequences flanking the gene to be deleted (*GAL1*) and an antibiotic resistance (*G418^R*). After transformation, the PCR product is exchanged for the gene by homologous recombination. Recombinant spores can then be selected for. (adapted from ergito.com)

Genome editing can be used beyond the verification of genes identified by gene mapping. It is also used to tag genes with fluorescent tags (e.g., with green fluorescent protein, GFP) or to tag them with affinity tags used for protein purification.

In higher eukaryotes, genome editing is more difficult since these organisms are much less efficient in homologous recombination. For these organisms, specific systems like the CRISPR/Cas9 system or the TALEN system have been developed, which are based on double strand breaks at the desired locus. We will discuss these systems in more detail later in this course. However, note that CRISPR/Cas9 is increasingly used in yeast as well, as it allows introducing chosen point mutations at a specific locus without having to introduce a selection marker.

A side note: Identification of open reading frames

How do researchers know which sequence of a DNA corresponds to a gene, e.g., which genomic regions are open reading frames? An open reading frame (ORF) has the potential to code for a protein or a peptide (but does not necessarily need to do so). An ORF can be detected within a DNA sequence, because it does not contain a stop codon. A start codon within an ORF can indicate the starting point of translation. However, the DNA contains two strands with opposite orientations, and there are three different ways to place a triplet sequence (a codon) on a DNA sequence. This results in six possible ways to read the genome, the so called reading frames (see figure 6-4). Start codons appear in the DNA with a frequency of 1/64 (since a codon is a sequence of three bases and

the DNA contains four different bases, there are $4^3=64$ possible combinations of codons), and stop codons occur with a frequency of $3/64$ (because there are three different start codons). Thus, many start and stop codons are present just by random.

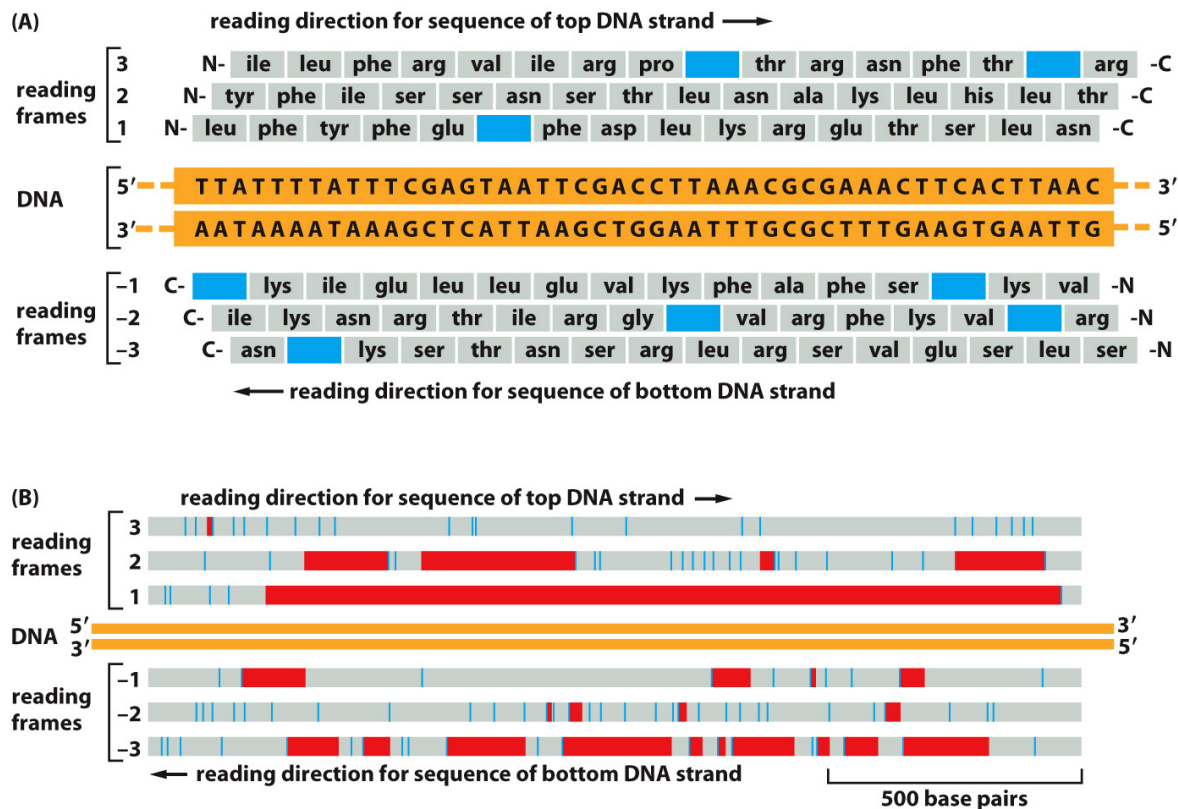


Figure 6-4 Finding the regions in a DNA sequence that encode a protein. (A) Any region of the DNA sequence can, in principle, code for six different amino acid sequences, because any one of the three different reading frames can be used to interpret the nucleotide sequence on each strand. Note that a nucleotide sequence is always read in the 5-to-3' direction and encodes a polypeptide from the N-terminus to the C-terminus. For a random nucleotide sequence read in a particular frame, a stop signal for protein synthesis is encountered, on average, about once every 20 amino acids. In this sample sequence of 48 base pairs, each such signal (stop codon) is colored blue, and only reading frame 2 lacks a stop signal. (B) Search of a 1700bp DNA sequence for a possible protein-encoding sequence. The information is displayed as in (A), with each stop signal for protein synthesis denoted by a blue line. In addition, all of the regions between possible start and stop signals for protein synthesis are displayed as red bars. Only reading frame 1 actually encodes a protein, which is 475 amino acids long. (adapted from figure 8-40, Molecular Biology of the Cell, Alberts, 6th edition, Garland Science)

ORFs were defined to be present if the distance between start and stop is large enough that the probability of having such a sequence by random is smaller than 5%. The distance required to fulfill this condition is about 300nt (= 100 amino acids). Therefore, shorter peptides are not identified, although we know now that they exist (the smallest known peptide proven to have a real function in vivo is 10 amino acids long).

Alternatively, one can also look at what part of the genome is expressed. This means purifying mRNAs from cells and analyzing them to deduce which parts of the genome is transcribed. In another approach, one can compare genomes of related organisms to identify conserved regions. These conserved regions are likely to correspond to ORFs.

Analysis of gene interactions and gene networks

After a gene has been identified, it is of great interest to identify other genes with which it might interact. Gene products can physically interact to perform a function; these interactions can be determined using biochemical methods. However, some interactions may not be detected using biochemical approaches, e.g., when the interaction is not strong enough or is transient. Furthermore,

some proteins may interact functionally, but not physically, for example if two different proteins regulate the same set of genes. Yeast genetics provides several methods to identify interactions that are impossible to find biochemically. Here, we will discuss how mutations in one specific process provide a starting point to identify other genes in the same pathway or genes in a different pathway that affect the process of interest.

Epistasis and suppressor analysis

We have already discussed previously that genes may show an interaction defined as epistasis. If two mutant genes in a haploid situation confer a phenotype that is quantitatively identical to that conferred by the single mutant gene alone, the two genes are defined as being epistatic with respect to one another. The phenotype caused by a mutation in one gene is masked by a mutation in another gene. Epistatic analysis requires the two mutants to have distinguishable phenotypes. It can be used to determine the order of gene function by testing whether the phenotype of the double mutant *ab* is similar to that of mutant *a* or mutant *b*. Thus, comparing the single mutant phenotypes with the double mutant phenotype can determine their relative order of function.

In lesson 2, we have discussed two examples of epistatic interaction where a suppressor mutation in a second gene reverts the phenotype of a mutation. Epistasis analysis is the analysis of the interaction between two or more genes that control a single phenotype. The phenotype of cells containing the *ade2* mutation (red colonies on medium without adenine) can be reverted by an additional mutation in the *ADE3* gene. Both *ade2-ade3* double mutants and *ade3* single mutants form white colonies on plates without adenine as the wild type. Thus, the *ade3* mutation suppresses the phenotype of *ade2*, and *ADE3* is epistatic to *ADE2* in the adenine synthesis pathway.

In another example, we discussed how the secretion phenotype of *sec61-1* mutants can be reverted by a secondary mutation in *ubc6*, a gene encoding an ubiquitin-conjugating enzyme that functions in the ER-associated degradation pathway (ERAD). By blocking the degradation of misfolded SEC61 protein, cells are able to survive. Thus, *ubc6* is epistatic to *sec61*.

The phenomenon of epistasis is used to identify interacting genes in a technique called suppressor analysis. This technique is used to look directly for other mutations that are epistatic to a mutation of interest. Suppressors of an interesting mutation are identified in two steps. The first step is the isolation of "revertants" of the original mutant – i.e., strains that no longer show the original mutant phenotype. Those revertants can be isolated by either a screen or a selection. A revertant strain could arise by either true reversion of the mutation or by a second mutation in an interacting gene, a so called extragenic suppressor, that compensates for the defect of the original mutation. Examples of extragenic suppressor were discussed above where we describe epistasis. In these cases, a mutation in one gene suppressed the defect caused by a mutation in another gene.

Whether a mutation leads to a true reversion (a mutation that reverts the original mutation) or an extragenic suppressor is determined by crossing the "revertant" with a wild-type strain, followed by tetrad analysis (see figure 6-5). For true revertants, all tetrads will be PD tetrads, since all spores will inherit a wild-type allele. In contrast, for extragenic suppressors, a different pattern can be observed, because there will almost certainly be recombination between the suppressor mutation and the original mutation, since they are located in two different genes. Therefore, some spores will inherit only the original mutation.

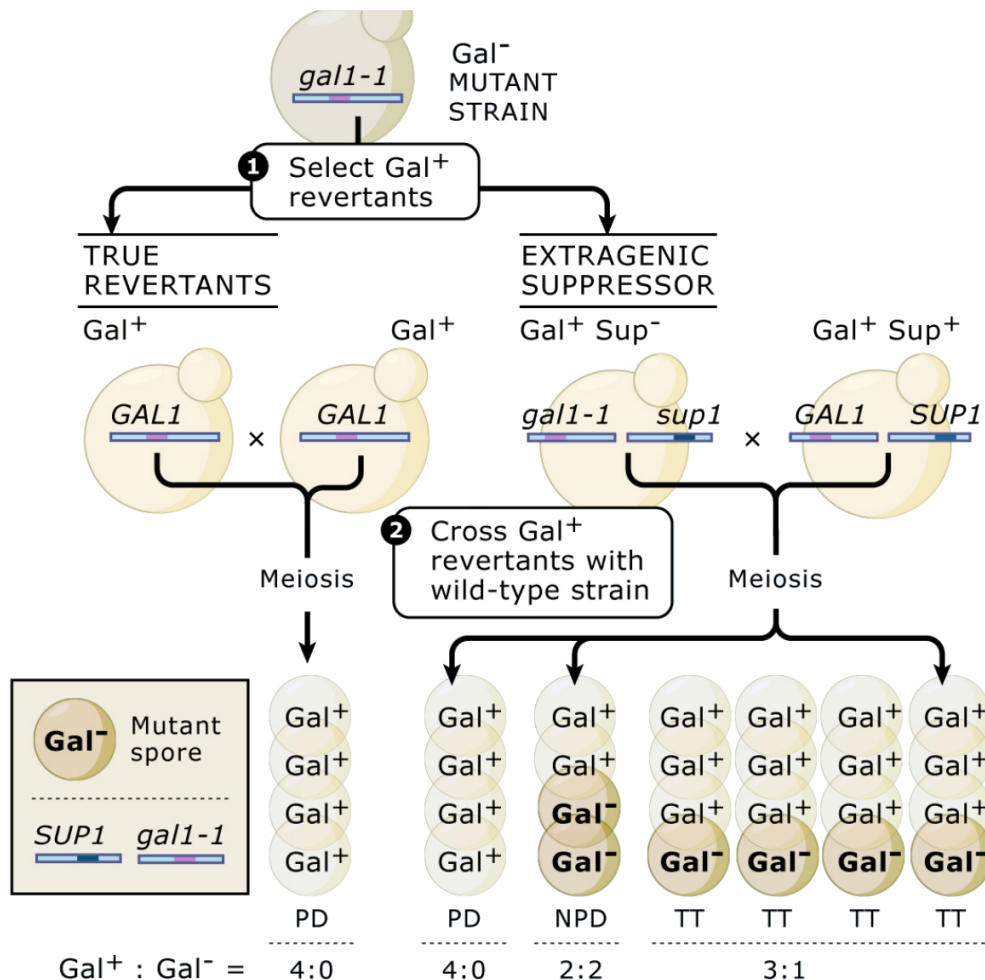


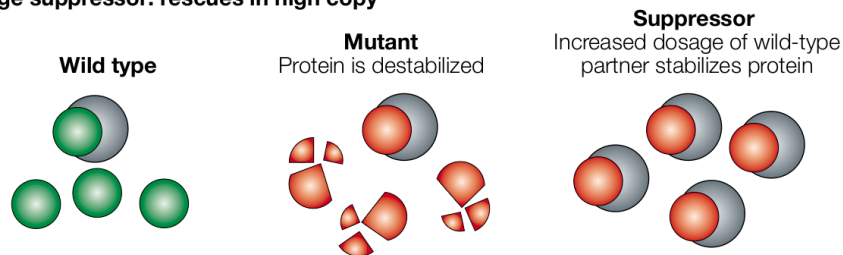
Figure 6-5 In order to test whether a cell contains a gene that is a true revertant, it is crossed with a wild-type cell. A true revertant will be completely linked to the initial mutation producing all PD tetrads. An extragenic suppressor mutation (here: *sup1*) will segregate from the initial mutation, resulting in PD, NPD, and TT tetrads. (adapted from ergito.com)

The type of mutation that is being suppressed will often determine the nature of the suppressor mutation. For example, a mutation altering the conformation of a protein might be suppressed by a compensatory change in an interacting protein. In contrast, a deletion mutation that removes the coding region of a gene cannot be suppressed by a change in an interacting protein. Only a suppressor that compensates for the complete loss of the initial gene product can suppress a deletion mutation.

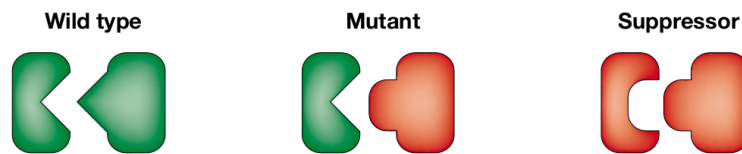
Suppressor analysis is a powerful tool for yeast geneticists. Often, the study of a gene identified as a suppressor will shed light not only on the function of the suppressor itself, but also on the function of the gene containing the initial mutation. Examples for how suppression analysis can reveal insights into the mechanisms of gene interaction are shown in figure 6-6.

In suppressor analyses, the double mutant shows a phenotype that is different from either single mutant. We call such interactions synthetic interactions. For lethality suppressors (synthetic viable mutants), the double mutant is viable when at least one single mutant is not.

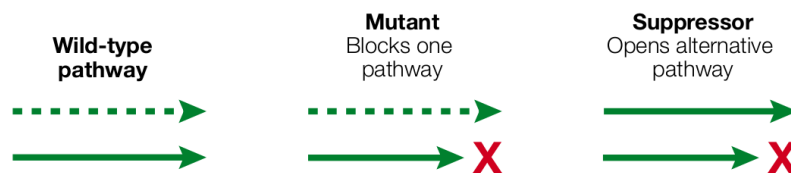
a Dosage suppressor: rescues in high copy



b Interaction suppressor: allele specific, gene specific



c Bypass suppressor: pathway specific, rescues null allele



d Nonsense suppressor: allele specific, gene non-specific

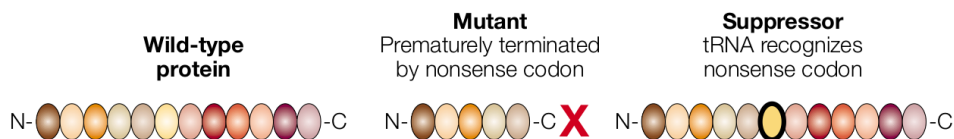


Figure 6-6 Suppressor mechanisms. Depending on the allele and gene specificity associated with suppressors, mechanisms can be inferred, as shown. **(a)** Dosage suppressors encode proteins that stabilize the mutant product when they are expressed at high levels. **(b)** An interaction suppressor restores the interaction between the mutant product and its partner(s). **(c)** A bypass suppressor activates an alternative pathway to the wild-type pathway. **(d)** A nonsense suppressor encodes a tRNA molecule that recognizes a premature termination codon and inserts an amino acid at that position. (Susan L. Forsburg, Nature Reviews Genetics, Sept 2001)

Synthetic lethality

The opposite of lethality suppression is synthetic lethality, where the double mutant is lethal under conditions in which both single-mutant parents are viable. Here, the approach is to search for synthetic enhancement to intensify the phenotype of the original mutation by a mutation in a second gene to the point of death. The rationale behind this is that if each of the two mutations impairs an aspect of an essential process, the double mutant should have a more severe phenotype than either of the single mutants. This approach can be useful to uncover redundant pathways or direct protein interactions.

Redundancy refers to the situation where more than one gene in the genome can perform the same function. In these cases, a mutation may not show a strong phenotype, because the presence of another gene can compensate for the mutation.

To determine whether a redundant gene or pathway is present, the mutation is crossed with other

mutations and the resulting phenotype is observed. If there is a single additional gene that is redundant, then mutating it in cells that already lack the first gene will confer the desired phenotype. Only the double mutant will show the phenotype, either single mutant would be normal.

Three possible relationships uncovered by synthetic-lethality screens are those among redundant genes in a regulatory pathway, genes whose products interact together in a complex, and genes whose products independently contribute to the same process.

An example for synthetic lethality for redundant genes in a regulatory pathway is shown in figure 6-7. Here, the two genes *TUB1* and *TUB3* each encode for a version of the protein α -tubulin. Most of the gene product is produced by *TUB1*. If *TUB1* is mutated, *TUB3* can still provide protein product, thus, the *tub1* mutant cells are viable. However, if an additional mutation in *TUB3* is present, the loss of *TUB1* cannot be compensated for and the cells die.

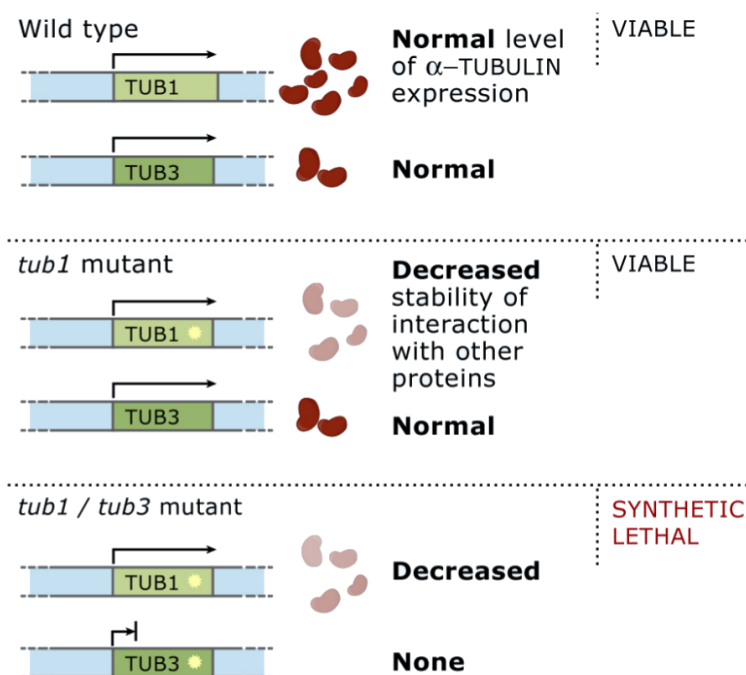


Figure 6-7 Shown are two genes, *TUB1* and *TUB3*; each encodes a version of the essential protein α -tubulin, with *TUB1* providing most of the product. Some point mutations in *TUB1* that destabilize the interactions of resulting protein are not lethal. However, in combination with a *tub3* null allele, there is not enough functional α -tubulin and the result is lethality. (adapted from ergito.com)

On the other hand, synthetic lethality can also be observed between genes whose products interact with each other. Here, a mutation in either of the genes that contribute proteins for the complex reduces the activity of the complex, but the activity is still sufficient for cell survival (see figure 6-8). If a double mutant that alters two proteins reduces the activity below a threshold that is not enough to ensure survival, the cells die. Synthetic lethality thus serves as a tool to identify direct interactions between gene products.

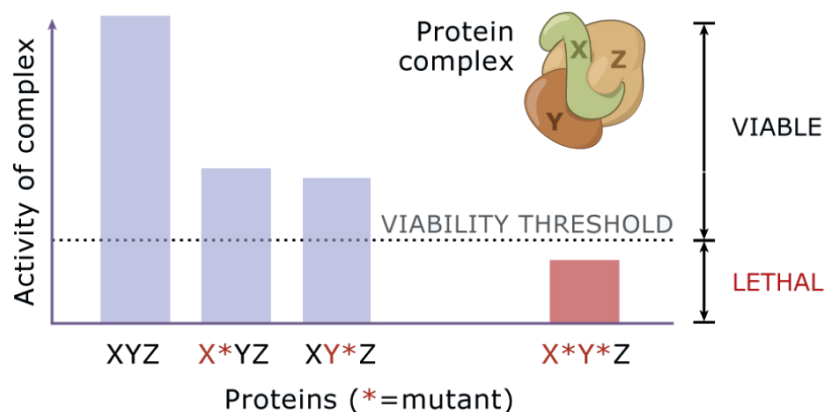


Figure 6-8 Several proteins interact in a complex that carries out an essential function. A mutation in X, Y, or Z reduces the activity of this complex, but still allows for viability. However, a double mutant that impairs two proteins reduces the activity below a threshold required for viability, resulting in synthetic lethality. (adapted from ergito.com)

Another scenario is synthetic lethality that inactivates two parallel mechanisms that achieve the same essential function. The large majority of synthetic lethal cases are due to mutations in genes that have overlapping functions in the cell. One example is arginine availability. Arginine can either be imported through the *CAN1* arginine transporter, or synthesized by a set of enzymes (*arg* genes). On medium containing arginine, *arg*-mutations have only mild effects. However, in combination with the *can1* mutation, any mutation in the *ARG* pathway is lethal.

Alternatively, the inactivation of quality control mechanisms in the context of a first mutation affecting the fidelity of a process can be deleterious. Many cellular processes are monitored by quality control pathways that ensure that the cell detects defects and rapidly reacts to them. For example, DNA damage leads to activation of the DNA damage checkpoint, which arrests the cell cycle as long as the damage is not repaired. *RAD9* is involved in this checkpoint. As long as the cells are not stressed, *RAD9* is not essential. However, the combination with a defect in chromatin organization becomes immediately lethal, since the cells cannot repair the DNA damage.

Remember that there are two special, although quite rare, cases of interaction between mutations that can provide specific information about the nature of these mutations. We have discussed them already in lesson 2. In rare cases, double mutants behave like wild types, although the two mutations are in the same gene. Thus, the two mutations complement each other. This situation is called intragenic complementation or allelic complementation. When observed, this situation indicates that the gene might code for a protein with two independent domains that can act independently of each other, such that an allele where the first domain is inactivated but the second functions normally may be complemented by another allele where the first domain functions normally but the second does not.

In the other case, double mutants with mutations in different genes that individually result in the same phenotype fail to complement each other in the heterozygous diploid, i.e., the heterozygous diploid still shows an abnormal phenotype. This is called non-allelic non-complementation. The fact that the diploid, which is heterozygous for each mutation, still shows the abnormal phenotype indicates that two genes encode physically interacting proteins. This may be the case if, e.g., two polypeptides build a multimeric complex. Here, heterozygosity for mutations in both genes can lower the concentration of the multimeric complex, which may lead to a phenotype.

Ordering genes in a pathway

Sometimes, multiple genes each play a distinct role in a common process. We have seen this in for the genes *ade2* and *ade3* in the adenine-biosynthesis pathway. Comparing the single-mutant phenotypes with the double-mutant phenotype can determine the relative functional order of two genes, e.g., to analyze which of the genes acts first (upstream) and which acts later (downstream) relative to each other. However, how is the order determined, e.g., if two genes are in an epistatic relationship, which gene is downstream and which one is upstream in the pathway?

To answer this question, we have to consider that there are two different types of biological pathways: biosynthetic (metabolic) and regulatory pathways. Both types of pathways involve multiple functional components (genes or gene products) that act in a defined sequence (see figure 6-9).

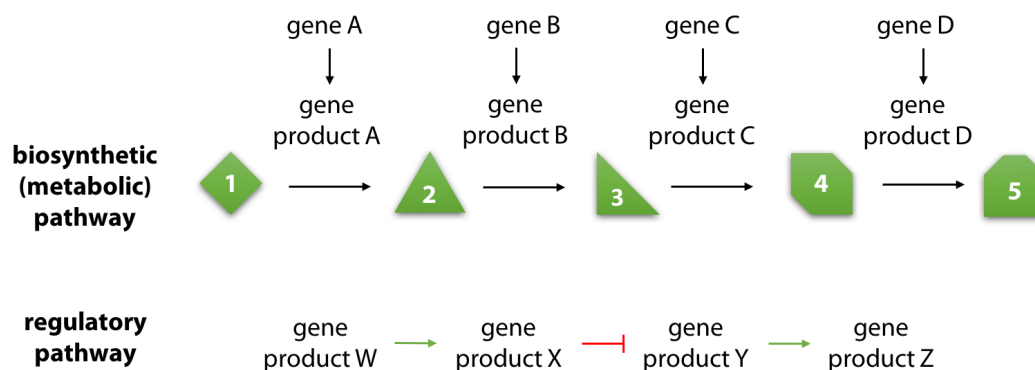


Figure 6-9 Difference between biosynthetic and regulatory pathways. In biosynthetic pathways, a precursor (metabolite 1) is converted via several chemical reactions into a product (metabolite 5) by the gene products of genes A through D. In regulatory pathways, the gene products activate or inactivate the other gene products of the pathway.

In the first pathway, a product Z is produced and a precursor Q is eliminated. In contrast, the second pathway regulates the final gene (D) in response to the initial gene in the pathway (A).

There are fundamental differences between these pathways:

- Each step in a metabolic pathway converts one intermediate to another: the "gene" corresponds to the catalyst (enzyme), rather than the intermediate(s).
- Each step in a regulatory pathway activates (+) or inactivates (-) a "switch", which is usually the next "gene" in the pathway (either the gene itself or its product).
- Each step in a biosynthetic pathway is required for the subsequent step(s).
- A given step (switch) in a regulatory pathway may act either positively or negatively on a subsequent step.
- In metabolic pathways, loss-of-function mutations that affect different genes in the same metabolic/synthetic pathway will have distinct phenotypes because they accumulate different intermediates in the reaction chain
- In regulatory pathways, mutations will affect the pathway in one of two opposite ways: either activation or inactivation.

Thus, if two genes with distinct phenotypes function in a simple linear pathway, analysis of the double mutant may allow the order of the two genes:

- For a biosynthetic/metabolic pathway, the "upstream" gene will be epistatic to the downstream

gene. This is because the intermediate upstream of the first gene/enzyme accumulates, but the blockade at this point will eliminate everything downstream.

Let's look at the *ade2-ade3* example again. *Ade3* is epistatic to *ade2*, and both genes act to affect a biosynthetic pathway (adenine synthesis). Thus, the upstream gene will be epistatic to the downstream gene: *ade3* acts upstream of *ade2*.

- For a regulatory pathway the order is reversed: a downstream mutant usually determines the phenotype, because any components upstream of it become irrelevant. Thus, the downstream gene will be epistatic over the upstream gene.

Let us look at an example where epistasis helped to analyze the functional relationships of genes regulating mitosis. The G2/M transition is controlled by the activity of the kinase Cdc28. Upon its own phosphorylation by another kinase, Cdc28 is inhibited and mitosis entry is delayed. To identify the kinase responsible for this phosphorylation, a mutation in *SWE1* was created, a gene encoding the ortholog of a kinase responsible for regulating the G2/M transition in *S. pombe*. Cells lacking Swe1 do not delay, but rather advances the timing of mitosis. Conversely, a mutation in the gene *mih1* shows a significant delay in G2 phase and delays the onset of mitosis. When double mutants of *mih1* and *swe1* were constructed, the double mutant showed a phenotype similar to the *swe1* single mutant.

Further analysis showed that *SWE1* encodes a protein kinase that phosphorylates and thus inhibits the activity of Cdc28. Mih1 was shown to be the phosphatase that removes the phosphates deposited by Swe1 on CDC28 to promote mitosis. Since the phosphorylation of Cdc28 delays mitosis, Swe1 is an inhibitor (a negative regulator) and Mih1 an activator (a positive regulator) of mitosis.

Summary

In this lesson, we have learned how mutants identified in a screen can be analyzed with respect to the nature of the mutation and the gene that is responsible for the observed phenotype. We have also seen that yeast has been particularly amenable for identifying and characterizing gene products that interact with each other, especially when two mutations enhance or alleviate each other's defects.