CONCEPTS IN MODERN GENETICS

Hallo zusammen :) Da es auf Vebis noch keine Zusammenfassung von diesem Kurs gibt, möchte ich euch gerne meine Zusammenfassung zur Verfügung stellen. Da dies alles meine Notizen sind, sind Fehler und auch Fehlüberlegungen nicht ausgeschlossen. Lest daher die PDFs unbedingt selber auch noch durch und gleicht sie mit der Zusammenfassung ab. Auch die Vorlesungen werden wohl jedes Jahr etwas anders sein. Ich wünsche Euch viel Glück!

Was ich noch von der Prüfung weiss:

- *Barral: Tetradenanalyse sehr genau anschauen,*
- *Bopp: Prüfungsaufgaben waren genau wie im Problem Solving, zudem wurden die RIDL Systeme abgefragt, Balancer wurden gar nicht abgefragt, man musste nicht erkläre*
- *Hajnal: 1:1 wie im Problem solving!*
- *Voinnet: Will wissen, mit welchem Experiment man was herausgefunden hat, alle DCL und Ago erklären, man musste Screen entwerfen mit welchem man die Komponenten wie Dicer etc, des RdDM oder VIGS (weiss nicht mehr welches von beiden) finden kann*

Allgemein: Ausser bei Voinnet wurden keine Auswendiglernfragen gestellt. Man muss auch daher auch keine Crossing Schmemas erklären oder Ähnliches. Es lohnt sich sehr, die Schwerpunktvorlesungen und die Problem Solving Lessons der Profs genau zu verfolgen, da diese hauptsächlich den Prüfungsstoff ausmachen.
Die beste Vorbereitung ist die Problem Solvings sehr genau anzuschauen und bei Herrn Voinnet die Experimente genau zu kennen.

VANESSA WÄLCHLI

ETH ZÜRICH HS2020

Week	Date	Lecturer	Topic	Format
1	14.09.20 (Mon) 15.09.20 (Tue)	No lecture all lecturers	Introduction	
2	21.09.20 (Mon) 22.09.20 (Tue)	Yves Barral	Genetic analysis in yeast	Introduction (online) Self-study (offline)
3	28.09.20 (Mon) 29.09.20 (Tue)	Yves Barral	Genetic analysis in yeast	On the search for mutants (online) Identifying the genes (online)
4	05.10.20 (Mon) 06.10.20 (Tue)	Yves Barral	Genetic analysis in yeast	Non-mendelian genetics (online) Problem-solving (online)
5	12.10.20 (Mon) 13.10.20 (Tue)	Alex Hajnal	Forward genetics	Introduction self-study (offline)
6	29.10.20 (Mon) 20.10.20 (Tue)	Alex Hajnal	Forward genetics	online tutorial 1 online tutorial 2
7	26.10.20 (Mon) 27.10.20 (Tue)	Alex Hajnal	Forward genetics	online lecture: quantitative genetics Problem solving
8	02.11.20 (Mon) 03.11.20 (Tue)	Daniel Bopp	Reverse genetics	Introduction (online) Self-study (offline)
9	09.11.20 (Mon) 10.11.20 (Tue)	Daniel Bopp	Reverse genetics	tutorial 1 (online) tutorial 2 (online)
10	16.11.20 (Mon) 17.11.20 (Tue)	Daniel Bopp	Reverse genetics	online Lecture: applications in medicine and agriculture Problem solving (online)
11	23.11.20 (Mon) 24.11.20 (Tue)	Olivier Voinnet	Small non-coding RNAs, epigenetic regulations	Introduction Self-study
12	30.11.20 (Mon) 01.12.20 (Tue)	Olivier Voinnet	Small non-coding RNAs, epigenetic regulations	Q&A miRNAs/siRNAs Resolving miRNA action in space
13	07.12.20 (Mon) 08.12.20 (Tue)	Olivier Voinnet	Small non-coding RNAs, epigenetic regulations	Transposon silencing-I Transposon silencing-II, transgenerational inheritance
14	14.12.20 (Mon) 15.12.20 (Tue)	all lecturers	General discussion topic to be announced Tutorial	debate exam preparations ONLINE

CONCEPTS IN MODERN GENETICS HS2020
GENETIC ANALYSIS IN YEAST

THE POWER OF YEAST (ONLINE LESSON)

- saccharomyces cerevisiae = first eukaryotic Organism with sequenced genome
- 12'000 kb of DNA, 16 Chromosomes, high gene density (5 mal höher als Mensch)
- 1 Gen alle 2 kb, 6600 Genes, mostly no introns

Technical and Biological Advantages

- 90 min Teilungsrate
- large scale production as well as single clone analysis (lipid cultures and plates)
- replica plating technique for selection and isolation of mutant clones
- Unlike mice or humans, yeast can be grown on defined media, which facilitates the reproducibility of experimental results.
- robustness towards environmental conditions: yeast can be maintained frozen or freeze-dried or be stored at room temperature for years.
- high rate of recombination → switch mating types and fix DNA damage and to ensure correct segregation of homologous chromosomes during meiosis
    - DNAs carrying homologies to the yeast genome are effectively integrated into the genome by homologous recombination
    - a high rate of homologous recombination is a distinguishing feature of a model organism, because it allows for straightforward targeted genome modifications.
- yeast chromosomes replicate via autonomously replicating sequences (ARS) that contain the origin of replication.
    - Pieces of DNA including ARS are replicated within the yeast nucleus even if the DNA is not part of the chromosomes
    - ARS to create plasmids that replicate within yeast cells, allowing them to easily move external DNA into yeast cells.
- Yeast cells can reproduce both asexually, which results in identical clones, as well as sexually, which allows for recombination
- The haploid stage is beneficial for genetic analyses because recessive traits manifest in haploid strains since only one chromosomal copy is present. This facilitates the isolation of mutations that result in recessive traits.
- On the other hand, diploids are useful in maintaining these mutations (since the wild-type copy of the gene is still present in heterozygous diploids) and to combine alleles of genes in order to determine recessive or dominant traits or to study gene interactions.
- Since yeast shares a high degree of conservation on the level of the amino acid sequence and protein function with more complex eukaryotic organisms, it has even served in screens for human diseases.

Specificities of yeast biology

- The most striking difference to animal cells is the presence of the cell wall in yeast
    - provides mechanical stability and protection, but also makes cellular migration much more difficult.
- unicellular, and single cells can easily be moved by water or wind.
- The thick cell wall of most yeast species also makes them immune to external viruses.
- Consequently, the yeast genome shows little modification by horizontal gene transfer and is fairly stable.
- recombination is therefore efficient in yeast, since the risk of inserting foreign DNA is lower compared to other species.
- symbioses
- some species of wood-ingesting beetles host xylose-fermenting yeasts in their guts, most likely to provide the enzymes to digest plant cell walls
- S. cerevisiae converts carbohydrates to carbon oxide and alcohol.
- **Candida albicans** is the most common cause of fungal infections worldwide.
- C. albicans morphology is affected by a range of environmental cues: For example, at low pH (< 6), C. albicans cells predominantly grow in the yeast form, while at a high pH (> 7), hyphal growth is induced

Mating type determination in yeast

- able to switch their mating type (change their sex). The haploid cells of *S. cerevisiae* exist in two different mating types: *MATa* and *MATα*.
- *MATa* cells produce '**a**-factor', a mating pheromone which signals the presence of an **a** cell to neighboring **α** cells.
- **a** cells respond to **α**-factor, the **α** cell mating pheromone, by growing a mating projection (known as a shmoo) towards the source of **α**-factor.
- Similarly, **α** cells respond to **a**-factor.

☞ The response of haploid cells only to the mating pheromones of the opposite mating type allows mating between **a** and **α** cells, but not between cells of the same mating type. Two cells of opposite mating type fuse to form a diploid cell.

☞ These differences between a and **α** cells are due to different genes being actively transcribed and repressed in cells of the two mating types.

☞ a cells activate genes which produce **a**-factor and produce a receptor which binds to **α**-factor and triggers signaling within the cell.

☞ a cells also repress the genes associated with being an **α** cell.

☞ **α** cells produce **α**-factor and a cell surface receptor which binds and responds to **a**-factor, and repress the genes associated with being an **a** cell.

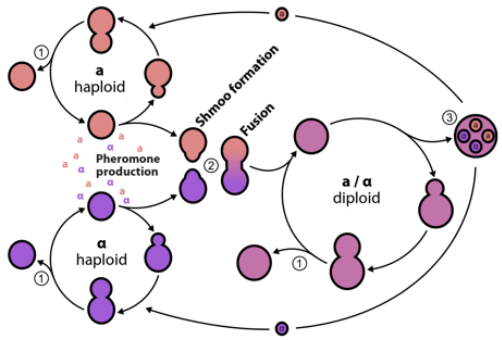


Figure 1-2 Yeast life cycle. The mating type, either MATa or MATα, of haploid yeast cells depends on the expression of certain sets of genes producing sex-specific pheromones and receptors. Haploid cells can either divide asexually by forming buds (1) or fuse to produce a diploid cell with another haploid cell of the opposite mating type (2). Upon sensing pheromones of the other mating type, a cell forms a protrusion termed shmoo. After fusion, the resulting diploid cell can either divide asexually by forming buds (1) or undergo meiosis and produce four haploid spores (3). (adapted from Wikipedia: Mating in fungi)

☞ The mating type is determined by genes present in the mating type locus (MAT). There are two different alleles of the MATlocus: MATa and MAT**α**.

☞ The two mating type alleles differ by 700 base pairs of sequences that encode regulators of the two different haploid mating types

☞ yeast can exist in a haploid (either MATa or MAT**α**) or a diploid state (MATa/MAT**α**)

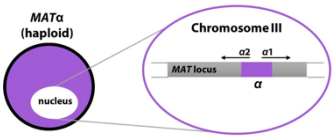


Figure 1-3 The sex-determining genes in yeast. The *MAT* locus consists of two open reading frames (arrows), leading to the expression of different genes. For *MATa*, these are *a1*, *a2*, for *MATα*, *α1* and *α2*.

Table 1 Overview over the abbreviations of the yeast sex-determining genes and their functions.

abbreviation	name	function
asg	**a**-specific genes	• α-factor pheromone receptor • a-factor pheromone
αsg	**α**-specific genes	• a-factor pheromone receptor • α-factor pheromone
hsg	haploid-specific genes	• budding-pattern genes • fusion genes • signal-transduction genes • repressor of diploid-specific gene
dsg	diploid-specific genes	• meiotic genes

☞ in haploids, the haploid-specific genes (hsg) must be expressed together with the respective sex-determining genes (asg or αsg) and diploid-specific genes (dsg) must be repressed.

☞ On the other hand, in diploids, haploid-specific genes must be repressed.

Function of the MAT genes: haploid

- The MAT**α** locus encodes two genes: **α**1 and **α**2. The gene **α**1 encodes for a transcriptional activator, the gene **α**2 for a transcriptional repressor.
- the MATa locus contains the genes a1 and a2.
- The product of the gene a1 by itself has no function; to be functional, it needs to be bound to **α**2. The complex of a1/**α**2 acts as a transcriptional repressor. a2 is a pseudogene without function.
- MATa is the default state of yeast; thus, in these cells, asg and hsg are always expressed and the expression of **α**-factor receptor and a-factor leads to the sex a (tab)
- MAT**α** cells express the **α**-specific genes due to their activation by **α**1. **α**2 represses the asg, thus, MAT**α** cells are of the sex **α** and produce a-factor receptor and **α**-factor pheromone
  - Therefore, **α**2 is the ultimate factor that determines the mating type.

Function of the MAT genes: diploid

- In MATa/MAT**α** diploids, the **α**2 protein has two functions: By itself, it represses asg. Together with a1, it forms a a1/**α**2-dimer that functions as a repressor for hsg.
- Since hsg normally represses dsg, the repression of hsg leads to an expression of dsg in diploids.
- The a1/**α**2-dimer also represses **α**1 (thus, **α**sg are repressed).
- That means that **α**2 can extend its function in MATa/MAT**α** diploids due to the ability to form heterodimers with a1

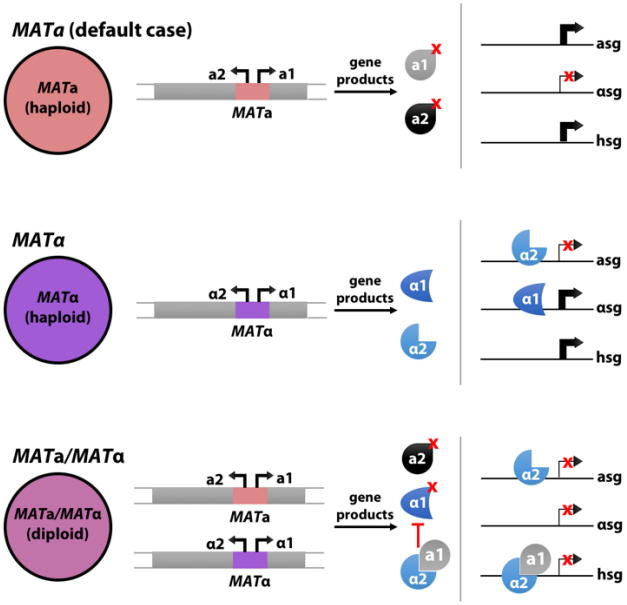


Figure 1-4 Molecular basis of mating type determination in yeast. The asg sex-determining genes have a strong promoter regulated by repression (by the α2 protein in *MAT*α haploids and the a1/α2 complex in *MATa/MAT*α diploids), while αsg genes have a weak promotor that requires activation (by α1).

How does the sex-determining system ensure that *MATa/MAT*α diploids don't mate?

The a1/α2 complex turns off haploid genes. It inhibits α1; thus, αsg is not expressed and no α-factor is produced (because α1 is required to activate the expression of αsg). On the other hand, the a1/α2 complex allows the expression of α2, the repressor of asg, thus, no asg are expressed. Thus, the repression of both asg and αsg ensures that the diploid does not produce mating pheromones and is non-mating.

Switching the mating type = homothallism

- es sind immer beide mating types in einer Kolonie present, auch wenn die Kolonie aus Klonen einer spez. Zelle mit spezifischem Mating type abstammt
- Since haploid cells have a strong tendency to mate with cells of the opposite mating type and form diploids, mating type switching and consequent mating will cause the majority of cells in a colony to be diploid, regardless of whether a haploid or diploid cell founded the colony.
- S. cerevisiae is heterothallic, because each yeast cell is of a certain mating type and can only mate with a cell of the other mating type.
- However, yeast cells within a colony (where all cells have the same mating type) can switch their mating type such that the subsequent mating of cells to the opposite mating type enables homothallic organisms to self-diploidize. (An organism is called homothallic if both male and female reproductive structures are present in the same thallus.)
- advantage diploid: the ability to undergo meiosis and form spores under nutritionally limiting conditions and to perform recombination to produce new genotypes.
- in research, mating type switching is altered (unterdrückt) → only diploids if crossed with cells of the other mating type => heterotallic
  - unlike the homothallic strains (which always return to the diploid state) they can be propagated as different types: haploid MATa, haploid MAT α, or diploid.
- **Haploid yeast switch mating type by replacing the information present at their MATlocus.**
  - For example, an a cell will switch to an α cell by replacing the MATa allele with the MATα allele.
  - ☞ **switching is possible, because yeast cells carry an additional silenced copy of both the MATa and MATαalleles: the HML (Hidden MAT Left) and HMR (Hidden MAT Right) locus**
  - The HML and HMR loci are silenced by proteins interacting with sequences flanking the HMR or HML locus, thus forming short regions of heterochromatin that is not transcribed = silent mating cassettes
  - these additional copies of the mating type information do not interfere with the function of the allele present at the MAT locus because they are not expressed

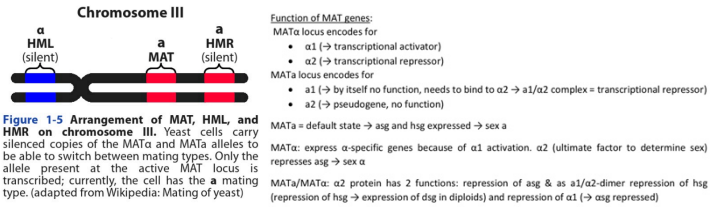


Figure 1-5 Arrangement of MAT, HML, and HMR on chromosome III. Yeast cells carry silenced copies of the MATα and MATa alleles to be able to switch between mating types. Only the allele present at the active MAT locus is transcribed; currently, the cell has the **a** mating type. (adapted from Wikipedia: Mating of yeast)

Process of switching mating type

1. The process of mating type switching is a homologous recombination event initiated by the HO gene, which encodes a DNA endonuclease that cleaves DNA specifically at the MAT locus.
2. Once HO cuts the DNA at MAT, exonucleases are attracted to the DNA ends and start to degrade the DNA on both sides of the cut.
3. This eliminates the DNA which encoded the MAT allele and the resulting gap is repaired by copying in the genetic information present at either HML or HMR, filling in a new allele of either the MATa or MATα gene.
4. the silenced alleles of MATa and MATα present at HML and HMR serve as a source of genetic

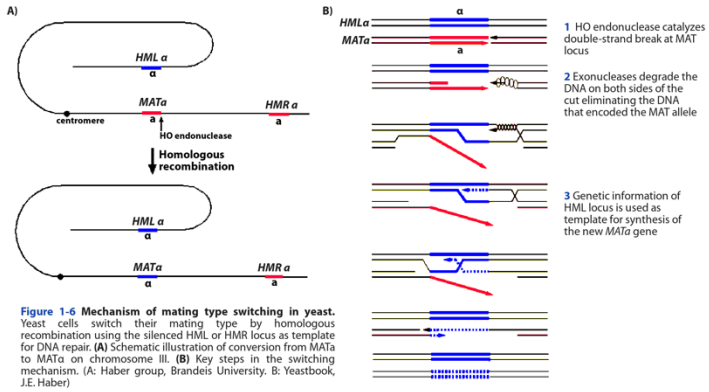


Figure 1-6 Mechanism of mating type switching in yeast. Yeast cells switch their mating type by homologous recombination using the silenced HML or HMR locus as template for DNA repair. **(A)** Schematic illustration of conversion from MATa to MATα on chromosome III. **(B)** Key steps in the switching mechanism. (A: Haber group, Brandeis University. B: Yeastbook, J.E. Haber)

information to repair the HO-induced DNA damage at the active MAT locus

THE RELATIONSHIP BETWEEN MUTATION AND PHENOTYPE (ONLINE LESSON)

- Reverse genetics erst möglich als whole genome sequencing startete

Types of Mutations

mutations can be defined based on the effect on the DNA sequence:

- point mutation: affect only a single base pair in the genome → nucleotide exchange or a deletion/insertion
- can also affect larger Pieces of DNA or whole Genes: Inversion, translocations

→ deletions often cause lethal or null mutations, whereas point mutations are the main cause for conditional, dominant-negative or suppressor mutations.

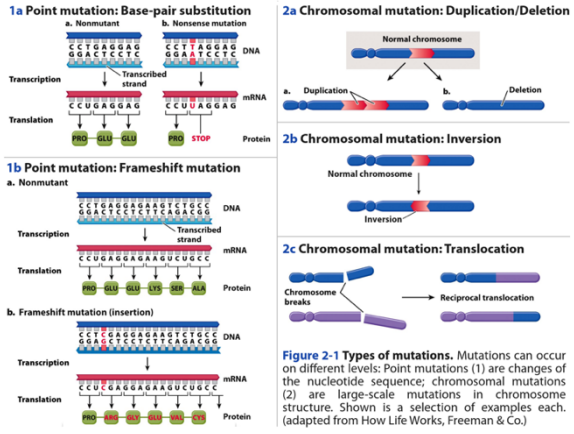


Figure 2-1 Types of mutations. Mutations can occur on different levels: Point mutations (1) are changes of the nucleotide sequence; chromosomal mutations (2) are large-scale mutations in chromosome structure. Shown is a selection of examples each. (adapted from How Life Works, Freeman & Co.)

we can define mutations based on their effect on the function of a gene:

- **lethal** mutation: organisms die prematurely → no normal development → mutation inactivates protein with essential function (in metabolism or development)
- **conditional** mutation: produces its phenotypic effect only under certain restrictive conditions
  - under other (permissive) conditions, the effect is not seen. For a temperature-sensitive mutation, the restrictive condition is typically high temperature, while at low temperature, the mutation does not show an effect.
  - Temperature-sensitive mutations generally destabilize the fold of the mutated protein, such that it loses function when exposed to higher thermal vibration
  - Other conditional mutations can be mutations required for the synthesis of amino acids.
    - In that case, the mutation has no consequence as long as the medium is a source of the amino acid.
    - it becomes lethal when the amino acid is depleted from the medium = auxotrophy mutations, meaning that unlike a prototroph, which does not need much more than a carbon and a nitrogen source (for example sugar and ammonium) to synthesize all other components needed to grow, these mutations require the addition of at least one additional nutrient.
- **loss**-of-function mutation: either reduces or abolishes the activity of the gene. (most common, normally recessive)
- **gain**-of-function mutation: increases the activity of the gene or makes it active in appropriate or inappropriate circumstances; these mutations are usually dominant.
- **Dominant-negative** mutation: dominant-acting mutation that blocks gene activity, causing a loss-of-function phenotype even in the presence of a normal copy of the gene. → This occurs when the mutant gene product interferes with the function of the normal gene product.
- **suppressor** mutation: suppresses the phenotypic effect of another mutation so that the double mutant seems normal. This can be due to a second mutation in the gene that is affected by the first mutation or due to a mutation in another gene, whose product interacts with the product of the first gene.

For genes that control pathways of development, gain-of-function mutations usually result in a dominant phenotype, while loss-of-function mutations mainly result in recessive phenotypes. Can you explain why this might be the case?

Diploid organisms have two copies of each gene. If one of those copies is mutated resulting in a loss-of-function mutation, the second copy can compensate for the loss. Therefore, loss-of-function mutations often result in recessive phenotypes. Gain-of-function mutations change the expression pattern of one gene copy such that it is expressed at an altered rate, in a different tissue or at another time. If a gene with such a mutation determines a certain developmental process, its expression is sufficient to activate this process. Therefore, gain-of-function mutations often show a dominant phenotype.

How mutations can cause a phenotype

Temperature sensitive (ts) mutations: Ts mutations are functional at low (permissive) temperatures, yet nonfunctional at high (restrictive) temperatures, and thus a rise in temperature quickly abolishes protein function. The mutant phenotype is usually due to a destabilization of the protein and the subsequent loss of protein function. (due to decreased melting temperature from the loss of a hydrophobic amino acid or due to a decreased ability to interact with DNA or other proteins, because the mutation removed an amino acid essential for interaction)

How would you use the replica-plating technique to screen for temperature-sensitive mutants?

Temperature-sensitive mutants can grow at low permissive temperatures (e.g., at 23˚C), but will die at high restrictive temperatures (e.g., at 37˚C). Therefore one would first mutagenize a cell population and grow the cells on plates without selective conditions. All cells - both temperature-sensitive mutants as well as 'normal' cells - will grow on this plate. In a next step the colonies are replica-plated onto two plates. One plate is kept at 23˚C (permissive temperature), the other plate at restrictive conditions, i.e., 37˚C. Colonies that grow only on the plate at 23˚C, but not at 37˚C are temperature-sensitive. The replica-plating technique is useful here for two reasons: First, the ts-mutants can only be identified by comparison of the two plates; second, the identification characteristic of the ts-mutants is non-growth at high temperatures. For further analysis of the colonies, one needs the other plate.

Dominant-negative (dn) mutations: If a *dn* mutation affects a transcriptional activator, the altered protein retains DNA binding activity, but lacks the ability to transactivate. It can complex with the DNA binding sites and displace the WT protein.

Gene Interactions: The concept of epistasis

Epistasis is the term that refers to the action of one gene upon another to affect a phenotype. **Gene *A* is epistatic over gene *B* if the phenotype of the double mutant is the same as the phenotype of mutants for gene *A*. (spielt keine Rolle ob im selben Pathway oder nicht)**

Epistasis occurs in the following scenarios:

- Whenever two or more loci interact to create new phenotypes
- Whenever an allele at one locus masks or modifies the effects of alleles at one or more other loci

Example 1: adenosine Pathway in yeast – epistasis between genes of same pathway

- Mutations in ade genes result in cells unable to synthesize adenine = adenine auxotrophs
- **ade2 mutant cells accumulate a red pigment = the substance AIR and appear as red colonies (while wild-type colonies are white.)**
- AIR accumulates because the Ade2 enzyme is normally required to carboxylate AIR on the way to adenine biosynthesis
- **BUT this red color phenotype can be reverted by an additional mutation in the ade3 gene.**

☞ Both the ade2/ade3 double mutants as well as ade3 single mutants form wild-type like white colonies. the ade3 mutation suppresses the red color phenotype of ade2. In genetics, we say that ade3 is epistatic to ade2

☞ and is therefore required for the formation of AIR. The ade3 gene is epistatic over the ade2 gene, which means that the Ade3 protein acts before/ upstream Ade2 in the same adenosine biogenesis pathway. In this situation, once a strain has lost the ade3 gene, the ade2 gene becomes irrelevant (the substrate for the Ade2 protein is anyhow not present) and can be mutated without the mutation having any phenotypic consequence.

☞ Importantly, the epistatic interaction between ade2 and ade3 only acts to reverse the red color phenotype. The adenine auxotrophy is not rescued in ade2/ade3double mutants, because both genes act in the same biosynthetic pathway to produce adenine.

Example 2: ERAD – epistasis between genes of different pathways

Randy Schekmann: developed a screening procedure to identify temperature-sensitive (TS) mutants that accumulate secretory vesicles intracellularly.

- Protein secretion starts with protein translocation into the ER where secretory proteins mature to adapt a functional 3-dimensional conformation before they are packaged into ER-to-Golgi transport vesicles.
- Proteins that fail to fold in the ER are not allowed to enter these vesicles, and are initially retained in the ER.
- Randy Schekmann identified a temperature-sensitive mutation in the gene SEC61 (the sec61-1 temperature-sensitive allele) that encodes the ER-embedded

translocation channel mediating protein import into the ER

- At 37°C, part of the Sec61 protein is misfolded, although the full protein still shows activity. The misfolded domain of the protein is recognized and leads the degradation of the entire protein by the ERAD machinery. Thus, sec61 mutant yeast cells do not grow at 37°C.
- In this case, it is not the mutation per se (the misfolded protein would still function, although at a limited rate), but the degradation of the misfolded protein that eventually kills the cell.

☞ the viability of sec61-1 mutant cells could be restored by introducing a secondary mutation that disrupts UBC6, a gene encoding an ubiquitin-conjugating enzyme that functions in the ERAD pathway. → Sec61-1 protein is no-longer degraded and can fulfill its function.

☞ Therefore, secondary mutations in ERAD suppress the phenotype of the sec61-1mutant strains.

☞ we can say that the lethality due to the sec61-1 mutation at 37°C is not caused by the fact that the mutation would inactivate the Sec61 protein but to how the cells react to the change in Sec61 organization, a change that they perceive as a stress.

What is the epistasis hierarchy in the case of sec61 and ubc6?

Gene A is epistatic over gene B if the phenotype of the double mutant is the same as the phenotype of mutants for gene A. Here, ubc6 is epistatic over sec61, because a mutation in ubc6 supresses the phenotype resulting from a mutation in sec61. The ubc6-sec61 double mutant shows the same phenotype as a ubc6 single mutant.

Gene interactions can lead to lethality

- Synthetic lethality arises when the combination of mutations in two or more genes leads to cell death, whereas a mutation in only one of these genes does not.
- One can use a mutation that confers a certain phenotype, but does not kill the cell, and systematically test other mutations to determine which ones confer lethality.

☞ Thus, synthetic lethality indicates functional relationships between genes or interactions between genes that function in the same biochemical process or between genes that act in pathways that appear to be unrelated, but affect the same ultimate process

Example synthetic lethality: arginine availability

- Arginine can either be imported through the CAN1 transporter, or synthesized by a set of enzymes (arg genes). The CAN1 transporter also transports the arginine analog canavanine, a component that is toxic for the cell. → mutations in CAN1 render cells canavanine-resistant.
- CAN1 mutant cells can survive as long as they can synthesize arginine.
- Mutations in any of the arg genes render cells auxotrophic for arginine, thus, these cells rely on the import of arginine from the medium to survive.

☞ However, when a mutation in any of the arg genes is combined with the can1 mutation, this leads to lethality even in medium containing arginine, since both pathways that assure arginine availability are disrupted.

GENETIC SCREENING: ISOLATION OF YEAST MUTANTS (OL)

forward genetic screen

Ablauf:

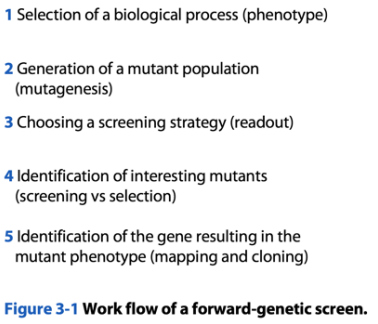


Figure 3-1 Work flow of a forward-genetic screen.

Mutagenesis

Mutationsfrequenz erhöhen:

- chemical mutagenesis: EMS = alkylating guanines to produce 6-O-ethylguanin → wird eingebaut bei DNA replikaiton → DNA Polymerase baut Thymin statt catosine ein → GC basenpaar wird zu AT Basenpaar → EMS macht Point mutation
- physical mutagenesis: irradiation: X-rays → DNA double strand beaks → deletions

Steps on the way to isolate mutants

- we want to do a screen to identify genes involved in regulating cell growth (1. Phenotype).
- We decide to use EMS as a mutagen due to its power to induce point mutations (2. Mutagenesis).
- Then, we have to decide for a screening method (3. Readout). → We could either measure the optical density of liquid cultures after a certain time point of growth or examine the size of yeast colonies using a microscope.
- Next, we have to identify the mutants for the desired phenotype (4. Screening versus selection), the most time consuming step in forward genetics.

- screening requires examination of large numbers of mutants and finding a few among them with the desired phenotype
    - in a screening procedure, one would screen for mutants that are sensitive to a specific drug, while in a selection experiment, one would select for mutants that are resistant to the drug and thus would survive.
    - Screening is time-consuming, but it allows the examination of every single colony, such that a wide range of mutations with different phenotypes can be recovered.
- a selection establishes conditions in which only the mutants of interest survive.
    - we discussed that mutations in the arginine transporter CAN1 cause resistance to the toxic arginine analog canavanine. Mutant cells are resistant to the drug because they cannot import it. Thus, mutants with impaired arginine uptake were isolated from yeast as canavanine-resistant strains simply by selection on canavanine-containing medium.
    - Selection saves a lot of time, because all the organisms that do not contain the desired mutation will not be examined; however, some mutations will not be recovered in a selection (for example, because they could survive, but grow too slowly).

Tetrades:

After the two haploid nuclei have fused (karyogamy) to form the diploid cell, this cell undergoes meiosis to form four haploid cells (spores) that are retained within an ascus (a kind of sac) and build a tetrad. This tetrad thus contains the four products of a single meiosis The tetrads can be dissected in order to separate the four spores and each spore is placed onto a specific position on an agar plate. The spores are then grown into colonies that can be tested phenotypically (e.g., checked if they show a mutant or wild type phenotype) or genotypically (e.g., sequenced).

The use of tetrad analysis in genetics = key tool in yeast genetics

- making predictions about the nature of the mutation; mapping mutations onto the chromosomes; determine whether two mutations are linked
- The power of tetrad analysis comes from the fact that all four meiotic products are contained within a single ascus.
- phenotypes that are not inherited in a Mendelian fashion must either be caused by more than one gene, or may follow a non-Mendelian type of inheritance, because the responsible gene is not located in the nucleus.
- If we follow how our mutation segregates relative to another known mutation (of which we know its genetic location), we can again see if the two genes are present on the same chromosome (then, they would be inherited together, because the two genes are segregated on the same chromosome during meiosis, and all spores would contain both genes, thus show the same phenotype)

Is the phenotype we observe in a mutant caused by a single gene or by several genes?

- If a single gene is the cause, the mutant allele should segregate in a 2:2 manner, such that half of the resulting spores will carry the mutant gene and show the mutant phenotype

Is the mutant phenotype dominant or recessive?

- the mutant cells need to be crossed with wild-type cells. If the phenotype caused by the mutation is recessive, the diploid cells resulting from the cross show the wild-type phenotype
- if the mutation results in a dominant phenotype, the diploid cells will show the mutant phenotype

Consider the mutation in the arginine transporter CAN1, which causes resistance to the toxic arginine analog canavanine. How would you find out whether the canavanine-resistance is recessive or dominant?

If the CAN1 phenotype was dominant, all diploid cells resulting from a cross of WT with CAN1 would be canavanine-resistant, thus grow on plates with canavanine. If the CAN1 phenotype was recessive, no cells would grow on canavanine.

How many genes has the screen identified all together? → complementation test

if one identifies the same mutations over and over again → screen is then called “saturated”.

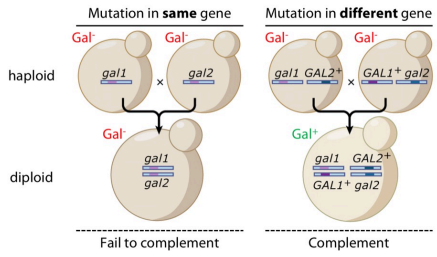
- By crossing one mutant with another, we can determine whether one mutation can complement the other one. If the mutants that were crossed contained a mutation in the same gene, both copies of the gene in the resulting diploid will be mutated and the diploid will therefore show the mutant phenotype. The mutations fail to complement each other.
- if the resulting diploid cell shows the wild-type phenotype, the mutations complement each other and are likely in two different genes. This is because the mutation causing the phenotype in the haploid can be

complemented for by the second wild-type allele of the gene in the diploid.

- A complementation group consists of mutations that do not complement each other. Thus, a complementation group is equivalent to a gene, and each mutation within a complementation group represents a mutant allele of the gene.

☞ Complementation tests can only be performed with recessive phenotypes, since mutations leading to a dominant phenotype will be manifested even in a heterozygous cell, such that whether the other mutation is on the same or a different gene does not affect the phenotype.

Example



How would the tetrads look like that arise when the diploid cells are sporulated after the complementation test shown in figure 3-8?

When the two mutations do not complement each other (left), all spores that result from the diploid would be Gal- (cannot grow without Galactose) → because each haploid spore would contain either the mutant allele gal1 or gal2. → Both mutations are in the same gene that is responsible for the Gal- phenotype.

If the two mutations do complement each other (right), the spores resulting from the diploid would be:
gal1 gal2
gal1 GAL2+
GAL1+ gal2
GAL1+ GAL2+

Thus, one spore (GAL1+ GAL2+) would be Gal+ (thus behave like wild-type cells) and grow on medium without galactose, while the other 3 would not (thus behave like Gal- mutants), because the presence of a single mutant allele (either gal1 or gal2) is sufficient to produce the phenotype. Thus, a ratio of segregation of 1:3 (wild type:mutant) tells us that the two mutations are on independent genes.

2 special cases where complementation tests do not show the expected results:

1) mutations in different genes may fail to complement each other if they act in the same pathway or are functionally connected.

→ A case of non-allelic non-complementation has been observed for example for tubulin: mutations in alpha- (A) and beta-tubulin (B) can lead to non-allelic non-complementation, because A and B bind to each other in a dimer (A/B). Point mutations in both genes can lead to the formation in the same cells of the dimers A/b, a/B, A/B, and a/b, of which only one (A/B) is functional. If, as it has been observed, the complex a/b is unstable and the other complexes are stable, then the complexes A/b and a/B become predominant and the cells show the mutant phenotype despite being heterozygous diploid. In that case, tetrad dissection shows a ratio of spores with a 3:1 ratio (mutant phenotype:wild-type phenotype), despite the two mutations not-complementing each other.

2) Mutations in the same gene may complement each other if the second mutation can reverse effect of the first mutation.
→ For example, when a multimeric protein is formed from subunits produced by different mutant alleles of the same gene. The resulting hybrid protein may exhibit greater enzymatic activity than is found in either of the single mutant proteins. Although each of the mutant genes produces defective gene products, the defective products may associate to produce a product that has sufficient activity to promote function.
This is particularly easy to understand in the case of a protein that assembles into a dimer, and where protein A is made of two functionally separable domains A' and A".
→ one mutant protein may be defective in A' (written a'-A") and the other in A" (A'-a"). In each haploid strain, the dimer is in the form a'-A"/a'-A" and A'-a"/A'-a" and is non-functional. In the heterozygous diploid, half of the dimers formed are a'-A"/A'-a" and are therefore functional.
→ As a consequence, we observe complementation despite both mutations being in the same gene. **Tetrad analysis would reveal that fact, because it leads to a 4:0 segregation, where all spores show a mutant phenotype, indicating that the two mutations are tightly linked (located on the same gene).**

☞ it is important to back up the complementation assays with segregation tests to confirm that mutations that do not complement each other are indeed in the same gene or vice-versa that mutations that do complement each other are indeed in different genes. → but, non-allelic non-complementation and allelic complementation are overall rare events.

EXAMPLES FOR GENETIC SCREENS IN YEAST (OL)

Screen to identify genes involved in the cell cycle

- first by Hartwell & Paul Nurse
- Hartwell assumed that mutations affecting the cell cycle would be lethal → they looked for conditional mutations that allowed growth at 23°C, but not at 36°C
- Upon increasing the temperature to restrictive conditions, cells mutant for cell-cycle genes continue to go through their cell cycle until they reach the point where the cell can no longer progress in the division process if the function of the mutated gene has not been completed. → arrest in cell cylce
- phenotypes: all unbudded cells (G1 arrest), all cells with a bud growing from the apex (elongated buds, S-phase entry), large budded cells with a single nucleus (defect in DNA replication, mitotic entry, mitosis), large budded cells with a divided nucleus (mitotic exit), chains of non-separated cells (defective in cytokinesis) => all cdc mutants
- The cell-cycle stage was identified both morphologically, but also by staining the DNA to quantify if the cells were haploid or diploid.
- example: mutant cells that had no bud and a single nucleus with a haploid DNA content indicated that they did not replicate their DNA and failed to enter the cell cycle (cdc28)
- mutants that formed large un-budded cells containing several nuclei failed to bud, but had entered mitosis leading to nuclear division
- +) mutants blocked at a certain stage of the cell cycle can be readily isolated by morphological screening, because yeast cells divide by budding, a process that can easily be followed using a light microscope.

How would cells look like that have a defect in DNA replication and thus fail to enter mitosis?

can not undergo mitosis and will arrest with a large bud, but a single nucleus with haploid DNA content.

How would cells look like that are unable to exit mitosis?

arrest with a large bud and divided nucleus.

Screen for genes involved in yeast metabolism

- Yeasts are able to synthesize all essential compounds, falls inorganische Quelle vorliegt
- screen aims at identifying mutants that are unable to synthesize the pathway metabolites and depend on importing these compounds from the medium (we say that these cells are auxotrophic for these compounds).
- Genes acting in such a metabolic pathway are conditionally essential:
  - In minimal medium the genes are essential, while in rich medium the genes are not required. → one can grow the mutants initially under permissive conditions (rich medium) and screen later for a lethal phenotype under restrictive conditions (medium lacking a specific compound).

example:

→ adenine biosynthesis, since this is a commonly used readout for several genetic screens.

→ already discussed: mutants defective in the gene ade2 accumulate an intermediate compound AIR in the adenine-biosynthetic pathway which turns the respective yeast colony red.

- The mutants that turn red are those in which either one of the two steps just after AIR in the pathway is blocked.
- . A mutation in either step causes AIR to accumulate in the cell. AIR itself is not red, but cells that grow aerobically oxidize it to a red pigment. → red mutant = ade1 & ade2 (ADE2-)
- In the presence of excess adenine, the biosynthetic pathway shuts down and ade mutant cells will remain white even in the presence of glucose and oxygen.
- When red mutants are grown in the **absence** of oxygen, they do not turn pink or red. **However, when the anaerobically grown colonies are exposed to air, they turn red within a couple of hours.**
- Some of these white mutants form smaller colonies and are called "petite colony mutants ", or just petites. Rather than being blocked in the adenine pathway they fail to make the red pigment, because they lack oxidative metabolism.

ZF: Genetically, any mutation that blocks the formation of AIR and therefore adenine, or any mutation that impairs oxidative metabolism, will block formation of the red pigment.

Screen for genes implicated in the secretory pathway

Randy Scheckmann: isolate mutants affecting the late secretory pathway

- Since the late secretory pathway is essential for many cellular characteristics (cell wall) mutations in this pathway are likely to be lethal. → developed a screening procedure to identify temperature-sensitive (ts) mutants that accumulate intracellular vesicles.
- cells **deficient** in secretion accumulate proteins and lipids inside the cells not being able to increase their

surface, and hence, their volume → cells are denser and can be separated by density gradient centrifugation

- → mutations they identified included those where the traffic from the ER (sec12), from the Golgi (sec7) or from the plasma membrane (sec1) were blocked (later steps of secretion)
- To isolate mutants in the early secretory pathway,they utilized the fact that in eukaryotes, proteins destined to be secreted have signal sequences at their amino terminus. → the signal sequence interacts with cellular factors that stop translation and target the nascent polypeptide chain to the ER. Once brought to the ER membrane, the new protein is usually co-translationally translocated into the ER.
- they fused a signal sequence to the DNA sequence encoding histidinol dehydrogenase (HIS4C), Histidinol is present only in the cytoplasm, and the ER membrane is impermeable to histidinol and histidine
- In yeast strains where the wild-type enzyme is replaced by this genetically engineered HIS4C-fusion enzyme, the **enzyme is directed to the ER where it cannot perform its function, because its substrate histidinol is not present.**
- Cells with mutations that affect the mechanism for translocating proteins into the ER should thus make ER localization of the modified enzyme less efficient and permit cytoplasmic retention of at least part of the enzyme; thus, these cells would survive even in the absence of histidine
- Note that a full inactivation of such genes, leading to a complete block of translocation, would not be viable.
- → such a screen **aimed** at identifying partial loss-of-function or temperature-sensitive mutations that already show a phenotype at a semi-restrictive temperature.

☞ The selection of yeast mutants that were able to grow on medium lacking histidine when containing the HIS4C-fusion protein allowed the identification of genes encoding proteins involved in the translocation of proteins into the ER membrane → Sec61 (channel forming of ER translocase)

mutant translocation apparatus

Alberts, 5th edition, Garland Science)

Screen for drug resistances

Cycloheximide resistance → inhibiting translation by inhibiting the translocation of the ribosome along the mRNA → no elongation of polypeptide-chain

- McCusker & Haber
- When S. cerevisiae cells **are exposed to high levels of cycloheximide, they stop both protein synthesis and growth.**
- Application of the **minimum inhibitory concentration of cyloheximide causes cells to stop growing, but protein synthesis is not completely inhibited.** → indicates that the cycloheximide-resistant temperature-sensitive (crl) mutants affect **structural genes encoding ribosomal proteins**
- crl mutants were also found to include mutations in proteasomal genes or in components that may act in the ubiquitination system → some of the phenotypes of crl mutants are resulting from defective protein degradation.
- cyh1R mutation and the gene CYH1 encodes the ribosomal subunit that binds cycloheximide,fully restores growth of the cells at all temperatures in the presence of high doses of cycloheximide.
- In heterozygous diploids cells, i.e., cells expressing both the wild type CYH1 gene (called CYH1S) and the cyhR allele, half of their ribosomes bind cycloheximide.

Assessing the information about the cyh1R mutation given above, can you deduce whether the cyhR allele is dominant or recessive?

Although only 50% of the ribosomes are blocked by cycloheximide in heterozygous diploid cells → only diploid cells homozygous for the cyh1R mutation are resistant to cycloheximide and thus thecyhR allele is recessive.

Rapamycin resistance

- S. cerevisiae cells treated with rapamycin irreversibly arrest in the G1 phase of the cell cycle
- Hall: made screen to isolate rapamycin-resistant yeast mutants in order to identify the cellular targets of rapamycin
- mutations found could be mapped to two genes that were named TOR (Target of rapamycin) 1 and 2. TOR1 and TOR2 were later identified as serine/threonine kinases controlling growth, survival and metabolism
- in mammal: mTOR is target of rapamycin, activated by PKC (protein kinase B)
- This study shows how findings from the model organism yeast can be transferred to mammals, because many yeast proteins have orthologs in higher organisms

GENE MAPPING AND IDENTIFICATION (OL)

Gene mapping using tetrad analysis

-) sequencing all mutations is too expensive → bcs we don't know where mutation is located → sequence always whole genome → also need to sequence more than one time for minimizing sequencing errors

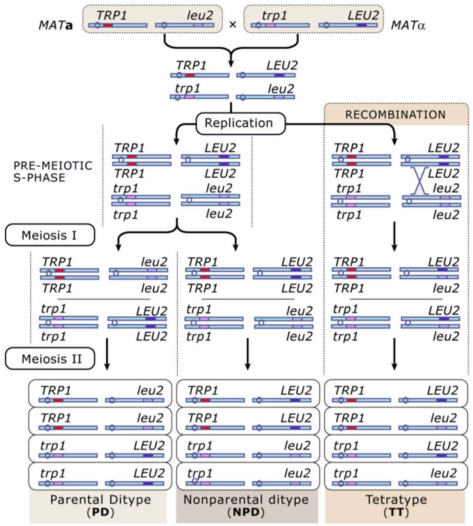
-) sequencing the genome of a mutant strain generally reveals many mutations when compared to a standard yeast genome; because the strain used in the screen keeps itself to evolve and already has a different genotype compared to the strain that was sequenced for the data base. → If more than one mutation is detected when comparing the mutant genome with a reference genome, one still does not know which mutation causes the phenotype.

genetic methods are used to narrow down the regions of the genome where the mutation is located → by linkage analysis (distance between two genes in the genome (the unknown gene affected in a mutant found in our screen and a known gene) is determined by measuring the frequency of recombination between the two genes occurring during meiosis) => linked = on the same chromosome

+) Yeast is well suited for these analyses, because the 4 spores in an ascus are the result of a single meiotic event → allows formulating specific predictions regarding the segregation patterns that will occur depending on whether two genes are linked or unlinked.

the 3 types of tetrads that can occur when two genes are segregating during meiosis: they arise as a result of the normal segregation of the markers: the parental ditype, the nonparental ditype, and the tetratype

example 1: the two genes are not linked to each other.



The two mutations are called a, on chromosome III, and b, on chromosome IV. (+ = WT); (The phenotypes are usually written as Trp- (a tryptophan auxotroph that requires tryptophan for growth) or Trp+ (a wild-type cell that can grow without tryptophan).
In the absence of recombination, there are two possible outcomes for the tetrads.

i) diparietal type (2 kinds of parental spores)
the two parental sets of markers segregate away from each other at meiosis I. → resulting tetrads have two spores bearing the genotype of one parent (TRP1-leu2) and the other two spores bearing the genotype of the other parent (trp1-LEU2)

ii) Non-parental ditype (NPD)
results again in a tetrad with two different classes of spores, AB (TRP1-LEU2) and ab(trp1-leu2). This class of tetrad that has two kinds of nonparental spores
→ Because segregation at meiosis I is random and each configuration of homolog pairs occurs at equal frequency, for two unlinked genes, the number of occurring PD tetrads and NPD tetrads is equal. **This prediction is one of the fundamental tests of genetic linkage in yeast.**

iii) Tetratype (TT)
arises as the result of crossing-over events or recombination between one of the genes and the centromere of the chromosome on which it is located → results in a tetrad that contains four spores, each with a distinct genotype. In our example of the unlinked A and B genes, the four kinds of spores will be AB (TRP1-LEU2), ab (trp1-leu2), Ab (TRP1-leu2) and aB (trp1-LEU2).
⇨ Such a tetrad that has four genotypically different spores is called a tetratype (TT).

→ Always first question tsting for linkage: "Is the number of PDs equal to the number of NPDs?" If the answer is yes, the mutations (and therefore the genes) are unlinked. If the answer is no, there is linkage.

summarize the outcome of linkage analyses as follows:
Unlinked gene pairs, and no crossing over between either gene and its centromere, yields
PD = NPD and TT = 0.

Unlinked gene pairs, with a crossing over between either gene and its centromere yields
PD = NPD and TT > 0.

example 2: the 2 genes A and B are located on the same chromosome and positioned relatively close to each other, that there is less chance for crossing over (A closer to the centromere than B)

- if we cross a wild type of genotype Ab with another of genotype aB, in >50% of the tetrads, there was no crossover at all between the two loci, such that after meiosis I, one nucleus contains the sister chromatids Ab/Ab and the other one contains the sister chromatids aB/aB.
  - Therefore, the A nucleus will contain one chromosome Ab and the other AB, while the a mother will have one chromosome aB and the other being ab.
  - Among the few cells that make **two** crossovers between A and B, some of them will make these crossovers across independent sister chromatids, such that after meiosis I, one nucleus will have two AB sisters, and the other will have two ab sisters. This results in the formation of a tetrad of the type NPD. (very rare)

- After meiosis II, we will therefore have two spores of genotype Ab and two spores of genotype aB. Hence, >50% of the tetrads resulting from the cross will be PD
  - After meiosis II, this combination produces a tetrad of type TT (Ab, AB, aB, ab).

→ As a consequence, for linked genes (less than 25% chance of crossover) PD>TT>NPD

example 3: two loci A and B are completely unlinked to each other and far from centromeres.

These loci can rearrange in all possible manners with the same probabilities. This can lead to 6 tetrads → all come with the same probability, this means that we if two genes are completely unlinked the proportions of the different tetrads will be 1:1:4 (PD:NPD:TT).

examining the tetrad patterns we observe across multiple asci, we can determine

i) whether two genes in the yeast genome are unlinked, tightly linked, or loosely linked to each other,

ii) whether both genes are tightly or loosely linked to the centromere, or iii) whether one or both genes are not linked to the centromere

Crossover frequency and genetic distance

how the frequency of tetrad types can be used to physically map the position of a gene by linkage analysis:

Use phenomenon of crossovers and the resulting frequency of tetrad types to determine the order of genes along a chromosome and the ~ distances between those genes.

This depends on the following basic principles:

- ☞ Crossovers are equally likely to occur at any point along the length of a chromosome.
- ☞ **The probability of a crossover between two genes is proportional to the distance between the two genes.** → the greater the distance between the two genes, the greater the probability that a crossover occurs between them during meiosis.

→ *The number of recombinants acts as a measure for the distance between the two genes that took part in the crossover event, measured in genetic map unit m.u. or centimorgan cM (distance between genes for which one product of meiosis in 100 is recombinant; recombination frequency (RF) of 1% = 1 centimorgan.)*

Bsp: Linkage is determined by the number of recombinants divided by the number of total progeny.
In our example of peas, the frequency of crossovers is 10 recombinants per 100 total progeny, thus 10%, which is equivalent to a distance between the genes of 10 centimorgans.

Calculate gene centromere linkage from tetrad frequencies

possible to map the position of a gene relative to its centromere, **provided that at least one known centromere-linked gene is present in the cross** → This is because the centromeres of non-sister chromatids (from homologous chromosomes) always segregate away from each other in meiosis I, and the centromeres of sister chromatids always segregate away from each other in meiosis II.

When tetratype (TT) tetrads are produced, you have seen above this indicates that a crossover has occurred between a gene and its centromere.

imagine two genes are on different chromosomes and therefore unlinked. The only case in which one does not observe tetratypes is when **both genes are very tightly linked to the centromere**.

studies to determine gene-centromere linkage are done with a gene of interest and a mutant in a known centromere-linked gene. → z.B trp1 (causes an auxotrophic phenotype requiring trp for survival, is close to the centromere of chr. 4) →trp1 is only one map unit from its centromere, meaning

that only one percent of the tetrads will show a recombination event between trp1 and its centromere.

example 1: two genes are very tightly linked to their centromeres (but genes not linked

- → *trp1 is tightly centromere-linked and is on chromosome IV.*
- → *A second centromere-linked gene is met14, which lies on chromosome XI.*
- → *Since these two genes are on different* ***chromosomes, they are not linked to each other.***

☞ *Therefore, in a cross of trp1 and met14, tetrad analysis* ***will show that PD = NPD.***

☞ *However, because there will be very little recombination between either gene and its centromere, the frequency of TT tetrads will be extremely low.*

By determining the frequency of TT tetrads, the possible linkage of any marker to its centromere can be determined by a genetic cross using the gene of interest and a known marker that is not linked to the marker of interest on a different chromosome, but known to be tightly centromere-linked, z.B trp1 gene

→ *Therefore, the TT tetrads in this cross will arise almost* ***entirely by crossovers between*** *the unknown gene and its centromere, as shown for leu2.*

for example:

cross between trp1 and a hypothetical mutation in gene B. We obtain the following numbers of tetrads:

40 PD; 40 NPD; and 20 TT.

What does this mean for the possible centromere linkage of B?

Because PD = NPD, we know that trp1 and B are unlinked to each other.

What about the linkage of B to the centromere? (Beispielrechnung)

linkage is determined by the number of recombinants divided by the number of total progeny. → *there are 20 TT tetrads and a total of 100 tetrads. Because each tetrad in this case has 4 spores, two of which are recombinant, there are a* ***total of 40 recombinant spores. The total progeny equals 400 spores (from the 100 total tetrads). Therefore, the linkage is calculated by 40 recombinants/400 total progeny = 10 map units. Thus, B is 10 centimorgan away from its centromere.***

For ***yeast****, this* ***mapping function is*** *usually expressed by an equivalent function in terms of tetrads: linkage = (½ · TT)/number of total tetrads. In our example, (½ · 20)/100 = 10%, which corresponds to 10 centimorgan (cM).*

How does one , in practice, determine whether a tetrad is PD/NPD/TT?

one isolates the spores by dissecting each ascus, aligns the four spores of each ascus in one row on an agar plate and grows them to colonies. Depending on the gene of interest, selection criteria are introduced by replica-plating in order to determine the phenotype of each spore. By comparison with the parent generation, the type of the tetrad can be determined.

example 2: two genes are linked to each other, but not to their centromeres (less than 25% crossing overs = less than 25 centimorgan)

- extreme cases when two genes are completely linked, with no detectable recombination between them, we expect that tetrad analysis **will reveal that all tetrads are PD tetrads**.
- As for centromere linkage, in the case of closely linked genes, we can calculate the degree of linkage based on the frequencies of the different types of tetrads produced.
- **If there are no NPD, than the distance is as above for the distance to the centromere**:
  → linkage = (½ · TT)/number of total tetrads in centimorgan.
- In case there are a few NPDs, then the calculation leads to the **Perkins formula**, which we will not demonstrate here: **linkage = ½ · (TT + 6 · NPD)/number of total tetrads.**

Here are some results of tetrads arising from crosses between the genes A and B. Using these data, determine for each case whether A and B are linked and whether both genes are centromere-linked.

PD	NPD	TT	Are A and B linked?	Are both A and B centomere-linked?
10	10	40		
30	30	0		
20	20	20		
50	0	10		

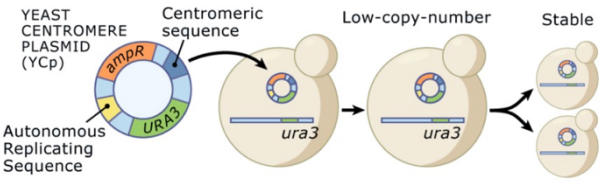
The correct answer is:

PD	NPD	TT	Are A and B linked?	Are both A and B centromere-linked?
10	10	40	No	No
30	30	0	No	Yes, because PD:NPD:TT is 1:1:<4
20	20	20	No	Yes, because PD:NPD:TT is 1:1:<4
50	0	10	Yes	Cannot be determined, because A and B are linked genes

GENE INTERACTIONS (OL)

Gene identification by complementation analysis

- Complementation analysis is based on the fact that reintroduction of a wild-type copy of the mutated gene should revert the mutant phenotype.
- <u>complementation analysis only works with recessive phenotypes, since diploids resulting from a cross of mutant and wild-type cells still display the phenotype</u>.
- We use a recombinant DNA library, a mixture of about 10^8 different plasmids with each containing a different fragment of the yeast genome.
- a selectable marker for transformation of bacteria (e.g., ampicillin resistance) and a selectable marker for transformation of S. cerevisiae that allows identification of the cells that have taken up the plasmid
    - Often, the URA3 gene is used as a marker that can complement an uracil auxotrophy. → the strains used are ura3- and that only cells containing the plasmid with the URA3(+) gene can grow without uracil.
    - the plasmids contain an autonomously replicating sequence (ARS) (such that it is replicated by the yeast cell) and a yeast centromeric DNA (CEN4).
    - yeast centromeric DNA required for the plasmid to symmetrically propagate to mother and daughter cell, such that it is maintained in all cells of the colony



- For the complementation system to work, the yeast **cells used to identify a gene must contain two particular genetic markers.**
- i) must contain the recessive mutation of interest (the one isolated in screen).
- ii) important marker is one used to select transformants, such as a ura3mutation. Cells lacking a functional ura3 gene cannot synthesize uracil and depend on uracil in the medium to survive (they are uracil auxotrophs).
- transformed cells are selected on plates lacking uracil in the media.
- those cells that survive 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

Problem:

- 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 → contains the gene that is causing the mutant phenotype. → 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.**

☞ <u>to determine the precise genomic segment</u> in the recombinant plasmid → sequencing the DNA at the ends of the complementing plasmid → searching a database of the entire yeast genome sequence, we can identify the exact genomic fragment in the plasmid → reveal the genes present in the insert that was able to complement a certain mutation.

☞ <u>which</u> gene on the plasmid is the one that complements <u>the mutation, 2 methods:</u>

1) 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.
2) perform classical complementation tests between the mutant of interest and mutations in each of the candidate genes. → 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

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.

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 → will also show whether or not the gene is essential for viability → yeast is especially suited to make

such deletions because of their efficient homologous recombination system

Erzeugung einer Deletion:

precise deletion of any yeast gene can be made using a DNA fragment synthesized by PCR. → a PCR reaction is set up to amplify a selectable marker (*) **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. → 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.

→ **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. (no essential genes)**

→ **if gene is not in the collection = mutation lethal**

→ every deletion can be made, because the whole genome of S. cerevisiae is known

→ genome editing also used to tag genes with GFP

(*: antibiotic resistance markers, here against the compound G418 or auxotrophic markers (lacking his3/leu2/trp1/ura3 gene – lack synthesis of these amino acids), erlaubt selection für PCR Produkt)

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.

Identification of open reading frames

- An open reading frame (ORF) has the potential to code for a protein or a peptide
- An ORF can be detected within a DNA sequence, because it does not contain a stop codon.
- 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 (man liest immer von 5' nach 3' aber kann an 3 versch. Positionen starten, dies mal 2 wegen 2 Stränge)
- Start codons appear in the DNA with a frequency of 1/64, and stop codons occur with a frequency of 3/64 (because there are three different stop codons).
- ORFs were defined to be present if the distance between start and stop is > 300nt = 100 AS (shorter peptides are not identified even though they exist, smallest = 10AS)
- Alternatively → purifying mRNAs from cells and analyzing them to deduce which parts of the genome is transcribed.
- conserved regions are likely to correspond to ORFs (compare genomes of related organisms)
- 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.

→ 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 (Gene interaction) and suppressor analysis

If 2 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

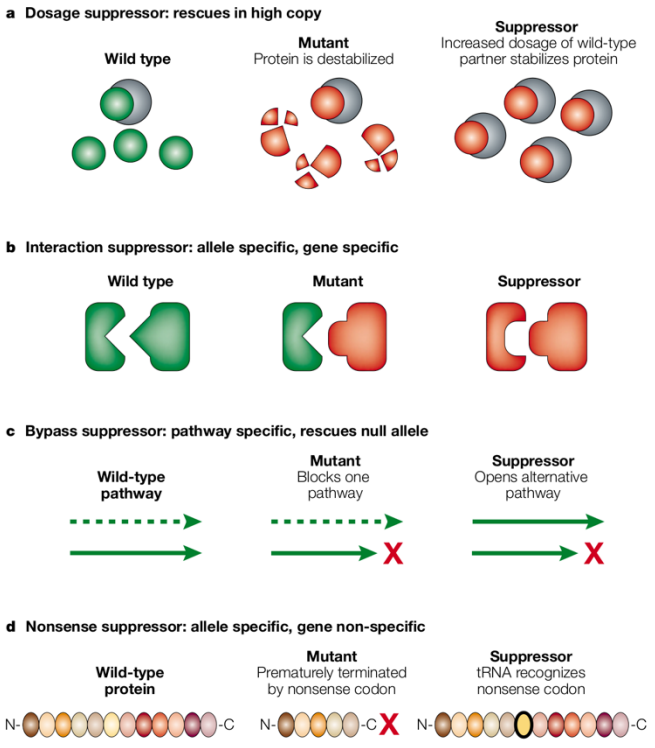


Figure 6-6 Suppresor mechanisms. Depending on the allele and gene specificity associated with suppressors, mechanisms can be inferred, as shown. **(a)** Dosage suppressors encode proteins that stabilize the mutant product when they are expressed at high levels. **(b)** An interaction suppressor restores 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)

epistatic with respect to one another = Epistasis analysis is the analysis of the interaction between two or more genes that control a single phenotype.

→ Epistatic analysis requires the two mutants to have distinguishable phenotypes

→ comparing the single mutant phenotypes with the double mutant phenotype can determine their relative order of function

⇨ epistasis is used to identify interacting genes in a technique called suppressor analysis.

⇨ Supressor Identification:

⇨ 1) isolation of "revertants" of the original mutant – strains that no longer show the original mutant phenotype.

- **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.**

⇨ 2) 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

- For true revertants, all tetrads will be PD tetrads, since all spores will inherit a wild-type allele.
- 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.

⇨ the type of mutation that is being suppressed will determine the nature of the suppressor mutation.

- a mutation altering the conformation of a protein might be suppressed by a compensatory change in an interacting protein
- 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.

⇨ 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.

Synthetic lethality

- where the double mutant is lethal under conditions in which both single-mutant parents are viable
- each of the two mutations impairs an aspect of an essential process -> double mutant has a more severe phenotype than either of the single mutants.

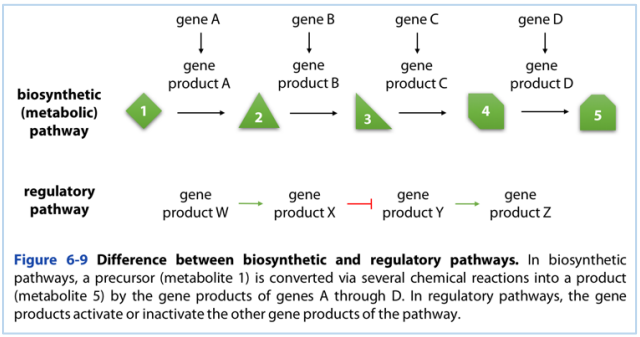


Figure 6-9 Difference between biosynthetic and regulatory pathways. In biosynthetic pathways, a precursor (metabolite 1) is converted via several chemical reactions into a product (metabolite 5) by the gene products of genes A through D. In regulatory pathways, the gene products activate or inactivate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

- This approach can be useful to uncover redundant pathways or direct protein interactions.
- 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,
  - genes whose products independently contribute to the same process.
- 1) Discover redundant genes:
  - the mutation is crossed with other mutations and the resulting phenotype is observed. → If there is a single additional gene that is redundant, double mutant will show the phenotype, either single mutant would be normal.
- → 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.
- 2) Synthetic lethality thus serves as a tool to identify direct interactions between gene products/ complexes:
  - synthetic lethality can also be observed between genes whose products interact with each other. → 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
  - If a double mutant that alters two proteins reduces the activity below a threshold that is not enough to ensure survival, the cells die.
- Complementation:
  - 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.
  - double mutants with mutations in different genes that individually result in the same phenotype fail to complement each other in the heterozygous diploid = non-allelic non-complementation. → indicates that two genes encode physically interacting proteins.

Ordering genes in a pathway

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?

→ two different types of biological pathways: biosynthetic (metabolic) and regulatory pathways. → Both types of pathways involve multiple functional components that act in a defined sequence

metabolic/biosynthetic pathway:

- Each step in a metabolic/biosynthetic pathway converts one intermediate to another: the "gene" corresponds to the catalyst (enzyme), rather than the intermediate(s).
- Each step in a biosynthetic pathway is required for the subsequent step(s).
- 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
- **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.**

  example: 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.

Regulatory pathway

- Each step in a regulatory pathway activates (+) or inactivates (-) a "switch", which is usually the next "gene" in the pathway
- A given step (switch) in a regulatory pathway may act either positively or negatively on a subsequent step.
- In regulatory pathways, mutations will affect the pathway in one of two opposite ways: either activation or inactivation.
- **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.**

example *where epistasis helped to analyze the functional relationships of genes regulating mitosis: The G2/M transition will occur when kinase Cdc28 is phosphorylated by another kinase by another kinase (made by SWE1) 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 show delay, but rather advances the timing of mitosis.***

→ *a mutation in the gene mih1 shows a significant delay in G2 phase (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. → so wirkt ein loss von Swe1 (swe1 mutation) aktivierend und ein Loss von Mih1 (mih1 mutation) inhibierend, wodurch Mitose nicht ausgelöst werden kann

→ *When double mutants of mih1 and swe1 were constructed, the double* ***mutant showed a phenotype similar to the swe1 single mutant.***

From these results, can you deduce the epistatic relationship of MIH1 and SWE1 in the pathway regulating mitosis?

Since the double mutant phenotype looks like the SWE1 phenotype, SWE1 is epistatic to MIH1. For both genes acting in a regulatory pathway, the downstream gene will be epistatic to the upstream one; thus, SWE1 is supposed to act downstream of MIH1.

NON-MENDELIAN INHERITANCE (OL)

A review of Mendel's Laws

→ to predict the probability that the offspring will have a particular genotype

- Law of Segregation, in a diploid organism that produces gametes by meiosis, each allele is segregated into a different gamete.
- The Law of Independent Assortment says that this segregation of alleles happens independently of other alleles. We now know that this is true for genetic loci that are not linked

Non-Mendelian inheritance:

genes that do not segregate in a Mendelian manner. This is true for genes that ***which are not located in the nucleus, such as mitochondrial genes, genes in plastids in plants or on other cytoplasmic structures, such as plasmids***

The key characteristic of non-Mendelian inheritance in eukaryotes is that the origin of the cytoplasm (and not the nuclear DNA) determines the phenotype.

- Cytoplasmic inheritance can be due to organelles or any other particles that can be passed on to the next cell generation and affect its phenotype.
- Most importantly for genetic analyses, mitochondria and chloroplasts have their own genomes and use their own transcriptional and translational machinery. A typical cell has **many** copies of the organellar genome.
- **In actively dividing cells, these organelles and their genomes must double, be partitioned at mitosis and passed on to the next generation → Thus, they will have special patterns of inheritance.**

Mitochondrial Inheritance

In most higher eukaryotes, the female parent will provide the organelles for the offspring.

→ the egg cell contributes much more cytoplasm to the zygote than the male gamete,
→ mitochondrial inheritance in animals and plants is **uniparental** and is passed to the next by **maternal inheritance** = little opportunity for genetic recombination (good for evolutionary history)
→ transmission of diseases caused by mitochondrial DNA occurs only via the mother.

→ In yeast, however, both haploid cells contribute equal amounts of mitochondria to the diploid zygote.
→ After a few generations, the mitochondria of a given cell generally contains mitochondrial DNA only from one or the other parent cell => for mitochondrial genes it is possible for all four of the cells that result from meiosis to inherit their mitochondrial genes **from only one of the 2 original haploid cells = non mendelian inheritance (all 4 haploid cells have the same mitochondrial genes)**

The use of yeast in studying mitochondrial genes

S. cerevisiae is an excellent model organism to genetically study mitochondrial function, because it is capable of satisfying its energy requirements with **ATP generated by fermentation = without respiration and without needing a functional mitochondrion.**
→ presence of the mitochondrial genome and oxidative phosphorylation are dispensable as long as fermentable carbon sources, are present in medium.

⇨ **yeast can survive mutations in mitochondrial genes that arrest oxidative phosphorylation**

⇨ **This means that mutations that impair mitochondrial function can be isolated and identified by their conditional growth: they grow on fermentable carbon sources but not on non-fermentable ones.** (mit carbon source kann auch ohne Mitochondrium ATP gezeugt warden, fehlt jedoch die fermentierbare Kohlenstoffquelle, stirbt die Hefe) (Eukaryot könnte das nicht da nicht zur Fermetation fähig)

⇨ Studieren von Genen auf mtDNA ist in Hefe gegenüber Eukaryot zusätzlich vereinfacht, da bei Hefe keine uniparentale Inheritance vorliegt = In yeast, however, mtDNA inheritance is biparental and thus, mtDNA transmission can be easily followed by crossing mutant and wild-type yeast strains.

⇨ The mitochondrial DNA of S. cerevisiae contains genes required for some mitochondrial functions → genes encode tRNAs, rRNAs, cytochrome oxidase, and an ATPase.

⇨ The remaining mitochondrial components are encoded in nuclear genes, including the genes encoding the subunits for mitochondrial RNA polymerase and mitochondrial ribosomes.

⇨ Mutants with lesions in the mitochondrial genome have so-called rho mutations. **These mutations segregate in a different fashion from nuclear pet mutations, indicating cytoplasmic inheritance.**

- **Rho mutations that are the result of the complete loss of mtDNA are called rho0 mutations.**
- Mutations where only part of the mtDNA is deleted are **called rho-.**
- If a rho0 strain is crossed by a wild-type strain (rho+) → The resulting four spores in each tetrad all contain wild-type mitochondria. Therefore, the **segregation is 4:0 rho+:rho0**
  - Rho- mutations can be used in complementation assays to detect whether a certain gene can complement and reverse the mutation, thus restoring mitochondrial function.
- When crossing a rho- mutant strain with a rho+ wild type strain, in most cases, 50% of the tetrads will show a 4:0 and 50% a 0:4 (rho+:rho-) segregation pattern.
  - In some cases, some rho- mutants show a bias towards the 4:0 (0:4>50%), indicating that the mutated genome proliferates faster than the wild type one and takes over.
  - **=> These mutant mtDNAs are called hyper-suppressor.** They generally corresponds to a mtDNA where the origin of replication of the

mitochondrial **genome is amplified through multiple tandem repetitions.**

⇨ For the identified mutants, it is straightforward to determine whether the affected gene is located on nuclear or mitochondrial DNA just by following the phenotypic segregation in tetrad analysis (4:0 when gene located on mitochondrial DNA)

QUESTION AND ANSWERING PART 1 (28/29.09.20)

Einblick in online lessons

Questions on powerpoint:

Gene-centromere linkage

- Group 1: Calculating gene-centromere linkage was unclear to us. Can you please explain the calculation of gene-centromere linkage again?
- Group 4: Gene mapping " How can you calculate the distance from gene to centromere, having the proportion of tetrads, if there are more than 2 genes of interest?"
- Group 8: Gene mapping: How to determine if two genes are centromere-linked (e.g. how do we get to 1:1:0<4)
- How do crossing-overs happen on the linked genes

Mating type switching

- Group 3: Our question is regarding sex switching. Can a cell change sex an infinite number of times and if so, what about gene stability/instability?
- Group 6: While changing a mating type how does the cell know which silenced locus it should use in order not to obtain the same mating type?
- Group 2: Yeasts can switch their mating type. How do they actually ensure that the mating type is changed since both copies are present in heterochromatin? Why can't it change from a to a again? And what triggers the change of mating type?

Epistasis

- Group 7: Could you please explain the concept of epistasis again using ubc6 and Sec61 as an example? Is it also called epistasis if the two genes regulationg don't interact with each other?
- Group 8: General concept of complementation could be explained a bit more.

Non mendelian heritage

- Group 5: We would like to have an introduction to non mendelian heritage. (Examples of M and non-M, and compare and contrast. Could you please explain the figure 6-6 in the part gene interaction?

Additional Question: Is yeast defined as heterothallic or homothallic on a single-cell or on a strain level?

Frage 1:

Chromosome 1 with centromere (CEN) and gene linked to centromere (Gene A = WT)

Chromosome 2 and Gene b (mutant) which is also close CEN

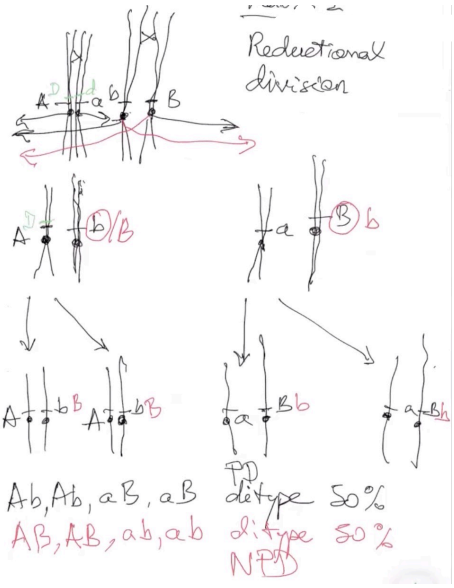
Chromosome 3 and Gene C which is far from CEN

Ab x aB -> how frequently are A & b recombined?

→ since A & b are very close to CEN = very little chance of crossing-over → Meiosis 1 = **reductional division (homologue separation)** → Meiosis 2: Mendelian division, da not reduced number of alleles, ditype

→ no parental ditype: Tetrade = AB, AB, ab, ab

→ parental ditype: Tetrade = Ab, Ab, aB, aB



Ac x aC -> Meiotic products - 2 crossing over events occur!

1) Möglichkeit: crossing over between genes:

Meiose 2 Produkte: Tetratype = AC, Ac, aC, ac (da Allel C und c crossing over gemacht haben! NPD ≠ PD

2) Möglichkeit: crossing over nicht zw. genes:

Meiose 2 Produkte: Tetratype: AC AC ac ac (parental) ODER Ac Ac aC aC (non parental) → PD = NPD

→ ½ = fc (recombination frequency) → (da only one of the 2 pairs also nur A oder C rekombiniert!)

PD = NPD; if:

→ when genes are on the same chromosome but very far away from the centromer

→ or genes are on different chromosomes

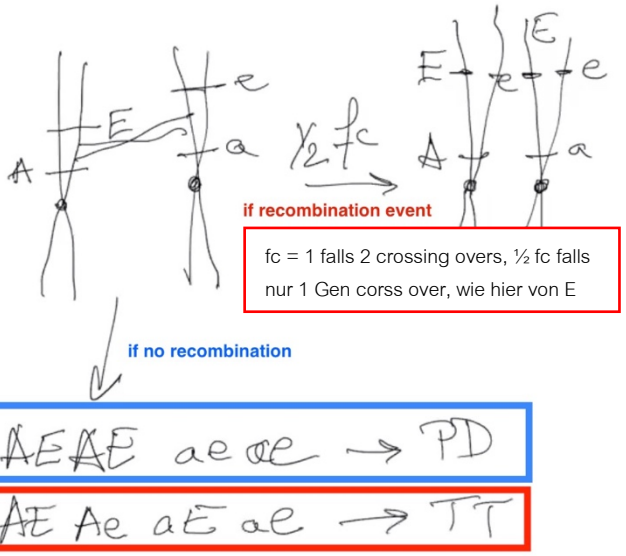
if Tetratype TT = 0 → Gene are on different chromosomes and close to their centromere

if NPD = 0 and PD = 100%: when 2 genes are very close to each other on the same chromosome

NPD = 0, PD = 95%, TT = 5%: very close to each other on the same chromosome, but there is still some cross over

Example for recombination event:

distance = ½ TT/(PD + TT) = 2.5 cM, since NPD = 0



Frage 2: Group 4

Never calculate between 3 points → every analysis is done by pairs → compare results between the pairs

Frage 3: Group 8

→ simply bcs PD = NPD and TT (Tetratype) = very low

→ tells you the 2 genes are centromere linked

Frage 4:

Example: how will we know if linked or not

→ if fc = 0 → NPD will be **0, PD = 100%, TT = 0**

1 - probability of more than 1 crossing over = 0

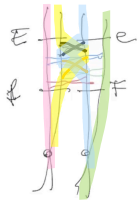
d = TT/2 (total)

→ if fc > 0 → ….

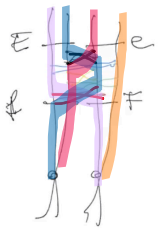
<u>cases of 2 crossing overs: **erster CO** ist schwarz, wenn der zweite die Farbe … hat dann:</u>

what can happen? 4 Möglichkeiten? (rot, gelb, grün, blau)

1) yellow: fE, fE, Fe, Fe → **gives you parental ditype PD** → the 2 cross overs are like there has been no CO (immer gelb und schwarzes Kreuz entlang fahren!



2) red: follow red line → fE, fe, Fe, FE → **Tetratype**!



3) blau: fe, fe, FE, FE → **NPD = non-parental ditype**

4) grün: fE, fe, FE, Fe → **Tetratype**

<u>Distanz berechnen: (falls 1 oder 2 crossing over möglich)</u>

½ single CO (= (TT – 4NPD)/ 2) + double CO (= 4 NPD) / TOTAL <u>= distance</u>

- ☞ <u>d = TT – 2 NPD + 8 NPD / 2 x TOTAL</u>
- ☞ <u>PD > TT > NPD</u>

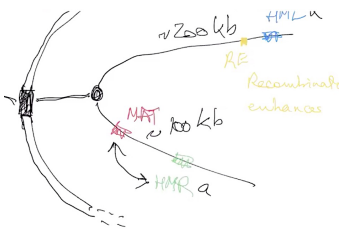
Frage 5: Group 3

strains that change sex every cycle → depends on endonuclease → no issue of gene stability

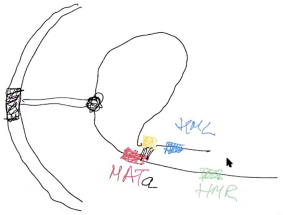
Frage 6: Group 6

→ *Chromosom wird von Spindel an Zellpol gezogen*

*** →



→ ***if recombination enhancer is active:*** *In a Mat A cell, MAT will recombine mit HML alpha und switch to Mat alpha*



→ *** ***if recombination enhancer is silenced (Mat alpha) – oberes Bild*** *(mating type dependent)* → *in a Mat alpha cell closest recombination partner of Mat alpha is HMR*

→ *bcs of recombination enhancer that is mating type dependent, in a MATa Cell the enhancer is active, brings HML near MATa, so when I have a break in MATa it will recombine mit HML alpha and switch to MAT alpha*

→ *in a MATalpha cell the recombination enhancer is silenced* → *closest to MATalpha is HMR for recombination* → *if break in MATalpha* → *it will switch to MATa*

→ ***all depends on recombination enhancer being active in MATa cell (bcs there is no alpha 2) and inactive in MAT alpha cell (since there is alpha2)***

QUESTION ANSWERING PART 2 (05.10.2020)

MEIOSIS:

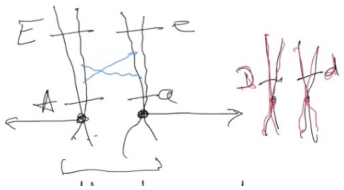
mit 1 Chromosomen

→ ohne Crossing over entsteht nach Meiose 2 ein Parental Ditype (AE, AE, ae, ae)

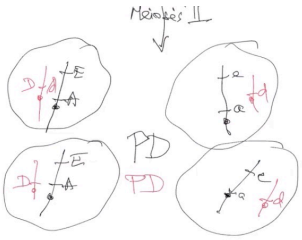
→ mit Crossing over in Meiosis I: es entsteht ein Tetratype (AE, ae, Ae, aE)

mit 2 Chromosomen (**ohne** Crossing over)

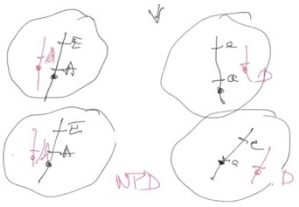
→ nach Meisosi II:



→ falls homologe Chromsomen mit Allel D und d so getrennt dass Chromatid mit Allel AE und Chromatid D zusammen und ae zusammen mit Chromatid d, dann ensteht ein parental ditype



→ Falls genau umgekehrt, also Chromatid d von Chromosom 2 zusammen mit Chromatid AE von anderem Chromosm, dann entsteht ein NPD = Non parental ditype



→ if there are 2 Gene (A und E)on the same Chromsomes near each other (näher als 50cM) = (50% chance for recombination of Alleles) sind 2 A und E <u>gelinked</u>

→ es können nur Gene auf dem selben Chromosom gelinked sein!

Frage 7: Group 2

Imagine having mutation on Sec61 (temperatursensitiv)
→ no phenotype at permissive temperature (24 Grad)
→ Dead at hight temperature > 37°, but grows mit 37°

SEC61 → (Wildtype Gene is therefore essential)
→ Sec61-I = point mutation → protein unfold at high temperature bcs of mutation = inactive = dead

sec61-1 & ubc6Δ double mutant → cells grow > 37° temperature → also ist ubc6Δ eine suppressor mutation für das sec61-1-Allel (mutiertes Allel)
→ UBC6 encodes ubiquitin conjugating enzyme that targets misfolded proteins for degradation

→ sec61 can function at 37° but a domain is misfolded due to mutation, what is recognized by Ubc6 and targeted to degradation → at 37° Sec6-1 protein is degraded → essentielles protein fehlt → dead

FORWARD GENETICS

INTRODUCTION (12.10.20)

forward:	Phenotype	→ Gene
reverse:	Gene	→ Phenotype

FORWARD GENETICS MODELS:

C. elegans (Fadenwurm) as an Animal Model

one step higher in complexity than yeast

- self-fertilizer or mating
- 1 Animal = 959 cells
- 800 minutes development

Drosophila

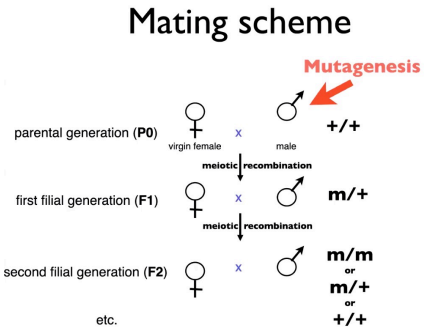
- embryogenesis with 3 larval stages
- segmentation
- Understanding embryogenesis = Nüsslein-Volhard and Eric Wieschaus

Zebrafish = vertebral model organism

REVERSE GENETICS MODEL:

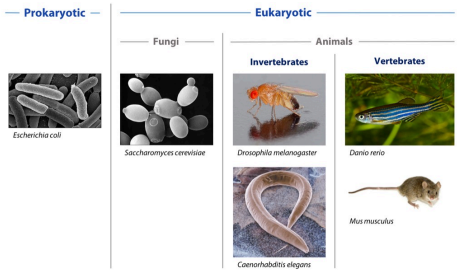
Mouse = long life cycle, good for reverse genetics
Huhn

Mutagenesis



dominant = screen F1 generation
recessive = screen F2 generation

CLASSICAL FORWARD GENETIC SCREENS IN DROSOPHILA AND C. ELEGANS (OL)



Caenorhabditis elegans / Fadenwurm

- study cell differentiation and developmental biology
- a transparent, non-parasitic nematode, lifespan: 2 weeks
- one of the most extensively studied organisms
- 5 autosomes and 1 sex chromosome, 100 mio. bp, 20'000 genes, 35% of genes have human homologs
- muscles, nervous system, gonads, erpidermis and gastrointestinal tract = like complex animals
- lack circulatory and respiratory tract – no bones or immune system
- unusual features:
  - 2 forms of sex
    - 1 sex chromosome XO
    - matched pair of sex chromosomes = hermaphrodite -> form of self-fertile female that can produce male and female gametes by internal fertilization
  - 4 larval stages, 800 minutes for embryogenesis, 959 cells in hermaphrodites, 1031 cells in males
    - eutely = adult of a species have genetically determined # of cells
  - can be kept as frozen stocks for many years
- developmental fate of every single somatic cell has been mapped, such that it is possible to predict the

differentiation pattern of each cell during embryogenesis.

What could be the mechanisms by which males are generated? And what are males needed for?

Male nematodes are generated by non-disjunction of X chromosomes during meiosis. Males are needed for moving mutations between strains to ensure genetic variability.

Why are hermaphrodite worms so handy for the genetic screen described above? Because they are self-fertilizing, F1 worms carrying a heterozygous mutation can produce offspring that is homozygous for the mutation. **No partner with the same mutation is required to produce a homozygous F2 generation.**

Forward genetic screens in C. elegans

The advantage of a forward screen is that it is unbiased and will reveal any mutation associated with the observed phenotype, regardless of the function of the respective gene.

→ *if the goal is to identify mutations in genes specific for* ***a certain function or tissue****, a genetic screen can be combined with fluorescent reporters. These reporters can be a GFP-marked protein, a GFP-marked cellular compartment, or even a GFP-marked cell, allowing to screen for changes in GFP expression*

Phenotypes:

- Unc: uncoordinated movement
- Lin: alterations of the invariant cell lineage
- dpy: small bodies
- lon: long bodies
- Rol: rolling locomotion

recessive F2 screen:

- population of WT hermaphrodites mutagenized
  - P0 contains recessive mutation "m" in sperm and egg cells → one animal contains both egg cells and sperm
  - F1 generation → result from fertilization of egg with mutated sperm → (m/+)
  - Bcs a heterozygous mutant worm of the F1 generation is a self-fertilizing hermaphrodite, it will produce both mutant eggs and mutant sperm: Thus, ¼ of the F2 progeny (F2) will be homozygous (m/m) for the mutation. → homozygous mutants show a phenotype different from the WT.

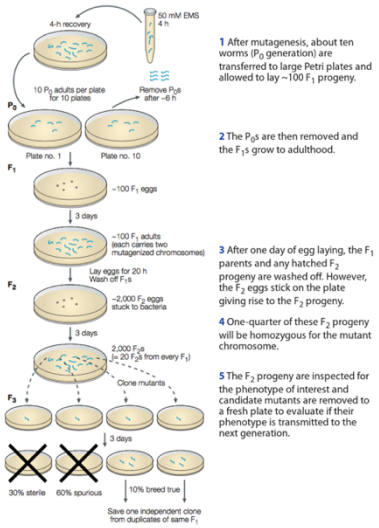
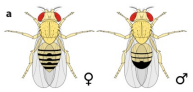


Figure 1-5 Protocol for EMS mutagenesis in worms. (adapted from E.M. Jorgensen and S.E. Mango, *Nat. Rev. Genet.*)

Drosophila melanogaster to study development

ADVANTAGES:

- 10 day generation time, inexpensive
- females (pointed abdomen) and males (darker backs and black patch on abdomen) can easily be distinguished



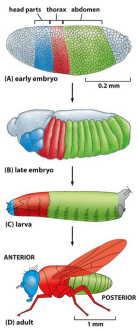
- 3 autosomes and 1 pair of sex chromosomes (XX and XY male)
- fully sequenced genome, 140 mio bp, 15'000 genes, 50% human homologs
- increasing repertoire of techniques and resources, such as mutant libraries, transposon insertions, reporter and RNAi collections, facilitate large-scale genetic analyses in the fruit fly.
- availability of phenotypic markers that can be used to select flies based on their phenotype. These can be dominant or recessive mutations that confer a specific phenotype (e.g. bend up wings, additional bristles) or transgenic markers, e.g. constructs where a gene or a regulatory region of a gene was fused to the coding sequence of GFP
- *NEGATIVE:* cannot be frozen

Forward genetic screens in Drosophila

The screen developed by Nüsslein-Vollhard and Wieschaus was revolutionary, because it was one of the first screens for phenotypes in the embryo rather than in the adult, which allowed to identify null or strong mutations in most of the essential patterning genes that are used throughout development.

Drosophila, because its **development** is especially suited to look for genes involved in embryonic development:

+ segmentatoin along its anterior-posterior axes → The number and types of segments are already determined in the embryo and are invariant in the wild-type → patterning of the embryo can be visualized already in the larval cuticle (the external shell), which serves as a straightforward morphological readout for a screen



+ in Drosophila embryogenesis, the mother loads the egg with most of the gene products that are required for the very early steps of embryogenesis. → maternal effect genes → involved in generating the four maternal signals that define the two main axes of the embryo → Thus, only very few zygotic genes are required to go through the early steps of embryogenesis, **which means that there are only few mutations that block early embryonic development, and most mutants in housekeeping genes complete** embryogenesis and produce a normal larva.

⇨ Therefore, the researchers reasoned that a study screening for patterning defects in the late embryo or larva would most likely **reveal transcription factors and signaling molecules regulating segmentation and patterning.**

<u>How can one screen for embryonic lethal mutations if re-crossing of mutants is not possible (because they die)?</u>

→ *isolate the heterozygous siblings and to keep these heterozygotes for further breeding.*

<u>How to distinguish heterozygous-mutant from homozygous wild-type flies, if the heterozygous flies phenotypically look like wild type (because the mutant phenotype is recessive)?</u>

→ *balancer chromosomes carries a visible marker mutation, such that the presence of this chromosome can be followed in crosses.*

Balancer chromosomes as tools to isolate and maintain lethal mutations

to maintain mutations that would cause lethality or sterility in a homozygous situation.

- carry a dominant phenotypic marker that allows selection of flies that carry this balancer (e.g. a marker that confers bend-up, curly wings or short bristles)
- they consist of one or more inverted DNA segments → it is no longer really homologous to its corresponding wild type chromosome → if one chromosome is a balancer, crossovers do not form in the vicinity of inversion breakpoints, because synapsis is inhibited. Thus, in most cases, crossovers do not form and recombination does not take place between balancers and normal chromosomes.
  - (recombination events that take place between the balancer and the homologous chromosome (in inverted regions), this leads to chromosomes with two centromeres or no centromere. If crossing over occurs between balancers and their homologous chromosomes, each resulting chromatid ends up lacking some genes and having multiple copies of other genes. Thus, cells resulting from meiotic cells where recombination between a balancer and its homologous chromosome has taken place simply do not survive)
- they carry a recessive lethal factor → 2 copies of balancer = not surviving

⇨ ***<u>only flies that are heterozygous for both the balancer and the mutation survive</u>***

Heidelberg Screen Setup

- *Nüsslein-Vollhard and Wieschaus made use of the tools (balancers, dominant and conditional markers).*
- *They used EMS to mutagenize male flies as the starting population.*
- *They decided to first concentrate on mutations that are on* ***the second chromosome*** → *important, because they needed to decide which balancer to use to "collect" the mutations.*

1) The mutagenized males were then crossed in bulk to virgin females that were heterozygous for a second chromosome balancer

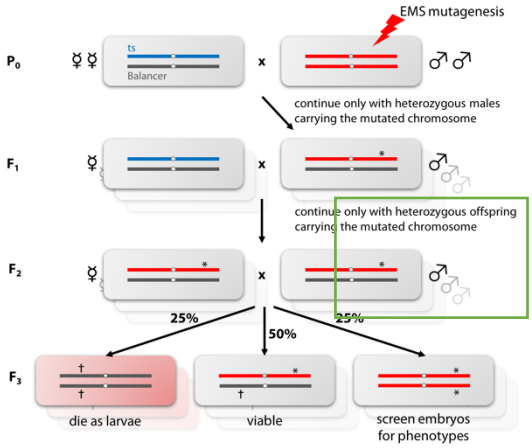
2) From the resulting F1 generation, each male carrying a balancer chromosome (and presumably a mutation on the other chromosome) **was backcrossed to virgin balancer females to form an F2 generation. This F2 generation is heterozygous for the mutation.**

3) Crossing males and females from this F2 generation will produce an F3 generation of the following composition: 25% of the flies will be homozygous for the balancer and thus die. Other 25% will be homozygous for the mutation. 50% of the

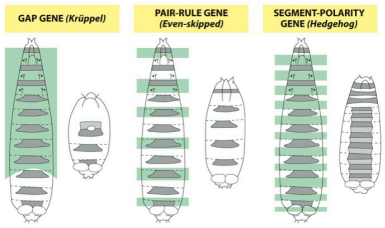
flies will contain a balancer and a mutated chromosome; these flies will be viable and can be used to set up a stock.

4) F3 generation was first screened if there were any flies that did not contain the balancer (these would be the ones that are homozygous mutant) → If there were flies without balancer, this indicates that the mutation is not lethal, and these fly lines would not be looked at further. (denn developmental mutationen sind homozygot lethal!)

5) If in the F3 generation only flies that carried the balancer occurred, all homozygous mutant flies must have died, **indicating that this line contains a lethal mutation that is of interest to the researchers. Those lines were analyzed further** to **see if 25% of the embryos (the homozygous mutant ones) showed a phenotype**, and the phenotype was further characterized.



6) Nüsslein-Vollhard and Wieschaus grouped the mutant phenotypes and saw that the genes they identified fall in different functional classes. These classes are the gap-, pair-rule and segment-polarity genes that work in a regulatory hierarchy to create a pattern in the embryo



And how can we know that the screen is saturated, e.g. all possible mutations have been identified?

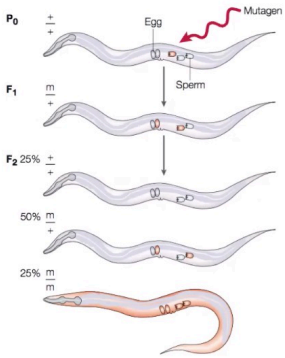
we can use complementation tests to group the mutations in complementation groups. A complementation group consists of mutations that do not complement each other, thus, these mutations are in the same gene. A complementation group is therefore equivalent to a gene. If the same mutations occur over and over again we know that the screen is saturated and we will not find any more new mutations even if we screen more individuals.

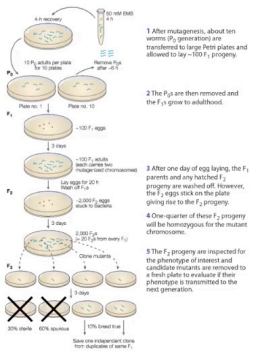
Which genes do you think Nüsslein-Vollhard and Wieschaus could not identify in their screen?

In the Heidelberg screen maternal effect genes, genes affecting patterning of internal structures like nervous system, and tissue-specific genes could not be detected. Further, small genes e.g. miRNAs, were not detected because small genes have low mutation frequencies (the chance of introducing a mutation is greater the larger a gene is).

VORLESUNG LESSON I-II: 19.10.20

Random mutagenesis screen





- mutagenize the male normally, random mutations
- 1th question: recessive or dominant mutation?
  - recessive mutations: screen F1 generation
  - dominant: screen F2 generation (25% of animals will be homozygous)
  - immer kreuzen mit WT, da wenn 2 mutierte kreuzen, dann anderer Phänotyp
  - evt. purification step to make sure each individual screened is only carrying 1 mutation in the genome (make selection and recross, repeat few times, to lose not-phenotype-linked mutations)

Using genetic markers to label specific chromosomes

- follow mutagenized chromosome
- transgene can be inserted in any chromosome to label it

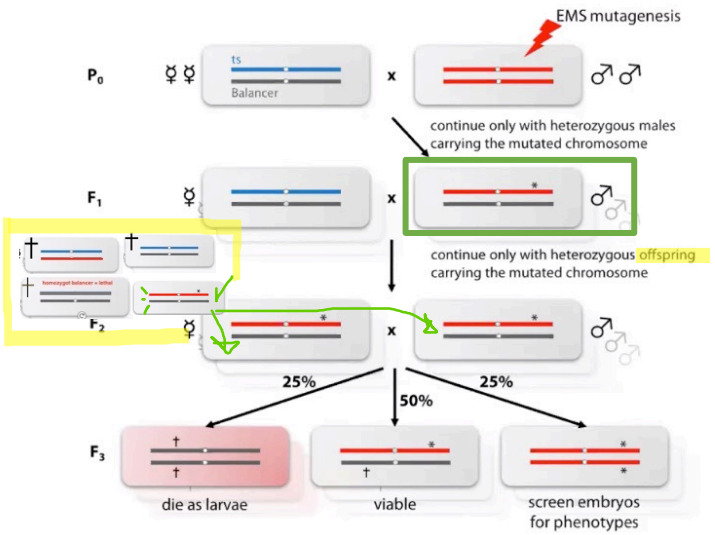
- mutations causing easily detectable phenotypes
- mutations in genes with precisely known chromosomal locations
- dominant or recessive mutations
- constitutive or conditional
- transgenic markers can be inserted at desired location

Using balancer chromosomes to maintain lethal mutations

- mutation can only be maintained in a heterozygous state → balancer chromosome to differ btw WT and heterozygous (WT dies, and homozygous also dies)
- falls individuals die kein Marker tragen, heisst das sie beistzen kein balancer, also sind sie homozygous, also ist mutation nicht lethal und uninteressant
- balancer chromosome cannot undergo recombination with homologue chromosome

The Heidelberg screen: Isolation embryonic lethal Drosophila mutations

- use of conditional marker on blue chromosome (ts) → high temperature → kill all animals that kill at least one conditional marker
  - alle Nachkommen wurden mit Hitze behandelt und so alle Nachkommen die conditional marker besassen und somit keine mutation, wurden ausgelöscht



wieso nicht einfach dieser mit einem gleich mutierten (Geschwister) kreuzen? Wieso zuerst namal mit blau grau kreuzen? DIES IST NUR MÖGLICH WENN JEDES ELTERNPÄÄRCHEN EINZELN GEHLATEN WIRD! DAS IST JEDOCH BEI GROSSEM SCREEN NICHT RATIONAL! WENN ALLE F1 IN GLEICHEM TOPF, DÜRFEN SICH DIE F1 NACHKOMMEN NICHT UNTERINANDER PAAREN, DA SONST MUTATIONEN GEMIXT WERDEN! MAN ENTNIMMT ALSO EIER NACH P0 KREUZUNG UND PAART SIE ALLI MIT EINEM VIRGIN BALANCER WEIBCHEN NOCHMALS, SO KANN AUCH BEI GROSSEM AQUARIUM AM ENDE FÜR HOMOZYGOT GESCREENT WERDEN! → SIEHE SPRACHMEMO ISIII 20.10,20 (* markiert)

3 von 4 Nachkommen sterben wenn Temp hochgefahren wird, da ts lethal bei hoher Temp, und auch tot wenn für Balancer homozygot!

Lesson II: FLP/FRT system to induce mitotic recombination

- a tissue specific recombinase system
- einfügen near centromere
- FLP will induce recombination between 2 FRTs
- study mutations in specific mutations
- Why this F1 screen allows to find genes in the late development:
  - untersucht mutations kill individuals as early embryo
  - now you want to know that is the function of the mutated gene in a specific tissue of an adult or larva, bcs most of genes, are not only used once in development (growth factors have different effect in different tissues)

FLP/FRT System combined with cell lethal mutations: give wildtype and heterozygous a growth disadvantage

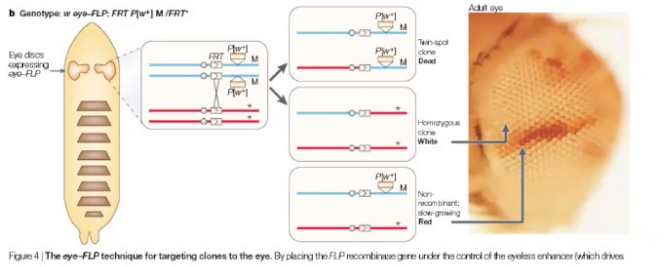


Figure 4 | **The eye-*FLP* technique for targeting clones to the eye.** By placing the *FLP* recombinase gene under the control of the eyeless enhancer (which drives

to increase likelihood of finding mutant patches of tissue, the trick of cell lethal mutations was used to give the non-mutant chromosome (blue) a disadvantage, by giving a lethal mutation M (w+ is a red marker!!!)

→ *if homozygous for blue chromosome (M) = non-mutated = lethal = cell apoptosis -> bcs genes important for basic metabolic functions are mutated → eliminate WT!!! (WÄREN RED)*

→ *even if heterozygous (RED) for lethal mutation M and wanted mutation m (blue, red), the cell will not grow that fast bcs they only can synthesize proteins at 50% rate = growth disadvantage!!!*

→ *homozygous red chromosome cells (WHITE) = can synthesize proteins at normal cells -> homozygous have a growth advantage over heterozygous in a tissue!!!!*

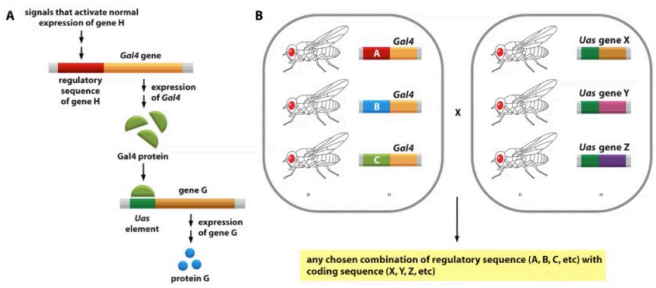
→ *most of eye cells would be white, they would be the offspring of this recombined homozygous red cell that do* ***not contain the red marker w+ and are therefore white***

Modifier screens: Isolating enhancers and suppressors to study pathways

- number of genes act in sequence → for an efficient way of study pathways <u>you must have a pre-existing mutation on one gene of the pathway</u>
- the chance that the secondary mutation which is a enhancer or suppressor that modify the preexisting phenotype (verstärken oder abschwächen des Phänotyps) is in the same pathway, is very high

USING THE GAL4/UAS overexpression system

- screen for phenotypes caused by overexpression
- chemicly mutations normally reduce gene expression, so reduce activity of a gene
- use overexpression of WT genes to create phenotypes we can screen for → mutagens that randomly overexpress gene
- Gal4 = transcription factor → to express it in the drosophila genome you use tissue specific promotors
- in WT, the TF has no effect, bcs WT has no Uas recognition sites
- to generate effect, you introduce Gene X you want to overexpress with a second transgene → such that X will be overexpressed in all cells that express the Gal4 Protein → bcs it will bind to the Uas sites which you located upstream of the Gene X you want to overexpress
- 2 transgenes are needed: 1 to express Gal4, and a second one that contains the gene you want to overexpress cloned downstream this Uas elements
- you can cross flies containing different Gal4 (= drivers) to flies containing transgenes expressing different target genes from Uas sites (i.e. you can express Gene Z in tissue A,B,C because A-C are tissue specific promotors)
- But how to turn this system to randomly overexpress gene, so not pick gene X, Y or Z
    - Trick: to package Uas elements, which are Gal4 binding sites, and put them between 2 transposonsable Elements



FORWARD GENETIC METHODS TO SCREEN F1 (OL)

<u>F2 screens:</u>

they have two major drawbacks:

- they are time-consuming and labor-intensive, since several generations are required for recessive mutations to show their effect.
- they do not allow to identify the functions of essential genes during later stages of development, because **mutations in these genes usually cause embryonic or early larval lethality**. → Thus, only the effect of the first essential function a gene exerts in development can be analyzed.

*In fact, genetic studies have shown **that only seven genes expressed by the Drosophila zygote are required** for the development of the embryo to the cellular blastoderm stage.*

→ *However, there must be many more genes acting during animal development. Here, we will discuss three strategies how these genes can be identified:*

1 – CLONAL SCREENS

- only cells in a specific tissue are homozygous for the mutation while the rest of the animal is heterozygous for a mutation
- recombination is induced artificially in mitotic somatic cells using the Flp recombinase that mediates site specific recombination between target sites, called FRT sites (target sites)
- recombination event can be used to make a mutagenized chromosome arm **homozygous** for the mutation of interest in clones of cells, and the resulting phenotype can be screened for
- FRT sites have to be present at identical positions on the homologous chromosomes and are artificially introduced, as well as the recombinase, which is then present as a transgene inserted somewhere else in the fly genome

Generate Clones:

1) First fly contains: 1 chromosome with mutation * **distal to an FRT site** and 1 chromosome carrying only the FRT site
2) Second fly contains: 2 chromoes carrying only the FRT site (marked with GFP)
3) cross both flies

4) 50% of progenies are heterozygous for mutation (interesting) , and 50% are homozygous for the blue chromosome, carrying no mutation but a FRT site (not interesting)

5) At the onset of mitosis in these heterozygous cells, the Flp recombinase mediates site-specific recombination between chromatids of homologous chromosomes at the FRT sites, leading to an exchange of DNA material that lies distal to the FRT site.

6) During cell division, the daughter chromatids segregate, and in some cases, this leads to daughter cells that are homozygous for the mutation. → contain two copies of **the maternal mutant allele**.

7) In the other cases, cells contain two copies of the paternal wild-type chromosome. These cells can proliferate and grow to clonal areas (clones) with identical genotypes.

→ homozygous and heterozygous cells can be distinguished by using visible markers that are located on the same chromosome arm as the mutation, so that both marker and mutation take part in recombination event

→ **homozygous will then be the only one not showing any markers**

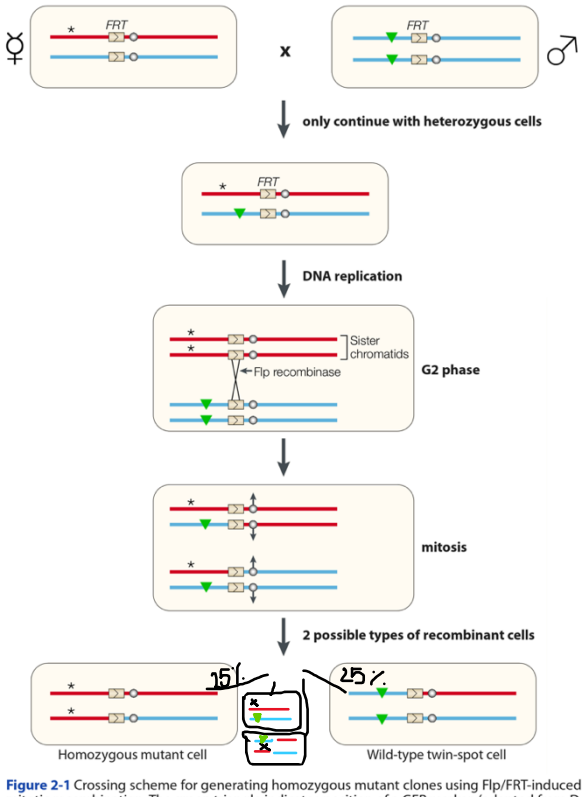


Figure 2-1 Crossing scheme for generating homozygous mutant clones using Flp/FRT-induced mitotic recombination. The green triangle indicates position of a GFP marker. (adapted from D. St Johnston, *Nat. Rev. Genet.*)

What is the probability of one cell inheriting two mutant (red) recombined chromatids? The chromatids segregate randomly, i.e. the chance of getting two mutant chromatids is 0.5 x 0.5, thus 0.25. (siehe gemalt)

advantages of clonal screens:

+ F1 screens can be carried out for recessive phenotypes, because no additional generations are needed to produce homozygous mutant cells

+ removes the risk of losing mutants due of the effect of mutant clones in other tissues.

+ by using a Flp recombinase that is expressed in specific cells or at a specific stage during development, one can control where and when recombination occurs

→ *thus, only the cells of interest are made homozygous, but the rest of the animal is heterozygous.*

→ ***Therefore, tissue-specific phenotypes of essential genes can be visualized regardless of their other functions in development, and embryonic phenotypes can be bypassed by specifically inducing mutations later during development, e.g., at late larval stages.***

What is the prerequisite for homologous recombination to occur in mitotic cells?

Homologous chromosomes need to be paired in mitotic cells! This is special in Drosophila and does not occur at a high frequency in most other organisms. Thus, mitotic recombination occurs with much lower frequency in other organisms

examples for clonal screens

eye-Flp experiment:

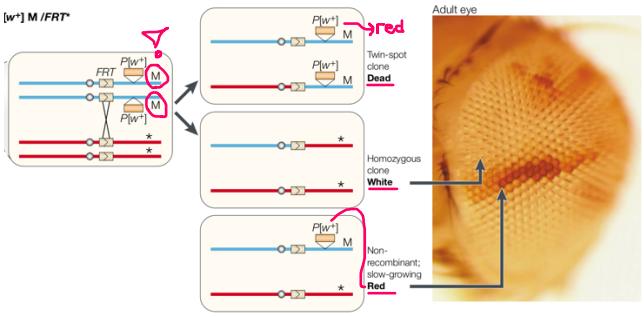
The Flp recombinase is fused to the regulatory region of the eyeless gene. → *Flp is expressed specifically in the developing eye disc, and homozygous mutant clones arise in the adult eye, but not in the rest of the animal*

→ to increase the # of homozygous mutant clones: one can select against the homozyghous WT clones and cells where NO recombination has taken place, by inserting a recessive lethal mutation M on the non-mutated FRT chromosome(blue)

→ chromosomes that carry Minute (M) mutations (recessive lethal) are used, when homozygous they lead to cell death, bcs the gene produces essential components of translation (z.b Ribosom)

→ *M is therefore present on the non-mutant FRT chromosome (blue)!*

*The mutant cells with no M, but the wanted mutation m (red arm with *) occupy almost the entire eye, since they outcompete the slowly-growing heterozygous cells, while the non-mutant clones die.*



tem can be combined with a *Minute* mutation to select for homozygous mutant clones. The *Minute* the non-mutant chromosome. The non-mutant twin-spot clones contain two copies of the recessive die, whereas the non-recombinant heterozygous cells grow slowly. Homozygous mutant cells therefore nd occupy a large area of the eye. (adapted from D. St Johnston, *Nat. Rev. Genet.*)

A screen for growth regulators using Flp/FRT system

example for screen where genes involved in growth regulation were identified using the eye-Flp/FRT Minute system: "Pinhead/Bighead" screen (HAFEN)

- mutants disrupting size regulation were identified by simply screening for flies with small or big heads
- The researchers used the ey-Flp/FRT system combined with a cell lethal mutation on the non-mutant chromosome
- starting population: fly lines carrying an FRT site for each arm of chromosome 2 and 3
- carried out four screens:
  - one looking for mutations on the left arm of chromosome 2,
  - one for the right arm of chromosome 2,
  - one for mutations on the left arm of chromosome 3
  - one for the right arm of chromosome 3.
- they used the corresponding FRT line as the starting population and mutagenized it by feeding with EMS.
- Then, they crossed these mutagenized flies with the corresponding line containing the non-mutant FRT line and the FLP recombinase
- looked in the F1 generation whether they would find any flies with bigger or smaller eyes than usual

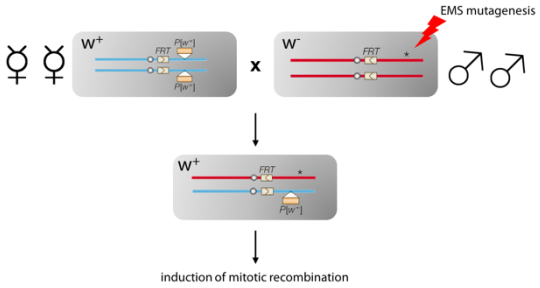


Figure 2-4 Genetic scheme to identify growth mutations using the Flp/FRT technique. Males carrying the Flp recombinase and the FRT sites that carried the w+ mutation (confers white eyes) were mutagenized by feeding with EMS. Mutagenized males were crossed to females to produce heterozygous embryos. Clones of cells homozygous for the induced mutations were generated in developing first-instar larvae by mitotic recombination at the FRT sites induced with the ey-Flp recombinase.

→ *They identified 580 small-head mutations that could be mapped to roughly 30 growth promoting genes, and 137 big-head mutations that were mapped to 17 growth inhibiting genes*

→ *This lead to the identification of main components of the insulin signaling cascade, such as the insulin-receptor substrate chico, the homologue of IRS in vertebrates.*

2- MODIFIER SCREENS

→ *synthetic screen in which we search for mutations that affect a given phenotype*

- *find components of pathways → carry out a screen for dominant enhancers or supressors*
- The idea behind this is that loss-of-function mutations for almost all genes are recessive → in a heterozygous mutant situation, **50% of the wild-type level of a protein is sufficient for normal development.**
- Reducing the activity of this protein **by a mutation in another gene** may lead to a phenotype because the protei n level is **now no longer sufficient for normal function.**
- Therefore, **mutations in genes acting in a pathway can be identified as dominant suppressors or enhancers** of a specific phenotype in this sensitized background.
- in yeast: synthetic interactions → occur when a double mutant has a phenotype different from either single mutant parent.
- If the double mutant shows a milder phenotype than the single mutant, we call these mutations suppressor mutations.
- mutations that worsen the phenotype of the single mutant are called enhancer mutations.
- Synthetic lethality is the most extreme case of enhancement

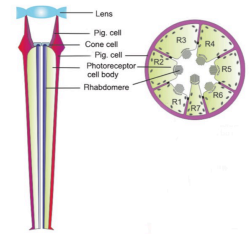
- Enhancers and suppressors are generally referred to as modifiers
- the same gene can act to suppress or enhance a phenotype, depending on the nature of the mutation
  - cyclin B is a cell cycle inhibitor
  - a deletion in cyclin B was found to enhance the cell-cycle phenotype.
  - Mutants containing extra copies of the cyclin B gene and thus had a higher expression level of cyclin B suppressed the cell-cycle phenotype

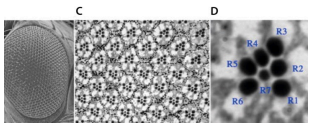
⇨ **it is important to keep in mind that the term "modifier" refers to a mutation rather than a gene**

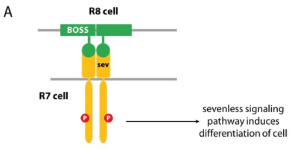
Example for a dominant modifier screen in Drosophila: playing with the sevenless threshold

screen that identified components of a signaling pathway downstream of a receptor tyrosine kinase (RTK) regulating cell differentiation in the Drosophila eye

- eye is an ideal organ as a readout for a screen, because its if mutations affect eye development, the eyes cannot form properly, but flies still survive and can reproduce normally (so eye function isn't essential)
- During eye development, the sevenless signal transduction pathway controls the **cell-fate choice between one of the eight photoreceptor cells, the R7 cell, and a cone cells**





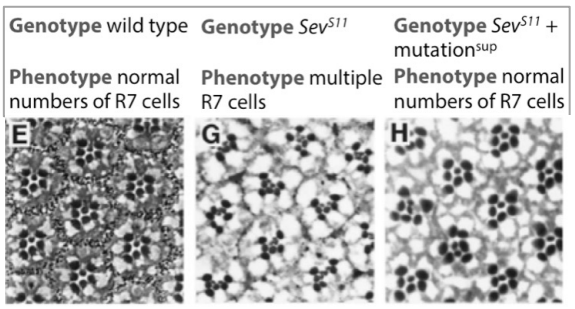


- Loss-of-function mutations in sevenless lead to a transformation of the central R7 cell into a cone cell.
- The receptor-tyrosine ki-nase Sevenless (Sev) in the R7 precursor cell becomes activated by binding to the bride of sevenless (**Boss**) protein, a seven-transmembrane protein that is expressed in the differentiated R8 cell
- In the absence of either Boss or Sev (sevenless), the R7 precursor cell develops into a cone cell → only 6 photoreceptors are present instead of 7 (bcs R7 is on top of R8)
- In the **wild-type situation**, the cone cells do not contact the signal-providing R8 cell; thus, they are unable to adopt the R7 cell fate

IDENTIFY THE COMPONENTS OF THE SEVENLESS SIGNALING CASCADE → PRESENTING 1 STRATEGY THAT RELIES ON REDUCING THE NORMAL LEVELS OF SEVENLESS SIGNALING TO INDUCE A PHENOTYPE:

- illustrates how genes acting in a pathway can be identified by creating a dosage-sensitive situation through modulating one of the pathway components.
- again-of-function mutation of Sev was used that is **constitutively active** (called Sev^{S11} and did not result from mutagenesis, sondern ist ein transgenic construct which was induced in the fly!) and signals independently of the binding of its ligand Boss.
- mutation leads to that Sev cannot be activated by a ligand!
- sevenless activity is necessary and sufficient for the determination of the R7 cell fate.
- using the Sev11 mutation, dosage dependent modifier mutations could be identified that decrease the strength of sevenless signaling and affect R7 differentiation

⇨ Expressing the Sev^{S11} mutation not only leads to an increased activity of the sevenless pathways in R7 cells, but also increases sevenless signaling in cone cells that normally do not activate the sevenless pathway because they do not receive a boss signal.



⇨ In the wild-type situation, sevenless activity is above the threshold in R7 cells, h owever, in cone cells (CC), sevenless signaling is below this threshold, thus, cone cells do not differentiate into R7 cells.

⇨ *If a second mutation (m) in the sevenless pathway is introduced, the sevenless activity drops, but still reaches the threshold required for photoreceptor cell differentiation in R7 precursor cells*

⇨ In cells expressing the constitutively active form of SevS11, sevenless activity is increased in R7 cells and at the same time increased in cone cells. → Sev is now hyper-activated in the cone-cell precursors, some of these cone cells develop into R7 cells, because sevenless signaling passes the threshold necessary for R7 differentiation in these cells. → *leads to ommatidia with **more than one R7** cell and therefore a rough eye phenotype*

⇨ *If a second mutation (m) in the sevenless pathway is introduced into the already Sev11 mutant, sevenless activity will drop below the threshold in cone cells. This suppresses the differentiation of cone cells into R7 cells*

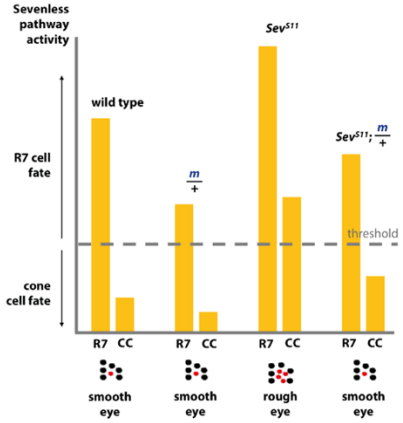
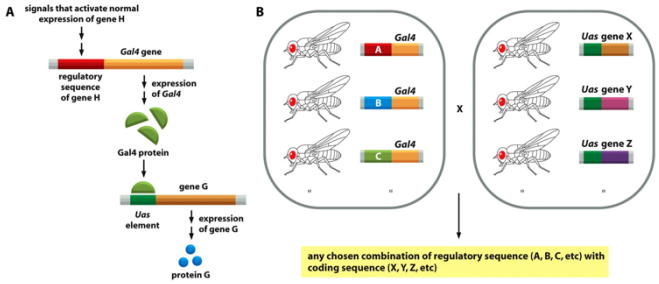


Figure 2-10 The cell fate in Drosophila eyes regulated by sevenless is dosage dependent. In wild type eyes, sevenless activity is high in R7 precurser cells, but low in cone cell precursers. Since sevenless activity does not reach the necessary threshold in cone cells, only R7 precursers differentiate into R7 cells. In eyes expressing the constitutively active SevS11 mutation, sevenless activity is increased in both R7 cells and cone cells. Thus, sevenless signaling reaches the threshold required for R7 differentiation also in cone cell precursers, resulting in the transformation of some cone cells into R7 cells and a rough eye. Introducing another mutation into this SevS11 sensitized background leads to a drop in sevenless activity below the threshold in cone cells. Thus, cone cell transformation is suppressed, resulting in a normal eye.

3- OVEREXPRESSION SCREENS

Genes with redundant functions might be missed, because the loss of this gene might not produce a visible phenotype to screen for. → *a useful complementation is the Gal4 system that allows the overexpression of genes and screen for phenotypes*

⇨ *the yeast transcriptional activator Gal4 is used to regulate gene expression in Drosophila by inserting the upstream-activating sequence (UAS) next to a gene of interest (gene X, Y or Z)*



This System can also be used in Drosophila to silence genes in certain cells using RNA interference:

1) flies containing the Gal4 gene under the control of a tissue-specific enhancer are crossed with flies containing a sequence coding for a hairpin-RNA that is specific for the gene to be silenced under the control of UAS sequence

2) In the resulting F1 generation, the hairpin RNA is expressed and induces silencing of the target gene by degradation of the target mRNA.

Screen:

- researchers make use of fly lines that contain transposons, so called P elements:
  - P elements can be used for **mutagenesis**, because the transposition of such an element from one locus to another may disrupt the reading frame of a gene such that the affected gene is inactivated and no longer transcribed correctly
  - Such EP elements can also be used in synthetic screens to search for modifiers of a given phenotype.
    - example: a mild rough-eye phenotype was generated by overexpressing the Drosophila myc gene specifically in the head.
    - Then, each line of the EP collection was crossed with flies overexpressing myc.
    - The eyes of the resulting flies from the F1 generation were then examined to identify those crosses where the rough-eye phenotype was suppressed
    - These lines contained EP elements inserted into genes that, when overexpressed, counteract Myc activity.
    - → Overexpression of one of the EP lines leads to a suppression of the rough eye phenotype
  - another scenario: transposons are used to **overactivate** rather than inactivate genes near the transposon insertion sites → *describe this case here in more detail:*

Overexpression Screens:

- *vectors containing a P element were constructed that carry UAS sites at one end (called EP element)* →

inject into Drosophila embryos to generate transgenic animals.

- *The EP-element vector will insert anywhere in the fly genome via the process of transposition.*
- *If the EP element containing an UAS sequence has inserted next to a gene, the UAS sequence can be used to overexpress this gene by activation through Gal4*
- *ideally, like to have a collection of lines with EP element insertions at any possible locus in the genome. To construct such a library, the transgenic fly lines were used, and the EP element present in these lines was induced to change its location by adding transposase → resulted in new lines where the EP element was inserted at different places in the genome,*

⇨ ***This collection can then be used in an overexpression screen by crossing each line with a line containing the Gal4 sequence under the control of a specific enhancer, resulting in tissue-specific overexpression.***

⇨ ***enhancer allows the expression of Gal4 in specific cells or at specific time points during development.***

⇨ ***example**: sevenless-Gal4 will be expressed in differentiating photoreceptor cells. → If flies carrying sevenless-Gal4 are crossed with the lines from the EP collection, one can screen for a rough-eye phenotype, a phenotype indicating that eye development has been disturbed .*

⇨ *EP lines that would result in a rough-eye phenotype when crossed with sevenless-Gal4 could be examined in more detail to see which gene is overexpressed by the EP insertion*

REVISITING GENETIC INTERACTIONS (OL)

in forward genetics, one tries to find the genes responsible for a certain phenotype, reverse genetics observes the changes in the phenotype depending on changes in the genotype

⇨ 'genotype' is used in two different ways: When looking at a single gene, it describes the allele combination of an individual with respect to this gene / the term also describes the genetic makeup of a cell, i.e., its entire set of genes

Allelic Interacitons

Sichelzellanämie:

- A point mutation in HbA gene (called HbS), the encoded oxygen-carrying protein complex hemoglobin results in an abnormal shape and decreased functionality
- However, people carrying the mutation are resistant to the disease malaria, since the pathogenic parasite normally spends part of its life-cycle inside red blood cells, but is unable to do so in individuals carrying the HbS allele.
- This example shows that the allele combination at a single locus can be responsible for at least three very different phenotypes.

Genotype	Shape of red blood cells	Anemia	Malaria resistance
Hb^A/Hb^A	round	No	No
Hb^A/Hb^S	few are sickle-shaped	No	Yes
Hb^S/Hb^S	sickle-shaped	Yes	Yes

- example of co-dominance: ABO blood group system
  - While individuals with IAi, IAIA, IBi, or IBIB show only one type of modification to the glycoproteins on blood cells, in individuals with genotype IAIB, both modifications are made, resulting in blood type AB.

Can you, as discussed for the previous example, list and describe the different levels of phenotypes that result from the allele combination at the ABO locus?

I^Ai or I^AI^A: The A allele-specific modifications of the H antigen are made, this phenotype is dominant.
I^Bi or I^BI^B: The B allele-specific modifications of the H antigen are made, this phenotype is dominant.
I^AI^B: Both the A- and B-allele-specific modifications of the H antigen are made, this phenotype is co-dominant.
ii: no modifications of the H antigen are made, this phenotype is recessive.

Penetrance and expressivity

*But how is it possible that a person carrying an allele that encodes a dominant phenotype does not show the trait? This phenomenon is called **incomplete penetrance***

Bsp: **The penetrance of an allele is the fraction of individuals carrying the allele that also expresses the associated trait.** For example, if an autosomal dominant trait such as breast cancer linked to the gene BRCA1 has a penetrance of 80%, 80% of females carrying the responsible allele will develop breast cancer and 20% not.

What is the molecular basis of this phenomenon?

- it may result from differential allelic expression, modulating influence of additional genetic variants (modifiers) or it may be age- or sex-dependent.

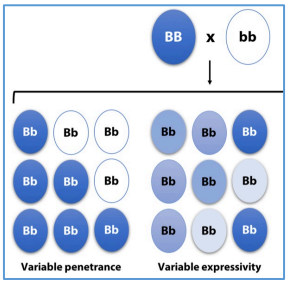
example: autosomal-recessive disease phenylketonuria:

- phenylalanine cannot be hydrolyzed and accumulates in blood → normal life when diet with no Phe

- determining factor whether phenotype is linked to genotype is the phenylalanine content of the diet

related is the phenomenon of variable expressivity of a trait:

- describes the variation of phenotypic expression across individuals that have a particular genotype
- Expressivity differs from penetrance, because it does not characterize phenotypic variation qualitatively and in a binary manner but in a quantitative manner



- **percentage indicating the penetrance** of a trait describes a fraction of individuals of a population (e.g., 65% of people with a given genotype express the phenotype),
- **a percentage for the expressivity** of a trait refers to the "intensity" of this trait in one individual.

Epistasis

Phenotypes are often not the result of the expression of a single gene, but of a combination of gene products that interact = the effect of one gene is influenced by one or more other genes = when the phenotype of the double mutant is the same as one of the single mutants

Example 1: cell cycle regulation by cyclins and CDKs (active when bound to cyclin)

- In mammals, the cell-cycle transition from G1- to S-phase is regulated by the cyclin D-Cdk4 and -6 complex
- pRB is an important regulator of genes promoting G1/S-phase transition by repressing transcription factors of the E2F family. When pRB is inactivated, E2F members facilitate the transcription of proteins involved in DNA replication.

different mutants in this regulatory pathway:

- Down-regulating mutations in proteins of the INK family (ink-mutants) or pRB (pRB-mutants) result in constitutive entry into S phase.
- down-regulation of cyclin D, CDK4/6, or E2F by mutation prevents entry into S phase

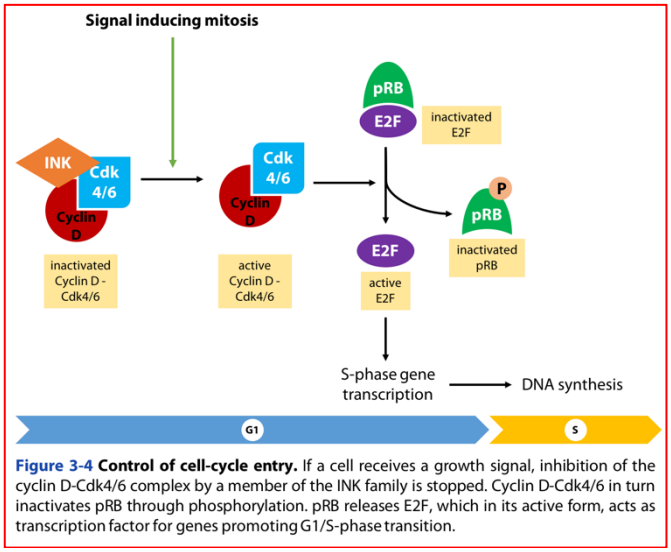


Figure 3-4 Control of cell-cycle entry. If a cell receives a growth signal, inhibition of the cyclin D-Cdk4/6 complex by a member of the INK family is stopped. Cyclin D-Cdk4/6 in turn inactivates pRB through phosphorylation. pRB releases E2F, which in its active form, acts as transcription factor for genes promoting G1/S-phase transition.

Can you predict the expected phenotype of the following double mutants?

- ink-cyclin D
- ink-pRB
- Cdk4-E2F
- pRB-E2F

ink-cyclin D: no entry into S phase
ink-pRB: constitutive entry into S-phase
Cdk4-E2F: no entry into S-phase
pRB-E2F: no entry into S-phase

What is the epistatic relationship of the genes involved in this signaling pathway?

In a regulatory pathway, a downstream mutant usually determines the phenotype and is therefore epistatic over the upstream gene.

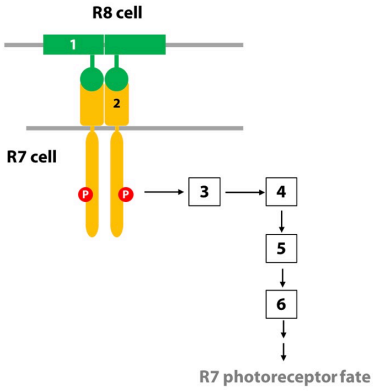
Cyclin D is epistatic over *ink*
pRB is epistatic over *ink*
E2F is epistatic over *Cdk4*
E2F is epistatic over *pRB*

Example 2: sevenless pathway in Drosophila

regulates differentiation of precursor cells into fully functional R7 photoreceptor cells in the eye

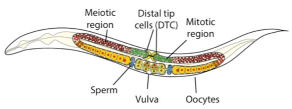
- evenless (encoded by sev) is a receptor tyrosine kinase that, when receiving a signal from its ligand Boss, dimerizes, self-activates through phosphorylation, and transduces the signal for cell differentiation into the nucleus
- downstream effectors (such as Ras1, GTPase, SOS..) have been identified in the dominant modifier screens using a sevenless sensitized background (constitutively active sevenless, Sevact)
- The phenotypic effect of the Sevact mutation can be reversed by loss-of-function mutations in sos, ras, raf, and the gene encoding MAP kinase, but not by a mutation in boss.

Can you, analyzing the information above, infer the epistatic relationship between the key players of the sevenless pathway and draw a regulatory pathway using the template below?



Example 3: Control of mitosis vs. meiosis

- *In wild-type C. elegans, the germ line proliferates starting from two primordial germ cells (called Z2 and Z3), which are set aside in the early embryo and only start proliferation after the worm has hatched*
- *continued rounds of mitosis take place in the two the somatic gonadal 'stem cell niches', which are each made up by one distal tip cell (DTC = does maintain germ stem cells)*
- *Once germ cells have left this niche, they start to differentiate and enter meiosis and therefore not proliferate anymore*
- *That means that there are mitotically dividing at the, and cells that are meiotically dividing as they move away from the DTC*



But what makes germ cells that divide mitotically switch to meiosis?

→ Notch signaling pathway controls transition

→ The DTC expresses the protein LAG-2 (wie Delta), which acts as a ligand for the Notch receptor GLP-1

→ GLP-1 is cleaved off and enters the nucleus, where it activates the transcription of genes that promote the mitotic and prevent the meiotic cell cycle.

→ If cells move away from the DTC, they lose their contact to the DTC and LAG-2 can no longer activate GLP-1 in these cells, thus, these cells enter meiosis

Interestingly, glp-1mRNA is present in all germ line cells, but GLP-1 protein is restricted to the mitotic region.

- This is accomplished, in part, by GLD-1, a translational repressor of glp-1mRNA.
  - GLD-1 binds directly to the glp-1 3'-untranslated region (UTR), thereby preventing GLP-1 translation in the meiotic region of the germ line.

But why does GLD-1 not inhibit GLP-1 translation in mitotic cells, preventing the receptor from being transported to the plasma membrane?

- accomplished by the action of the FBF proteins
- These proteins are expressed in the mitotic region and their expression is activated by GLP-1 signaling.
  - the FBF proteins are sequence-specific RNA-binding proteins that bind to the 3'-UTR of the target mRNAs gld-1 and gld-2 mRNAs, which encode proteins required for meiotic differentiation and prevent so their translation

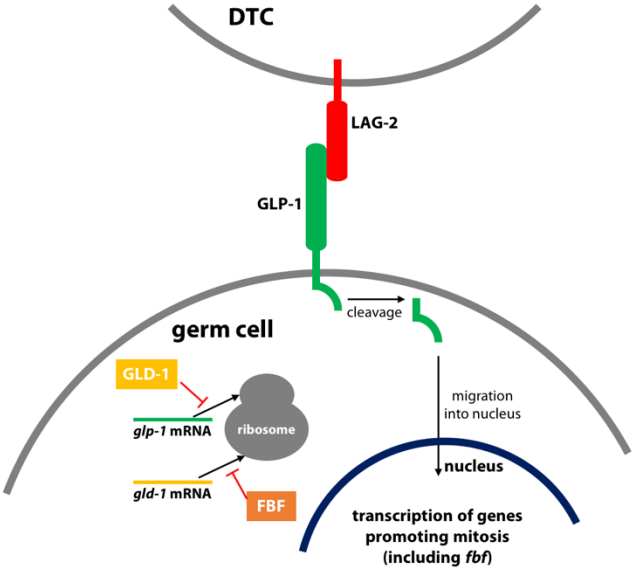


Figure 3-7 Signaling pathway controlling entry into meiosis of germ cells in *C. elegans*.

- *When GLP-1 signaling is turned off in cells moving away from the DTC, FBF abundance is reduced and GLD-1 and GLD-2 proteins can be translated where they promote the entry of germ cells into meiosis by activating pro-meiotic genes.*

Examination of mutants:

- *Loss-of-function mutations in glp-1, lag-2 or fbf-1/2 result in immediate entry of germ cells into meiosis without prior mitotic expansion of the stem cell pool.*
- *double loss-of-function mutations in gld-1 and gld-2 prevent germ cells from entering meiosis; instead, the germ line stem cells continue to divide mitotically, causing a germ cell tumor.*

- *germ cells in glp-1(lf); gld-1(lf); gld-2(lf) triple mutants do not enter into meiosis and form a mitotic germ cell tumor. Thus, the default state of a germ line stem cell in the absence of any input is to divide mitotically.*

What phenotype does the current model predict for the following mutant combinations?

- glp-1(lf); gld-1 (lf)
- glp-1(gf); lag-2(lf)
- lag-2(lf); gld-1(lf); gld-2(lf)

glp-1(lf); gld-1 (lf): germ cells enter meiosis without previous mitosis because gld-2 is redundant with gld-1in promoting meiosis acting downstream of glp-1.

glp-1(gf); lag-2(lf): germ cells form a mitotic tumor because glp-1 is epistatic to lag-2.

lag-2(lf); gld-1(lf); gld-2(lf): germ cells form a mitotic tumor because gld-1 and gld-2 are epistatic to lag-2.

RECONSIDERING GENE IDENTIFICATION USING NOVEL MAPPING STRATEGIES (OL)

to determine which genes are affected by the mutation

→ *position of these loci is determined on the basis of recombination frequencies*

Complementation groups and mapping

before mapping:

- complementation tests to group mutations into complementation groups → *A complementation group contains all mutant alleles that were discovered in a screen*
  - complementation group consists of mutations that do not complement each other; thus, these mutations are in the same gene.
  - For mutations that are in the same complementation group we already know they are on the same location on a chromosome

- allelic series:
  - alleles of a gene can affect the same phenotype, but they may differ in their strength and from moderate to high activity = if multiple alleles are present → dominance hierarchies
  - revealed by observing the phenotypes of each possible heterozygote offspring
  - Bsp blood types: controlled by 3 alleles: I^A, I^B and *i*.
    - *i* is recessive, and $I^A$and I^B are co-dominant, but both are dominant to *i*.
    - allelic series for the alleles determining the blood group is: $I^A = I^B > i$

- hierarchy of alleles can determine which allele from one complementation group is chosen for mapping: this is usually, the allele exhibiting the strongest phenotype that may be caused by a complete loss of function.

There are some general rules that apply to the strategy of mapping that we have already discussed, but will refresh here:

- Mapping is always a relative term. The position of gene can only be mapped relative to the position of a gene or marker for which the genomic location is already known (the reference points on a genetic map).
- Genes that are close together on the same chromosome are linked and do not segregate independently. They are inherited together.
- Linked genes lead to a larger number of progeny that resemble the parental class than expected if the two genes were to assort independently (i.e., unlinked genes),
- The underlying mechanism is recombination during meiosis and crossing over.
- The farther away genes are from each other, the greater is the chance that a crossover will occur between them.
- Recombination frequencies reflect the physical distance between genes (assuming that recombination is random).
- Recombination frequencies of two genes vary between 0 and 50%.

- Linkage:
  - describes the fact that genes that are nearby on the same chromosome tend to stay together during the formation of gametes = inherited together

Mapping by recombinant frequency

to determine the linear order of linked genes along the same chromosome

- accomplished by carrying out a testcross = mating between an individual that is heterozygous for two or more genes and one that is homozygous recessive for the same genes.
  - Genes located on different chromosomes show a recombination frequency of 50%
  - Genes located far apart on the same chromosome show a recombination frequency of 50%, because crossover events are frequent enough that it is likely for recombination to occur
  - the closer two genes are to each other on a chromosome, the smaller the recombination frequency will be (approaching 0% if the genes are very close).

⇨ Thus, a recombination frequency of less than 50% indicates linkage.

⇨ recombination frequency is a measure for the distance between linked genes, and recombination frequencies can be used to establish linkage maps showing the relative position of several marker genes.

In what cases is the probability of a crossover between two linked genes NOT proportional to the distance between them?

An exception is if two genes are at or close to the centromere. Also, if the distance between two genes is more than 30 m.u., the distance will be underestimated because the frequency of double crossovers is increased.

First chromosomal linkage map

- that frequency of crossover is a function of the genetic distance between genes was already realized in the early 19thcentury by Thomas Morgan and Alfred Sturtevant
- they could use the frequency of recombination between the mutations to draw a chromosomal linkage map for the genes involved. They defined the unit of genetic distance as:

- Recombination frequency = Nr. of recombinants/total number of flies x 100 = Map distance

- *One map unit is equivalent to what percentage of recombination frequency?* → *1%*
- results from Sturtevant showed that the probability of crossover occurring between two genes that are relatively close on the same chromosome were additive. → Thus, the map distances between genes are additive, which allows the construction of a linear chromosome map.

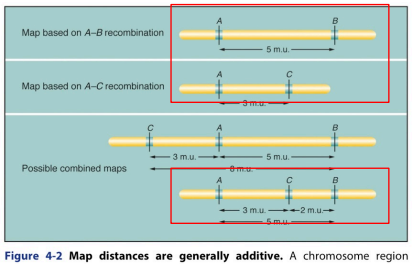


Figure 4-2 Map distances are generally additive. A chromosome region

Mapping with molecular markers

so far: *mapping genes using recombination frequencies by counting visible phenotypes that were produced by the different mutations*

Disadvantage:

- *phenotypes are often laborious to score and may interfere with the phenotype of the mutant of interest*
- ***because mutations with easily scored, viable phenotypes are relatively infrequent, the mapping resolution available using this approach is limited.***

Solution*:*

- *A much higher degree of differences in the DNA between two chromosomes is available at the molecular level, but not producing visibly different phenotypes*
- *sequence differences can be thought of as molecular markers. Their position can be mapped by recombination frequencies in the same way as mutations producing visible phenotypes.*

Types of molecular markers

- SNPs and Indels:
  - The alleles of such markers are co-dominant and usually phenotypically neutral
  - high density of SNPs in the genome and the availability of rapid scoring methods make for an attractive tool to analyze recombinant chromosomes

- VNTR (simple-sequence-length-polymorphisms) and repetitive DNA sequences:
  - Bsp für VNTR sind copy number variations = CNV
  - CNV is a DNA segment that is > one kilobase (kb) and that is present at a variable copy number in comparison with a reference genome.
  - CNVs are relatively frequent (e.g., between 5- 9% of the human genome are estimated to be CNVs).
  - While some of them have no apparent influence on the phenotype, while others have been definitively linked with a disease.
  - best-defined CNVs are the trinucleotide repeats (TNRs), which consist of three nucleotides repeating in tandem.
  - CNVs resulting in a massive increase of TNR copy have been shown to be a cause of Huntington's disease

General principle of mapping a mutation using a genetic marker considering (only a single chromosome & haploid)

Organism: Neurospora Crassa

The 2 strains, Mauriceville and Oakridge, have five different genetic markers (location known) that each has two different alleles, M (for the Mauriceville strain) and O (for Oakridge).

- In Mauriceville, all markers have the M allele, and in Oakridge all markers have the O allele.
- Mauriceville shows the wild-type phenotype (+),
- Oakridge has been mutagenized and now contains the mutation of interest, m, whose location is unknown.

1) The mutant strain Oakridge is now crossed with the wild-type strain Mauriceville and the progeny analyzed.

2) For every marker, the allele present is compared to the resulting phenotype in order to identify the marker location where the marker allele correlates strongest with the mutant phenotype

3) Here it is marker 4: bcs marker 4 always present as the O allele in the progeny showing the mutant phenotype, and always present as the M allele in the progeny showing the wild-type phenotype.

*5) Since the **allele present at marker position 4 correlates strongest** (and shows therefore the **strongest linkage**) with the resulting phenotype, this marker must be closest to the mutation m and thus allows pinpointing the mutation to this location in the genome.*

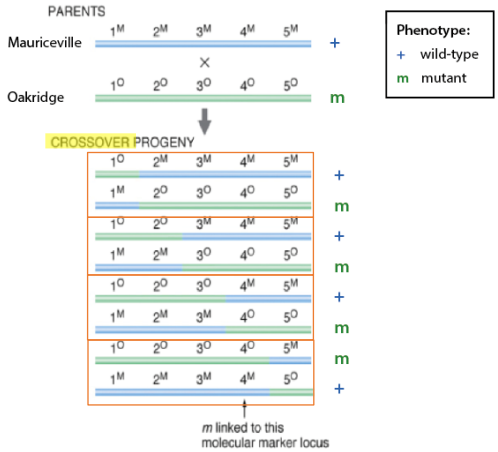


Figure 4-4 Principle of marker mapping. Two strains of the fungus *Neurospora crassa*, Mauriceville and Oakridge, are represented in blue and green, respectively. Both strains contain five different genetic markers (1-5) with known location, with each having two different alleles *M* or *O*. Mauriceville is a wild-type strain (+), while Oakridge shows a mutant phenotype (m). Mauriceville and Oakridge are crossed and the progeny analyzed for the occurrence of the alleles markers in respect to the resulting phenotypes (+ or m) in order to map the mutation *m* to one of the molecular markers.

Principle of SNP mapping

SNPs are DNA polymorphisms at a specific position in the genome that differ in only a single nucleotide, with each variation of the nucleotide being present to some degree within a population.

→ *SNPs are thus well suited to be used as molecular markers for gene mapping*

→ *allow for gene mapping to a few kilobases = at a very high resolution*

- In order to map a mutation of interest to a SNP, the SNP has to be detected and identified molecularly with the use of so-called SNP microarrays.
- SNP microarrays contain many different nucleotide sequences on a chip allowing to simultaneously bind and identify thousands of SNPs (up to 500'000) over the entire genome

Would you expect a higher or lower than average SNP density in gene coding regions?

- A lower than average SNP density, because these SNPs are most likely selected against if they produce a phenotype that is harmful for the organism
- In non-coding regions, the effect of SNPs is likely to be less harmful

Example of a SNP mapping approach in C. elegans:

- DNA sequence polymorphisms between the wild-type C. elegans strain (Bristol) and a closely related strain (Hawaiian) are used as genetic markers.
- these two strains differ by around one SNP or Indel every 800 base pairs (like humans) → every gene in the genome is marked by multiple SNPs or Indels.
  - Bristol strain contains the recessive mutation to be mapped and a visible, recessive marker as reference to identify recombinants on the chromosome of interest
  - only Hawaii-Strand contains SNP markers! (polymorphic line!

The first phase, chromosome mapping, is similar to traditional two-factor mapping and seeks to identify the relevant chromosome and rough position of the gene of interest.

The second phase, interval mapping, seeks to place the gene of interest in an interval between two SNPs, and can be used iteratively to fine map the gene.

1) the two strains are crossed and the mutant F2 progeny of such a cross are analyzed for their distribution of SNP markers

2) leads to a heterozygous F1 generation. In these animals, recombination can take place, leading to an F2 progeny that can be screened by looking at the visual marker.

3) In F2 progeny, recombination leads to

- animals containing the mutation, but that do not show the marker (because they are heterozygous for the marker, left).
- animals that show the marker, but not the mutation (right case in picture)

4) Genomic regions close to the mutation of interest show a small incidence of Hawaiian SNPs, because **no recombination has taken place** at this position. (man schaut wo nicht mit SNP rekombiniert wurde, den dort wo keine recombination stattfand heisst das, dass dort die Mutation liegen muss.)

5) Thus, in the mutant animals (left), **the piece of Bristol sequence that is present in every mutant animal must be close to the mutation.** Thus, this region must be "free" of Hawaiian SNPs.

On the other hand, unlinked regions (regions that are far away from the mutation) contain an even distribution of Hawaiian and Bristol SNPs.

6) genotyping of SNPs by PCR of single recombinants across the region reveals which of the SNPs are always Bristol → these must be linked to the mutation and gives therefore the position of the mutation

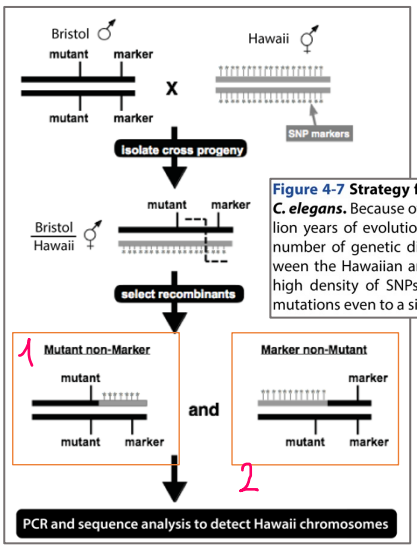


Figure 4-7 Strategy for high-resolution SNP mapping in *C. elegans*. Because of geographical separation, several million years of evolutionary drift have led to a considerable number of genetic differences (DNA polymorphisms) between the Hawaiian and Bristol *C. elegans* populations. The high density of SNPs in the genome allows to fine map mutations even to a single gene.

To identify the mutated genes, we first need to cross our isolated mutated animal with a polymorphic line. This line contains genetic material with many SNPs which will help us to which helps us to detect the approximate location of the mutation in the zebrafish genome.

After the whole genome sequencing of the mutated fish, we simply need to look where the recombination frequency appears the lowest and look for the mutation in this location.

which genomic region has picked up Hawaiian genotype (polymorphic) -since It is a recombinant a specific part of genome has to be hawaiian

-> now you can determine SNPs and determine where the chromosome is grey and where black → map the location of your mutant

when you have genotyped all 200 mutant non-makrer recombinants:

→ you will find 1 recombinant where the grey chromosome comes closest to the mutation, **you will find the marker closest to the mutation only once**

→ all the markers that are further away from the mutation, you will find them multiple times (frequency how often you find them, gives you information how far the marker is apart)

How can this SNP mapping approach be used to map dominant mutations?

For dominant mutations **pick the non-mutant progeny in the F2 and look for regions where the Hawaii genotype is enriched.** This works well if the allele is completely dominant. If semi-dominant, one needs to breed to the F3 generation to identify recombinants that have really lost the semi-dominant allele.

Combining whole-genome sequencing and SNP mapping

"bulk segregant analysis":

- instead of analyzing each individual animal of the F2 generation for the SNP allele to be present, **the F2 animals carrying the mutant phenotype were singled out to fresh plates.**
- These F2 animals were left to produce progeny by **self-fertilization**
- **this F3 progeny of around 200** F2 recombinants were pooled, **DNA was prepared and the pool was subjected to whole genome sequencing** (WGS).
- In regions unlinked to the mutation = far away, the parental chromosomes will recombine in a largely non-biased manner.
  - unlinked loci will appear in even ratio of Hawaiian vs. Bristol SNPs → the sequencing will show a

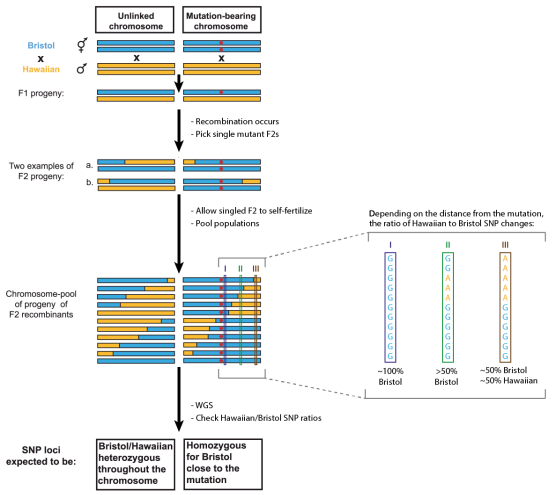
in 1) you look for regions **where you never find the hawaiian genotype (polymorphic)**
in 2) you look for regions where you never find the homozygous black genotype (only find Hawaiian)

50/50 ratio of Hawaiian versus Bristol SNPs

- the closer a SNP (of one individual) is to the mutation (which is on the other individual), the more rare it is to

find a recombination event between this SNP and the mutation.

- In this case, the sequencing will detect less Hawaiian SNPs than Bristol SNPs in regions that are closer to the mutation.
- In the areas very near to the mutation, the sequencing will detect only Bristol SNPs, since the SNP and the mutation are too close for recombination to take place.



Which factors determine the mapping resolution one can expect in such a bulk segregant analysis? How can mapping resolution be increased?

The number of recombinants determine resolution as long as the number of markers is not limiting. Resolution can only be increased by looking at more recombinants or by selecting for informative recombinants in the region of interest using phenotypic markers.

Integrated genome maps

A recombination map thus represents the arrangement of genes on a chromosome and distances are indicated in centiMorgans. However, ***these maps are hypothetical constructs***

- ***Physical genome maps*** *are constructed by identifying the base-pair sequence of genes as well as their position along the chromosome,* ***and thus provide exact DNA sequence information.*** *The units of distance on a physical map are numbers of DNA bases*
- ***Whole genome sequencing*** *provides the most thorough tool to construct a complete physical genome map* ***at single base-pair resolution*** → *large numbers of small genomic fragments are sequenced and then assembled into a whole sequence*
- *later assign genes to the sequence, a process called annotation*
- *Physical maps are now available for most of the genetic model organisms.*

⇨ ***the phenotype mapped on the recombination map can be linked to a function deduced from the physical map.***

⇨ *Thus, both maps contain information that is complementary:*

⇨ *the physical map shows a gene's possible action on the cellular level*

⇨ *the recombination map reveals the effect of the gene on the phenotypic level.*

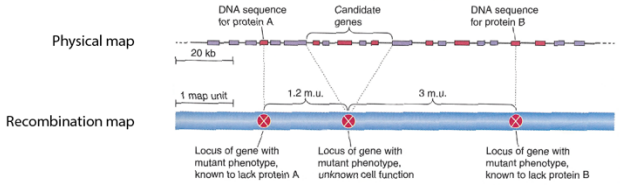


Figure 4-9 The alignment of physical and recombination maps allow to connect a phenotype with the function of a gene that is currently unknown (purple), or suspected (red) from another organism.

alignement builds up an integrated genome map

- *integrated genome maps are crucial for determining the genetic origin and reason for a disease, for identifying surrogate genetic markers for the disease, as well as for finding correlations between diseases and between disease and genotype*

Integrated genetic maps for any organisms are nowadays easily accessible throughout the internet, for example:

- Human: www.ensembl.org
- *C. elegans*: www.wormbase.org
- *D. melanogaster*: www.flybase.org

In which regions of the genome are the gene distances shown on physical and on recombination maps out of proportion?

- In regions where recombination occurs more frequently (recombination hot spots) and less frequently (cold spots) than on average.
- Examples are centromeric regions: here, recombination is lower.
- Hot spots are for example CNVs, e.g., trinucleotid repeats cause "fragile sites" within the DNA sequence that are more prone to recombination.
- Chromosome ends are also hotspots for recombination

VORLESUNG LESSON II-IV: 20.10

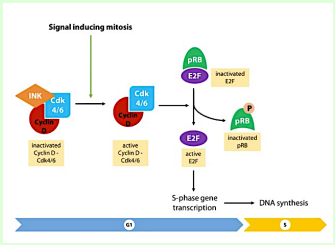
Lay focus on Epistasis!

- power of epistasis: even though you don't know proteins of pathway, you can bring ordnung into the mutants, by producing double mutants by combining opposite phenotype individuals

Now you have the model (picture below), answer this question:

You take mutation in pRB (always entering S phase) and combine it with mutation of CDK4 (never entering S-Phase) – what will be the expected phenotype?

- CDK mutant = not enable phosphorylative of pRB → (but pRB only stays **inactive** if it gets phosphorylated) → so now pRB will **always be active** → **now it can bind and inactivate E2F** → **NO TRANSITION INTO S PHASE**
- ***for erfolgreiche Transition into S-Phase, muss pRB von der Kinase phosphoryliert und damit inhibiert werden, da sonst E2F inhibiert wird, was dazu führt, dass keine Transkription der S-Phase Gene erfolgt***



- since it is a double mutant: there is also a loss of pRB!!!
- That means pRB can't even be active since it is not functional → that means, even pRB is not inhibited by the Kinase, pRB is not functional and therefore cannot bind to E2F! Therefore E2F will always stay active and able to indicate S-Phase transcription

⇨ **DOUBLE MUTANT WILL TRANSIT INTO S-PHASE!**

⇨ IMMER DIE MUTATION DES MOLEKÜLS DASS AM WEITESTEN DOWNSTREAM IM PHATWAY LIEGT, IST ENTSCHEIDEND FÜR DEN PHÄNOTYP!

⇨ IN THE DOUBLE MUTANT YOU WILL THEREFORE SEE THE pRB SINGLE MUTANT PHENOTYPE!

neue Frage: pRB und E2F double mutant – phenotype?

- pRB will not be able to inhibit E2F → E2F will always be active
- BUT: E2F is also mutated and therefore not functional, also bringt es E2F gar nichts immer aktiviert gekonnt zu sein → also können keine S-Phase Gene transkribiert werden

⇨ THE DOUBLE MUTANT WILL STAY IN THE G1 PHASE AND WILL SHOW THE SAME PHENOTYPE AS THE E2F SINGLE MUTANT

⇨ SO YOU CAN CONCLUDE THAT pRB ACTS UPSTREAM OF E2F AND CDK UPSTREAM OF pRB

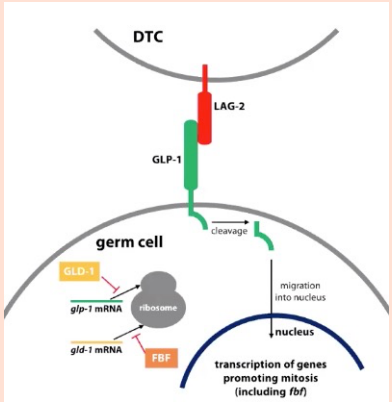
if double mutant forming is not efficient – bcs the various loss of function mutations lead to the same phenotype – how to ordnen die Reihenfolge of those mutated molecules? (z.B of MAP-Kinase pathway)

⇨ YOU NEED TO COMBINE GAIN AND LOSS OF FUNCTION MUTATIONS ALLELES! IF ALL LOSS OF FUNCTION ALLELES SHOW THE SAME PHENOTYPE!

- **you cannot only analyze loss of function phenotypes that cause the same phenotype**
- you need to construct at gain of function mutations too!

NOW HIS FAVORITE EXAMPLE:

- LAG2 (expressed by mitotic somatic cell) binds to receptor GLP1 expressed by germ cell
- germ cell stays as long meiotic, as LAG2 is bound to it's receptors → cannot enter meiosis because FBF blocks translation of specific mRNAs that promote Meiosis, like GLD1
- GLD1 mRNA encodes protein, extracts cells to enter Meiosis → also leads that GLP1 Receptor cannot be produced anymore and cell cannot get into mitosis anymore
- as soon as LAG2 or GLP1 is mutated, cells cannot stay mitotic
- mut GLD1 mutated → cells cannot enter meiosis
- what if double mutant: GLD1 (promeiotic) and GLP1 (promitotic) – what is the Phenotype?



- GLP1 Receptor mutated: is not receiving Mitotic signals

- GLD1 mutated: cell not receiving meiotic signals

→ <u>cell with no input? DEFAULT FATE IS MITOSIS! - SCHUTZMECHANISMUS (NO PREDICTION WAS POSSIBLE HERE)</u>

⇨ EPISTASIS ANALYSIS IS POWERFUL BUT PREDICTION IS NOT ALWAYS CLEAR FROM THE MODEL – LIKE HERE ONLY EXPERIMENT CAN TELL ANSWER

Novel mapping strategies

Elucidate genetic linkage: measure recombination frequency

- same chromosome = linked: when recombination frequency is low (< 50%, 50% wenn auf selbem chromosom aber chromosom sehr gross, also sind gene einfach sehr weit auseinander auf demselben Chromosom)

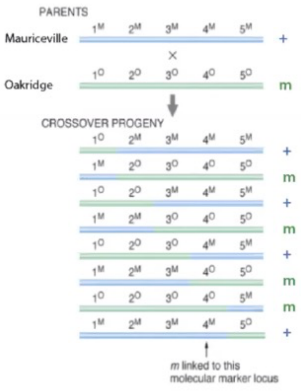
<u>map mutation "m":</u>

- you need to have a strain with recessive anatomical marker (1-5) and mutation and another polymorphic strain, containing many SNPs (select recombinants by marker)
- you need to know the genotype of both strains!
- both strains know cross and undergo recombination
- is the position in the genome of individuals showing the mutant phenotype green?
- all individuals showing the WT phenotype have marker 4 in oakrige form
- all individuals with mutation: the marker 4 is always blue, also the Mauriceville form

⇨ all other markers you find mutants where marker #4 is blue

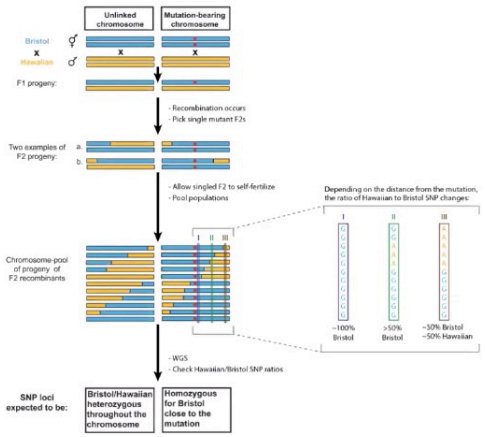
⇨ so you look for the marker that always segregates with mutant phenotype – will be most closely linked to the mutation

⇨ do this in large numbers



Marker free Mapping by whole genome sequencing

- only got possible when WGS got less expensive



- pick many mutants of F2 generation = hybrids of blue-orange phenotype
- extract DNA and mix → sequence whole pool of DNA in bulk (all DNA together in one tube)
- look at aligned sequence and compare with the blue reference sequence and ask how often do I find the Hawaiian genotype in percentage at this location in comparison to the blue Bristol genotype
- <u>where the percentage of Hawaiian (polymorphic strand) is =0, there your mutation must be located</u>
- <u>only limiting factor is the number of recombinants</u>
- you get 2 answers at the same time: position of the gene AND point mutation/variation that causes the phenotype you are mapping

Trick Question:

use same approach for mapping of dominant alleles – how?

→ point mutation generated (one red dot) is dominant

- WT will be recessive – we don't map dominant mutated allele sondern WT allele

⇨ to map dominant mutations, you map the recessive WT!

⇨ pick recombinants that do not show mutant phenotype sondern WT phenotype!

⇨ and gene will map where we always find the orange genome (Hawaiian polymorphic)

AN INTRODUCTION TO QUANTITATIVE GENETICS VORLESUNG (26.10.20)

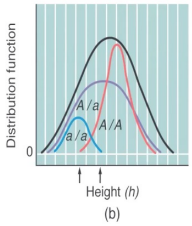
Measuring the genotype to the phenotype contribution

→ *There is no either WT or mutant in nature! There are only different expressivites!*

→ *human genetics is the prime example for quantitative genetics (Eltern generieren sehr untersch. Nachkommen)*

→ *How much is actually genetic/heritable, and how much of the variability is due to environment changes*

How to QTL (quantitative trait loci) analysis to identify the genetic heritability underlying quantitative phenotypes:



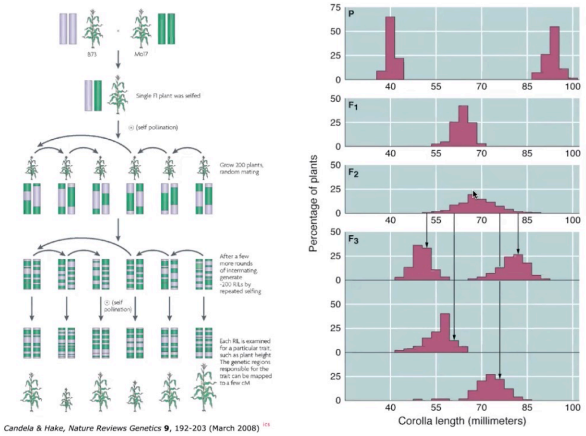
- single locus showing co-dominance of A and a (heterozygous in between of homozygous phenotypes)
- quantitative trait show continuous phenotypes
- Heritability = "regression of the mean"
- Environment = body size did increase in the last 100 y, an manchen Orten wie Appenzell waren Menschen auch Kleiner als in Genf, heute gibt es diese Unterschiede nicht mehr
  - amount of food you take in is the main determinant of the body size! → genetic variation is nowadays very small since we can eat as much we want

Genes and environment: Norm of reaction studies

- phenotype = size of drosophila eye
- with higher temperature, the eye size decreases in WT
- in mutants with small eyes, they will respond differently to the temperature! → eye size goes sogar up!
- nearly impossible to predict how specific alleles correspond to changing environmental factors!

Recombinant inbred lines (RILs) as a tool to study quantitative traits

1) find strains homozygous inbred strains with different phenotype (B73 and Mol17, both homozygous for all alleles) – you can also do multi-dimensional RILs with more than 2 strains
2) cross them, growing hybrid F1 with 50% grey and green genotype, and start inbreeding
3) going trough multiple generations and always inbred them to generate a inbred strand, which is like a puzzle made of green and grey genotype, until you have around 200 inbred lines
4) each inbred line has a different genetic composition in comparison to others, but within each lines, all the individuals have the same genotype caused by inbreeding
5) measure phenotype and plot those values
6) in the end you simulate **a natural population**



How can we get at the genes that cause one recombinant inbred line?

Principle of: QUANTITATIVE TRAIT LOCI (QTLs)

→ ***continuous phenotype (like human size or length of a plant) are not caused by single genes but rather by multiple gene interaction that causes the phenotypic spectrum of variation***

- *single genetic loci* ***that contribute in combination*** *with other loci to the phenotypic variation are called* ***quantitative trait loci (QTLs)***
- *A single* ***QTL never determines*** *the entire degree of variation → it's always the combination of all QTLs together that cause the phenotype*
  - *it is not known how much QTLs for example influence i.e the human body size*
- *Some QTLs may show* ***epistatic interactions btw. each other (one QTL only matters in a certain context)***
- *in the end, every QTL corresponds to a gene that exists in different polymorphic states/ has different alleles!*
- *How can we identify and map QTLs?*

QTL mapping

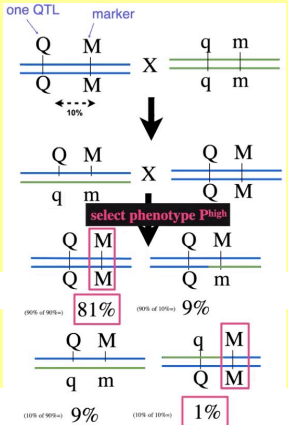
We can do totally normal recombination mapping, but the recombination frequencies we get are not proportional to the gene distances!!!

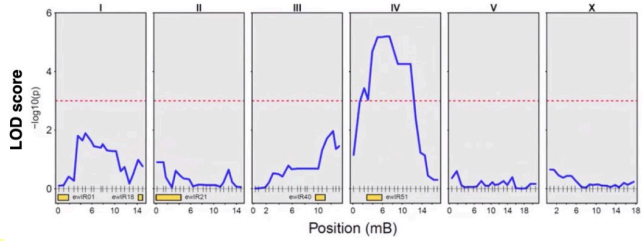
-> **recombination frequency** is nämlich nur 1 Factor! (has experimentally to be found out)
-> 2nd Factor is the **relative contribution of the measured QTL to the whole phenotype**! (has to be found out)

The linkage of a marker to a QTL is determined by its distance from QTL and by the relative contribution of the QTL to the phenotype.

hypothetical example: just look at 1 QTL that contributes to a specific phenotype

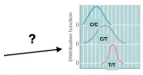
- **Two lines** are selected that differ in a parameter P (P= size, weight or behavior etc.). The line with high P is probably homozygous for the Q allele of one (of many) QTLs, the line with low P is probably homozygous for the q allele.
- We assume that:
  (1) The QTL contributes to 90% to the phenotype (meaning individuals with a phenotype of P^{high} have a 90% probability of being homozygous Q/Q).
- (2) A polymorphic marker 10 m.u. away from a QTL is tested → we measre if 10 m.u = 10% is also experimentally confirmed
  - A theoretical **difference value (z-score)** of 82% can be predicted for this marker. (90% out of 90% + 90% out of 10% = 81% + 100% - 90% - 9% = 82%
  - so 10% map distance is expected → but we measure 100% - 82% = 18% map distance, so we think Marker M is further away from QTL than it actually is!
  - conclusion: gene distances when dealing QTLs are not proportional to the distance of the marker to the QTL → 10% ≠ 18%!
  - so in fact we don't know how far the marker is away from the QTL AND we don't know how much the QTL relatively contributes to the phenotype! (so 2 parameters we don't know: distance, and relative contribution!)





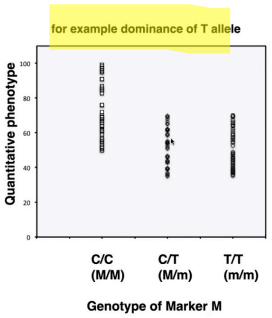
→ all we can do is measure this difference values = a theoretical z-score = LOD score = -log10 (p)

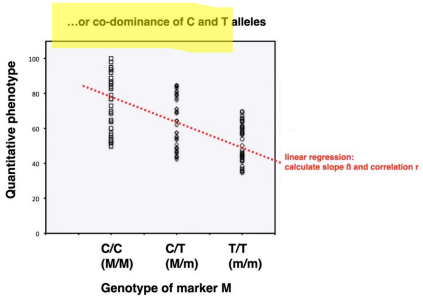
- therefore we measure the phenotype of 3 different genotypes of marker M (M/M, M/m, m/m)
- look at distribution of phenotypes in the different genotypes
- do the values show a normal distribution?



- do statistical tests to calculate probability p of a significant difference between the genotypes (e.g. using ß and r values from a linear regression analysis together with and sample size).
- calculate the LOD score (likelihood of disequilibrium) as LOD=-log10(p)
- repeat this for many markers distributed across the entire genome
- plot LOD score of the markers along chromosome positions

QTL Mapping





Now measure for each marker the likelihood for non-random distribution (that two populations are different for each other and do this for many markers) = for each of markers of all individuals you measure the quantified phenotype/trait and calculate the LOD score, that shows the probability that one or the other population of a particular genotype is different

for each marker, you measure LOD score and plot LOD score on y axis and marker position on chromosome on X-Axis -> efficient way to find many QTLs at the same time

Curve of QTL mapping → peak = a QTL = one or other genotype is enriched → here strong peak on chromosome 4: multiple markers that show a lot a high LOD score => that means there is 1 QTL on chromosome 4 that contributes a lot to this phenotype → must be gene around position 6 on chromosome 4

maybe also a weak QTL on chromosome 1 at position 3, but does not contribute much to the phenotype

Threshold for LOD = 3 → nur drüber = QTL

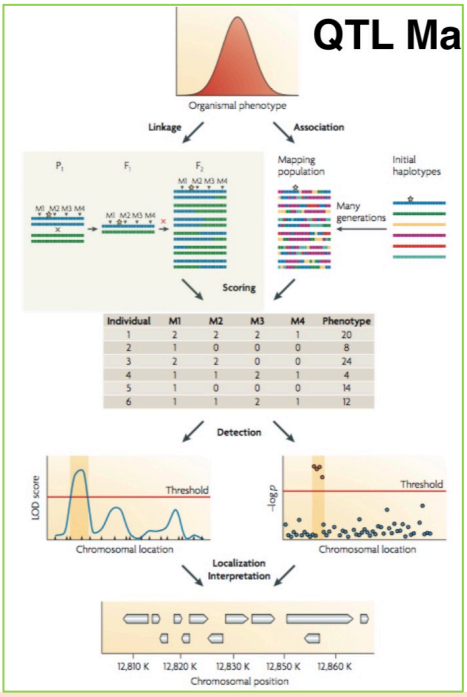
a quantitative phenotype is z.B lifespan, body size…

Summary QTL Mapping:

- Create RILs (recombinant inbred lines) between two polymorphic strains: gain hybrids in F1 generation and in F2 (puzzle of green and blue) you create recombinant lines and inbreed those recombinant lines and for each line you quantify average phenotype of all individuals from one specific RIL and determine the state blue or green for many markers (markers are genotyped) along the chromosome and calculate the LOD scores for each marker and plot the LOD scores vs. the position of the markers → LOD score represent a QTL
- we look for regions on the chromosome where the LOD scores are higher than expected = a disequilibrium = QTL
- In the end we want to find the QTL corresponding gene(s) that contribute to the quantified phenotype
- number of markers is nearly unlimited since there is whole genome sequencing -> so limiting factor is the number of recombinants you can isolate and genotype for its markers

in shorter words:

- Create RILs (recombinant inbred lines) between two polymorphic strains displaying variation in a quantitative phenotype (left)
- ... or use natural variation in strains isolated from the wild population (right) ***
- Order individuals according to strength of phenotype.
- Do genotyping, calculate and plot LOD scores for each marker along the chromosomes.
- *zusätzliches Bsp bei 42:00 min in Vorlesung*



Systems genetics = different QTL approaches to build up models

Also possible to make QTL analysis for molecular phenotype, such as gene expression level (eQTLs) that correlates with a specific phenotype or protein levels (pQTLs) or tall or short body size (QTTs)

→ man kann diese Varianten miteinander verbinden = integrated approach = Systems genetics

→ i.e you can molecular variants that are associated with changes in gene expression and correlate them with an atomical classical phenotype you can measure (verbinden von eQTL and QTT)

Example of his own lab:

- how does activity of oncogene change when going into different environment
- Generate recombinant inbred lines btw Bristol line (carrying RAS gene of function allele) and Hawaii → each RILs carrying the oncogenic mutation, but rest of individual is puzzle of genome mix
- → are there regions in the genome, are there QTLs that either enhance or suppress the activated Ras gene of function phenotype?
- → find multiple QTLs on different on chromosomes that influenced RAS activity
- then elucidate Gene that corresponds to QTL with introgression lines → gene amx2 was polymorphic

Human Genetics

in comparison, in GWAS, you need to use a natural population and don't have RILs (recombinant lines) → so you only can measure each individual once, in recombinant lines you can measure the phenotype of all individuals of the same RILs and take the average phenotype

- the power/resolution is higher with RILs = less noise
- bei QTL analysis & mappingreichen 200 recombinants und in GWAS brauchts mind. 1000 individuals to find QTLs
- GWAS is more reserved for human studies
- is able to identify relatives = long range familial searches

The 1000 genomes project (A QTL mapping strategy where you use the natural variation in strains isolated from the wild population ***)

- multiple SNPs are inherited together
- human genome behaves like blocks of lego
- single nucleotide polymorphisms microarrays: genotype millions of SNPs = markers gleichzeitig
- markers are so chosen that they cover all the endoding sequences → all 25'000 genes are covered by multiple SNPs and CNVs that are used as markers

important terms to know in human genetics:

human genetics is in fact the same as QTL analysis but the model organism isn't C.elegans sondern the human

- linkage disequilibrium, LOD score, odds ratio (all 3 signify, non random distribution of specific genotypes at a specific marker location with a phenotype => statistically significant enrichment of one genotype over the other with respect to certain phenotypes you can quantify)
- haplotypes (human genome is inherited in haploblocks where is little recombination)
- allele frequency (different SNPs have different frequencies in various populations around the globe, take that into account when calculate LOD score)
  - is not 50/50 allele frequency how it is in lab strains
- susceptibility genes (QTLs that render you more likely to develop a specific disease)

REVERSE GENETICS

INTRODUCTION

While **forward genetics** seeks to find the **genetic** basis of a phenotype or trait, **reverse genetics** seeks to find what phenotypes are controlled by particular **genetic** sequences

GENOME SEQUENCING (OL)

- *starts with a cloned segment of DNA, or a sequence, which is used to introduce programmed mutations back into the genome to investigate gene product and function. This strategy relies on engineering the genome through adding, modifying, and replacing genes.*
- *A prerequisite for reverse genetics is the knowledge of the DNA sequence of an organism*

independence of information and physical representation in DNA

uncoupling of information from the physical properties of the storage medium has two important consequences:

1. The same DNA-sequencing methods can be applied to virtually any piece of DNA, regardless of the specific information stored in it. A DNA strand encoding the amino-acid sequence of a protein can be read just as a DNA strand encoding a promoter region of a gene or even a completely random sequence.

2. Determining the base sequence of a piece of DNA captures the most relevant biological information about this piece of DNA.

Sanger Sequencing = first generation sequencing

= "dideoxynucleotide chain-termination method"

Sanger assembled all the components that are needed for a DNA-synthesis reaction in a test tube: a single-stranded template DNA, a primer, a polymerase, and triphosphate deoxynucleotides (dATP, dGTP, dTTP, and dCTP). When mixed together, the primer anneals to its complementary site on the template, the polymerase binds to the formed duplex and begins to extend the primer by incorporating complementary nucleotides according to the sequence of the template strand.

Sanger added one additional ingredient to the synthesis reaction: A dideoxynucleotide (ddNTP, figure 1-1 right). This

is a chemically synthesized nucleotide analog, in which the 3'-hydroxyl group is replaced by a hydrogen atom → prevents the addition of any further nucleotides to this growing strand

the polymerase can either incorporate a natural dATP in the growing strand and synthesis will continue with the next nucleotide; or, an artificial ddATP is incorporated, in which case the synthesis of this particular strand is terminated. Which of the two happens is determined by chance and by the relative concentration of dATP and ddATP.
Those strands where a "normal" dATP was incorporated will continue to grow until the polymerase encounters another thymidine, in which case there is again a chance that synthesis is terminated by addition of a ddATP

- The result of this synthesis reaction is a population of DNA strands of different lengths (fragments)
- fragments contain information about the positions where the template strand contained a base complementary to the dideoxynucleotide used in the reaction
- By performing the synthesis in four separate reactions, each containing a different dideoxynucleotide, one obtains four populations of newly synthesized strands. Together, the fragments of these strands encode the complete base sequence of the DNA strand
- 2nd step of Sanger sequencing is to read out this information by separating the newly synthesized strands by gel electrophoresis
- By stepping along this staircase from the bottom of the gel (shortest strand) to the top, the sequence of the newly synthesized strand can be read in the 5'-to-3' direction
- radioactivity-based version of the Sanger method was soon replaced by a fluorescence-based version
- The maximal read length of the Sanger method is 1000 bases. It is limited by the relative size resolution obtainable by electrophoresis
- the gel electrophoresis needs to separate two strands that are 999 and 1000 bases long. This brings the required relative resolution to 1/1000, which approaches the current technical limit
- error rate extremely low = 1/10'000 bp wrong

Advantages (why sanger is still used today):

- First, the Sanger method employs sequencing primers that can be chosen by the experimenter. This allows the targeted sequencing of a specific portion of a larger DNA molecule while the 2nd and 3rd generation workflows will sequence the entire DNA in the sample.
- Second, the Sanger method can read long stretches of DNA in a single reaction with very low error rates. In medical diagnostics it is therefore still common to verify candidate mutations found by 2nd and 3rd generation sequencing methods via the Sanger method.

Next generation sequencing = 2nd generation sequencing

SBS (Sequencing by synthesis) by Illumina

- method is based on a sequencing reaction that resembles the natural DNA-synthesis process.
- The main difference between the Sanger and SBS reactions is in the type of nucleotides: In contrast to Sanger sequencing where a mix of natural and modified nucleotides is employed, SBS uses only modified nucleotides (sanger = permanently attached fluorophore (ddATP), SBS = chemically cleavable groups (dCTP)

The basic biochemical process of SBS

Template molecules are attached to a solid surface (called a flow cell), a primer is annealed to the template strand and a polymerase is bound. The individual cycle of the sequencing by synthesis reaction then proceeds as follows.

1) The flow cell is flooded with a solution containing the four different types of fluorescently labeled, chain-terminating nucleotides.
2) The polymerase uses the template strand to extend the primer with the complementary nucleotide. The terminating group on the newly incorporated nucleotide prevents the incorporation of additional nucleotides.
3) Any non-incorporated nucleotides are washed away.
4) The fluorescence signal of the sample is measured and reveals, which of the four nucleotides has just been incorporated.
5) Both the fluorophore and the terminating group are removed by chemical cleavage and washed away. This prepares the reaction for the next cycle.

→ SBS reaction shares many components with the Sanger reaction (use of a polymerase, fluorescently labeled nucleotides…)

grosser Vorteil von SBS gegenüber Sanger Sequencing:

SBS method can be parallelized and miniaturized. Instead of improving the speed or accuracy of the individual reaction, SBS makes it possible to perform billions (!!!) of parallel sequencing reactions in a flow cell the size of a microscope slide → reading out the sequence information while the molecules stay attached to the surface. No physical transfer of reaction products from one instrument to another is necessary

→ the signal from a single fluorophore molecule would be too weak to be detected. Therefore, before the actual sequencing-by-synthesis reaction can be performed, the individual DNA molecules bound to the flow cell need to be amplified in a PCR-like process with the use of primers

→ SBS instrument simply has to record a high-resolution image of the flow cell's surface after each reaction cycle and the series of color changes at a given xy-coordinate reveals the sequence of the corresponding DNA molecule

Limitations of the 2nd generation sequencing technologies:

The need to synchronize reactions across molecules limits the speed and read-length of second-generation sequencing technologies

But, since there is no way for the molecules that have fallen out of lockstep to get back into lockstep in SBS, the fraction of out-of-lockstep molecules increases with every cycle → read length limited to 200-300, bcs nur so viele Runden möglich bis random fluorescence signale die Ergebnisse stören

Third generation sequencing: sequencing individual molecules

idestep the synchrony-based problems of second-generation sequencing techniques by performing the sequencing reaction on individual molecules

PacBio: fluorescence-based single molecule real time Sequencing (SMRT)

Advantage:

no need to synchronize the reaction of multiple molecules in a cluster -> the reaction no longer has to be stopped and reagents do not need to be exchanged at each step of the reaction cycle

→ all necessary reagents are added to the reaction mix and nucleotides are observed in real-time as they are added, one after another to the growing DNA strand using fluorescent labeled nucleotides

→ The polymerase cleaves the attached fluorophore during incorporation into the growing DNA strand, allowing it to diffuse away from the sensor area before the next labelled dNTP is incorporated

Disadvantage:

the signal that can be obtained from a single fluorophore molecule is very limited → problem when fluorescent noise

→ **solution**: keep the volume around the polymerase, in which the fluorescence is monitored to the extremely small size of a few zeptoliters (10-21 liters) → Nucleotides will briefly diffuse into the monitored volume, but will diffuse out again very quickly on the time scale of a few microseconds. Only when the nucleotide recognizes and binds to the template strand's complementary base in the polymerase active site, the nucleotide and its attached chromophore remain in the monitored reaction volume long enough to generate a significant fluorescent signal.

Nanopore technology: reading DNA sequences directly

DNA strand directly by passing it through a narrow protein pore in a membrane and measuring how the sequence of the passing DNA molecule influences the ionic current through this pore (similar to a porin protein)

The Pore is embedded into a polymer-based membrane that separates two chambers containing electrolyte solutions

By applying a small electrical voltage across the membrane that separates these two chambers, the electrolyte ions flow through the pore and generate an electrical current that can be measured

→ Bcs the DNA molecule is also charged, it is pulled through the pore from one chamber to the other where it restricts the flow of the electrolyte ions passing through the pore, which is reflected in the electrical current measured between the two chambers.

→ Due to their slightly different size, shape and polarity, the different bases have a slightly different effect on the electrolyte current.

Disadvantages:

The challenge in interpreting this signal is that this current depends on the base itself, but also on the identity of the neighboring bases.

Advantages:

Just as the PacBio technology, the Nanopore techology has no fundamental read-length limit and the quality of the reads also stays constant over the entire read length.

→ Read lengths of up to 200'000 bases have been reported and the error rate appears to be slightly higher than the 10% achieved by the PacBio technology.

→ Using this technology, the whole readout process in the Nanopore instrument takes place on a cheap-to-manufacture semi-conductor chip - none of the lasers, fiber optics, optical grating etc. of the fluorescence based instruments are needed (nur bei Nanopore! Bei PacBio immernoch grosses Gerät nötig!)

Zusammenfassung:

First-generation technologies are marked by a clear separation of the biochemical sequencing reaction and the read-out process.

In second-generation technologies this separation no longer exists. Instead the biochemical reaction and the readout are part of one integrated process that takes place in the same reaction vessel. Another feature shared by second-generation sequencing technologies is the need for an amplification step, in which individual molecules from the sequencing library are multiplied in a PCR-like process prior to sequencing.

Third-generation technologies have found ways to circumvent this amplification step and sequence individual molecules from the sequencing libraries directly, thus avoiding potential pitfalls from this amplification step and making much longer sequence reads possible.

How to sequence longer DNA strands

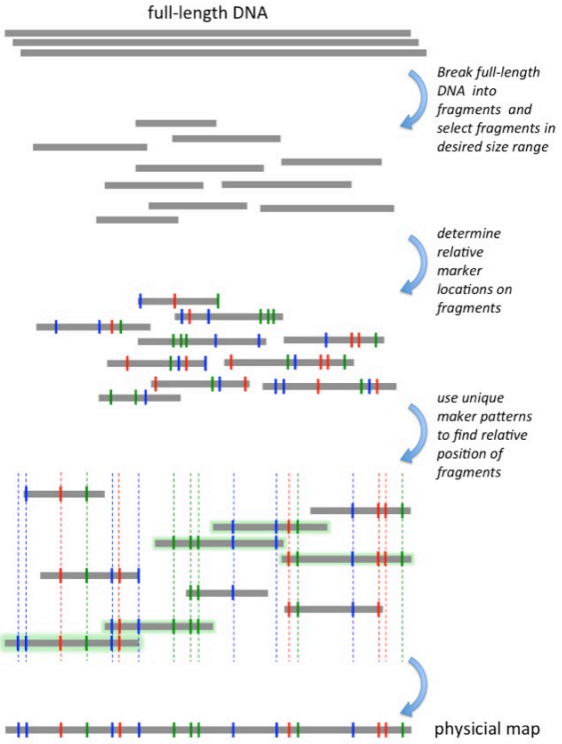
the initial full-length stretch of DNA has to be broken up into parts that are of the appropriate size for a given sequencing technique = **referred to as library generation**.

After the sequence of each of these smaller pieces has been obtained, these stretches of sequence have to be put back together in the correct order to obtain the sequence of the full-length stretch of DNA = **is called sequence assembly**.

library generation and sequence assembly are intimately linked parts of what is called a **sequencing strategy**.

There are two strategies for sequencing long sequences (e.g., whole genomes), approaching the task in very different ways: map-based sequencing and shotgun sequencing. In practice most large-scale sequencing projects currently use a hybrid of the two strategies.

Map-based sequencing



i) The full-length DNA is broken into large fragments of about 100'000 to 200'000 bp length using restriction enzymes.

ii) These large fragments are inserted into bacterial artificial chromosomes (BAC), circular DNA constructs that allow the insertion and propagation of long stretches of DNA (like bacterial plasmids do).

iii) **The position of each of these large fragments in the initial full-length DNA sequence is determined by physical mapping**.

iv) Then, a subset of these BACs are selected that spans the entire sequence.

v) The selected BACs are each digested further into fragments of about 10'000 bases.

vi) Again, physical mapping is used to determine the position of each of these smaller fragments within their respective BAC.
vii) These fragments in the plasmids are now sequenced using a primer that is complementary to a known portion of vector sequence directly adjacent to the 5' end of the inserted fragment.
iix) The sequence generated in this reaction will then reveal a substantial stretch of the sequence of this insert. ix)This sequence can then be used to generate the primer for the next sequencing reaction and so forth (this process is called primer-walking).
x)The process is repeated until the appearance of sequence stemming from the plasmid indicates that the entire sequence of the insert has been obtained.

Disadvantage:
process represents a very substantial logistical challenge. A diploid human genome, for example, contains 6 billion base pairs, which corresponds to a minimum of 600'000 fragments that each needs to be stored in a separate container and for which separate primers need to be designed and synthesized.

Shotgun sequencing

Here, the full-length piece of DNA is sheared into fragments of a size that can be sequenced in an individual read. These fragments are then sequenced individually.

Advantages:
In this approach, all of the subcloning, mapping and primer-walking steps are eliminated → But, the resulting library is completely unordered

How can the sequences of the individual fragments in this library be assembled to yield the full-length sequence?
The answer lies in the fragment sequences themselves. Because multiple copies of the initial piece of DNA were sheared randomly, most of the fragments will overlap partially with several of the neighboring fragments → One can use these overlapping sequence sections to find the neighbors of a particular fragment and then the neighbors
→ Assuming random sequences, an overlap of just 16 bases (4^{16} = 4.3 billion combinations) would be sufficient to uniquely identify a neighboring fragment in a library derived from an entire haploid human genome (3.2 billion bases).

Disadvantage:
many of the DNA sequences of interest contain duplicated genes or long stretches of repeat sequences → it is typically not possible to accurately assemble entire mammalian genome sequences using a pure shotgun approach and the shotgun sequencing data needs to be complemented by traditional mapping data.

HOW TO CREATE TRANSGENIC ANIMALS (OL)

Reverse genetics relies on engineering the genome through adding, subtracting, and replacing genes.
three major approaches to reach these goals: ***the generation of transgenic animals (adding genes), generation of knockouts (deleting genes), and genome editing (replacing or deleting genes)***

Transgenesis:
process where modified versions of a gene or a foreign gene – what we call a transgene - can be introduced into animal cells

Genome Engineering

How a gene can be altered:

- gene **deletions**/ knockouts are useful if the gene is not essential
- **replace** gene of interest by one that is expressed in the wrong tissue or at the wrong time during development
- genes can also be **engineered** so that they are expressed normally in most cells and tissues, but deleted in certain tissues. This is especially useful when a gene has different roles in different tissues

Reportergenes

Das Reportergen wird dabei meistens nach dem zu untersuchenden Gen, unter demselben Promotor eingeführt

- **create new types** of proteins in an animal, which allows following where, when, and how a gene of interest is expressed in an organism. → reporter constructs are created and introduced into the genome.
  - A reporter gene (often just called reporter) consists of a gene that is attached to a regulatory sequence of another gene of interest → scientists place the reporter gene and the gene of interest in the same DNA construct and insert it an organism.

- It is important to use a reporter gene that is not natively expressed in the cell or organism under study, since the expression of the reporter is being used as a marker for successful uptake of the gene of interest. (reporter ≠ gene of interest)
- Each reporter differs in the amount of information it provides about the expression of a gene

Translational reporter: gene fusion (antother gene is fused to target gene)

- Commonly used reporter genes (induce visually identifiable characteristics involve fluorescent or luminescent proteins.
- In such a reporter construct, the gene of interest is fused to the coding region of GFP
- When this reporter is introduced into the genome, the **GFP-tagged protein of interest can be tracked inside cells by monitoring** its fluorescence. We call this a translational reporter, because here, the translated fusion protein is the readout.

Advantages:

- Translational reporters can provide information about a gene's expression pattern, because additional regulatory information that may be present in introns or 3′-UTRs is included in such reporter constructs.
- In addition, translational gene fusions can also provide information about subcellular localization and the temporal aspects of gene regulation.

Disadvantages:

- **the fusion of a gene to GFP can sometimes disrupt protein function or even lead to toxicity of the chimeric product.**
- the **zu untersuchende protein (to which GFP is fused) may fold differently than the "normal" one, which may affect its binding to interaction partners**, mask potential phosphorylation sites or affect protein stability, which will lead to a false interpretation of the expression pattern.

Transcriptional reporters (promotor fusion – a promotor of another gene is fused tot he target gene)

(man fügt Promotor von anderem Gen noch ein, sodass immer wenn Zielgen abgelesen, wird das andere Gen auch exprimiert

- **consist of a promoter fragment from a gene** of interest that drives the expression of **another** gene, i.e. the gene **lacZ**, which encodes the bacterial protein beta-galactosidase = blue color
- Thus, the blue color indicates **in which cells the gene of interest is expressed within a tissue or animal**.
- Alternatively, transcriptional reporter constructs can also be created that drive the expression of GFP
- Typically, promoter fragments, immediately upstream of the start codon, contain a significant portion of the cis-regulatory information, necessary to provide a tentative expression pattern of the endogenous gene under study.
- Transcriptional reporters can be used to visualize the expression pattern for a gene of interest by fusing 5′-upstream sequences to GFP can be done in a number of ways and usually presents no technical challenge.

Advantages:

- Compared to translational reporters, however, promoter fusions may not give a complete representation of the real expression pattern of a gene, both spatially and temporally.

Disadvantages

- Since not the original gene but a different one (e.g., lacZ or GFP) is expressed, modifications of the expression pattern of the gene by external influences or by developmental clues cannot be studied if these modifications affect the protein level

How can you test whether your engineered fusion-protein faithfully reflects the localization and action of the wild-type protein?

You can cross animals expressing the fusion protein with animals that lack the function of the protein of interest (a mutant). Assuming that your mutant has a specific phenotype, if the fusion protein can rescue the mutant phenotype, you can be pretty sure that the fusion protein localizes and functions like the wild-type protein.

How transgenic animals are generated

→ the altered (verändertes) gene, once it is introduced into a cell, must recombine with the cell's genome to be stably maintained after cell division.

→ Furthermore, the altered gene must be integrated into the germ line so that it can be inherited to the next generation.

→ to control which animals contain the transgene, the transgenic construct contains a marker gene (erst wenn Individuals that got transgene in their germ cells are mated, the marker will be visible (in the F1) since erst dann the marker is vorhanden in the somatic cells and the corresponding phenotype is visible

Electroporation is one method used to introduce DNA into bacteria or cells in culture. → a brief electric shock renders the cell membrane temporarily permeable by shifting lipid molecules in the bilayer, allowing foreign DNA to enter the cytoplasm

- Electroporation allows cellular introduction of large highly charged molecules such as DNA which would never passively diffuse across the hydrophobic bilayer core

---

Transgenes can also be brought into cells by **viral transfer**, a technique frequently used in plant and mammalian cells, which we will not discuss here.

In plant cells, genes are frequently introduced by a technique called **particle bombardment**: DNA particles are shot through the cell wall into the cell with a special modified "gene-gun", since plant cells have a starre/rigide cell wall and electroporation does not work

- mechanical method to deliver DNA into cells using a gene gun (small heavy metal particles are coated with DNA and shot into the target cells under high pressure – is able to transform almost any cell type and is not limited to transformation of the nucleus, it can also transform organelles such as plastids
- common tool for labeling subsets of cells in cultured tissue

---

Microinjection is used to introduce DNA into mammalian cells, fly embryos or worms.

In C. elegans, the transgenic DNA is simply injected directly into the worm:

- injection strategy is determined by the worm's reproductive biology: the gonads of the worm are syncytial, meaning that there are many nuclei sharing the same cytoplasm
- The DNA solution is injected into the syncytial region of one of the arms, thereby exposing the DNA to more than 100 nuclei
- By chance, a few of these nuclei take up the DNA (because during cell division, the nuclear membrane breaks down).
- The transgenic DNA can now be present in the nucleus in two different ways:
  - first, the DNA can form an array with multiple copies of the DNA that exists as independent units that are not part of the chromosome = **extrachromosomal array**, (not wanted!)
  - or the transgenic DNA can be integrated into a chromosome and therefore allows **stable inheritance of the transgenic construct** → Since this integration is random, the incorporation of transgenic DNA may disrupt genes present on the chromosome, which may lead to lethality if an egg or a sperm with such an integration take place in fertilization **(integrated arrays)**

Microinjection in drosophila:

- during the syncytial stage of Drosophila development, the fertilized egg cell contains many nuclei that are not separated by cell membranes
- cells that will become the future germ cells are located at the posterior position of the egg → Injecting the transgenic DNA at this end of the embryo will result in a few nuclei that incorporate the DNA (hopefully these nuclei are destined to form the future germ cells, since then the transgene will be inserted into the germ line)
- Often, the marker used for transgenesis is a gene conferring red eyes, because transgenic flies can easily be detected by simply looking for red eyes (white (+) genotype = Marker gene that is available in transgene) in a microscope → implies that the embryos used for injection must have white eyes (white(-) genotype) since they act as a reference
  - Markers are erst in F1 generation ersichtlich, daher wird *Individuum 1 (white (-) mit Transgen und Markergen white(+))* gekreuzt mit *Individuum 2 (white (-) mit Balancer (Balancer damit herausgefunden wird wo Transgen inseriert)*
- the insertion of the transgene into the DNA **is random, thus, the transgene can be inserted into any of the four chromosomes.** → Thus, for further analysis of these

flies, it is important to determine on which of the chromosomes the transgene has inserted ***

Once you have obtained transgenic flies, how can you determine where the transgene is inserted? Do you know a trick how chromosomes can be marked in Drosophila? Hint: How can heterozygous-mutant be distinguished from homozygous wild-type flies in an F2-recessive screen?

Balancers are useful tools, because **they carry a visible marker mutation, such that the presence of this chromosome can be followed in crosses.**
Furthermore, **if flies are heterozygous for the balancer and the mutation (or a transgene in our situation), there is no danger of recombination to occur between homologous chromosomes** (which, for example, could separate the marker gene from the transgene if crossover occurs between those), **because the presence of the balancer chromosomes prevents recombination between the chromosomes**

*** Determining the insertion site of transgenes

1) the flies containing the transgene (TG) in their germ line would be crossed with flies that are white- and also carry a balancer for one of the chromosomes.

- The balancer (in this case for the 2nd chromosome *) carries the mutation Cy, which confers bent up wings, and the X-chromosomal cis-homozygous mutation white -, which confers white eyes.
- From the phenotype we know that they contain the balancer and the transgene (TG); however, these flies could have different genotypes:
  - If the TG is on the second chromosome, these flies would have the genotype TG / Cy
  - If the TG is NOT on the second chromosome, these flies would have the genotype + / Cy; TG (the TG is now on another chromosome, with the ; separating the different chromosomes)
- Therefore, we need another cross to determine where the TG is inserted.

2) We therefore cross the red-eyed Cy flies with balancer flies (white eyes, Cy). For the progeny of this cross, there are two possibilities:

- Three phenotypes occur if the TG inserted on the 2nd chromosome (left in figure 2-10);
- four phenotypes occur if the TG inserted on the 3rd chromosome (right in figure 2-10).

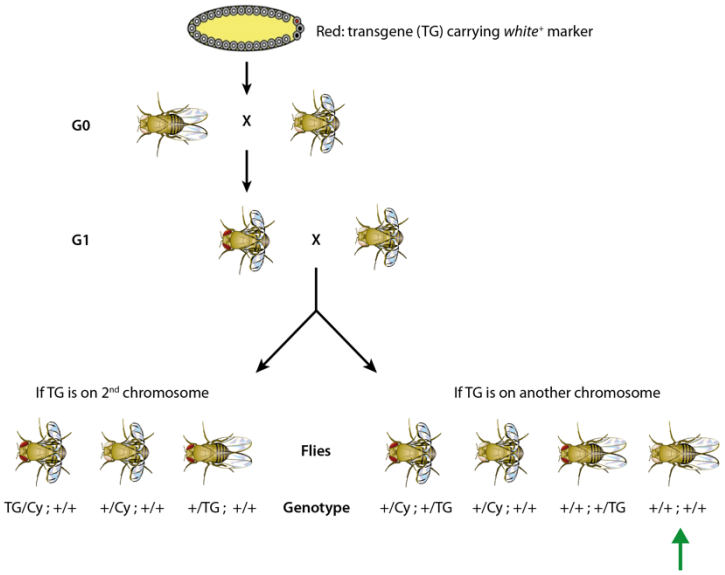


Figure 2-10 Crossing scheme to determine the insertion site of transgenes in *Drosophila*. After injecting the transgene into embryos, these embryos develop into adult flies (G0). G0 flies are crossed with flies containing a balancer chromosome, here, one for the 2nd chromosome, carrying the dominant marker *Cy* that confers bend up wings. In the G1 generation, flies with red eyes will appear, indicative of transgenic animals. To determine whether the transgene is on the 2nd chromosome, G1 flies are crossed again with flies containing the balancer. There are two different possibilities for the phenotypes of flies in the G2 generation (left or right), with only one phenotype that is different between both cases (green arrow). This is the one determining the location of the transgene. In both cases, both chromosomes (second and third) are indicated.

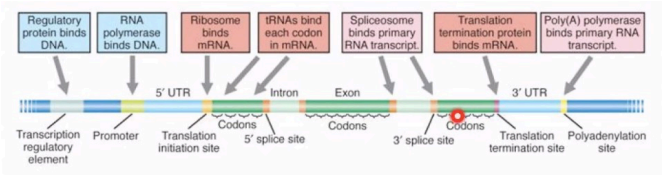
  - Thus, **if white-eyed, non-curly flies (Transgene has dominant marker red eyes, so there is no TG)** occur in the progeny, we know that the TG is not on the 2nd chromosome, because in all other flies, either the TG or the balancer are present.
  - If the Cy marker and the transgene segregate independently, they are on different chromosomes

* Why don't we start with a balancer for the X chromosome?
Transgenes that are inserted on the X chromosome can, in many cases, be identified straightforward in a cross with any balancer or even with no balancer at all. If the G0 fly was a male and the transgene was inserted on the X, the progeny of any cross with such a male would result in only female transgenic animals (having red eyes). All males would not contain the transgene (and would have white eyes), because they inherited their X chromosome from the mother (who was not transgenic). If the G0 fly was a female, we can't say from the results in the G1 whether it the transgene is X-linked.

VORLESUNG: TOOLS & TECHNOLOGIES "TUTORIAL 1" – DANIEL BOPP

1) Sanger sequencing – alles schon gewusst
2) SBS (dominant second generation)
3) Nanopore (mobile application!) and PacBio → combine with Illumina for higher accuracy

Aufbau of a gene:



gene function studies:

- gene prediction algorithms
- homology finding algorithms (BLAST)
- EST database (partially sequenced cDNAs)
- RNAseq data

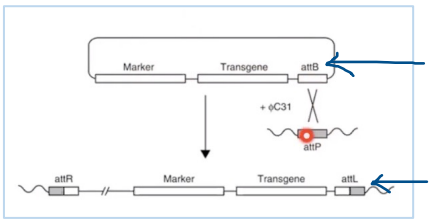
Metagenomics:

is the study of genetic material recovered directly from environmental samples. Traditional microbiology and microbial genome sequencing rely upon cultivated clonal cultures. This relatively new

wichtige Stichworte:

- optical mapping of single DNA: BioNano Irys System
- Bioinformatics → gene annotation and structure → predicted proteins structure and function
- bioballastic delivery = Gene gun → bombardment with DNA coated gold or tungsten beads

Site-directed transegenesis: man kann selbst entscheiden wo das gewünschte Gen (inkl. Marker) eingefügt wird, mithilfe einer attP sequence (Crispr Cas9 alternative)

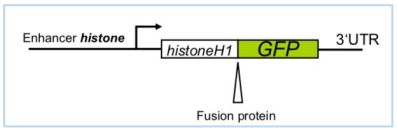


Gen konstrukt zum einfügen

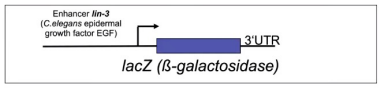
Virengenom

- introduce this construct that contains gene of interest = Transgene, with a marker
- recombine construct at one specific site → attB and attP = viral recombination system with use of an integrase that can only catalyse integration but no excision
- → genaue Insertion des Konstrukts an der attP site = attP sequence serves as a fixed docking site in the genome

Translational reporter transgene: Tag a protein of interest, to look at it's intracellular distribution



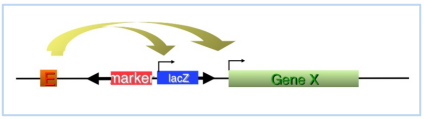
Protein of interst is fused to GFP (in vitro) and reintroduce (promotor stays the same)



to find out at what time of development this gene is expressed → take and modify promotor sequence to drive any type of marker (mostly lacZ) → hier keine Fusion mit Marker!

Enhancer Trap Methods:

- use fact that P-Elements can randomly insert in the genome → p-Element has preference for 5' regions of a gene (not jumping into genes = no interruption of gene function)
- if they jump at 5', they also become under control of the nearby promotor elements = Enhancers, that normally drive the expression of the Gene, which P-Element flanks
- Enhancer will therefore not only drive expression of gene X, but also of reporter (Marker or lacZ (Promotor of LacZ will respond to any enhancers nearby))

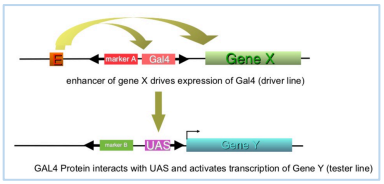


- you produce many enhancer trap lines and pick out the ones with interesting patterns (novel expression patterns for example)

for this example you have to introduce 2 constructs: Gal4 and Gene of interest that should be expressed in this specific tissue with a UAS sequence in this case you use interesting enhancers in a genome to drive the transcriptional activator Gal4 (specificly interacts with upstream UAS element)

- below = responder construct to Gal4 that contains the UAS, that has neben Gal4 also have to be introduced into the drosophila (Gal4 kommt nur in yeast normally vor) → Gene Y (artificial) wird nach dem selben expression pattern wie Gene X (schon im Genom) exprimiert

- → can help you to have gain of function → misexpress a particular gene or express it in other tissues and look what consequences are



other application: misexpression can be very informative, what the function of a particular gene might be
→ example: "ectopic expression of hedgehog"
→ what happens when gene is expressed at another location

Interference by RNA: drive misexpression of hairpins to misexpress a gene with hairpin transgenes in Dropsophila (also a approach to silence your gene of interest)

- very powerful method: with this 2-component system, you have very strong control spatial and temporal expression of gene or here hairpins to silence genes
- available: one UAS hairpin transgene specific for each gene
- one single hairpin construct is not enough to look at silenced phenotype bcs → hairpin is not necessarly specific to your gene of interest, bcs of off target effects
- to control: use 2 hairpins that target your gene to validate your phenotype

"TUTORIAL 2" – DANIEL BOPP

Comparison of pronuclear injection and homologous recombination

	TRANSGENIC	ES DERIVED
Genetic action	dominant	Recessive or dominant
Insertion site	random	Targeted to endogenous
Copy number	Variable, 1 to >100	Heterozygotes 1, homozygotes 2
expression	Usually integration site dependent	Usually same as endogenous
time	6-9 months	1.5 -2 years
cost	>$3000	>$12,000
other	Insertional mutation	Effects on neighboring genes

Transgenesis vs. homologue recombination in totipotent stem cells

1)
- transgenesis is random!
- variable: often more than one insertion in the genome!
- random insertion of any modified sequence → vary about the fact that insertion is random, gene expression can be altered (influencing nearby promotor sequences

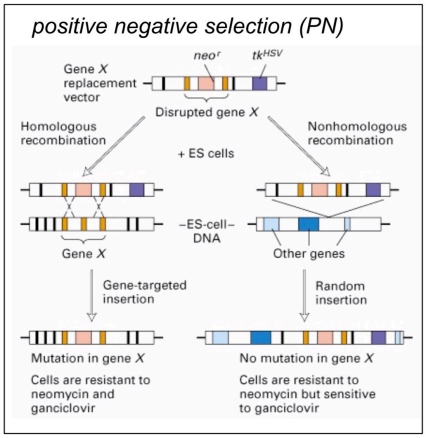
2) much higher cost!
- with positive or neg. selection you can target specific gene
- heterozygote: normally only one allele is replaced
- replacement of sequences: same promotor is acting on newly introduced sequences → less problem of position effects

---

Gene targeting in the mouse by homologous recombination: approach to replace sequences instead of integrate/add them later! f.e Replace a mutated gene with a healthy WT gene

- Embryonic stem cell approach
  - homologous recombination event needed (but very rare) → test many individuals to see this recombination
  - method developed to increase homolog recombination by make it in culture and not in vivo
    - need cells that are totipotent!
    - look for recombination events in culture in vitro and not look at animals
- positive negative selection: how to do selection for those rare events such as homolog recombination (how to find the cells in culture that have undergo it)
  - look for integration of antibiotics resistance gene (Marker) by growing them on antibiotics → positive selection
  - but also rare nonhomologous recombination can happen! (does not recombine at location where the homologous sequences are) → you want to negatively select for those events bcs you want replacement and not integration of sequence elsewhere
    - add a negative selecting gene like thymodin kinase (tkHSV) and negatively select for it (cells are resitant to neomycin but sensitive to ganciclovir, bcs they produce toxic component)

- in summary: look for those totipotent embryonic stem cells that grow on antibiotics and did not die when put on ganciclovir and reintroduce them in vivo in the mouse! → give rise to mosaic animal = newborn is chimeric (host genotype and modified genotype)



- look for those mice, where the stem cells have contributed to germ lines, bcs offspring is then also chimeric

---

Interpreting knock-outs: what are limitation of inserting gene with transgenesis or replace gene by homologus recombination

Problem 1: "Null" mutations → no phenotype! (maybe redundant gene function???)

- solution: use sensitized genetic background (needs hypothesis about gene function)

Problem 2: early embryonic lethality

- solutions: mosaic analysis → conditional knock out with cre-loxP in the mouse, and with FRT system in Drosophila

---

why using genetic mosaics? *to test autonomy (Gene expressed in cell where gene product is used) versus nonautonomy (gene product is used in other cells where gene is not expressed)*

→ *is the function of gene required in the cells where it is expressed or in neighboring cells? (which tissues need to carry the WT gene function?*

use recombination tools (somatic):

Cre / loxP system (*phage P1*) "lambda integrase" family
- integrates bacteriophage genome into host genome
- very efficient, when high recombination rate required
- preferentially used in mammalian systems

FRT / FLP system (*S. cerevisiae*) "lambda integrase" family
- maintains copy number of 2µm plasmid
- less effective, when low recombination rate required
- dominant tool in Drosophila

- asymmetric spacers give direction
- catalyses integration and excision

in both cases you need short sequences composed of 2 inverted repeats (recognized by Cre-Recombinase or Flipase) and a spacer in between

Problem in Genome Engineering including homologous recombination is the very low absolute frequency of recombination!

→ ***Solution = induced double stranded breaks substantially increases frequency of local recombination → guide your sequence-specific endonuclease (ZFN to specific locations (cleavage requires dimerization)***

Zinc-Finger nucleases:

two events can happen:

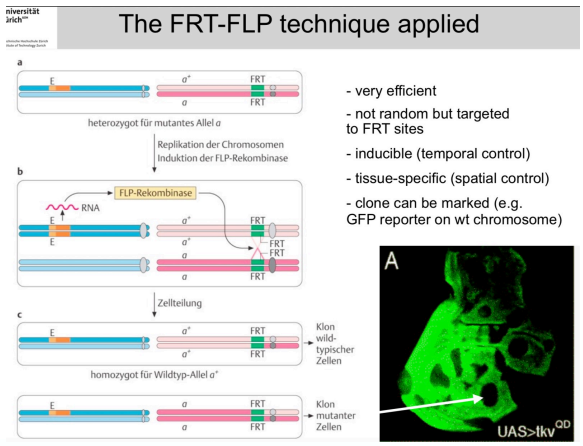
- **nonhomologous end joining event:** doubles strand break gets repaired by cellular machinery → but repair system has no specific template (repair sequence differs from WT sequence = disrupt gene function) → **use for creating disrupted genes**
- **Homology driven Recombination:** DNA Template that is injected → repair System has no template and use it for homologous recombination = editing/replacement event is taking place
- **ZFN used when/for:** model organisms with no effective gene-targeting/disrupting procedure; for non model organisms, **alternative for ES based homologous recombination**, genome therapy
- -) target capacity is very low (limited in which sites can be targeted; neighboring fingers can alter specifity; off target effects, expensive and time consuming assembly

Crispr/Cas System

- CRISPR: Clustered, regularly interspaced, short palindromic repeats = adaptive immune system in bacteria
- 3 types of CRISPR known (type II for genome engineering)
- 3 component System:
  - Cas9: provides enzymatic activity
  - Transactivating CRISPR-RNA: tracrRNA → triggers Cas9 activity
  - CRISPR-RNA: crRNA, short guide that targets the genomic sequence
  - tracrRNA and crRNA can be combined into one precursor

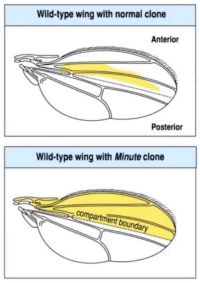
FRT-FLP System

to introduce recombination in somatic cells:



use FLP-FRT System to….

to generate large region of mutant tissue:

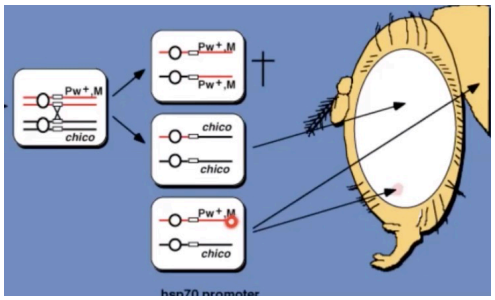


Minute technique expands clonal area → *expansion of mutant clones (mutiertes Gewebe vergrössern)*

to introduce minute mutations on the chromosome that carries the WT allele: → induced minute mutation gives disadvantage to chromosome with WT allele and advantage to chromosome having the mutated allele → large mutant clone when introduce minute mutations! (to increase the tissue that is mutant to look for a clear phenotypes

to genearte tissue specific clones:

- FLP Recombinase that is driven by Enhancer, which is specificly acting in the progenitor cells in the eye
  - only these cells can undergo recombination
  - you also start here with heterozygous condition
    - 2 mal chico (white cells since no marker) or 2 mal marker (marker homozygous lethal → die) = in both cases recombination was successful
    - heterozygote für chico und marker = marker dominant, also rote Zellen → homologous recombination was not successful

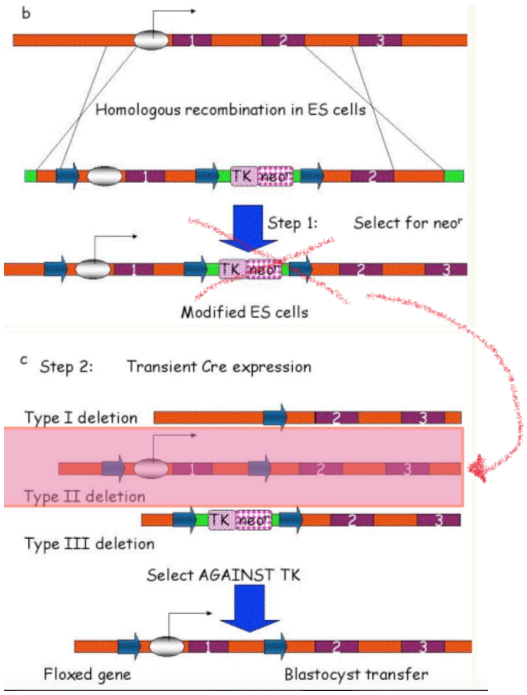


to generate mutant germline clones

- mutate maternal activity of a gene in the zygote
- also start point = heterozygosity

Cre-LoxP mediated recombination

- "floxing a gene"
- cre, a recombinase catalyses recombination btw. loxP sites (introduce those sequences into genome)
- loxP contains two 13bp inverted repeats flanking a core of 8bp
- asymmetry of the core gives directionality
- select for type II deletion and transfer those ES cells into blastocysts
  - if loxP sites are in same direction = fragment in between/the gene of interest will be deleted (= typ II)
  - loxP just prepares for deletion, but Cre must be induced for deletion event



use Cre-loxP for…

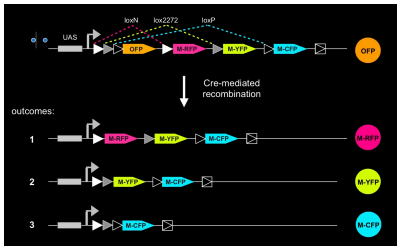
conditional gene targeting (by gene disruption)

- spatial control = tissue-specific cre-expression
- temporal control = inducible cre-expression (hormone dependent)

multicolor labeling in vivo

- no disruption of gene anymore!
- proteins labeled to look for connectivity during development
- transfer vector with different fluorescent proteins

- different loxP sites for recombination (3 fluorescent proteins = 3 different pairs of loxP sites that flank protein) → 3 different outputs after recombination upon Cre induction
- it's just a labeling method! with multicolor staining you can follow connections of one particular cell bcs the cells have different colors! → per zufall erhält eine Zele dadurch eine der 3 Farben und kann von den anderen unterschieden warden = random recombination (Cre ist nicht LoxP Paar spezifisch, sondern homologe Rekombination findet bei Cre Induction einfach zwischen irgeneinen der 3 LoxP paare statt



-

VORLESUNG 3: APPLICATIONS IN MEDICINE AND AGRICULTURE

applications in pest control:

1) sterile insect techniques

- agricultural pest: "Medfly" → fly lies eggs in citrus fruits
- disease vector: anopheles stephensi → Mosquito causing Malaria
- Aedes albopticus = Tiger mosquito

→ what can we use against pest insects?

pesticies and isecticides

- **Problem**: nonspecific → pertuberance of ecosystem (Bienengefährdung)
- strong selection for resistance

***why genetic control** strategie s are preferable to insecticides:*
→ *genetically modified insects*

- GM individuals can self-disperse and are highly effective in seeking conspecific WT mates (+)
- species specific (+)
- vertical transmission of modifying agent minimizes off target effects

2 outcomes:

i) population suppression (ecological balance impact)

→ *S(terile) I(nsect) T(echnology) = SIT*

- Males are genetically modified (gamma irradiation sterilization

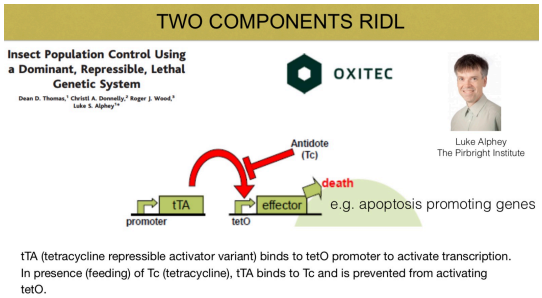
→ *Derivative of SIT:*

Release of Insects carrying a Dominant Lethal (RIDL)
get a dominant lethal (RIDL) = dominant, repressible lethal genetic system)

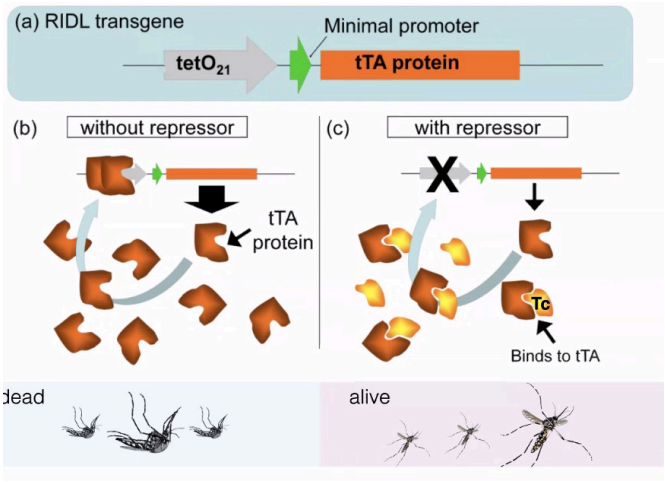
- RIDL only becomes active when released in nature! → all of progeny = dead
- only modified males are released bcs female would sting humans!
- → **males have to be separated from females (should die) before release!** ***
  - when female WT mated with sterilized or modified (RIDL) male, the female offspring won't survive (*** die Männchen überleben um die genetische Version weiterzugeben) = complete elimination in the end

→ Formen von RIDL:

- bi-sexual RIDL (males have to be separated from females before release) → but very effective! → 80-90% suppression
  - GM (geneticall modified) mosquitos release in Florida and Brazil
  - *how get rid of females before release? ****
    - ***sex specific promotors*** *(yolk promoter only active in female fatbody)*
    - ***transformer gene**: conserved sex-specific switch (sex specific expression based on alternative splicing) → active state of gene only occurring in females, inhibiting their development*
    - *both sexes express transcript, but in females it is alternatively transcriped =* ***female killing RIDL = female specific RIDL***
- two or one component RIDL
  - Two component system: → female specific insect lethality engineered using alternative splicing (z.B female-killing RIDL is a 2 component system)



- one component system: tTA protein in large level is toxic (case wenn no repressor vorhanden)
  - with repressor, tetracyclin leds animals survive (suicidal transgene product tTA is inhibited)



2) gene drive systems

Gene drive Systems targeting sex determining gene:

Gene drive refers to ability of a gene to be inherited more frequently than Mendalian genetics would dictate, thus increasing in frequency, perhaps even to fixation

allele transmission >> 50%

Driving system are superior to conventional SIT methods which rely on continuous productions and releases of sterile males

- endonuclease based gene drive system targeting the sex gene (with CRISPR)
- nicht mehr davon abhängig, nur noch männchen zu releasen
- Masculinization by gene drive possible → transgene that promotes maleness or confers female lethality
- transmission by males more effective than by females

left: in transgenic males Cas9 only expressed in germ cells

right: in transgenic females Cas9 is expressed in germ cells but also **maternally** inherited through cytoplasm to zygote

ii) population replacement (no elimination, but replacement by genetically modified version and niche cannot be taken by another pest)

applications In human health

1) creating disease models

Importance of knockout mice in biology & medicine :

Knocking out the activity of a gene provides valuable information about the function(s) of that gene. Since many genes are common between humans and mice, knockout mice gives researchers information about the function of that gene in humans.

Knockout mice have been useful in studying and modeling different kinds of cancer, obesity, heart disease, diabetes, arthritis, anxiety, aging and Parkinson disease. Knockout mice also offer a biological context in which drugs and other therapies can be developed and tested.

current issues in mouse genome engineering:

- *replacing mouse genes with human genes = humanising*
- *generation of alleles with subtle modifications (point mutations) or large chromosomal abberations to create models for human pathologies*
- *towards a phenome (phenotype of every gene in the mouse) KOMP (Knock-Out Mouse Project) EUCOMM (European Conditional Mouse Mutagenesis)*

Bsp: multiple lentiviral expression (MuLE) system

- *cancer model*
- *A lentiviral system is very effective at delivering genetic material to whole model organisms and almost all mammalian cells, including non-dividing, inactive or growing, and difficult-to-transfect cells including neuron, primary and stem cells.*
- ***The efficiency of lentiviral transduction is close to 100%, making it ideal as an expression vector system***

engineered tumors containing up to 5 different genetic alterations:
- knock down of 2 tumor suppressor genes (Rb and Pten)
- activation of oncogenic gene Ras
- Cre recombinase for "floxing" (tamoxifen dependent)
- pyromycin resistance

2) therapeutical applications

- *repertoire of targeted manipulations:*
  - *inducing lesions: NHEJ*
  - *replacing sequences: HDR*
  - *transcriptional activation/ repression*
  - *DNA modifications*

Bsp: using CRISPR to replace areas

A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice

- *Intravenous infusion of two AAVs, one expressing Cas9 and the other expressing a guide RNA and the donor DNA, into newborn mice with a partial deficiency in the*

urea cycle disorder enzyme, ornithine transcarbamylase (OTC).

- *This resulted in reversion of the mutation*

In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy

In dogs, CRISPR fixes a muscular dystrophy

Delivery: adeno-associated virus (AAV) is used as a vector for delivery and expression of the components of the CRISPR- Cas9 system to skeletal and cardiac muscle

Bsp: Gene editing in human iPSCs

- *how to get Crispr in the right affected cells → make use of pluripotent stem cells! (differentiated cells are juvenized back to pluripotent stem cells and perform gene editing (so no editing in vivo but rather in vitro)*
- *→ modify certain genes and sites in genome → modified versions are in vitro differentiated and reintroduced to patient!*
- *for candidate gene validation*
  - *z.B neurological disorder → harvest cell from patient → introduce pluripotente (redifferentiate) → test what happends when you modify your candidate gene*

REVERSE GENETICS IN INVERTEBRATES (OL)

The main goal of transgenesis is germ-line transmission of the inserted sequences → ***the altered gene must recombine with the cell's genome.***

In bacteria and yeast, this occurs with high frequency using the cell's own recombination machinery.

In more complex organisms, the procedure of genome integration is more complicated due to a lack of efficient recombination mechanisms.

Integration of the transgenic DNA: Transposons

Transposons are transposable Elements that can relocate from one genomic location to another.

- *believe that TEs make up 44% of the human genome*
- *TEs might carry out some regulatory biological function*
- Each group of transposable elements contains autonomous and non-autonomous elements. Autonomous elements encode the products required for transposition. Non-autonomous elements that are able to transpose have no significant coding capacity but retain the sequences necessary for transposition.

There are two broad classes of transposable element:

1) **DNA transposons:** Most DNA transposons move as pieces of DNA, cutting and pasting themselves into new genomic locations. The transposition of DNA transposons is catalyzed by an enzyme called transposase

2) **retrotransposons**: relocate by a copy-and-paste mechanism: first, they are transcribed into RNA, which is reverse-transcribed to DNA. This copied DNA is then re-inserted in the genome at a different position. Thus, the original transposon is maintained at its original position

Retrotransposons are further divided into two groups:

i) long terminal repeat (LTR) retroelements: contain two long terminal repeats (LTRs; black arrows) and encode the genes for the enzymes r everse transcriptase and integrase, both crucial for retrotransposition.

ii) non-LTR retroelements: lack LTRs and encode genes for reverse transcriptase and an endonuclease that cuts the target DNA.

In most organisms, a considerable number of "active" transposons exist. The host organism must have evolved mechanisms to regulate transposon transcription and transposition. How do you think organisms live with transposable elements?

Suppression of transposon transcription: One mechanisms that suppress transcription are chromatin modifications, including modifications of histone tails, DNA methylation and alterations in chromatin packing and condensation. For example, DNA methylation on cytosine residues (CpG islands) can function to repress transcription (remember for example gene silencing in genetic imprinting).

Transposon-mediated transgenesis in Drosophila

In Drosophila, transgenesis mainly relies on the P-element transposon = P-element mediated transgenesis

disadvantages:

the size of the DNA that can be integrated is limited and the location of integration cannot be controlled.

- P elements are transposable elements, which were originally ident ified within the fly's own genome

- researchers tried to mate females from laboratory strains with males derived from wild populations → The progeny showed **sterility, a high mutation rate and a high frequency of chromosomal aberrations** = These flies are **dysgenic**, and the phenomenon is therefore called hybrid dysgenesis.
- Investigators hypothesized that **the dysgenic mutants are caused by the insertion of transposable elements into genes, thereby rendering them inactive and causing the mutation**. => This transposable element was the P element.
- **Interestingly, P elements were inserted frequently in wild-type strains of Drosophila, but not in laboratory strains.**
- *Why do the P elements present in wild-type strains not cause trouble, e.g., are inserted into genes and cause mutations?*
  - *It is now thought that the transposase genes in P elements are silenced in wild-type strains, thus, P elements cannot relocate in wild-type flies.*

P-Element = a TE

Drosophila genetics is dominated by the use of P elements. P elements served to develop different transgenic methods (e.g., enhancer trap or misexpression libraries) or technologies (e.g., generation of Flp/FRT insertions and RNAi) and they are used for gene disruption to create mutants (insertional mutagenesis and imprecise excisions)

- P elements, like other transposons, contain two terminal repeats, including inverted-repeat sequences of 31 base pairs and other internally located sequence motifs absolutely required for their mobilization or transposition.
- Mobile (also autonomous) P-element transposons **encode an enzyme called P transposase that catalyzes transposition of the transposon**.
- **To use P elements for transgenesis, the P transposase was separated from the P element transposon backbone** (called non-autonomous P element,
  - A plasmid that encodes P transposase, a so-called helper plasmid, is provided together with the transgene that contains the transposon backbone necessary for integration, the sequence of interest and a marker.
  - In general, transposons are injected into fly strains that are not possessing the same transposon, **avoiding unwanted mobilization events of transposons present in the genome**, thereby ensuring the stable integration and maintenance of the injected transgene.

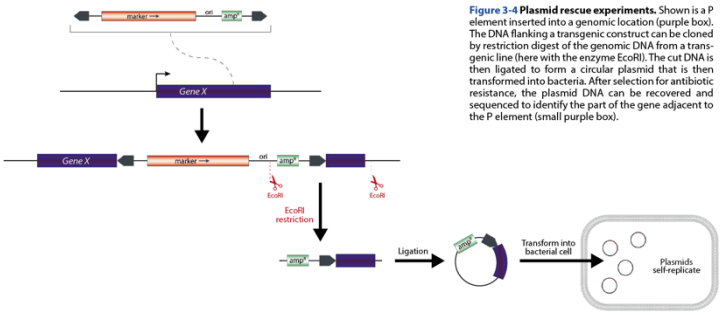


Figure 3-4 Plasmid rescue experiments. Shown is a P element inserted into a genomic location (purple box). The DNA flanking a transgenic construct can be cloned by restriction digest of the genomic DNA from a transgenic line (here with the enzyme EcoRI). The cut DNA is then ligated to form a circular plasmid that is then transformed into bacteria. After selection for antibiotic resistance, the plasmid DNA can be recovered and sequenced to identify the part of the gene adjacent to the P element (small purple box).

- Different transposons have unique insertion-site characteristics.
- Integration events of P elements are strongly biased towards the 5′ end of genes.
- These include the PiggyBac transposon identified in the cabbage looper moth Trichoplusia ni., which can be used in a variety of hosts, ranging from insect to mammals.

Application of transposon mediated transgeneis

two main groups of experiments:

1. gene-disruption methods: occurs when a transposon insertion interferes with the function of a gene
2. transgenic technologies: usually involve introducing the different components of novel techniques or to insert exogenous genes into the genome

→ Flp/FRT system to create mutant clones by inducing mitotic recombination would not have been possible without use of P-elements.

→ Special P-Elements (EP Elements) also allow overexpression of genes

Enhancer Traps

One transgenic technology that relies on P element transposition is the so-called enhancer-trap assay.

→ ***An enhancer trap is a transgenic construct that is used to identify genes that are expressed in a specific tissue.***

- *The enhancer trap construct contains a transposable element (that allows random integration into the genome) and a reporter gene (that allows visualization of the spatial regulation by the enhancer).*

- *For selection, the construct usually includes a genetic marker, e.g., w+, producing red-colored eyes in Drosophila.*
- *The most common enhancer trap constructs are P[lacW] that carries a lacZ gene fused to the the 5` end of the P element (see figure).*
- ***As P elements frequently insert near the promoter of a gene (near 5' end of genes = upstream), lacZ will be expressed in the same spatial and temporal pattern as the gene into which the element is inserted.***

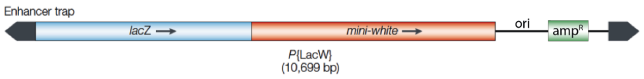


Figure 3-3 The enhancer trap construct P[lacW]. This construct consists of a P element that contains the *lacZ* gene and a *white+* gene (here called *mini white*). Further, it contains elements that are necessary for propagation in bacteria (ori and ampicillin resistance). The black arrows represent the P element ends containing the inverted repeats.

→ ***When the transgenic construct (enhancer trap) inserts near a tissue-specific enhancer, the weak promoter present on the construct comes under the control of this enhancer, resulting in tissue-specific expression of the reporter gene.***

→ *enhancer trap allows hijacking of an enhancer from another gene, and so, identification of which enhancers regulate the expression of which gene*

→ *By mobilizing (aktivieren von Transposition) the transposon to other random chromosomal sites,* ***one can detect diverse enhancers at those sites and study their tissue-specific gene control or identify flanking genes*** *displaying their expression patterns.*

→ *Enhancer-trap transposons have been engineered to permit cloning of genomic DNA adjacent to the site of any particular insertion, enabling the isolation of the enhancer and its target gene from genomic DNA.*

→ *P[lacW] also carries sequences that are necessary for propagation as circular plasmids in bacteria, such as the ori (origin of replication) and the ampicillin-resistance gene*

- these elements are useful for a method called "plasmid rescue" that serves to clone DNA that flanks a transgenic construct.
- This is useful in all cases where the insertion site of the P element is not known, e.g., to identify the original insertion site of a P element or after a P element has been mobilized to integrate into another unknown location in the genome.
- Genomic DNA isolated from flies containing a P[lacW] insertion can be cut with restriction enzymes, the cut sequence is ligated, and the circular DNA is transformed into bacteria.
- After selection for ampicillin resistance, the plasmid DNA is recovered and sequenced to identify the gene. → **Like this, the gene that is associated can easily be identified**

Gene disruption methods using P-Elements (a Transposon)

- P elements inserted into genes disrupt genes at random locations, creating mutants with different phenotypes.
- **Several gene-disruption projects** have generated stocks that each contain a single P element construct **inserted at a known position of the genome.**
- **These stocks are available to all researchers and one can simply order the stocks that carry an insertion in the gene of interest.**

how the stock collections they generated can be used for reverse-genetic approaches:

Gene misexpression by P elements

- We have already discussed the use of special P elements, **so called EP elements, that carry UAS sequences** that allow **activation by the transcription factor Gal4** (see lesson 2 in "Forward genetics")
- If a P element is inserted **near the 5' end of a gene**, the **expression of this gene can be induced by Gal4.** → There exist a few thousand P(EP) insertion lines that can be used for overexpression screens.

Gene disruption

- P-Elements are inserted such that they suppress the expression of a gene by disrupting its sequence. → **transposons can act as a mutagen to cause insertional mutations in proximally located genes**
- P elements have been used in different projects with the aim to introduce a disrupting insertion into every gene in Drosophila = **insertion mutants**
  - They are used **for gene identification in complementation analysis** for mutants identified in forward-genetic screens → they serve to validate a mutation discovered in forward-genetic screens by comparing the phenotypes of the insertional mutant with the one obtained in the forward screen.
- However, for many genes, insertion mutations are still **lacking**.

- If there is a P element **near the gene** of interest it can be remobilized providing the transposase enzyme.
- The P element **will often transpose within a few kilobases of the original location, hopefully affecting your gene of interest.**
-

P-element excision (the fly knockout)

Figure 3-6 P element excision mutagenesis. A P element (orange) is inserted into a genomic location near a target gene (green). Transposase catalyzes the excision of the P element, which leaves a double strand break (DSB). Repair of these DSB can lead to either a precise excision of the P element (above) or to a deletion of sequences adjacent to the DSB, leading to a small deletions of the gene.

- P elements can also be used to create gene-specific deletions (gene knockouts) in the fly.
- if a P element is **inserted in close vicinity** of a gene of interest, a null allele can be created fairly easy using the "imprecise-excision" method in which the **excision of the P element is induced by temporarily providing transposase activity** (crossing into a transposase-expressing line).
- Because transposase is only present temporarily in these organisms, its activity is mainly restricted to mediating the excision of existing P elements → excised **P elements usually do not again integrate at another locus.** (they are simply excised)
- **The method actually relies on the fact that excision of the P element results in double-strand breaks (DSB) at the insertion site, which have to be repaired by the cellular repair machinery.**
- DNA repair mechanisms can process the DSB in different ways, generating different products:
  - i) the ends of the DSB **are used to induce new DNA synthesis,** using a homologous sequence as a template (called homology-directed repair).
    - can lead to **either a restoration of the P element at the insertion site** (if the P element on the sister chromatid served as a template for synthesis) or **the restoration of the original chromosome without the P element** (if the homologous chromosome without the P element served as a template for synthesis).
    - The latter is also called a "precise excision", because the P element is excised without damaging the chromosomal region where it was inserted
  - ii) before repair, **the DSB can be enlarged to a gap by the cellular repair machinery**, and repairing the gap can lead to deletions that can extend in one or both directions of the original insertion site = imprecise excision experiment
    - Those events are relatively frequent such that imprecise-excision experiments will usually give a **few different deletions** for the gene of interest

Do you know another technology that has not been discussed in this course yet, with which one can study the function of a gene where no deletion is available?

To study the function of genes where no insertional mutant can be obtained, other tools can be used that allow a first, straightforward analysis of a gene function, for example, libraries of transgenic insertions that mediate gene knockdown by RNA interference, which we will discuss later in this lesson.

How could you phenotypically detect whether a P element has been excised and is therefore absent from the genome or not?

If the P element carries a dominant marker (e.g., white+ that confers red eyes), the flies can be screened for the presence of red eyes. Red-eyed flies still contain the P element and therefore, no deletion is to be expected in these flies.

What experiment could you perform to detect on a molecular level whether a P element has been excised or not, and what kind of deletion has been induced by the excision of the P element? In case your first thought is to sequence the whole genome, ask yourself why this might not be the simplest approach

Since you know the original position where the P element was inserted, you can design primers that flank the P element: the forward primer would bind in the genomic region that is 5' from the P element insertion and the reverse primer would bind 3' of the P element insertion, in the gene. Then, a PCR is performed on DNA extracted from flies where the transposon has been mobilized. If the P element has not been excised, the PCR product will include the P element sequence and thus be big (let's s say, 3 kb). If the P element has been excised, but the DNA has not been deleted, this will give a fragment that is smaller and includes only the genomic sequence of this region. If a deletion has occurred and this deletion has removed part of the gene, then the PCR product would be even smaller, or- in the extreme case if the deletion has removed the part where the reverse primer would bind- no PCR product would be obtained.

Post transcriptional regulation by RNA interference - RNAi

a different method of gene-expression regulation that does not alter the DNA sequence, but regulates on a post-transcriptional level -> to study a genes function

→ *RNA interference (RNAi) is a* ***naturally occurring*** *cellular process in eukaryotes, in which* ***short regulatory RNA molecules interfere with mRNA translation*** *and thereby affect – usually suppress -* ***protein production*****.**

General mechanism:

I. double-stranded RNA fragments are cleaved into smaller pieces of 20 to 30 base pairs in length by an endonuclease called Dicer
II. these shorter RNA fragments are separated into single strands, of which one strand is incorporated into an enzyme complex known as the RISC complex; and
III. the RNA-enzyme complex recognizes mRNA molecules with a sequence that is complementary to the RNA fragment in the complex and blocks or cleaves it, preventing translation.

where does an RNA fragment come from that provides sequence specificity to the inhibitory enzyme complex?

In principle, there are two answers to this question:

- First, the RNA fragment was transcribed from <u>non-protein-coding regions of the cell's own genome</u> → the genome of all animals and plants <u>encodes RNA molecules that do not code for proteins, but instead have a regulatory role</u>
- Second, RNA can also be <u>introduced</u> into the cell

☞ What makes this mechanism so attractive for biological research is that **is allows sequence-specific inhibition of a gene <u>without altering the genome/DNA Sequence permanently</u>**.

☞ RNAi can be used to infer the function of a protein-coding gene by knocking it down and observing the effect, or, using libraries of RNA molecules, a gene encoding a known function can be identified.

RNAi in model organisms

☞ *The double-stranded RNA has to be processed inside C. elegans or Drosophila cells by* ***dicer*** *to produce siRNAs that can be incorporated into the RNA-induced silencing complex (RISC).*

- In C. elegans, double-stranded RNAs are introduced simply by feeding the worms with bacteria that express these RNAs
- In human and mouse cell lines, siRNA can be generated in vitro (there are even companies that synthesize siRNAs) and is introduced into cultured cells by transfection. <u>These siRNAs can be directly incorporated into the RISC to mediate gene silencing</u>
- In Drosophila cultured cells, dsRNA molecules that can be easily generated by in vitro transcription of plasmids

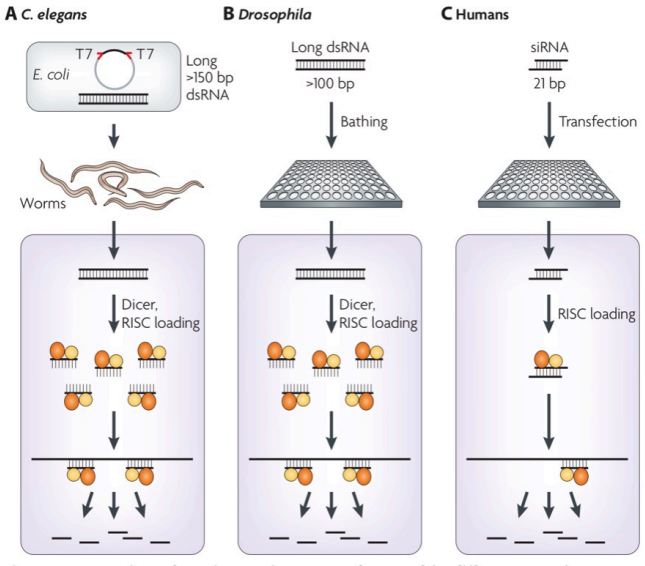


Figure 3-9 The Gal4-UAS system for directed gene expression is used to downregulate genes in *Drosophila* by RNA interference. (A) Flies containing the *GAL4* gene under the control of a tissue-specific promoter are crossed with flies carrying inverted repeat sequences to produce a hairpin RNA. **(B)** Double-stranded DNA (dsRNA) that is homologous to the target gene is delivered by expression of an inverted repeat RNA. The dsRNA is processed into 21-25-nucleotide small-interfering RNAs (siRNA) by the Dicer ribonuclease. The siRNA is used to guide the sequence-specific degradation of mRNA, leading to post-transcriptional silencing of the target locus. (A: adapted from D. St Johnston, *Nat. Rev. Genet.*, 2002; B: adapted from M.D. Adams and J.J. Sekelsky, *Nat. Rev. Genet.*, 2002)

are added to the culture medium

- adding dsRNA to Drosophila cultured cells is easy and rapid (silencing occurs with 72 h after adding the dsRNA) and therefore well suited for large-scale screening.
- However, the goal of most reverse genetics studies is to determine phenotypes that are associated with the loss of gene function in the organism:
- Therefore, **transgenic flies have been generated that express an RNA with a long inverted repeat that can fold back on itself to become double stranded (a so-called hairpin RNA).**

- This hairpin RNA is then processed **into siRNAs by dicer to mediate RNA degradation.**
- By placing the **RNA hairpin construct under the control** of UAS sequences, it is **possible to express the dsRNA in specific tissues or at specific times during development using the Gal4-UAS system**

☞ → *In brief, **the binding of the transcription factor Gal4 to UAS sequences activates the expression of the gene downstream of the UAS sequences**.*

☞ *Since many **different** Gal4 lines exist that drive the expression of Gal4 in different tissues, or those where the Gal4 is under the control of an inducible promoter, **both tissue-specific as well a temporal expression of the dsRNA can be achieved**.*

- A library of transgenic flies expressing different RNA hairpin constructs was constructed, targeting over 88% of all protein-coding genes in the fly.
  - This library is publically available and can be used to systematically silence genes and analyze their functions in any tissue and any stage of Drosophila lifespan.
  - ☞ By crossing flies containing a UAS-RNA hairpin construct targeting a gene of interest with flies expressing Gal4 under a tissue specific promoter, the expression of the gene of interest can be inhibited
  - ☞ the generation of the transgenic fly library carrying the hairpin RNA constructs was only possible due to the existence of P elements.
  - ☞ The hairpin constructs were cloned into plasmids containing inverted repeats from a P element, which mediated the insertion of the hairpin construct into the genome of the flies

What problems can you think of to be associated with gene silencing using RNAi?

The biggest problem are off-target effects, because RNAs are small (20-25 nucleotides) and can bind also other genes than the desired one, leading to the downregulation of a non-related gene. Another problem is that some genes are not effectively downregulated by RNAi. Also, in some tissues, like the brain, RNAi does not work efficiently.

GENE TARGETING IN THE MOUSE (OL)

the mouse is the model closest to humans that is small (share 95% of genes with humans) and cheap enough for efficient handling in the laboratory → model of choice for human physiology and disease for many years

Germline transformation in mouse

two reverse approaches are applied to mice: transgenics and knockouts.

Transgenesis:

Transgenic mice can be designed for the identification and characterization of specific regulatory sequences or the overexpression of a protein of interest, either in the whole animal or in specific tissues.

There are two main strategies for creating transgenic mice: i) ectopic insertion by microinjection:

- a solution of transgenic DNA is injected into the nucleus of fertilized oocytes with a micropipette → oocytes are inserted into the oviduct of a foster mouse, where some of the oocytes will develop into baby mice.
- During embryonic development of the modified fertilized oocytes, the transgene integrates at random locations in the genome, typically as multi-copy arrays (random integration ist leider relative selten!)(**Lösung)
- hopefully the cells where the transgene has inserted are part of the germline → transgene will then be inherited and be in all somatic cells of the progeny, nicht mehr nur in germ line
- (+) advantages:
  - the ectopic insertion technique is less laborious than gene targeting, the second strategy that will be described below.
- (-) disadvantages:
  - the **expression pattern** of the integrated construct may be **abnormal**, because **regulatory elements present on the chromosome next to the integration site** may affect the expression of the transgene.
  - The random integration approach using microinjection is rather inefficient, because only 5-10% of the injected oocytes lead to a transgenic mouse → because **the insertion of the transgene into the genome occurs only rarely!** (**Lösung)

** Transposon mediated integration in mice

researchers have searched for ways to increase the insertion rate

- transposons have been used successfully in plants and invertebrates to insert foreign DNA into cells.
- However, the usage of transposons in vertebrates has been limited → Nearly all transposons compatible to and present within vertebrate genomes are inactive.
- Nevertheless, it is possible to identify dormant transposons and **artificially recreate them as active agents.** → This transposon, called Sleeping Beauty
- The Sleeping Beauty (SB) transposon system is composed of **a transposase and a transposon**.
- The SB transposase inserts a **transposon into a TA dinucleotide base pair in a DNA sequence** (In mammalian genomes, there are 200 million TA sites.)
- The insertion site can be somewhere in the same DNA molecule from where it was excised or in another DNA molecule (or chromosome).
- The transposase can be supplied by another source, in which case the transposon becomes a non-autonomous element → Non-autonomous transposons are most useful as genetic tools, because they cannot independently continue to excise and re-insert after the initial insertion.

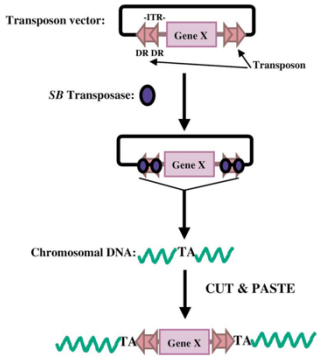


Figure 4-2 Cut-and-paste mechanism of Sleeping Beauty transposons. Sleeping Beauty transposase binds to two direct repeats (called DRs) in each of the inverted terminal repeats (ITRs) of the transposon (shown as arrows), precisely cuts the transposon out of the plasmid, and inserts the transposon into a target DNA, shown as chromosomal DNA. SB transposons integrate into TA dinucleotide base pairs. (adapted from Hackett *et al., Mol. Ther.*, 2003)

ii) targeted insertion: The stem cell method

targeted transgene insertions based on homologous recombination → the transgene contains sequences that are homologous to sequences present in the genome → ***the transgene replaces its genomic homologous counterpart****. This technique can be used to add genes, replace genes, e.g., to replace a genomic locus with a transgene, or to delete genes and produce knockouts.*

1. the transgenic DNA **cannot simply be injected** into eggs, because homologous recombination takes place only rarely

2. Lösung: the **first step of the targeted technique is carried out in cultured mouse embryonic stem cell (ES cells) that can be transformed more rapidly.**
   a. These cells can **be efficiently screened for integration events due to selectable markers present on the transgene**
3. ES cells in which integration was successful are then injected into a blastocyst-stage embryo, which is **implanted into a foster mother**
4. Some of the ES cells may be incorporated into the developing embryo.
5. If this **happens, the mouse developing from this embryo will be chimeric**: it will contain cells that derive from the injected ES cells and those that derive from the embryo that was used for injection.
6. If the chimeric mouse **takes up ES cells into its germ line, some of the resulting offspring will inherit the transgene.**
7. These mice will be heterozygous for the transgene. These can then be mated to produce mice homozygous for the transgene.

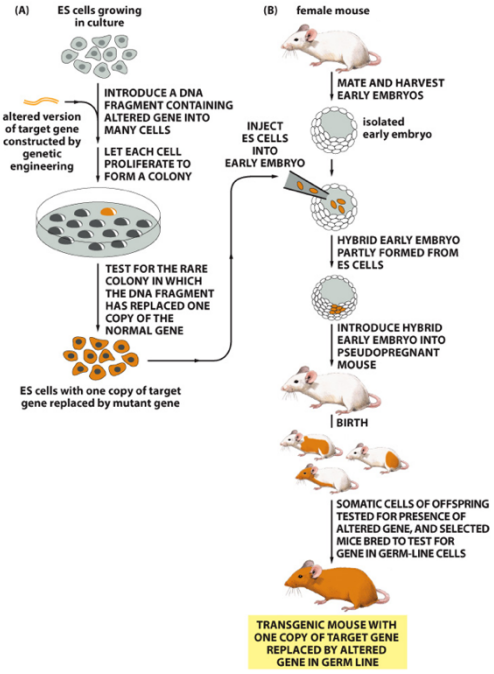


Figure 4-3 Overview of the targeted-insertion strategy to produce transgenic mice. (A) When the altered gene is introduced into ES cells, only a few of them will have their corresponding normal gene be replaced by the transgene via homologous recombination. Cells containing the targeted insertion are selected for using markers present on the transgenic construct (not shown). **(B)** The altered ES cells are injected into early mice embryos (blastocysts) and implanted into surrogate mothers. The mouse produced by such an injected embryo is chimeric and contains cells derived from ES cells and cells from the donor embryo. If ES cells were taken up into the germline of the developing mouse, some of the offspring of such a chimeric mouse will contain one copy of the altered gene. (adapted from figure 8-53, Molecular Biology of the Cell, Alberts, 6th edition, Garland Science)

The generation of knock-out mice

The targeted insertion method *described above can be used to either add, replace, or delete genes of interest.* → *now consider how this method is used to delete genes and create so-called knockout mice*

Positive-negative method to select for targeted integration

☞ the targeted insertion is based on the process of homologous recombination

☞ the **transgene to be inserted needs to be flanked by sequences that are homologous to the sequence to be deleted** → a special vector is created (called targeting vector).

Herstellung v. Targeting Vecto & einschleusen in ES Cells:

1) After the gene of interest is cloned, **the coding sequence of the gene is disrupted by the insertion of the neomycin-resistance gene** (used as marker to indicate vector integration into chromosome) **into the protein-coding region of the gene** by the use of the targeting vector (e.g., into exon 2 in figure).
2) This targeting vector having the cloned gene and the marker gene is **then introduced into mouse embryonic cells (ES cells) by transformation.**
   - **ES cells** derive from the inner mass of an embryo in the blastocyst stage, an early stage of embryonic development → They are undifferentiated pluripotent stem cells that **can develop into any cell type in the body**, including the germ line.

3) If an ES cell takes up the targeting vector, **this vector can integrate into the genome of the cell into three different ways:**
   - First, **homologous recombination** between the DNA sequences present on the targeting vector and the corresponding homologous genomic sequences in the ES cells can take place as desired. → Since the targeting vector includes the **neomycin-resistance** gene between the homologous DNA sequences, it will be inserted into the chromosome
     ☞ Thus, the chromosome will then carry the targeted insertion of the neomycin resistance, disrupting the coding sequence of the gene; in most cases, this leads to a deletion of the gene function, and thus, a knockout.
   - In the second scenario, the **targeting vector might insert randomly into the DNA**, just as it happens in the ectopic insertion described above. **(passiert nur sehr selten, darum v.a uf recombination setzen)**
     ☞ This will result in an integration anywhere in the genome, **but will not lead to a disruption of the gene of interest.**
   - In the third scenario (this is the most frequent one) there will be no insertion at all, and all chromosomes will be unchanged.

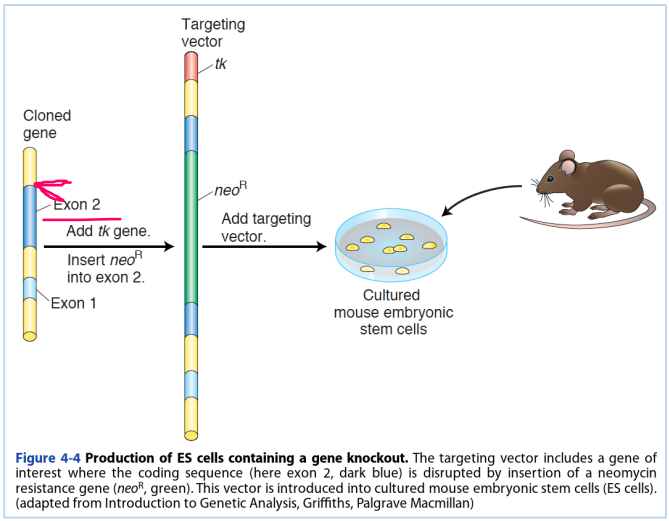


Figure 4-4 Production of ES cells containing a gene knockout. The targeting vector includes a gene of interest where the coding sequence (here exon 2, dark blue) is disrupted by insertion of a neomycin resistance gene (*neo*^{R}, green). This vector is introduced into cultured mouse embryonic stem cells (ES cells). (adapted from Introduction to Genetic Analysis, Griffiths, Palgrave Macmillan)

How can these three events be distinguished to select only those cells that contain the targeted insertion of the targeting vectors?

→ two selection markers come into play:

- The first one, the neomycin-resistance gene, is used to select for cells that have integrated the targeting vector, because these cells will survive in medium containing the antibiotic neomycin
- The targeting vector carries a second marker, the thymidine-kinase gene (tk, shown in red in figure)
  - This gene is used in a negative selection assay:
  - If an antiviral agent, in this case the drug ganciclovir, is added to the cells, all cells containing the tk gene die.
  - If ganciclovir is phosphorylated by the herpes simplex thymidine kinase, the phosphorylated form of ganciclovir causes premature DNA chain termination and apoptosis.
  - Importantly, the tk gene is present on the targeting vector outside of the homologous regions!
  - i) Thus, when homologous recombination takes place, the tk gene does not take place in this recombination and is lost !!!→ wenn homologe Rekombination also erfolgreich ist, dann kann dies

mit 2 Experimenten überprüft warden → die Zellen müssen auf Antibiotika Platten überleben, andererseits müssen sie bei Zugabe von ganciclovir ebenso überleben, da das tk gene bei erfolgreicher Rekombination verloren geht!

- ii) In contrast, when the targeting **vector integrates randomly**, any part of the vector can be integrated, including the tk-HSV gene (middle panel in figure) → Consequently, the combination of a positive and a negative selection identifies the desired cells → they have no recombination, also bleibt tk Gen drin, also stribt die Zelle bei Ganciclovir Zugabe

☞ i) cells that **grow on neomycin and ganciclovir must have had a homologues recombination event.**

☞ ii)Cells that are neomycin-resistant but ganciclovir-sensitive (so die because having the tk Gen) most likely have had a **random recombination event**

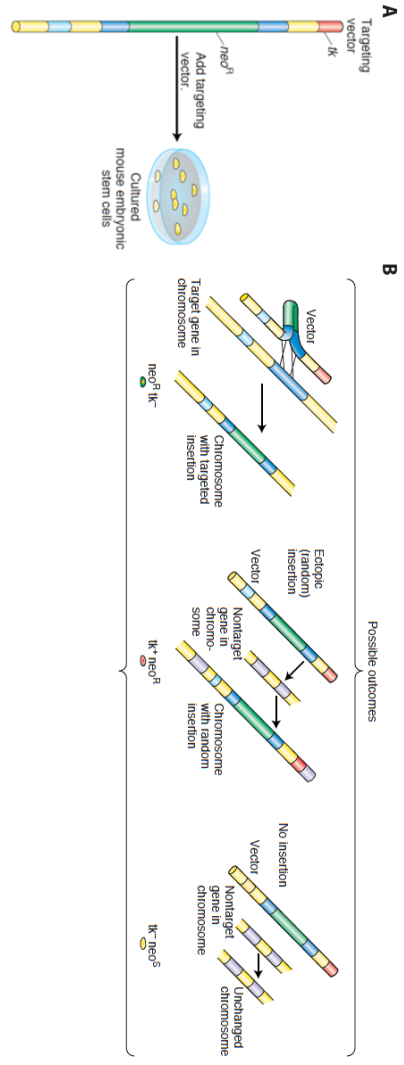
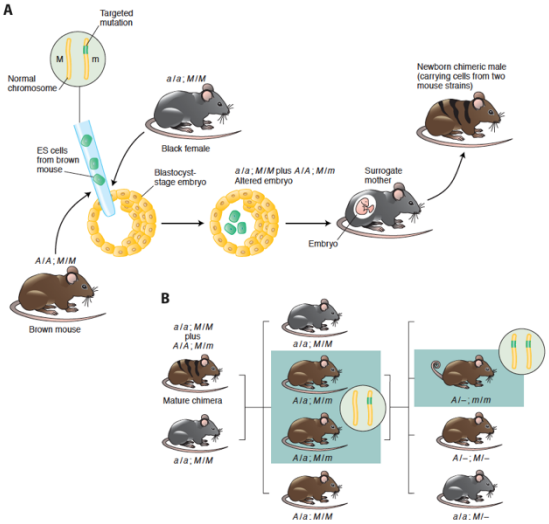


Figure 4-5 The principle of positive-negative selection. (A) Two genes have to be present on the targeting construct that can be used for selection: the neomycin-resistance gene (*neo*^{R}) and a thymidine-kinase gene (*tk*). **(B)** The presence of the neoR gene renders cells resistant to neomycin antibiotics (positive selection). The *tk* gene confers sensitivity to ganciclovir. Only the cells that underwent homologous recombination will be neomycin-resistant and ganciclovir-sensitive. (adapted from Introduction to Genetic Analysis, Griffiths, Palgrave Macmillan)

Injecting ES cells with successful targeted insertion (by homologous recombination)

Now that we have identified ES cells that contain the targeted insertion (these ES cells were the only ones surviving when adding ganciclovir and antibiotika), these cells have to be injected into blastocysts from a foster mother:

1) To be able to distinguish between the injected ES cells and the cells that derive from the host embryo, **the ES cells are taken from a strain of brown mice** (also called agouti mice) and the embryo used for injection is derived from a black mouse.
2) The injected/transformed embryo is then transplanted **into a surrogate mother that is also a black strain.**
3) If the injected ES cells become incorporated into the host embryo, the developing mouse will be chimeric - it will contain cells from two different mouse strains: the brown (derived from the ES cells) the black (derived from the embryo, see figure 4-6A).
4) **The newborn mice can be screened for chimeric animals simply by assessing fur color**
   - solid black animals **indicate that there was no insertion of ES cells,** and these animals are excluded from further analysis.
5) The chimeric (black-brown)mice are then mated with black mice and the progeny is screened.
6) If in the chimeric mice, the ES cells have been taken up into germ line cells, then some of the resulting mice will inherit the targeted insertion in all their cells (brown mice in figure 4-6B) → These mice will be heterozygous for the insertion.
7) Heterozygous mice are mated to produce **mice that are homozygous for the insertion.** → These represent knockout animals that can be analyzed to study the effect of the gene deletion.



How can heterozygous mutant mice be distinguished from their littermates that are brown, but do not contain the targeted insertion?

One big disadvantage in mice is that they need to be genotyped to detect whether they contain the targeted insertion. Unlike in Drosophila, where balancer chromosomes carrying dominant visible markers are used to "follow" chromosomes and therefore molecular genotyping is not necessary, mouse genetics uses PCR to analyze the DNA for the presence of a transgene or a knockout.

Conclusion

This method to create knockout mice with a deletion of the gene function of interest is very powerful and has led to important conclusions about the role of genes.

It is also used to **insert a desired gene into a specific locus in the genome, a situation called knock-in.**

- In this scenario, a mutant version of gene present in a genome can be replaced by its wild-type counterpart, which is the ultimate goal of gene therapy.
- Conversely, a wild-type gene can be replaced by a mutant version to create a pathogenic situation in the mouse to study a disease.

Disadvantage:

- However, these mutations are constitutive mutations **that are present in every cell of the organism during its entire life.**
- Such an integral modification often leads to embryonic lethality and thus prevents the analysis of these mutations.
- **To overcome this limitation, a system has been developed to express mutations conditionally, either in a specific tissue or at a specific time.**

Conditional knockouts with the Cre/loxP system

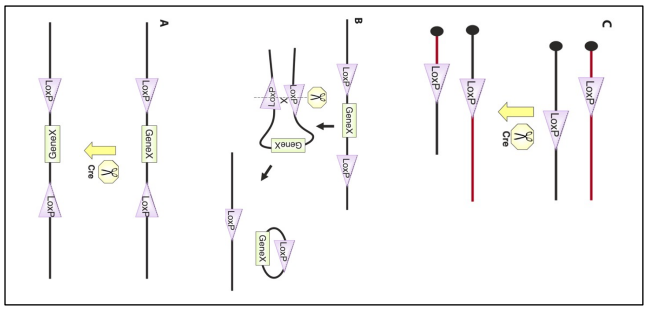
- needed when studying embryonic-lethal mutations
- when test the effect of a mutation at a specific stage of development or in a specific tissue

Flp/FRT system:

- If animals carry a site-specific recombination enzyme together with a target DNA that has recognition sites for this enzyme, the target DNA sequence can be rearranged at an appropriate time by activating the gene encoding the enzyme.
- a system that is widely used forward genetics in Drosophila to induce homozygous mutations in certain tissues, e.g., the Drosophila eye.

Cre/loxP system:

- In mice, a similar system, the Cre/loxP system, has proven to be very efficient and is therefore the most widespread tool among mice geneticists.
- **The system originates from a bacteriophage and can be used to promote recombination in different organisms including plants,** yeast, Drosophila and C. elegans.
- consists of a site-specific recombinase (Cre) that catalyzes the recombination specifically between two loxP sequences.
- The loxP site is a 34-base-pair DNA sequence and its location and orientation determines the result of the recombination.
  - A) If the loxP sites are **on the same DNA** strand but **face opposite directions**, the recombination event **inverts** the DNA sequence between the two loxPsites.
  - B) If the loxP sites are on the same DNA strand and have **the same orientation, the sequence between the loxP sites is excised** as a circular piece of DNA (most frequent)
    - ☞ Most frequently, loxP sites are introduced in the same direction around the gene of interest (also called "floxing" a gene), which results in a complete knockout of this gene when recombination takes place
  - C) If the loxP **sites are on different DNA strands,** the recombination **results in translocation of the two strands.**



Cre-mediated gene activation.

Alternatively, the system **can be used to activate gene expression.**

- In this case, two loxP sites are introduced between the promoter and the gene to be activated.
- Here, an intermediate sequence that prevents expression, e.g., a **polyadenylation sequence, is placed between the two loxP sites**.
- The gene of **interest is thereby kept silent until the polyadenylation sequence is excised by the Cre recombinase.**

how we can make the Cre/loxP system conditional

The Cre recombinase can - just as the Flp recombinase in Drosophila – be placed under a tissue-specific promoter to restrict its expression to this tissue (see figure 4-9).

Temporal control: (man entscheidet, wenn Cre in den Nucleus kann)

For temporal control, the expression of the Cre recombinase is induced **by external factors, such as heavy metals, hormones, ligands, etc.**

- For example, Cre can be fused **to a ligand-binding domain for the human estrogen receptor (ER)**.
- In the presence of its ligand, the ER is able to enter the nucleus (man bestimmt zu welchem Zeitpunkt man ER Ligand Tamoxifen zugibt)
- Thus, upon stimulation with the ligand tamoxifen, **a Cre** (containing the ligand binding domain of ER) **will translocate into the nucleus, where it can induce site-specific recombination on DNA sequences containing loxP sites**

How many crosses are required for a Cre/loxP conditional knockout described above?

In order to perform a Cre/loxP conditional knockout, the mice have to be homozygous for the loxP-flanked allele of interest and contain a cre gene that encodes the Cre recombinase. If the mice contained only one copy of the loxP-flanked allele, the gene function of interest could not be completely knocked out. –One starts by generating mice that are heterozygous for a loxP-flanked allele and heterozygous for the cre transgene. For this, you have to mate a homozygous loxP-flanked mouse of interest to a cre transgenic mouse strain (see Figure A below). Approximately 50% of the offspring will be heterozygous for the loxP allele and heterozygous for the cre transgene. This is your first cross. Then, you mate these mice back to the homozygous loxP-flanked mice (see Figure B below). Approximately 25% of the progeny from this mating will be homozygous for the loxP-flanked allele and hemizygous/heterozygous for the cre transgene. These will be your experimental mice in which you can observe the effect of the Cre/lox conditional knockout of your gene.

Figure A

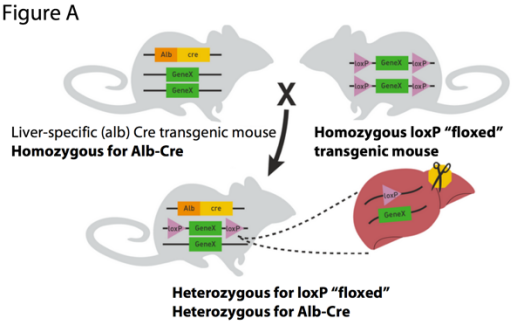
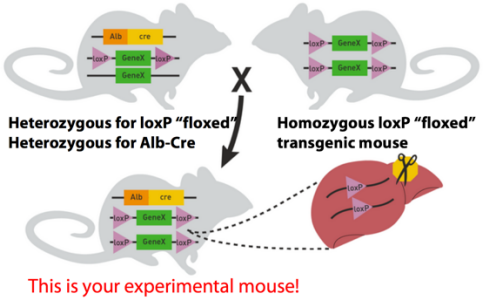


Figure B



For most cases, you **do not have** to create your "floxed" mouse by starting with the positive-negative selection of ES cells, but you can simply order it from companies that are specialized in the generation of transgenic mice.

The use of mosaics in mice

The Cre/lox system generates genetic mosaics, in which the animal consists of wild-type and mutant cells.

- Such mosaics allow studying genes that are important for very early events in development, since constitutive mutations in these genes would be lethal and not produce adult organisms.
- Furthermore, mosaics are informative for determining the tissue or cell type in which a given gene is required.
- **Mosaics can be used to test whether a gene function**
  - **is required in cells where it is expressed** (we say, the gene has a cell-autonomous function)
  - or in **neighboring cells** (then, the gene has a non-autonomous function).
- Thus, in mosaics, one can determine whether only the mutant cells exhibit the mutant phenotype (autonomous)

or if the phenotype is also observed in other wild-type cells (non-autonomous).

Can you think of a scenario how a mutation can affect a neighboring wild-type cell?

If the mutation affects a secreted protein that acts on the neighboring cell, then the loss of this molecule in one cell can affect the cell fate of the other cell. Another example would be the situation when a cell expresses a transmembrane protein that is required to activate a signaling receptor in the neighboring cell. For example, in the case of Notch signaling, an interaction between the Notch receptor on one cell with delta on another cell is required to activate Notch signaling. If one cell does not produce delta due to a mutation, the adjacent cell cannot receive the activating signal; thus, Notch signaling is not activated in this cell, although the signaling pathway is intact in this wild-type cell.

Visualization of cellular events with Cre/loxP

- The Cre/lox system can also be used to create mosaic animals in **which differently labeled cells can be created within an animal that can serve to visualize cellular events.**
- One strategy uses the Cre/lox system to visualize synaptic circuits by genetically **labeling neurons** with distinct colors.
  - The so-called brainbow transgenes allow the expression of multiple fluorescent proteins from a single transgene by Cre recombination.
  - The trick is to use incompatible lox sites, such that Cre-mediated excisions between pairs of incompatible lox sites creates mutually exclusive recombination events and therefore results in differently labeled cells.
  - In these experiments, up to **three alternative lox sites were used.**
  - Excision between either pair of identical lox sites removes one site of the other pair, thereby preventing further recombination
  - The promoter present on the brainbow transgenes is used to drive the expression of the fluorescent proteins (here: RFP, YFP and CFP).
  - Since the brainbow constructs are inserted **in tandem in many copies (as an multicopy array) in transgenic mice, the differential expression of** multiple copies of these constructs generates **fluorescent mixtures, allowing the labelling of individual neurons and glia cells with as many as 90 distinguishable colors.**
  - By placing the brainbow transgenes under any promoter of interest to direct their expression to a subset of neurons, it is **possible to follow these neurons and to visualize their synaptic interactions**

Biomedical applications of the Cre/loxP system

Gene-targeting strategies have also been successful in testing specific functions of mutations that are associated with various types of cancer.

- One example is the loss of the Retinoblastoma gene (Rb), which has been related to different types of cancer.
- Mice with a "floxed" Rb gene and tissue-specific Cre expression have been used to test the role of the Rb gene in the development of cancer.
- It was found that the knockout of Rb in the prostate was sufficient to induce extensive cell proliferation.
- However, in many other tissues, such as the retina, the mammary epithelium or the cerebellum, a second mutation of either an Rb-related gene or p53 was necessary to induce tumor growth.

Current projects in mouse-genome engineering try **to replace mouse genes with human genes** ("humanizing") **or to introduce specific point mutations ("knock-ins") or large chromosomal aberrations to create more accurate models for human pathologies.**

→ Since mouse models can faithfully recapitulate the phenotype of human diseases and allow studying the disease on a uniform genetic background, these projects are very promising in helping to understand the molecular mechanisms underlying the disease.

However, mice are still not humans, and in some cases, the etiology of a disease is diverse in different mammals.

→ however, p53-knockout mice develop a different spectrum of tumors than humans do: While mice often develop lymphomas and sarcomas, humans tend to develop epithelial cell-derived cancers.

☞ **Such phenotypic differences limit the utility of knockout mice as models of human disease.**

GENOME EDITING (OL)

Until now, we have discussed ways to alter a genome of an organism that rely on either random integration or on targeted integration by homologous recombination.

With the development of targeted mutagenesis (as discussed in the mouse knockout approach) it was possible to specifically alter a gene of interest.
However, all gene-targeting methods rely on homologous recombination, and recombination events are rather rare: while still relatively frequent in yeast (one in 10^4 cells undergoes homologous recombination), only one homologous recombination event happens in 10^7 mouse cells.

For yeast and mouse embryonic stem cells (ES cells), selection in cell cultures allows the recovery of the rare cells with the desired event
→ *but for many model organisms, **ES cells are not available**, screening or selection procedures are not adequate, and and the development of useful gene-targeting approaches is hindered by the low frequency of recombination.*

→ ***Thus, one of the big challenges was to increase the frequency of recombination**. Here, we will present three approaches that researchers developed to tackle this problem:*

Increasing the frequency of homologous recombination

- the key process is the **induction of DNA double-strand breaks (DSB) at the target site.**
- **natural pathway of DSB repairs is copying from a homologous template.**
- the approach of gene targeting by introducing DSB simply **provides an exogenous template** for a naturally occurring repair process
- Alternatively, DSB can be repaired by joining the broken ends directly without a need for a homologous **template, a process called** non-homologous **end joining.**
    - Sometimes, this **ligation is carried out imprecisely and can create deletions, insertions, or substitutions** at the break site and is therefore a tool for targeted mutagenesis

☞ **While non-homologous end joining stimulates targeted mutagenesis, the repair of DSB by homologous recombination can be utilized for targeted gene replacement by supplying a donor DNA sequence.**

To achieve efficient gene targeting, it is desired to increase the number of DSB, since this increases the likelihood of homologous recombination.

APPROACH 1: ZINC-FINGER NUCLEASES (ZFN)

☞ DNA is cut by DNA endonucleases → by targeting endonucleases to a specific genomic sequence, DSB can be induced at the desired location, followed by the generation of desired modifications during subsequent DNA repair.

☞ For this purpose, **a nuclease was created by fusing the DNA-cleavage domain of a bacterial endonuclease with a DNA-binding domain**, a zinc-finger domain.

☞ Zinc fingers are small structural motifs (domains) present in proteins → the domains contain multiple finger-like protrusions that make tandem contacts with their target molecule.

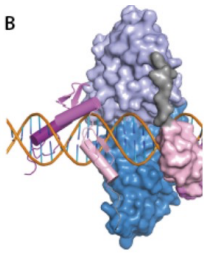
☞ Their binding properties depend on the amino-acid sequence of the finger domains and on the linker between fingers, as well as on the number of fingers.

☞ **Zinc-finger proteins typically function as interaction modules that bind DNA, RNA, or proteins.**

☞ The zinc-finger nucleases (ZFN) created for gene targeting are based on the structure of the natural restriction endonuclease FokI (from Flavobacterium okeanokoites).

☞ **The DNA-binding domain can be engineered to target a specific DNA sequence such that the zinc-finger nucleases can be directed to the desired sequence where it induces DSB.**

☞ The most useful **design for a synthetic DNA-binding domain proved to be a set of three Cys2His2 zinc fingers, with each finger primarily binding to only 3 base pairs of DNA.**



☞ It is important to note that that **two** ZFN molecules need to bind **simultaneously** with their DNA-binding domains to their recognition sequence on the genome

Why is the requirement for dimerization a great advantage of ZFNs? And what challenge does this dimerization pose?
Because it provides high specificity. A monomer is not active, so cleavage does not occur at single binding sites. Also, the

cleavage reagent is assembled, and thus cleavage occurs only at the target if the fingers have adequate specificity. Palindromic recognition sequence are needed for two ZFN to bind.

How are the engineered zinc-finger nucleases introduced into animals to generate gene deletions?

- ZFN-based genetic engineering does not require the use of ES cells, **because ZFNs can be injected directly into early-stage embryos** → fertilized single-cell embryos are injected with plasmids containing the
- In the nucleus, **the ZFN locates the target sequence and creates a DSB.**
- The DSB stimulates the cellular process of non-homologous end joining and results in the mis-repair of the DNA sequence. → knockout mutation
- The embryos are then further cultivated (Drosophila) or implanted into a foster mother (mice) and allowed to divide and grow into whole organisms.
- At birth, the animals are screened for mutations and the animals containing the targeted knockout are identified.

Advantages:

- ZFNs have successfully been used for targeted gene replacement in a variety of organisms
- ZFNs also provide an alternative strategy to embryonic-stem-cell-based homologous recombination, and can, thanks to their low immunogenicity, be used for gene therapy.
- Since ZFN-based genetic engineering does not require the use of ES cells, this allows targeted gene disruption in a wider spectrum of organisms and in a shorter time.
- Further, the ZFN method **results in efficient germline transmission of targeted genetic mutations without incorporation of foreign DNA sequences (unlike in the ES-based knockout approach)!!!**

Disadvantages:

- the design and production of ZFNs are **expensive** (several 1000 US$) and time-consuming (>1 month), since context-dependent effects, e.g., neighboring fingers can alter the specificity, and off-target effects have to be taken into account.
- The three zinc-fingers themselves **pose rather inflexible molecules, and can therefore only be targeted to sequences of about 500 base pairs.**
- **a substantial portion of ZFNs fails in vivo due to the limited capacity of target cells towards the ZFNs.**
- not every nucleotide triplet has a corresponding zinc finger, and interactions between zinc fingers within an array can reduce their specificity.

APPROACH 2: TRANSCRIPTION ACTIVATOR-LIKE EFFECTOR NUCLEASES (TALEN)

method that is also based on engineering the FokI nuclease to direct it to specific sequences by fusing it to different DNA-binding domains.

- Instead of zinc fingers, **they used transcription activator-like (TAL) effectors, proteins that are secreted by the bacteria of the genus Xanthomonas when they infect various plant species.** → **can bind promoter sequences in the host plant and activate the expression of plant genes that aid bacterial infection.**
- They recognize DNA sequences through a central **repeat domain** consisting of a variable number of ~34 amino-acid repeats.
- The sequence of each repeat is similar **but not identical: it differs at amino-acid positions 12 and 13.**
  - This amino-acid pair is called repeat-variable di-residue (RVD), and each RVD recognizes a specific base.
  - Upon folding of the TALE protein, **the different RVDs are exposed at the protein surface within a sequence.**
  - This RVD sequence allows the TAL effector to specifically bind the sequence of corresponding bases, and thus binding to the target DNA sequence => it's the RVD that determines which single nucleotide the TAL effector will recognize:
    - ☞ HD targets cytosine, NI targets adenine, NG targets thymine, and NN targets guanine
- Knowing how certain RVDs "code" for certain bases, TALEs can specifically be designed to bind a particular DNA sequence of interest.
- Like ZFNs, TALEs are fused to a DNA-cleaving domain, such as FokI, and thus provide a tool for genome editing through targeted gene replacement

Advantages:

- they have a **potential cleavage sites approximately every 35 base pairs in the genome, while potential target sites**

for zinc fingers are roughly every 500 base pairs, because zinc-finger targets are confined to sequences composed of triplets with corresponding zinc fingers.

- They are context-independent and appear to show little off-target effects and low immunogenicity.
- Furthermore, the designed TALEs can be fused to another protein in order to achieve a targeted effect.
- For example, by fusing a designed TALE to transcriptional activation domains (AD) or to repression domains (RD), the designed fusion protein serves as an artificial switch for gene regulation in vivo

APPROACH 3: THE CRISPR/CAS9 SYSTEM

Method 1 and 2 use the principles of DNA-protein recognition to induce site-specific double-strand breaks to increase homologous recombination.

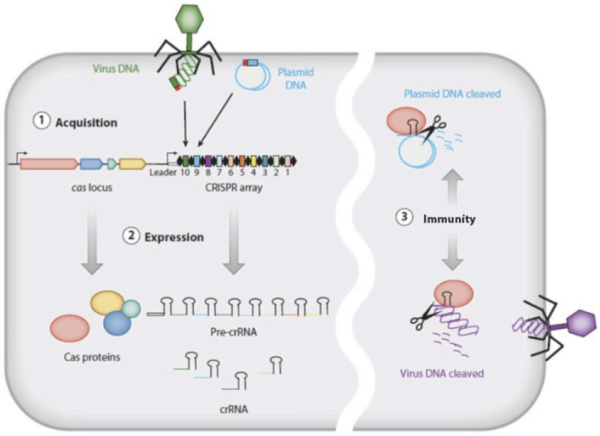
Crispr-Cas9 technology: a system that uses an RNA-programmable system to introduce double-strand breaks

- *originally provides bacteria with adaptive immunity to viruses and plasmids.*

The CRISPR/Cas9 immune defense has two important features:

- the host cell can **specifically** incorporate short sequences from invading genetic elements (virus or plasmid) **into a specific region of its genome, the so-called CRISPR-array locus.**
  - This locus contains **repeat sequences, which are called clustered regularly interspaced short palindromic repeats (CRISPR),** and pieces of the foreign DNA integrated between these repeats (called protospacer sequences).
- when these sequences on the CRISPR-array locus are **transcribed** and precisely **processed into small RNAs,** they guide a multifunctional protein complex (Cas proteins) to recognize and cleave incoming foreign genetic material.

- In bacteria and archaea, **three different** CRISPR/Cas9 systems exist that act similarly: here, we will discuss the type-II system of the bacterium Streptococcus pyogenes, which served as the basis to develop the CRISPR/Cas9 genome-engineering system.
  - shortly: transcription of the CRISPR array leads to the formation of a long RNA that is subsequently cleaved into short crisper RNAs (crRNAs).
  - These crRNAs contain the sequence information about the invader (since they are transcripts of the foreign DNA).
  - **These RNAs bind to another RNA called the transactivating RNA (tracrRNA), and this RNA hybrid is able to build a complex with a cellular endonuclease, the Cas9 protein.**
  - If this RNA-Cas9 complex encounters invasive DNA present in the cell, **the crRNA directs the Cas9 nuclease to the invasive DNA by complementary base pairing.**
  - The Cas9 nuclease then cuts and destroys the incoming foreign DNA form a Plasmid or a virus

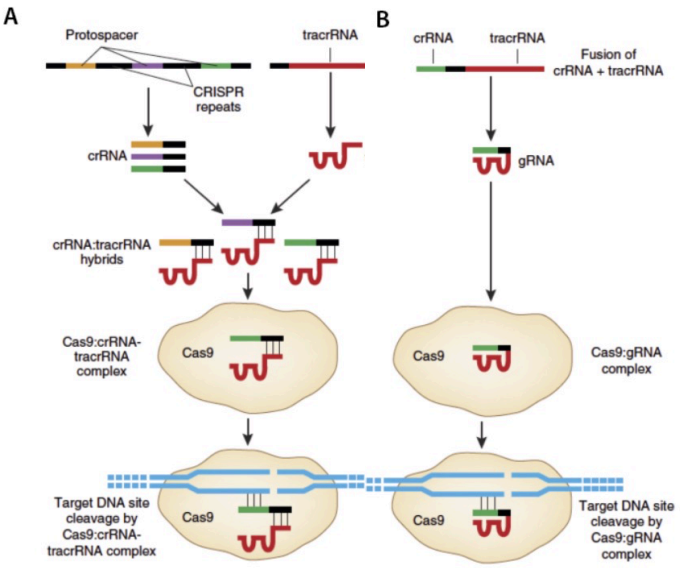


Hot to generate immunity against this virus/foreign DNA/ genauer Mechanismus erklärt (fast gleich wie oben):

- Transcription of the CRISPR array leads to a long RNA that is subsequently processed into crRNAs, each harboring a variable sequence transcribed from the invading DNA (called the protospacer sequences, marked in different colors in figure 5-7) and part of the CRISPR repeat (black in figure 5-7).
- Each **crRNA hybridizes with a second RNA, the tracrRNA, through a short region that is complementary to the CRISPR repeat present on the crRNA.**
- **These two RNAs together complex with the Cas9 nuclease.**
- The tracrRNA and Cas9 form a complex with each unique crRNA.
- **Accordingly, each crRNA:tracrRNA:Cas9 complex seeks out the DNA sequence complementary to the crRNA.**
- The protospacer sequence (contain forteign DNA) present **on the crRNA directs the complex to the**

complementary sequence on the DNA, where the Cas9 nuclease cleaves the target DNA.

- After the crRNA-tracrRNA:Cas9 complex binds, the Cas9 separates the double-stranded DNA target and cleaves both strands (see figure 5-7A).
- The crRNA:tracrRNA:Cas9 complex unbinds after inducing the DSB.



Crispr-Cas in genome ingeneering:

- In the CRISPR-Cas9 genome-engineering technology, the dual tracrRNA:crRNA was engineered as a single-guide RNA (sgRNA) that retains two critical features:
  - a sequence at the 5'-side that determines the DNA target site (the protospacer, green in figure 5-7B)
  - and a duplex-RNA structure at the 3'-side that binds to Cas9.
  - By specifically designing the protospacer sequence to target a genomic sequence of interest, only the synthetic sgRNA and Cas9 have to be provided to a cell in order for the system to target and efficiently induce genomic DSB (5-7B)
- This development created a simple two-component system in which changes in the guide sequence of the sgRNA program Cas9 to target any DNA sequence of interest.

a) -

b) After the Cas9-induced double-strand break occurred, the DNA can either **be repaired by non-homologous end joining (NHEJ), which can induce insertions or deletions, or by homology-driven repair (HDR).**

c) HDR can be used **to insert desired sequences through recombination of the target locus with supplied DNA donor templates** (see figure 5-8B).

d) If pairs of guideRNA (gRNA) are used, the Cas9 nuclease can **induce large deletions or chromosomal rearrangements, such as inversions** (see figure 5-8C).

e) by fusing the **Cas9 nuclease to protein domains that can activate the expression of proteins, the expression of proteins can be specifically induced** (figure 5-8D).

f) The fusion of Cas9 with other effector domains may be used to alter DNA modifications (see figure 5-8E).

g) if Cas9 is fused to fluorescent proteins, specific genomic loci are marked and analyzed by imaging (F).

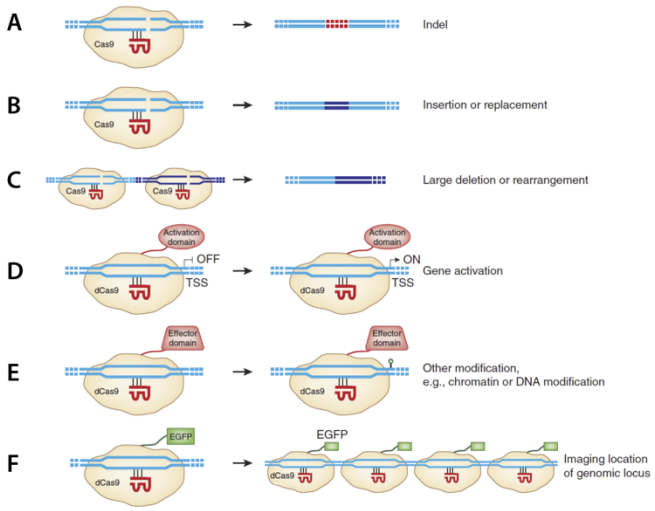


Figure 5-8 Overview of various Cas9-based applications. (A,B) gRNA-directed Cas9 nuclease can induce

Comparing the three methods for targeting DSB to specific genomic loci

All three methods presented here aim at inducing double-strand breaks by employing molecular scissors using endonucleases from bacterial systems.

- In the ZFN and TALEN approach, the endonuclease is directed to a cleave at a specific site with the help of protein domains that bind DNA (zinc-finger domains or TALE).
- In the CRISPR/Cas9 system, an RNA is used to guide the endonuclease to the desired DNA sequence.

Imagine you want to study the effect of gene dosage of a certain gene X. Gene dosage is the number of copies of a particular gene present in the cell of an organism. Usually, a diploid cell has two copies of a gene. An increase or

decrease in the gene dosage can result in different levels of gene product. Which approach would you choose: RNAi or CRISPR? Can you explain why?

CRISPR would be the approach of choice, because organisms with one (heterozygous mutant) or two modified alleles (homozygous mutant) can be made. On the contrary, gene dosage cannot be controlled by RNAi. The degree of RNA interference (and thus, the degree of gene downregulation) depends on many parameters that are not easily controlled by the experimenter, such as how much dsRNA enters the cell, how much of it is processed and loaded onto RISC, how much mRNA of the target gene is present, how many mRNAs are targeted, how much protein was already made from the mRNA before RNAi was induced, etc.

Comparison of RNAi and CRISPR

RNA interference and the CRISPR/Cas9 system are both important tools in molecular biology. What do these two methods have in common and how do they differ?

- Consider the following subquestions:
- What are the consequences on a molecular level?
- Which molecular factors are involved?
- What is the mode and place of action?
- Which genome functions can be studied?
- Where do these technologies originate from?
- How labor-intensive is each method?

Common points:

- Both are mediated by small noncoding RNAs with a target specificity of roughly 20 nucleotides
- Both work in combinations with a ribo-nucleoprotein complex to target specific nucleic acids sequences
- Both originate from defense mechanisms against foreign DNA

Differences:

- Technological aspects:
  - RNAi: fast, easy to use (transfection, feeding, etc.), suitable for large-scale genetic screens
  - CRISPR: labor intensive, requires transgenic animals
- Molecular consequences:
  - RNAi: Transient gene downregulation, does not change genome sequence
  - CRISPR: permanent gene knockout, changes genome sequence
- Mode and place of action:
  - RNAi: in the cytoplasms, target nucleic acid is mRNA
  - CRISPR: in the nucleus, target dsDNA
- Genome functions that can be assessed:
  - RNAi: Protein coding sequences
  - CRISPR: Protein coding, Introns, Promoter
- Origin:
  - RNAi: is an endogenous eukaryotic pathway that uses cellular machinery (RISC)
  - CRISPR: derives from a bacterial defense mechanism

SMALL NON-CODING RNA'S

STEPPING INTO SMALL RNA BIOLOGY (OL)

- *However, the function of a gene can also be affected by changes that are not associated with a change in the DNA sequence, in which case the gene is said to undergo epigenetic regulations*
- *RNA interference is the process of RNA silencing where the down regulation of gene expression is mediated by short noncoding RNAs (small RNAs) derived from long, perfectly complementary dsRNA molecules.*
- *RNA silencing via small RNAs suppresses gene expression post-transcriptionally by base pairing to mRNAs, but we will see that in some organisms small RNAs can also anneal to nascent transcripts of target DNA and promote epigenetic modifications of the DNA via cytosine methylation.*
- *In the past: RNA was mostly considered as a mere mediator, but not a direct actor, in the genetic information flux*
- *cellular functions: First, the linear sequence of RNA makes it a simple source of genetic information. Further, the single-stranded nature of RNA allows the formation of secondary structures to facilitate interactions with other molecules. RNAs can even assume tertiary structures that present surfaces for interactions and contain internal environments that create binding sites for metal ions, such that some RNAs can promote catalytic reactions.*

- *Ribozymes are catalytic RNAs that function within the ribosome to link amino acids during translation.*
- *The viroid hammerhead ribozyme cleaves and linearizes the genomic RNA to allow subsequent replication in a rolling-circle mechanism.*
- *small, non-coding RNAs can control gene expression by recognizing the expressed mRNAs of genes, either directing their degradation or inhibiting their utilization in translation = **post-transcriptional gene silencing or PTGS***

Classes of small non-coding RNAs (sncRNA)

three main categories, only present in eukaryotes:

Short interfering RNAs (siRNAs) – (normally exogenous)

- siRNAs act in the biological gene-silencing process called RNA interference (RNAi).
- The stimulus that triggers RNAi is the presence of **a long dsRNA** in the cell, and this molecule was originally introduced **artificially** into organisms such as plants, worms or flies using injection or transgenic expression.
- RNAi does not exist in cells for the sake of silencing transgenes or experimental dsRNA!
- In fact, it was found that RNAi is an evolutionarily conserved response to exogenous dsRNA, which reflects an ancient endogenous defense against foreign or invasive nucleic acids, such as viruses or transposons
- Recently, it has become clear that there are also **endogenous sources of siRNAs**
- siRNAs, on the other hand are produced by consecutive cuts by Dicer along a perfect, long dsRNA molecule (remember the movie in the introduction lecture). Therefore, siRNAs, unlike miRNAs, are always part of a population that is distributed along their long dsRNA precursor.

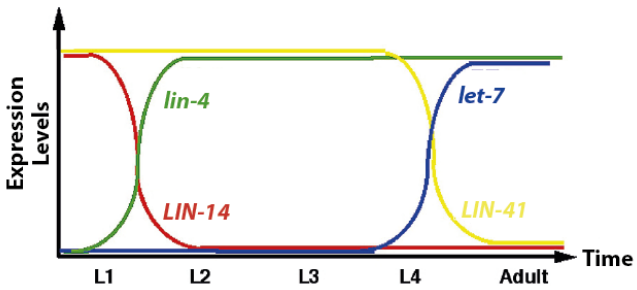
microRNAs (miRNAs)

- are encoded by the cell's own genome and serve to regulate the organism's gene expression.
- Both miRNAs and siRNAs start out as precursor molecules that are processed within the cell into short, double-stranded RNA fragments of 20-30 nucleotides.
- While miRNAs are processed from stem-loop or hairpin-like precursors (they arise due to mismatches between the two not completely complementary strands), siRNA precursors are usually longer and fully complementary dsRNAs.
- miRNAs are excised by Dicer from their short (roughly 70 nucleotide long in mammals) and imperfect stem loop precursor as a single small RNA species, which may accumulate to tremendous levels in the cell (up to 50,000 molecules/cell).
-
- Processing of the precursor is, in both cases, mediated by the enzyme Dicer. One of the two strands of the small dsRNAs is loaded onto so-called RNA-induced silencing complexes (RISCs) and the target mRNA is recognized through Watson-Crick base pairing.
- Once the target is recognized, its expression is silenced
- Both siRNAs and miRNAs are double-stranded sequences produced by Dicer RNase III enzymes
- both are present in many cell types of **eukaryotic** organisms.

piwi-interacting RNAs (piRNAs)

- are Dicer-**independent**, single stranded and found **exclusively in metazaons** where they exert their function in the germline.

Discovery of small regulatory RNAs



- lin-4 being the first microRNA to be discovered
- Worms with lin-4 loss-of-function mutations recapitulate cell-division patterns that are specific to the first larval stage (L1) at inappropriate later stages of development, and lin-4 activity was shown to be required for the transition from the L1 to L2 stage
- lin-4 transcripts of only 61 and 22 nucleotides in length. Thus, Ambros and his colleagues reasoned that lin-4 did not encode a protein and was perhaps exerting its regulatory role as RNA
- lin-4 RNAs were partially complementary to a repeated sequence in the 3' UTR of the lin-14 gene. Thus, the non-

coding transcript of lin-4 regulates lin-14 mRNA through binding to its 3' UTR

- lin-4 is partially complementary to seven different sites in the lin-14 3' UTR (see figure 1-2B). This feature is common; many miRNA target transcripts carry multiple miRNA target sites.

The phenotype of a lin-14/lin-4 double mutant looks like the lin-14 single mutant. What is the epistatic relationship between lin-14 and lin-4, and how would you place them into a pathway?

Remember that in a regulatory pathway, a downstream mutant usually determines the phenotype and is therefore epistatic over the upstream gene. Here, the double mutant shows the lin-14 phenotype, thus, lin-14 must be downstream and therefore epistatic over lin-4:
lin-4 ---l lin-14 ⟶ larval-to-adult development

- another forward genetic screen in C. elegans identified the second miRNA, let-7
- in let-7 that causes the reappearance of larval cell fates during adult development. let-7 encodes a small, 21-nucleotide long RNA
- let-7 is complementary to sequences in the 3' UTR of lin-41, and exerts its function by binding to this region to inhibit the translation of lin-41

☞ In summary, during C. elegans development, the accumulation of the two miRNAs lin-4 and let-7 is inversely correlated to the expression of their targets, lin-14 and lin-41, respectively (see figure 1-4). This allows to perfectly control the timing of their expression and therefore an adequate larval-to-adult transition.

The connection to the siRNA pathway revealed the mode of biogenesis of miRNAs

- In 1999, plant scientists discovered that in all cases of transgene silencing studied, RNA complementary to the transgene mRNA was detected.
- they had sense and antisense strands, suggesting they were short double-stranded RNAs most likely derived form a longer dsRNA precursor
- most of the plant transgenes that triggered RNA silencing were organized into inverted repeats as a result of aberrant transgene insertion into the genome, a feature that allows the production of long dsRNA molecules
- Strikingly, the size of these sRNA was the same as the ones of the miRNAs lin-4 and let-7. This led to the idea that the siRNA and miRNA pathways could be connected via Dicer
- a loss of function mutationDicer showed similar phenotypes as lin-4 or let-7 mutants, suggesting that a common dsRNA processing machinery is used by in the siRNA and miRNA pathways

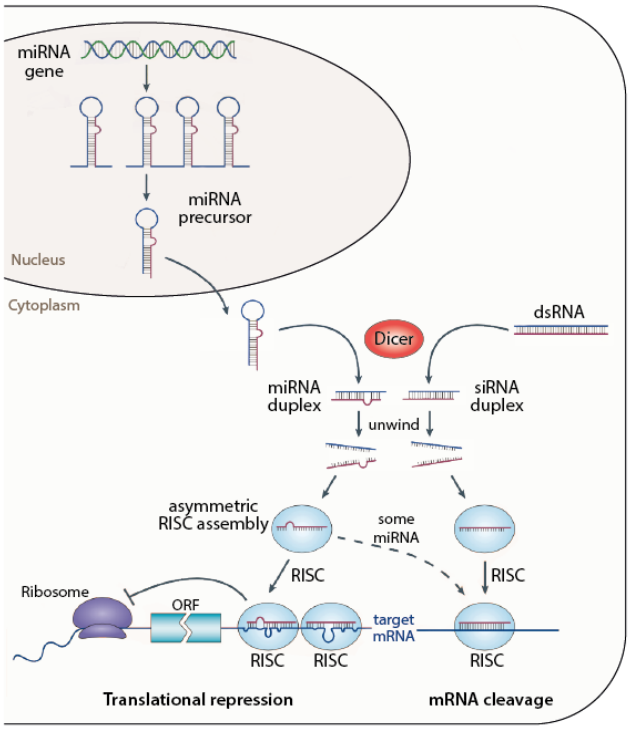


Figure 1-6 Mode of action of microRNAs (miRNAs) and small interfering RNAs (siRNAs). Note that, although not represented on the scheme, some imperfectly matched siRNAs may also trigger translational repression. (Adapted from L. He and G.J. Hannon, *Nature*, 2004)

Modes of action of miRNAs and siRNAs

Gemeinsamkeiten miRNA und siRNA:

- Both siRNAs and miRNAs depend on the enzyme Dicer for their maturation from precursor molecules into the 20-nucleotide long, small fragments, and both small RNAs need to be incorporated into the RISC to mediate their gene-silencing function.

Unterschiede:

- siRNAs were found to bind to their mRNA targets with **perfect complementarity** and direct the cleavage of the mRNA at the site of complementarity
- miRNAs were found to act as translational repressors

- In many cases, miRNAs bind to the 3' UTR of the target mRNA through multiple **imperfect matches** and therefore negatively regulate expression by repressing translation of the target mRNA, as originally found for lin-4 and lin-14

Ausnahmen:

- These rules are not always true → in plants, for example, most **miRNAs anneal to their target with nearly complete complementary** and therefore mediate cleavage of their target mRNAs just like siRNAs, but the remaining non-cleaved mRNA fraction may also undergo translational repression.
- Likewise, some miRNAs in animals promote target cleavage via **extended miRNA-target complementarity while**, conversely, some mammalian siRNAs imperfectly paired with their target can trigger **translational repression rather than mRNA degradation**.
- Thus, it is important to remember that both mechanisms of silencing genes exist, and that both mechanisms can be used by either siRNAs and miRNAs. (siehe Bild)
- **In cases where gene expression has to be adapted rapidly**, e.g., as a response to an external stress, **a regulation that is based on translational inhibition is more flexible, because it is quickly reversible**, enabling translation to resume as soon as the stress is over.
- **By contrast, mRNA degradation** is **irreversible**, and therefore some delay will occur before the cell has produced new mRNA that can be used for translation of the protein.

Biogenesis of miRNA and RISC assembly

SUMMARY - HOW MATURE MI-RNA IS GENERATED:

The two processing events that lead to mature miRNA:

i) the miRNA transcripts (called pri-miRNA) are processed into precursors of 70 nucleotides (called pre-miRNA) that contain the typical stem-loop

  - In animals, this first process is mediated by a nuclear ds-RNA-specific nuclease called Drosha.

ii) The second processing step cuts off the loop from the pre-miRNA to create a mature miRNA of approximately 22 nucleotides.

- In plants, the two steps take place in the nucleus, while in animals the pre-miRNA is exported into the cytoplasm where the enzyme Dicer carries out the final cleavage reaction to produce a single, discrete species of miRNA.
- Plants use the Dicer homolog Dicer-like 1 (DCL1) to carry out both cleavage processes in the nucleus
- Another specificity of plants is that miRNAs, like all plant sRNAs, undergo 2'-O-methylation on both strands via the action of the RNA methylase HEN1.
  - This modification protects the end of the miRNAs from undergoing oligouridylation and subsequent degradation.
- Animal miRNAs are not methylated, in contrast to endogenous siRNAs and piRNAs found in these organisms.

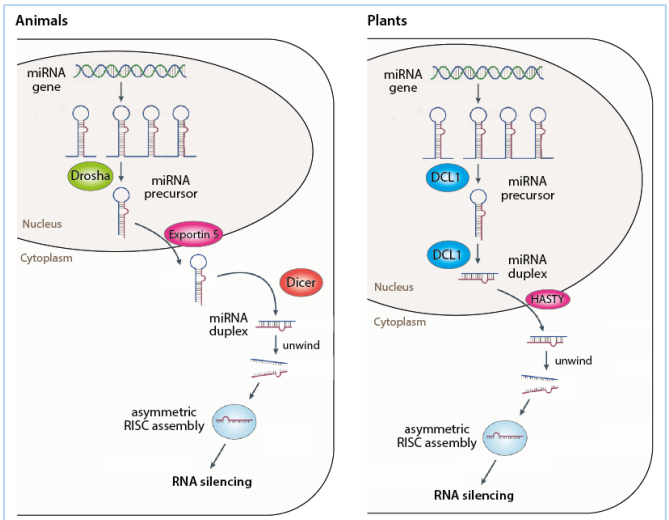


Figure 1-7 Biogenesis and RISC assembly in animals (left) and plants (right). (Adapted from L. He and G.J. Hannon, *Nature*, 2004 and E.J. Chapman and J.C. Carrington, *Nat. Rev. Genet.*, 2007)

LONG EXPLANATION - HOW MI-RNAS ARE GENERATED:

1) The pre-miRNA (in animals) as well as the miRNAs (in plants) are exported into the cytoplasm by Exportin 5 or HASTY, respectively, a Ran-GTP dependent transporter.
2) In the cytoplasm, the miRNAs associate with the RISC. To be able to specifically bind the target mRNA, it is important **that only one strand** (the so called guide strand) **is incorporated into the RISC, while the other one (so called passenger strand or miRNA* strand) is discarded.**
   a. The **separation of the two miRNA** strands and the selection of one specific strand to be incorporated into RISC is based on thermodynamic properties of the duplex as the unwinding begins at the extremity displaying the highest free energy (i.e., the weakest base-pairing) → **The process does not require energy**

b. On the contrary, **strand separation from perfect siRNA duplexes requires energy** and is catalyzed by **AGO-mediated "slicing"**

3) The selected miRNA guide strand then becomes incorporated into AGO, where it acts as a guide
4) AGO exposes the most 5'-end nucleotides of the miRNAs used for initial target scanning.
   a. The nucleotides 2-8 in this region, **called the seed region, are essential for scanning** and **initial nucleation of the base-paring between the miRNA and its target.**
   b. **If the strength of pairing is not sufficient, then the AGO-miRNA complex might release the target and move on**

- There are multiple AGO homologs in each animal species suggesting that the **composition of RISCs might be different to act in a tissue-specific or developmentally regulated fashion.**
- Some of the **AGO proteins have endonuclease activity and can catalyze the cleavage of target base pairs.**
  - This initial cut, called **"slicing" represents the critical first step in some of the RNA-silencing events that proceed through RNA destabilization**

Post transcriptional gene silencing by miRNAs

- Almost all animal miRNAs bind to the 3' UTRs of their target genes with <u>mismatches</u>.
  - the degree of miRNA-mRNA complementarity **determines the regulatory mechanism:**
    - <u>perfect complementary</u> allows **AGO-catalyzed endonucleolytic cleavage** (slicing) of the complementary mRNA strand.
    - **zu 99%: mismatches at the active catalytic site of AGO** preclude/ausschliessen cleavage and promote **repression of mRNA translation**
    - It is believed that translational repression might be the default mechanism to repress gene expression in animals.

- In contrast, most plant miRNAs bind with <u>near-perfect complementary to their target</u> site within the coding region of their targets.

<u>If 99% of animal miRNAs display imperfect mismatches to their target in order to avoid slicing, why have the catalytic residues of their corresponding AGO proteins been conserved in evolution?</u>

Perhaps the 1% sliced miRNA targets is very important for growth and development, or, alternatively, maybe the slicer residues of AGOs are used for other processes. Indeed, the slicer activity ofAGO was shown to process precurser miRNAs without the requirement of Dicer. This is a rare event, however, it was shown to be important during early embryogenesis in zebrafish.

<u>If most animal miRNAs have binding sites in the 3' UTR of transcripts, what is the implication for the size of the 3' UTR of house keeping mRNAs (i.e., mRNAs whose levels are mainly subjected to very limited or no fluctuation)?</u>

Most of the transcripts not regulated by miRNAs in animals are indeed housekeeping mRNAs with short 3' UTRs that avoid miRNA-mediated silencing and thus fluctuations of gene expression.

Mechanisms of translational repression by miRNAs

How and at which step do miRNAs repress translation?

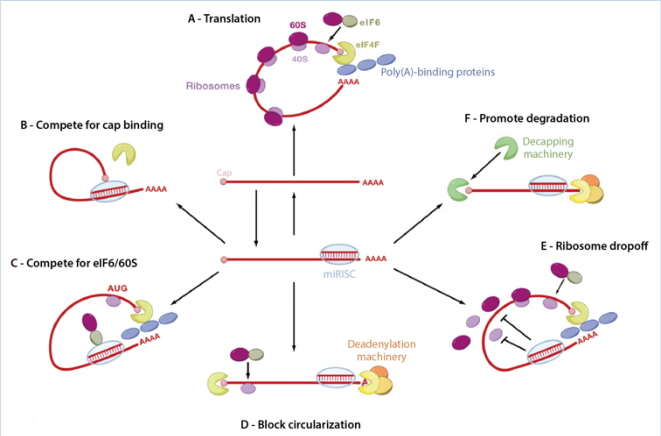


Figure 1-10 Possible mechanisms of translational suppression by miRNA-containing RISCs (miRISCs). (A) Non-repressed mRNAs recruit initiation factors and ribosomal subunits and form circularized structures that enhance translation. **(B)** When miRISCs bind to mRNAs, they can repress initiation at the cap-recognition stage or **(C)** the 60S-recruitment stage. **(D)** miRISCs can induce deadenylation of the mRNA and thereby inhibit circularization of the mRNA. **(E)** miRISCs can also repress a post-initiation stage of translation by inducing ribosomes to drop off prematurely. **(F)** Finally, miRISCs can promote mRNA degradation by inducing deadenylation followed by decapping. (Adapted from R.W. Carthew and E.J. Sontheimer, *Cell*, 2009)

1) Translation of an mRNA molecule is initiated by the protein complex eIF4F (figure 1-10A, green open circle) that recognizes the 5'-cap structure of the mRNA.
   a. eIF4F is the protein complex that regulates translation initiation.
2) After binding to the 5' cap of the mRNA, proteins of the **initiation complex recruit the 40S** subunit of the ribosome (figure 1-10, dark purple) together with another initiation factor, eIF6 (see figure 1-10, olive circle).
3) At the AUG start codon, the 40S subunit joins the 60S subunit (light purple in figure 1-10) and elongation begins.

a. The initiation complex also interacts with the 3' end of the RNA through interaction with proteins that specifically bind to the poly(A) tail of the mRNA (blue circles in figure 1-10).
b. **This simultaneous interaction of the initiation complex with the 5' and the 3' ends of the mRNA leads to a circularization** of the molecule, which **greatly enhances translation efficiency**.

There are several ways how post-translational repression is mediated by miRNAs (see figure 1-10, B-F):

- In all cases, however, translational repression is coupled to a form of accelerated mRNA decay that is induced by deadenylation and decapping of the target mRNA.
- This is mediated by a **deadenylation and decapping (yellow and green in figure 1-10, respectively) machinery, containing 3'-to-5' (for deadenylation) and 5'-to-3' (for decapping) nucleases**.
- Thus, in animals, the action of miRNA ultimately results in decreased mRNA levels, although this effect is usually very modest (1.5 to 3 fold reduction) for any single miRNA.
- It is very important to understand that mRNA decay coupled to translational repression is completely distinct from mRNA endonucleolytic cleavage (slicing) mediated by AGOs in the context of perfect miRNA:target base pairing.

The net effect of most animal miRNAs is to reduce target mRNA levels by approximately two folds on average. However, most genes function normally in a hemizygous condition, where a single allele can produce enough gene product to maintain the wild-type phenotype without the other allele. How, then, is the miRNA effect achieved in cells?

- The 3' UTRs of many animal mRNAs contain multiple targets for a single mRNA and, often, for distinct miRNAs. It is the bulk effect of these molecules that counts here.
- There might be "elite target" mRNAs displaying much stronger effects for a single miRNA, while the other mRNAs might be futile or even used to buffer out the action of the miRNAs. In that scenario, currently under experimental investigation, targets would become regulators of the miRNA itself!

IDENTIFICATION OF MI-RNA'S AND THEIR TARGETS (OL)

*Due to their small size and their high degree of homology, identifying and quantifying miRNAs in the lab experimentally is challenging. Here, we will **discuss the methods commonly used to identify and experimentally detect miRNAs themselves, but also the mRNA targets they regulate***

- miRbase repository (http://www.mirbase.org), the main miRNA resource, which now includes thousands of miRNAs from all organisms.
- raditional methods for detecting miRNAs include Northern blotting, RT-PCR, microarrays and sequencing

Northern Blotting

*to study gene expression by the detection of RNA, e.g., miRNAs → one can determine whether a certain RNA of interest is present in a sample, and at which level → requires the **researcher to know the sequence of the investigated RNA***

☞ ***Northern blotting is the only technique that allows for the (semi-) quantitative visualization of miRNAs.***
☞ *Because gel electrophoresis separates RNA sizes, mature as well as pre-miRNAs can be analyzed*

1) iRNA can be better detected if the total RNA samples are separated by size using electrophoresis (15-18% polyacrylamide gel)
2) After electrophoresis, the RNA is transferred onto a membrane, a process called blotting (hence the name).
3) This membrane is incubated with a hybridization probe (a synthesized miRNA-complementary oligonucleotide), a nucleic acid that is complementary to the target sequence (the RNA to be detected).
   - To detect the hybridization between the probe and the target RNA, **the probe has to be labeled**, e.g., with radioactive phosphorus (32P) that is incorporated into the nucleic-acid probe.
4) After hybridization, **an X-ray film** is applied to the dried membrane to detect the radioactive signals in those samples where hybridization took place

Disadvantages:

- compared to other methods, e.g., qRT-PCR, northern blotting **suffers from low sensitivity** (RNAs with low abundance cannot be detected), **low throughput**, and **high input RNA requirements** (typically in the order of 5–50 µg total RNA per sample have to be loaded onto the gel).
- Radioisotopes (32P) are still the most commonly used labeling system for northern-blot detection, even **though they pose several safety concerns** for the researchers using them and the environment where they are disposed.
- the use of **radioisotopes greatly increases the amount of time required to conduct** the experiment, and in some cases, 32P labels must be exposed for days to detect weak signals.
- **The biggest problem of northern blotting is the low sensitivity of this method**, such that low abundant RNAs, such as some miRNAs, are difficult to detect.
  - The sensitivity of detection can **be increased by using locked nucleic acids (LNA) in the probes.**
  - LNAs are RNA nucleotides, in which the **ribose moiety is modified with an extra bridge connecting the 2' oxygen and 4' carbon. The bridge "locks" the ribose in the 3'-endo conformation.**
  - LNAs hybridize with DNA or RNA according to Watson-Crick base-pairing rules, where the locked ribose conformation increases the hybridization properties of the probe.

Advantages:

miRNA detection via radioactive labelling and hybridization remains the number-one method of miRNA analysis in plants and many other species, because it is comparatively cheap, easy to perform in every laboratory, and does not rely on high-tech equipment, such as sequencing machines.

LNA oligonucleotides are extremely stable in vivo and highly specific. So, what other application would LNA be useful for in terms of miRNA detection?

They can be used as very efficient probes to detect miRNA expression in a variety of tissues or whole organisms, using a method called in situ hybridization

Quantitative reverse-transcriptase PCR (RT-PCR)

popular technique for validating and accurately quantifying miRNAs → you also aim at identifying whether a specific transcript is present in your sample (the RT-PCR part) and additionally determine to which amount this RNA is present (the quantitative part).

RT-PCR part (Process for mRNAs, not miRNAs)

1) starts with a population of RNA molecules isolated from a tissue or cell culture, which you would like to amplify.
   a. in order to amplify the RNA using PCR, **the RNA must somehow serve as a template.**
   b. However, **PCR uses DNA polymerase**, and most DNA polymerases cannot use RNA as a template.
   c. Thus, RNA has to be converted into DNA first. → **unique type of polymerase known as reverse transcriptase (RT, encoded by retroviruses), which is able to synthesize DNA using an RNA template**
2) RNA is first reverse-transcribed into DNA using a primer that binds to the RNA of interest.
   a. The resulting DNA is called copy DNA (cDNA).
   b. how to choose the primers used for reverse transcription and for the PCR reaction → 2 possible strategies:
      1. If you want to analyze the **expression of a single gene**, you can use primers that will anneal specifically with the RNA of interest. → Thus, you use gene-specific primers.
      2. In many cases, however, **you would like to analyze the expression of several genes within the same sample.** → To be able to reverse-transcribe different RNAs at the same time, there are two options:
         I. First, you can **use primers that randomly bind to RNAs.** These are usually short primers of six nucleotides in a random arrangement (also called random hexamers), e.g., ACGGCA.
         II. Second, you can use a **primer that relies on the fact that mRNA molecules carry a poly(A) tail at their 3' end, to which primers containing multiple T nucleotides can bind.** These primers are called poly dT primers or oligonucleotides.

3) In the next step, the second strand is synthesized by PCR. **This double-stranded DNA molecule now serves as a template for amplification in the subsequent PCR cycles.**

a. the PCR reaction, the primers are chosen specifically for each RNA to be amplified.
b. Thus, for each analyzed RNA, a PCR reaction is set up with a forward and a reverse primer that specifically bind to the target

The workflow presented above applies to **mRNAs** (NOT miRNAs!!!) that contain a poly(A) tail!

- However, if we wish to analyze miRNAs, we face a problem:
- miRNAs are tiny → Thus, to reverse-transcribe miRNAs into DNA, the size of the miRNAs have to be increased!!!
  - This can be achieved by **adding a poly(A) tail using** the poly(A) polymerase.
  - This added poly(A) tail can then be used for universal priming of miRNAs with oligo dTs.
  - These oligo dTs contain a universal adapter, a nucleotide sequence that can later on be recognized by a primer in the PCR reaction.
  - Once the miRNA has been converted to cDNA, it can be assayed using the same approach as a conventional PCR experiment.
  - In the PCR reaction, a **miRNA-specific forward primer together with a universal reverse primer** (that binds to the adapter at the 3' end and was introduced in the RT reaction) is used

How can this system be turned into a technique to measure RNA quantities?

→ use the rate of the PCR reaction as a measure of the RNA amount present in the sample, because the rate at which the PCR product is generated is directly related to the amount of RNA present in the sample.

- Thus, at **each cycle of the PCR, the amount of PCR product is determined**.
- This is achieved by **adding chemical dyes to the PCR that fluoresce only when bound to double-stranded DNA**.
- Thus, by simply measuring the fluorescence after each cycle → The intensity of the fluorescence signal reflects the amount of RNA amplified.
- If a reaction reaches a high fluorescence intensity already after a few cycles, this indicates that the RNA amount in this sample was high.
- To be able **to actually quantify the amount of RNA present, the RNA of an internal reference gene is also amplified by RT-PCR**.
- Here, genes are chosen that are not supposed to change their expression, so **called "housekeeping genes", such as ribosomal genes**. Then, the relative amount of each RNA to this rRNA is calculated.

Advantages:

- qRT-PCR is relatively **fast to perform in the lab**,
- allows direct quantification of the RNA present in a sample.
- **As well as being sensitive and quantitative, qRT-PCR is also relatively inexpensive and flexible, making it the preferred choice for validating novel miRNAs.**

Disadvantages/ Limiatations:

- unlike in conventional qRT-PCR, to detect miRNAs, only one flanking primer can be specific to the miRNA, so care must be taken to ensure only one product is being amplified.
- The **short template length** can prove a particularly problematic issue when trying to distinguish miRNA isoforms that may only differ by a few nucleotides, because the primers may anneal to, and amplify several distinct miRNA species at once.

Ultra-deep sequencing (RNA-seq)

*RNA can, just as DNA, be sequenced, with a method called RNA-Seq, **to reveal the presence and quantify the amounts of RNA present in a sample** → it is also applicable for identifying de novo small RNAs, such as microRNAs.*

The workflow for this technique includes the isolation of RNA, the conversion form RNA into cDNA, and then sequencing of the cDNA using a next-generation-sequencing platform

why RNA into cDNA and not sequence RNA directly?

- First, RNA is **considered highly unstable**, because it is a lot more prone to **hydrolysis** due to the nucleophilic 2' hydroxyl groups.
- Second, in most cases, **you will have small amounts of RNA, and you will have to amplify it before you can use it for sequencing**.
- PCR amplification only works on DNA, and you therefore have to reverse-transcribe it to cDNA.
- Third, many sequencing techniques still rely on **DNA polymerase** for sequencing.

Can you think of a way how you could sequence RNA directly?

In theory, if you had enough material to start with, you could be using Nanopore sequencing to directly sequence RNA.

Sequencing process of small RNAs:

1. **the RNAs to be analyzed are isolated through size selection to enrich for those small RNA molecules** (e.g., with a size-exclusion gel).
2. the RNA **must be converted to double-stranded complementary DNA** (reverse transcriptase RT that uses RNA as the template!)
   - In order to convert RNA to DNA, a reverse transcription reaction as described in the qPCR assay is performed.
   - RT, like other polymerases, requires a primer annealed to either DNA or RNA to initiate polymerization.
   - In the ligation step, the RNA 3' adaptors are ligated to the 3' and 5' ends of the miRNA with T4 RNA ligase, **which act as primer binding sites during reverse transcription and PCR amplification.**
   - This primer is designed to capture miRNAs that **have a 3' hydroxyl group resulting from the enzymatic cleavage by Dicer-like enzymes**
   - Similarly, **a 5' adaptor is added** that **captures the 5' phosphate group in a second ligation step**.
   - The 5' adaptor is also designed to capture small RNAs with a 5' phosphate group, characteristic for all Dicer products, including microRNAs, rather than RNA degradation products with a 5' hydroxyl group.
   - The next step is the reverse transcription that converts the small adaptor-ligated RNAs into cDNA clones used in the sequencing reaction.
2. PCR
   - PCR is then carried out to amplify the pool of cDNA sequences
   - This creates cDNA constructs **that contain the 5' and 3' adapter sequences**. The PCR is performed using primers that anneal to the end of the adapters.
3. adding adaptors for Illumina Sequencing
   - Finally, massive parallel sequencing is done using a next-generation-sequencing platform, such as Illumina.
   - The currently available next-generation-sequencing platforms **are able to perform massive parallel sequencing,** which allows sequencing a genomic region or a miRNA molecule **multiple times (100-1000) in a short time frame**
   - **This procedure to aim for a high number of replicate reads for a DNA or RNA molecule is referred to as "deep sequencing".**

☞ Generating the complementary DNA, adding the adaptors, and amplifying the DNA for sequencing (PCR) are part of a process called "library preparation".

☞ This approach allows researchers to detect rare molecule species comprising as little as 1% of the original sample.

Advantages of RNA-Seq:

☞ Deep sequencing of RNAs **thus not only provides the sequence, but also the frequency of RNA molecules that are present at any particular time**

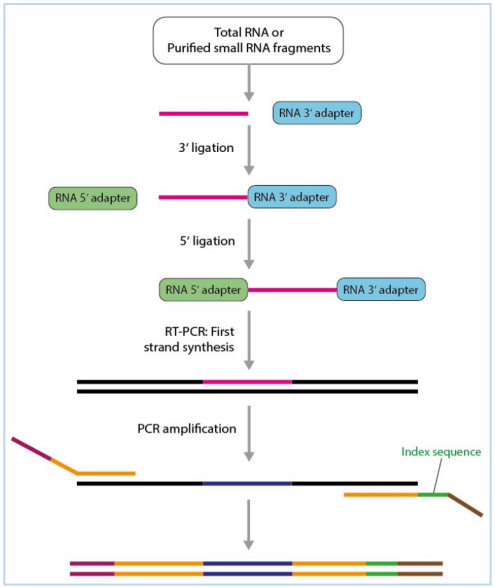
☞ it can identify **both known and novel miRNAs** !!! (unlike Northern blots, for example, where knowledge of the sequence of the miRNA is required).

Disadvantages:

☞ However, a limitation of this technique is the medium-high input quantity required (500 ng - 5 µg total RNA).

☞ **the library preparation and PCR amplification can potentially induce errors** in the sequence and some bias toward preferential sequencing of particular sRNA species to the detriment of others may occur due to preferential ligation.

☞ **the library preparation is time-consuming and expensive,** and the costs for deep sequencing are still relatively high although prices continue to be reduced steadily.



Identification of miRNA targets: Target predictions

To be able to understand the function of miRNAs, the identification of their target genes is essential.

- Generally, computational predictions using miRNA characteristics are difficult, because miRNAs are only 21-25 nucleotides long.
- However, in plants, bioinformatics has served to identify many miRNA targets, **mainly because in plants, most miRNAs are nearly perfectly complementary to their targets.**
  - Interestingly, many predicted plant miRNA targets are **transcription factors**, highlighting possible miRNA functions in regulating diverse developmental processes in plants.
- The major part of animal miRNA targets has now also been identified through computer-based predictions.
  - The computational prediction of animal miRNA targets is less straightforward, **because the miRNA-target pairings are not entirely complementary and mostly rely on the seed-complementary region.**
  - computer-assisted predictions initially had to rely on rules that were built on a few known miRNA-target interactions; as targets became experimentally confirmed and new miRNA:target pairs discovered, **these rules were considerably refined throughout time.**
    - For example, starting from three animal miRNAs, lin-4, let-7, and Bantam, the targets of which have been experimentally validated, the fly genome was searched for miRNA targets on the basis of the following three criteria:
    - i) **perfect complementarity between the target 3'-untranslated region** (UTR) **and the seed sequence of the miRNA**;
    - ii) favorable structural and thermodynamic heteroduplex formation between miRNA and its putative targets
    - iii) evolutionary conservation of miRNA target sites and of the seed in particular between closely related species
- There exist several software packages, the algorithm of which take into account many parameters including conservation across species, multiplicity of target sites within a predicted target, thermodynamic data for the free energy of the miRNA-target hybrid, and mRNA accumulation data if available in the cell type, tissue, or organ under consideration.

Identification of miRNA targets: Validations

After targets have been computationally predicted, **they have to be experimentally validated to verify that they actually constitute a real target.**

- The GFP-sensor approach is one way to monitor how the expression of the target gene depends on the activity of a certain miRNA in vivo both in animals and plants
  - **the researchers fused a GFP reporter open reading frame upstream of the predicted target's 3' UTR** (this generated a construct known as a "miRNA sensor") and **examined GFP expression in Drosophila tissues with and without overexpression of the corresponding miRNA or in animals carrying loss-of-function mutations in the miRNA of interest.**
- Another type of reporter frequently used for miRNA target validation in cultured cells uses luciferase, an enzyme that produces luminescence by converting the substrate luciferin.
  - In the so called dual-luciferase assay for miRNA-target validation, the 3' UTR of the gene suspected to be a miRNA target is cloned downstream of the translational stop codon of a luciferase gene.
  - For this purpose, **commercially available vectors carrying the luciferase gene are used**; and they usually also **contain a second type of luciferase gene (producing a protein that emits in a different light spectrum) that is devoid of the artificial miRNA target site.**
  - This second luciferase **is used for normalisation for the light emission quantification.**
  - By transfecting the vector construct into cells, a fusion mRNA between the luciferase gene and the 3' UTR of the potential miRNA target is transcribed.
  - If the predicted target is the actual miRNA target, then the endogenous miRNA will cause a reduction in activity of the miRNA-reporter luciferase compared to that of the control luciferase, indicating that the miRNA can bind to the 3' UTR target gene under investigation
  - The light emission by the luciferase reaction is measured in a luminometer, providing a convenient

way of monitoring and also precisely quantifying the effect of the miRNA on its target.

All biological assay require a negative control. Which negative control would you use in the dual luciferase assay for the activity of any given miRNA?
A luciferase mRNA carrying a scrambled miRNA target site in its 3' UTR. This should provide 100% luciferase activity

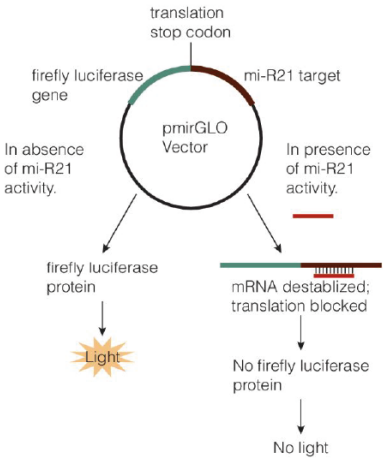


Figure 2-7 The luciferase reporter assay to validate an miRNA target. The 3′ UTR sequence of a suspected target gene is cloned downstream of a *luciferase* gene. After transfection of this construct in cells, a *luciferase*-3′ UTR fusion mRNA is made from this construct. If the target represents a real target, the presence of the respective miRNA will silence expression of this fusion mRNA, thus, luciferase activity is reduced. Note that the second *luciferase* gene used for normalization is not represented in this scheme. (Adapted from promega.com)

- These miRNA-target interactions are deposited in a widely accessible database, the miRTarBase
- miRTarBase has accumulated more than 360'000 miRNA-target interactions derived from extensive literature search
- Generally, the collected interactions are validated experimentally by reporter assays, western blot, microarray, and next-generation-sequencing experiments.
- The database also contains information about expression levels and integrates data from other databases, e.g., the cancer genome atlas (TCGA).

MI-RNA'S IN ANIMAL DEVELOPMENT AND DISEASE (OL)

Characteristics of animal miRNAs

- In animals, most miRNAs bind to the 3' UTR of their target gene
- The vast majority of animal **miRNAs are imprecisely complementary to their target, resulting in translational repression coupled to target mRNA decay**; however, there are **exceptions where animal miRNAs bind with perfect complementarity, allowing Argonaute-catalyzed endonucleolytic cleavage (slicing) of the** complementary mRNA strand (see figure 1-9).
- Many animal mRNAs (**miRNA-target transcripts) carry multiple miRNA** target sites for several distinct miRNAs in their 3' UTR. This allows **gene-expression control at the post-transcriptional level.**
- Interestingly, the location of some of these miRNAs is evolutionary conserved, and they are similarly co-expressed with their host genes in different animals. → This suggests that these miRNAs have important, evolutionarily conserved roles. Many miRNA genes are located at genomic regions that are involved in different cancer types.

But how can (novel) miRNA genes arise in a genome?

- **the only requirement for a gene encoding a functional miRNA is that it produces a transcript that can form the correct secondary structure that can be recognized by Drosha and Dicer to form mature a miRNA.**
- Because RNAs form **non-perfect folded, hairpin-like structures**, it seems more likely that **a miRNA gene might emerge more easily than a novel protein-coding gene.**
- Another favorable characteristic of miRNA function **is that they recognize short sequences within their target gene, for interaction**
  - thus, miRNA **target sites can easily be acquired and/or lost in the transcriptome of the cell.**
  - Here, remember that the core site of animal miRNA-target recognition is only six nucleotides long (the seed sequence) → a few base-pair changes in the miRNA-gene sequence are very likely to alter the repertoire of possible target genes.

The role of miRNAs in vertebrate development: lessons learned from zebrafish

One problem in studying the function of miRNAs during animal development is the lack of miRNA loss-of-function mutations.

- Many miRNAs have been missed in classical forward genetic screens, **because miRNA genes are relatively small (roughly 50 times smaller than the average protein-coding gene), and it is statistically less likely that a mutation will be induced into a small gene during mutagenesis.**

How would one go about studying the very early role of miRNAs in development if loss-of-function mutants for most miRNA genes are not available?

- organisms with defective miRNA biogenesis have been extremely useful, where the enzyme Dicer was knocked out or knocked down.
- In animals lacking Dicer function, pri-miRNAs are transcribed, but the precursors cannot be processed into mature miRNAs → miRNAs are required for normal development and are required in adults to maintain tissue integrity.

Here, we will discuss the results obtained in vertebrate models with a focus on research done in zebrafish:

- In zebrafish, loss-of-function mutations in zygotic Dicer (maternal dicer was not knocked out) lead to a developmental arrest at a rather late developmental stage (larval stage, day 8 after fertilization, indicating that the maternally provided Dicer is sufficient to reach this point)
- At this relatively late stage, all the major organs have already been formed, and the fish has no obvious abnormalities.
- Thus, researchers generated mutants lacking maternally provided Dicer in addition to the Dicer knockout → These animals are deficient in both maternally derived as well as zygotic Dicer and are therefore called MZ (for maternal-zygotic) mutants.
- Although MZ animals do not process pre-miRNAs into mature miRNAs, they only have mild defects during early development.
    - They show an intact axis formation and differentiate into different cell types, indicating that embryonic patterning is not affected
    - However, at later stages, they show some morphogenesis defects, mostly during brain formation, neural differentiation, somite formation, and heart development.
    - In the MZ-Dicer-mutant embryos, neurulation is severely affected.

Here, we will discuss two of the abnormalities that occur in Dicer-mutant animals:

- i) The formation of the brain ventricles and the formation of the midbrain-hindbrain boundary.
    - The ventricles are cavities within the brain that produce and transport the cerebrospinal fluid which bathes the central nervous system.
    - When looking at the MZ-Dicer mutants, it was found that the brain ventricles are not formed correctly.
- ii) The developing brain is subdivided by several constrictions into distinct regions, and the boundary between midbrain and hindbrain is very prominent
    - the constrictions that divide the brain into distinct regions did not form in MZ-Dicer mutants, and the midbrain-hindbrain boundary was not developed
    - one specific family of miRNAs, miR-430, were highly expressed between fertilization and the first 48 h of development.
    - They therefore found out that injection of a pre-processed, mature miR-430 (a duplex miRNA that resembles the Dicer-processed form of a miRNA) could rescue the brain morphogenesis defects in the Dicer mutants.
    - Amazingly, a single miRNA is sufficient to restore near-wild-type brain development, indicating that miR-430 alone is responsible for a major switch in brain development.
    - This also indicates that miR-430 most likely controls many different target-mRNA transcripts that are involved in brain morphogenesis.

☞ These findings in zebrafish indicate that miRNAs are not essential for cell-fate determination and early patterning during the first 48 h of development, but that they are essential for later steps of embryogenesis.

☞ interestingly, the expression of miRNAs is highly tissue-specific in fully developed tissues

☞ Even within an organ, different miRNAs are specifically expressed within distinct regions.

☞ These results suggest that the role of miRNA during development is not only in tissue-fate establishment, but also (and in this example, mainly) in differentiation or maintenance of tissue identity and integrity.

☞ another striking phenomenon: many miRNAs that are expressed during development stay highly expressed in the adult. → This again suggests that many miRNAs have specific roles other than developmental patterning and that they act as guardians of established cell fates.

☞ The failure to maintain an acquired cell fate is one of the causes of cancer.

miRNA in human disease

miRNAs play a major role in a wide range of developmental processes including cell proliferation, cell cycle, cell

differentiation, metabolism, apoptosis, developmental timing, neuronal cell fate, neuronal gene expression, brain morphogenesis, muscle differentiation, and stem-cell division

Remember that animal miRNAs ***can repress the gene translation of hundreds of their targets****.* → *Thus, it is easy to imagine that* ***changing the miRNA composition*** *and action* ***might be leading to disease.***

Mechanisms that change miRNA action

ways how miRNA action can be changed:

First, let's consider changes that affect **a small number of miRNAs or their targets**:

- Here, a miRNA **may acquire a mutation resulting in loss of its function.**
- there may be **a gain-of-function mutation** in a miRNA,
- and **overexpression** by amplification of the **miRNA locus** may work like overexpression of an oncogene, resulting in cancer.
- On the other hand, the binding **sites of miRNAs** within their target mRNAs might be **mutated** such that the miRNA cannot interact with its target any longer
- Or, a gene **acquire a new and undesired miRNA target sequence that results in its aberrant silencing**.
  - Here, remember that the seed sequence is only **six nucleotides long**, and that many miRNAs are expressed simultaneously in the same tissue.
  - If an important gen is only one mutation away from becoming a target for one of these miRNAs, a single mutation in the gene will result in an undesired reduction of this gene's activity, which may be the cause of disease.
  - global changes in miRNA and/or target accumulation can occur. → **mutations could render the miRNA biogenesis** pathway non-functional or lead to an increase in miRNA biogenesis.

☞ Thus, **miRNA action may be altered both by mutations or by gene amplification**.

how single point mutations can affect miRNA-mRNA interactions and lead to a different miRNA response eventually causing disease:

different steps of miRNA biogenesis and processing necessary for their sequence-specific gene silencing

1) miRNA genes are transcribed by RNA polymerase II to produce a 500–3'000-nucleotide transcript, called the pri-miRNA
2) which is then cropped to form a pre-miRNA hairpin of 60-100 nucleotides by a multi-protein complex including Drosha.
3) **This double-stranded hairpin structure** is exported from the nucleus to the cytoplasm by Exportin 5.
4) Finally, the pre-miRNA is cleaved by Dicer to produce a double-stranded miRNA of approximately 20 nucleotides in length.
5) The guide strand is incorporated into the RNA-inducing silencing complex (RISC), while the passenger strand is discarded.
6) The single-stranded miRNA that was incorporated into RISC guides it to the 3'-UTR mRNA sequence of the target to facilitate translational repression coupled to mRNA decay, or, in rare cases, mRNA endonucleolytic cleavage

Can you think of a specific example for an alteration that a) changes miRNA function on a global level or b) affects a small number of miRNAs?

- Global changes: loss-of function mutations in Pol-II, Drosha, Dicer, Exportin-5, AGO; duplication of a dicergene, gain-of-function mutation in Drosha or Dicer that enhances enzymatic activity
- Specific changes: duplication of a miRNA gene, mutation of a miRNA gene that disturbs hairpin structure formation, mutation in seed sequence disturbs interaction with target, mutation of miRNA target sites

Mutations affecting miRNA biogenesis

SNPs in pri-mRNA sequences;

possibilities how SNPs can affect miRNA functions:

- on the level of transcription of the primary transcript,
- on the level of miRNA biogenesis,
- or by affecting miRNA-mRNA interactions.

how SNPs in some miRNA genes were shown to affect miRNA biogenesis or processing:

- The first evidence that mutations (SNPs) in miRNA genes can have a functional effect came from studies on

chronic **lymphocytic leukemia**, where a mutation **in mir-16-1 led to low levels of miR-16-1 expression** → change in the sequence of the pri-miR-16-1 (see figure 3-4A) → this **SNP changes the processing and the levels of mature miRNA.**

- **SNPs in the pri regions (sequences present only in the pri-miRNAs) lead to decreased mature miRNA levels, associated with a higher risk to develop cancer.**

- In figure 3-4B, the case for miR-196a-2 is depicted, **in which a SNP has been found to affect the 3' sequence of the passenger strand.**
  - **This SNP is suspected to affect miRNA maturation,** since the presence of this SNP reduces the levels of mature miR-196a-2.
- SNPs can also lead to an up-regulation of miRNA processing and therefore an increase in mature miRNAs!

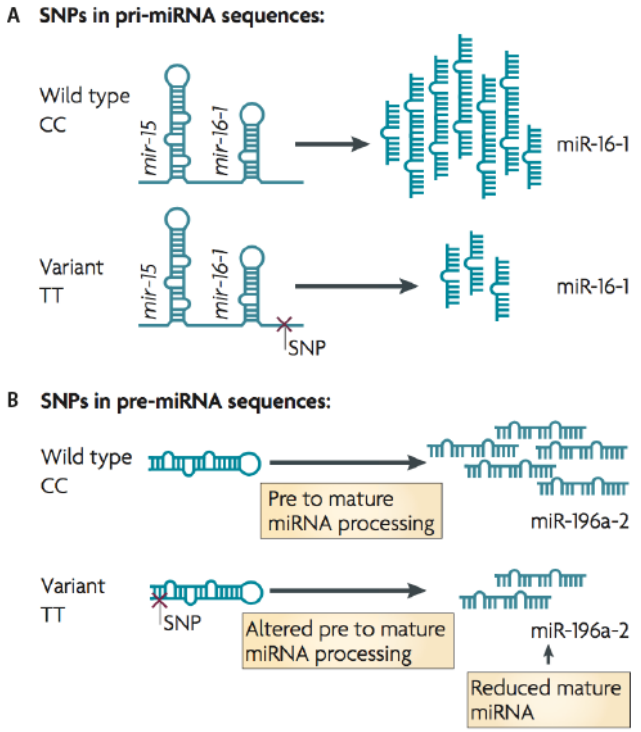


Figure 3-4 SNPs affecting miRNA processing lead to changes in mature miRNA levels. SNPs can occur at different positions in pri-miRNA **(A)** and pre-miRNA sequences **(B)**. Such SNPs can lead to either an increase or decrease in processing (here, examples for a decrease are depicted). (Adapted from B.M. Ryan *et al., Nat. Rev. Cancer*, 2010)

Mutations affecting miRNA-target interactions

mutations can also affect the binding of miRNAs to their targets.

This can either be due to mutations in the seed sequence of the miRNA (1) or due to mutations in the target genes (2), so that the miRNA can't bind to the target anymore

- F

(1): mutations/SNP in the miRNA encoding sequence

- For the miRNA miR-146a, a SNP was identified that **affects the seed sequence in the passenger strand.**
  - Interestingly, individuals that are **heterozygous** for this mutation showed **a greater risk of developing cancer** than homozygous individuals.
  - Why would a heterozygous situation pose a greater risk to develop the cancer?
    - This can be explained **by the effect of this mutation for the action of miR-146a:** the SNP falls within the seed region of the passenger strand, thus creating a new target binding site.
    - In heterozygous individuals, there are now three different mature miRNAs: one from the guide strand and two different ones from the passenger strand, the latter two differing only in the SNP in the seed sequence.
      - Thus, each of these three mature miRNAs can **target different mRNAs.**
    - In homozygous individuals, however, only two species of mature miR-146a are produced (the original 5p strand and the 3p strand containing the SNP);
      - **thus, the repertoire of putative target mRNAs is smaller** than in heterozygous individuals.
    - In patients with miR-146a-associated tumors, researchers have found that somatic mutations exist that change miR-146a-homozygous cells into heterozygous ones.

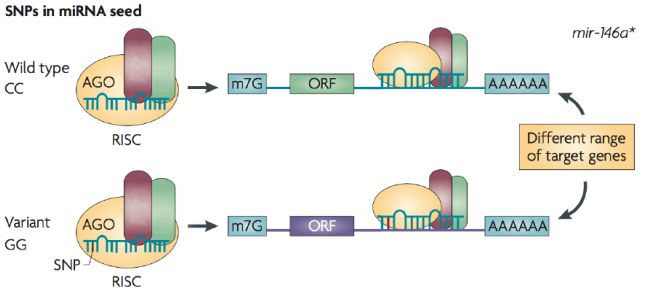


Figure 3-5 Single nucleotide polymorphisms (SNPs) in mature microRNAs (miRNAs) within the seed sequence can strengthen or reduce binding between the miRNA and its mRNA target. Moreover, such SNPs can create or destroy target-binding sites, as is the case for *mir-146a** (* denotes the passenger strand). AGO-associated proteins are shown in red and green. ***Abbreviations***: AA, poly(A) tail; m7G, 7-methylguanosine cap. (Adapted from B.M Ryan *et al., Nat. Rev. Cancer*, 2010)

(2): mutations/SNPs in the targeted gene, so that miRNAs binding site is altered

- SNPs can also occur **in the miRNA binding sites within target genes (snp in target gene statt in miRNA gene locus).**

- Analogous to mutations in the seed sequence of miRNAs, SNPs in the 3' UTR of a gene may create or destroy a miRNA binding site.
- found in many cancer types. → let-7binds to the 3' UTR of KRAS, a proto-oncogene that is mutated in many cancers.
  - Over-activation of KRAS leads to several malignancies.
  - Interestingly, a SNP was found in one of the let-7 complementarity sites within the 3' UTR of KRAS
  - Thus, by destroying the target site for let-7, KRAS expression was up-regulated (because it was no longer repressed by let-7), leading to increased cell growth and proliferation, two hallmarks of cancer cells
  - On the other hand, SNPs can also create a target site for a miRNA, a scenario that was shown for the 3' UTR of CD86

Mutations in the miRNA pathway affect animal breeding

how SNPs have been shown to affect miRNA function in three different areas: Animal breeding, inheritable diseases and the development of cancer

→ mutations affecting miRNA can function during natural selection in evolution, **these mutations also function in artificial selection in animal breeding**.

how a miRNA-target-specific mutation ***resulted in an advantageous outcome*** *in* ***sheep breeding by a*** *SNP that occurs in the target gene:*

- A point mutation that creates an **illegitimate mir-1 target site in the 3' UTR of the myostatin** gene was found to **inhibit myostatin expression**, which contributes to **increased muscle mass of sheep of the Texel breed.**
- Texel is a breed that is now a popular lean-meat sheep all over the world, because it is heavily muscled.
- **Myostatin acts to inhibit muscle-cell growth** and differentiation (myogenesis).
- **Thus, inhibition of myostatin expression by the mir-1 miRNA through a mutation as it occurred in the Texel breed increases muscle mass.**

☞ (miRNA kann nun neuerdings durch Mutation auch myostatin binden und inhibieren, da myostatin so verändert durch mutation/SNP dass es ein target für miRNA darstellt)

☞ Keep in mind the different ways how miRNA action can be affected by mutations: **through alterations in miRNA biogenesis, through loss or gain of a seed sequence, or through loss or gain of a target sequence.**

Mutations in the miRNA pathway are involved in cancer development

Taking into account that 30% of human genes are regulated by miRNAs, it is not surprising that miRNAs have been shown to be implicated in many different human diseases.
→ *miRNAs can act as oncogenes and tumor suppressors and are involved in a variety of pathways deregulated in cancer.*

Genes that act as negative regulators of growth and differentiation are referred to as **tumor-suppressor genes**, while those **acting to enhance growth and differentiation are called oncogenes**.

Mutations in these protein-coding genes are selected for in cancer cells, and there are key proteins (e.g., KRAS or PTEN) that are mis-regulated in various cancer types.
→ the case of KRAS, where a mutation in its 3'-UTR target site for let-7 leads to an up-regulation of KRAS, resulting in increased cell growth and proliferation, hallmarks of cancer cells → **Recently, miRNAs were also shown to act as tumor suppressors or oncogenes → Cancer-associated miRNAs are also called "oncomirs", and they can be used in diagnosis and treatment of cancer.**

50% of annotated human miRNA genes are located in areas of the genome that are known as **fragile sites**, sites within a chromosome that are susceptible for chromosome breakage, amplification and fusion with other chromosomes.

Figure 3-8 summarizes the current view **on how miRNAs can act as tumor suppressors or oncogenes in cancer progression.**

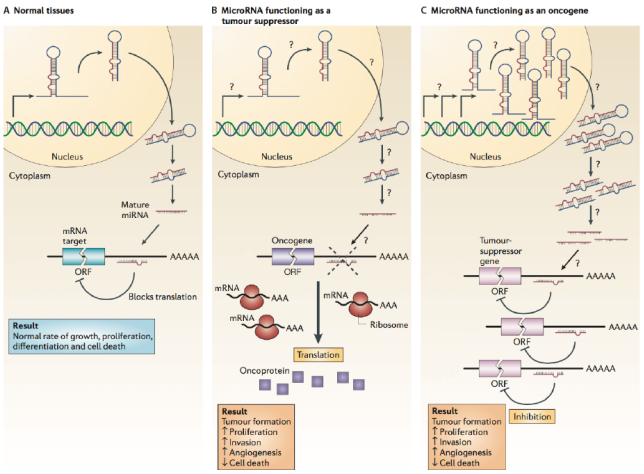
miRNAS act as tumor supressors:

- If the loss of a miRNA function leads to tumor formation, **this miRNA is supposed to act as a tumor suppressor**.
  - **The loss or reduction of a miRNA** can be due to mutations that affect the **biogenesis of miRNAs** as it was described for the case of mir-16-1 above

- **miR-15 and mir-16-1 negatively regulate the expression of an oncogene**, the anti-apoptotic gene BCL2; thus, reducing the amount of miR-15/miR-16-1 up-regulates the expression of BCL2, resulting in cell-death inhibition.
- **Hence in this case, the normal function of mir-15/mir-16-1 is to suppress tumor formation**, i.e., mir-15/mir-16-1 act as tumor suppressor genes (see figure 3-8B).

miRNAS act as oncogenes:

- Here, alterations/mutations **that enhance the expression or processing of miRNAs lead to increased amounts of miRNA that inhibit the expression of a target tumor suppressor gene**, leading to cancer progression (see figure 3-8C).



What could be the alterations that lead to increased levels of mature miRNA?

The levels of mature miRNA can be increased due to amplification of the *miRNA* gene (e.g., through duplications), due to a mutation in the promoter of the *miRNA* gene that renders this promoter constitutively active or due to increased efficiency of miRNA processing (as we have seen in figure 3-4). Alternatively, through translocations, the *miRNA* gene could be placed under the control of a different promoter, which might lead to a stronger expression or even an expression in a tissue or cell type where this miRNA is normally not expressed

There are several examples where miRNAs can act both as tumor suppressors as well as oncogenes.

- This reflects the fact **that a single miRNA can control many unrelated targets, such that the same miRNA can control opposing cellular processes**.
- This ability seems **to depend on in which cell types the miRNA is expressed and** which target mRNAs are present in this cell type.
- how this works in the case of miR-125b, which has opposite roles (oncogene and tumor suppressor) in different cancer types or cell lines:
  - miR-125b targets a number of genes, including transcription factors, growth factors, and members of the BCL2 family and plays important roles in cell differentiation, proliferation, and cell death.
  - **miR-125b functions as a tumor suppressor in ovarian**, thyroid, or breast cells, **it targets oncogenes that control cell proliferation or cell-cycle progression to reduce tumor formation.**
    - Accordingly, **miR-125b was found to be downregulated** in ovarian, breast, or thyroid cancer (see figure 3-9, right).
  - In other cellular contexts, however, **miR-125b acts as an oncogene to promote tumor formation by increasing cell proliferation and inhibiting apoptosis.**
    - Thus, in cancers of the prostate or the thyroid as wells as in glioblastoma and neuroblastoma, **the expression of miR-125b is often upregulated**

Predicting tumor origins by miRNA profiling

miRNAs are expressed with high tissue specificity, and they are important regulators of cell proliferation and their malfunction is involved in tumor formation.

→ **Consequently, it is possible to use the differential expression profiles of miRNAs in cells to aid in diagnosing and classifying cancers.**

→ **This is useful, because it is often difficult to trace from which tissue a tumor arises, either because it metastasized or because the tumor cells are undifferentiated.**

- Researchers noticed that within the clusters, the tumors were grouped according to their embryonic lineage,

e.g., tumors of endothelial origin such as colon or liver were clustered together.

- Additionally, **that miRNA signatures within tumors reflect their developmental history is also compatible with the evidence that miRNAs direct tissue-specific developmental functions**

miRNA circulating in plasma or serum can be used as biomarkers

- Current techniques for cancer diagnosis commonly involve a **biopsy of the cancer tissue**
- have been made to search for biomarkers in human fluids, such as blood plasma or urine.
  - biomarkers are substances that indicate the presence or severity of a disease.
  - miRNAs that circulate in the blood plasma or serum and can be utilized as biomarkers for disease, because their level is significantly different in patients compared to healthy people.

There are several hypotheses about the origin of miRNAs in the blood serum → One theory is that they are present in apoptotic bodies, cell fragments produced during cell death that are taken up by phagocytes.

Summary:

In this lesson, we have seen that **miRNAs are regulators of cell or tissue differentiation.**

→ Undifferentiated (or poorly differentiated) cells do not require miRNAs to survive, as we have seen from the example of early zebrafish development.

→ miRNAs show a highly tissue-specific expression at later stages when most of the cell types have been formed, indicating that miRNAs function to maintain an already established cell fate.

→ A de-regulation of miRNAs often results in a failure to maintain cell fates, and this failure often results in cancer.

→ In line with this, many types of human cancer cells have reduced miRNA expression compared to their fully differentiated tissue of origin.

→ Due to their tissue-specific expression, miRNA profiling can predict the cancer tissue origin, an information relevant for cancer treatment.

MI-RNA'S IN PLANTS (OL)

- small non-coding RNAs (sRNAs) as important regulators of cellular functions, focusing on miRNAs in animals.
- As their animal counterparts, plant miRNAs are 20- to 24-nucleotide RNAs that regulate the expression of their targets post-transcriptionally.

several differences btw plant and animal miRNAs:

- the location of miRNA genes in the genome
- the length of pre-miRNAs
- the degree of sequence complementarity
- the resulting modes of action.

Plant-miRNA biogenesis

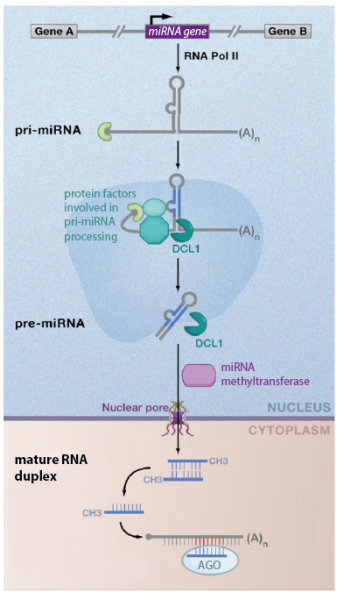


Figure 4-1 Biogenesis of miRNAs in plants. miRNA genes, which are usually located intergenically, are transcribed by RNA Polymerase II (RNA Pol II) and processed into mature miRNA duplexes by Dicer-like 1 (DCL1) and other processing factors in the nucleus. They are exported into the cytoplasm, methylated to prevent degradation, and the guide strand binds to Argonaute (AGO) forming the RISC. In plants, 10 Argonaute proteins are predicted to exist. (Adapted from O. Voinnet, *Cell*, 2009)

mi-RNA Herstellung:

1) **Like in animals, plant miRNAs are transcribed from genomic DNA:**
2) **while in animals, many miRNA genes are found within introns or even exons** → miRNA genes of plants are in their vast majority intergenic, i.e., located in non-coding regions between the protein-coding genes
3) Like animal miRNAgenes, **they are transcribed by DNA-dependent RNA-polymerase II (**the same polymerase that also transcribes protein-coding genes into mRNA) producing **a primary miRNA (pri-miRNA) that is capped and poly-adenylated (as many pri-miRNAs in animals)**.
4) In the nucleus, the pri-miRNA undergoes two processing steps catalyzed by the RNase III Dicer-like 1 (DCL1),
   1) with the first one being the formation of the stem-loop precursor (pre-miRNA).
   2) The second excision by DCL1 forms the mature miRNA duplex.

3) Plant pre-miRNAs are much more variable in size than their **animal counterparts (usually about 70 nt long)**.
4) This is different to miRNA biogenesis in animals, where nuclear processing of the pri-to-pre-miRNA is mediated by Drosha and processing into the mature miRNA duplex is catalyzed by Dicer in the cytoplasm.

5) In plants, the mature miRNA duplex is exported into the cytoplasm and both strands are 2'-O-methylated at the 3' end by HEN1 to protect the duplex from oligo-uridylation and degradation by exonucleases.
6) One of the two miRNA strands, **the so-called guide-miRNA strand, which contains the sequence complementary to the target sequence, is then incorporated into Argonaute (AGO) proteins.**
   1) In principle, both strands can contain the guide sequence necessary for target recognition.
   2) The mechanism by which the correct guide-miRNA strand is selected is not yet formally resolved in plants but occurs most likely as for animal miRNAs (see lesson 1).
7) AGO loaded with the guide-miRNA forms - together with other proteins - the RNA-induced silencing complex (RISC) to mediate target-gene specific silencing.
   1) AGO1 is the main effector of miRNAs
   2) Some miRNAs are also effected by AGO10, which is in the same genetic clade as AGO1.

Slicing or translational inhibition - two different modes of silencing

how miRNA function:

Perhaps the most prominent difference between animal and plant miRNAs is the degree of complementarity to their targets:

Complementary miRNA to target mRNA:

- Typically, animal miRNAs are thought to regulate their targets through imperfect complementarity to sites found in 3'-untranslated regions (3' UTR), which mostly leads to translational inhibition and only rarely to slicing, i.e., endonucleolytic cleavage of the target mRNA.
- Plant miRNAs, in contrast, largely regulate their targets by binding to highly complementary target sites in coding regions (see figure 4-2).
  - This has two consequences for plant miRNAs:
    - i) they often function **through mRNA degradation**, clearing the target mRNA **irreversibly**
    - ii) they are predicted to have **a much more limited number of mRNA targets than their animal counterparts**, because they require near-perfect matching with their target.
  - An advantage **is that computer-based target identification is very straightforward**, because it is much less prone to false positives than in animals, in which conservation of the seed-complementary sequence **(which is just six nucleotides long)** is the main criterion.

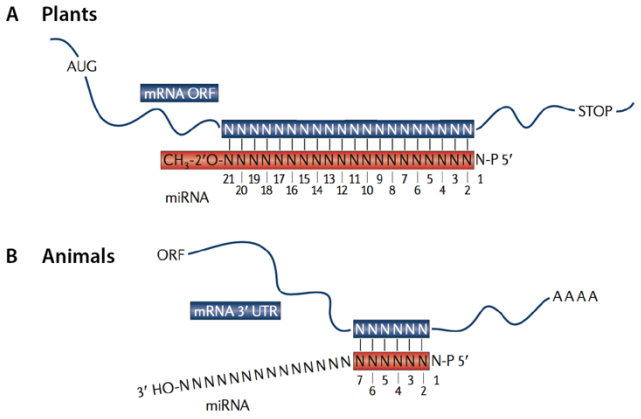


Figure 4-2 Difference in sequence complementarity between plant and animal miRNAs and their targets. (A) Plant miRNAs mostly bind to target sites in coding regions with high complementarity, often leading to slicing (RNA degradation). **(B)** Animal miRNAs typically bind to sites in 3' UTRs of their targets with imperfect complementarity, resulting in translational inhibition. (Adapted from E. Huntzinger and E. Izaurralde, *Nat. Rev. Genet.*, 2011)

how target cleavage is achieved mechanistically:

1) After binding of the RISC to a target, AGO1 can cleave the target mRNA between nucleotide 10 and 11, opposite the guide-miRNA strand.
2) The mRNA fragments resulting from the endo-nucleolytic cleavage are degraded from the newly generated 3' and 5' ends (see figure 4-3B).

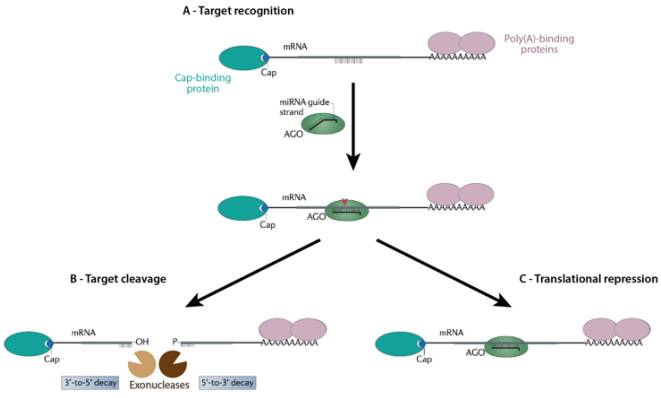


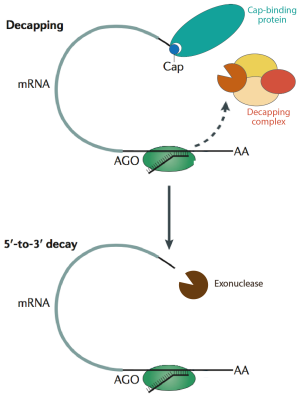
Figure 4-3 Mechanism of miRNA-mediated gene silencing in plants. (A) Plant miRNAs that are bound to Argonaute (AGO) recognize mRNA targets containing fully or nearly complementary binding sites, which are predominantly located in the ORF. mRNA ends are protected by a 5' cap and a 3' poly(A) tail, which are recognized by cytoplasmic proteins (in blue and pink) that facilitate translation. AGO can cleave the mRNA in the base-paired region (between nucleotides 10 and 11, opposite the miRNA strand, at the red arrowhead). **(B)** The mRNA fragments resulting from the endonucleolytic cleavage are degraded from the newly generated 3' and 5' ends. **(C)** Alternatively, cleavage by AGO is prevented and the mRNA target is translationally repressed by an unknown mechanism. (Adapted from E. Huntzinger and E. Izaurralde, *Nat. Rev. Genet.*, 2011)

Why are the cleaved mRNA fragments not degraded from the original 5' and 3' end?

During mRNA processing, the 5'-capping structure and 3'-poly(A) tail are added to the pre-mRNA to prevent exonucleolytic degradation. When AGO cleaves the target mRNA in the middle, the ends resulting from this cleavage are not protected anymore from RNA nucleases.

Can you think of another mechanism by which target mRNA could be degraded?

As discussed in lesson 1, animals deploy another mechanism of target-mRNA degradation as a consequence of RISC binding: Most animal miRNAs do not promote slicing due to a conserved central mismatch that hinders catalysis by AGO. They thus promote translational repression which is linked to mRNA decay. Hence, the RISC complex activates a decapping complex, which removes the 5'-capping structure of mRNAs, rendering it accessible for 5'-to-3' exonucleases (see figure and see also lesson 1).



Adapted from E. Huntzinger and E. Izaurralde, *Nat. Rev. Genet.*, 2011

Will the target mRNA be sliced or translationally inhibited?
→ the *MAD mutants experiment*

However, despite high sequence complementarity to their targets, plant miRNAs were found to often function by translational inhibition:

forward-genetics screen in Arabidopsis: looked for mutants that have defects in miRNA-induced silencing processes:

- used a transgenic strain that constitutively expressed (GFP) reporter construct, where **GFP was fused upstream of the sequence containing a certain miRNA (miR-171) target site** (figure 4-4A).

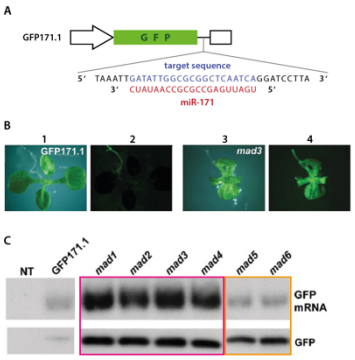


Figure 4-4 Discovery of miRNA-action deficient mutants (*mad* mutants). **(A)** Schematic representation of the construct GFP171.1 encoding GFP containing a miR-171 target site (blue). The miR-171 target site is perfectly complementary to miR-171 (red). **(B)** Images 1 and 2 show a transgenic plant (GFP171.1) with normal miRNA-mediated silencing (1: transmitted light. 2: blue light excitation). Here, GFP mRNA is silenced by miR-171 and no fluorescence is visible under blue light. Images 3 and 4 show plants mutant for the miRNA *mad3* after 15 days of growth (3: transmitted light. 4: blue light excitation). Since GFP expression is not silenced, the plant fluoresces under blue light. **(C)** Gel electrophoresis analysis of GFP mRNA (top) and GFP protein (bottom) in non-transgenic plants (control, NT), GFP171.1 (transgenic plant before mutagenesis), and *mad* mutants *1* to *6*. The mutants *mad1* to *4* (pink) are deficient in silencing, while *mad5* and *6* (orange) are defective in silencing by translational inhibition. (Adapted from P. Brodersen *et al., Science*, 2008)

- Upon expression of the miR-171 GFP sensor construct, a fusion mRNA between the GFP gene and the miRNA target sequence is made.
- Because the researchers knew miR-171 **is expressed in virtually all parts of the plants, they predicted that these plants would, under normal conditions, not show green fluorescence, because the presence of the miRNA target site in the fusion mRNA would prevent translation of this RNA due to miR-171 binding and silencing.**
- Now, after mutagenizing **the researchers could identify Arabidopsis mutants that exhibited bright-green fluorescence, indicating either a defect in miR-171 biogenesis or activity** (see figure 4-4B).
  - In some of the **mutants** they identified, the lack of miRNA **function could be traced to defective miRNA biogenesis** (e.g., mutation in dcl1) or stability (e.g., mutation in HEN1) since analysis by gel electrophoresis showed **that little mature miRNAs of 21 to 24 nucleotides in length, including miR-171, accumulated in these mutants.**
  - However, the researchers also identified six mutants, in which i) mature miRNAs were produced and ii) green fluorescence could be observed indicating that these plants were miRNA-activity deficient (mad mutants).

Characterization of these mad mutants led to a surprising result:

*While in all of them, GFP protein was strongly expressed (see figure 4-4C, lower panel), the **GFP mRNA was only detectable in four of the mutants** (see figure 4-4C, upper panel).*

- From this observation, we can conclude that mutants mad1 to mad4 were unable to degrade GFP mRNA, explaining the fluorescence.
- However, in mad5 and mad6, miRNA-guided mRNA degradation appeared to occur normally as indicated by the fact that the decrease in GFP mRNA was comparable to that seen in the parental transgenic line before mutagenesis (see figure 4-4C, upper panel);
  - **both mutants failed to silence the GFP-mRNA fusion** at the translational level, since the GFP protein could still be detected in these mutants (see figure 4-4C, lower panel).

This indicates that silencing of GFP expression is due to two distinct mechanisms, which can both be mediated by the same miRNA (here miR-171):

- i) mRNA slicing and degradation leading to reduced mRNA levels
- ii) inhibition of protein production from the remaining mRNA, which was not sliced and degraded.
    - Apparently, mad5 and mad6 are defective in the second process.

Summary

- First, in plants, **miRNA-mediated silencing functions by target-mRNA slicing and degradation as well as translational inhibition.**
- Second, one single miRNA (e.g., miR-171) **can result in a combination of the two types of silencing.**
    - true for most plant miRNAs, but with varying miRNA:target pair
- Third, since the miRNA used in this study showed **perfect complementarity to its target and triggered both silencing mechanisms**, one cannot infer from the sequence complementarity, whether the target will be sliced or translationally inhibited.
- Moreover, neither the position of binding to the target (coding region or 5'/3' UTR) nor the degree of pairing appears to be predictive of the prevalence of one silencing process over the other.
- There is recent evidence **that translational inhibition is promoted by protein factors that interact with AGO**

Which mode of action is especially well suited for the regulation of which cellular process?

- The slicing and degradation of mRNA targets is **irreversible** and could provide a switch that is required for **permanent decisions such as cell fate acquisition during embryogenesis or in adult stem cell niches found in plant meristems.**
    - One argument in favor of this hypothesis is that mutations perturbing the silencing process by mRNA slicing and degradation have severe consequences for plant **development as do mutations in factors required for miRNA biogenesis** (e.g., dcl1 mutant).
- Translational inhibition, on the other hand, is **reversible**
    - well suited to coordinate stress-responsive gene expression.
    - There are several examples of miRNAs that inhibit negative regulators of stress responses by binding to their mRNA.
    - **Adaptation to stress is therefore achieved by repressing the stress-response repressor.**
    - **Reversibility of this process would ensure that the translation of the negative regulators resumes immediately after the stress is gone.**
    - **This would reduce the fitness costs** of a prolonged stress-response activation.
- Ultimately, processes that are regulated by only target degradation or translational inhibition are expected to be rare, since the flexibility that arises from combining both mechanisms is very advantageous for an organism.

Modulating plant miRNA activity

How is it possible that, despite all cells having the same genetic makeup, different cell types show different activities of genes and their products?

→ *The activity of various miRNAs must therefore be modulated differentially in every cell!*

→ *In principle, we can expect this to occur on three different levels: Transcription of miRNAgenes, miRNA processing, and activity of mature miRNAs as part of the RISC complex.*

Regulation of *plant-miRNA* gene transcription

In contrast to some of their animal counterparts, plant miRNA genes usually are independent transcription units with their own promoters and regulatory elements!

- Depending on the **combination of transcription factors** that bind to the regulatory regions of miRNA genes, **highly specific expression pattern in different cell types can thus be achieved.**
- example: **miR-166** is an important regulator in plant development. **It is encoded by seven distinct genes (mir166a-g) producing different miRNA primary transcripts (pri-mRNAs)**
    - that, however, all produce nearly identical mature miRNAs after processing, only **differing by one or two nucleotides.**
    - studies showed that not all mir166 paralogs are expressed equally in different parts of a plant.

- high degree of cell specificity is likely due to different regulatory elements in the promoters of the different paralogs.

Can you imagine how an organism benefits from having multiple genes that encode the same, or very closely related, gene product, as in the case of the mir166 gene family?
Most of the evolutionarily conserved plant miRNAs are encoded by gene families. This provides redundancy and protection from losing the gene function through individual mutations. miR-166 is involved in plant development and its loss would have severe consequences for the organism. It is therefore beneficial for the plant to have multiple genes encoding the miRNA. A high degree of redundancy (as in the mir166family with its 5 paralogs) would eventually enable some level of diversification among the miRNA genes, possibly granting the acquisition of new targets by diversified mature miRNAs only divergent by a few nucleotides; being involved in new genetic regulations, the paralogs would eventually be possibly regulated by distinct transcription factors.

Which part of a miRNA would the most sensitive to mutations enabling a change in or acquisition of target sequences in mRNAs?
The six-nucleotide seed sequence used for initial scanning by AGO and target recognition, which nucleates the miRNA binding to its target sequence

Regulating miRNA processing and activity

- *Dicer-like 1 (DCL1) is the main enzyme that processes pre-miRNAs to yield mature miRNAs in plants.*
- *AGO, which incorporates guide miRNA strands and carries out the silencing reactions, is the key miRNA effector protein.*

→ regulation of miRNA processing and activity involves regulation of DCL1 and AGO → is mediated by two miRNAs, miR-162 and miR-168, that specifically target the DCL1 and AGO mRNA → provide an important feedback regulation

Another possibility to modulate miRNA action temporarily involves ***their binding to inert targets, which makes them unavailable for silencing their cognate targets.***

This mechanism is used, for example, as a ***response to*** phosphate starvation:

- If a plant has too little phosphate, a complex **stress response is turned on that promotes phosphate uptake**, but also closely monitors that the phosphate level does not rise above toxic levels.
- Therefore, positive as well as negative regulators of phosphate uptake are expressed as a response to phosphate starvation.
- The protein PHO2 is such a negative regulator.
  - PHO2 is inhibited by a certain miRNA (miR-399), **which is expressed in response to phosphate starvation until phosphate levels have risen again.**
  - The activity of **miR-399 itself is regulated via another mechanism:**
    - IPS1, a non-protein-coding RNA is also transcribed as a response to low phosphate.
    - It has a target site for miR-399, → this target site displays a central mismatch that **prevents the slicing reaction and therefore sequesters miR-399-bound RISC complexes for as long as IPS1 is transcribed in the cell** (see figure 4-7).
    - IPS1 therefore modulates miR-399 activity temporarily through so-called "target mimicry".
    - The ways of transient miRNA regulation described here can also be used in genetic studies by engineering RNAs that mimic a target and temporarily knock down miRNAs.

Plant miRNA-directed regulation

how miRNA in turn regulate cellular processes.

keep in mind the mechanisms by which different levels of miRNA expression and activity can be modulated, because they are the basis of the regulatory circuits we will discuss.

In multicellular organisms such as plants, ***it is often necessary that certain gene products are only active in a specific tissue but not in another*** *(spatial restriction)* ***or at a specific time (temporal regulation).***

How can such patterns be achieved by miRNA-mediated silencing?

☞ *Keep in mind that in both scenarios, silencing can be achieved by either cleavage of target mRNA,*

translational inhibition, or a combination of the two mechanisms.

- ☞ *Because plant miRNA usually show very high sequence complementarity to their targets, silencing by both translational inhibition and slicing occurs*
- ☞ *miRNAs in plants is regulated on different levels: the transcriptional level, on the level of biogenesis, and regulation of activity.*
- ☞ *plant miRNAs can also mediate the spatial restriction or temporal regulation of target expression.*

let's consider two hypothetical domains - A and B - within a developing plant organ:

In spatial restriction, **we want a certain protein target to be translated in domain A, but not in domain B.**

- An easy way to achieve this is **to express a miRNA that specifically binds the target mRNA and silences it in domain B, but not in domain A** (figure 4-8 A)
- In this scenario, if the organism was unable to produce the miRNA, the target protein would be present in both domains.
- Let's look at one example of how spatial restriction is achieved by miRNA-mediated silencing in Arabidopsis:
  - The protein CUP-SHAPED COTYLEDON2 (CUC2) influences the serration of leaves and how the sepals of flowers are joined during development.
  - **CUC2 is regulated by the miRNA miR-164.**
  - In wild-type Arabidopsis flowers, **miR-164 is expressed in the central part**
    - → normally, miR-164 constrains CUC2 expression to the outer part of the flower.
  - In plants **containing a mutation in cuc2** that is not **responsive to miR-164 (miRNA kann nicht mehr an mRNA von mutierte Sequenz binden) the CUC2 protein is also now present in the central part of the flower** (4-9B).
    - In miR-164-resistant cuc2 mutants, however, CUC2 is produced even in the presence of miR-164.
  - → You can see from this example **that the resistance of a target to its miRNA can result in the same phenotype as a loss-of-function mutation** of the miRNA itself

In temporal regulation, we look at only **one domain, in which a certain target protein is not produced at one point in time, but is actively translated at another:**

- To achieve this, **a gradient of miRNA expression generates an opposing gradient of its target over time** (see figure 4-8B). Although target transcription stays constant, decreasing miRNA expression results in increased target translation, and vice versa. In the absence of miRNA expression, the target would be translated all the time.

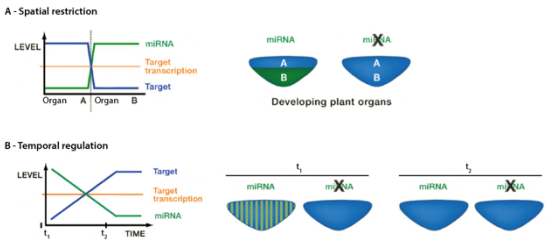


Figure 4-8 Diagram of regulatory circuits mediated by plant miRNA. (A) Spatial restriction of target (blue) accumulation in two developing plant organs A and B. Although target transcription is the same in both organs, the target accumulates in organ A where the miRNA is not present and is depleted from organ B where miRNA is expressed. In absence of miRNA, the target is translated in both domains. **(B)** In temporal regulation, a difference in target accumulation at two points in time can be achieved by generating a gradient of miRNA expression that results in an opposing gradient of its target. In the absence of miRNA, the target is translated at both time points. (Adapted from O. Voinnet, *Cell*, 2009)

SI-RNA-MEDIATED GENE SILENCING IN PLANTS (OL)

The stimulus that triggers RNAi is the presence of a long double-stranded RNA (dsRNS) in the cell, which was introduced artificially into organisms such as plants, worms, or flies using injection or transgenic expression.

siRNAs were observed during transgene- and virus-induced silencing in plants, which is consistent with their role in the defense of the genome against foreign or invasive nucleic acids, such as viruses or transposons.

However, are there also endogenous siRNAs? If so, what is their origin?

Comparing siRNAs and miRNAs

they are similar in both their biogenesis and activity in the cell

- siRNAs induce the degradation of homologous mRNAs in a similar way as miRNAs (that are perfectly matched to target transcripts) → **via "slicing" followed by exonuclease-mediated degradation of the cleaved mRNA fragment.**
- In RNAi, the long dsRNA is converted into a population of siRNAs of about 20-25 nucleotides in length by the enzyme Dicer.
- The siRNAs are incorporated into AGO proteins as part of a RISC involved in target RNA recognition.

- siRNA-mediated gene silencing is mainly associated with RNA degradation by slicing; however, recent findings indicate that they may also function in translational inhibition as well as in other sequence-specific gene-silencing mechanisms (which we will discuss later).
- the main criterion to differentiate these molecules is the following:
  - miRNAs are excised by Dicer from their short (roughly 70 nucleotides long in mammals) and imperfect stem-loop precursor as a single small RNA species, which may accumulate to tremendous levels in the cell (up to 50'000 molecules/cell) → nur 1 miRNA aus stem loop structure
  - siRNAs, on the other hand, are produced by consecutive cuts by Dicer along a long perfectly double-stranded dsRNA precursor. → siRNAs, unlike miRNAs, are always part of a population that is distributed along their long dsRNA precursor. → mehrere siRNA aus dsRNA

Experimental evidence for endogenous siRNAs

To identify small RNAs, they used deep sequencing that allows determining the sequence and the size of an RNA.

- Deep sequencing of Arabidopsis showed that the majority of small RNAs present in plant cells are actually siRNAs. → largest part were 24 nucleotides long
  - These 24-nucleotide-long RNAs could not be miRNAs, because a similar distribution pattern was found in Arabidopsis mutant for DCL1 (remember that DCL1 is the Dicer homolog in plants that produces miRNAs).

What are all these small, 24-nucleotide long RNAs if they are not produced by DCL1 and are therefore not miRNAs?

Except DCL1 **every DCL produces small RNAs of a very specific size.**

- DCL2 produces short RNAs (siRNA) of 22 nu length
- DCL3 produces short RNAs of 24 nucleotides length
- DCL4 produces short RNAs of 21 nucleotides length.
- for DCL1, which produces 20- to 24-nucleotide-long miRNAs,

On a molecular level, how can the specific sizes of DCL2 (22 nt), DCL3 (24 nt) and DCL4 (21 nt) products be achieved by the different Dicer enzymes?

It is the middle domain of Dicer called the "molecular ruler" that determines the size of Dicer products. The ruler represents the distance between the PAZ domain that Dicer uses to anchor itself to the long dsRNA extremity, and the RNase-III catalytic residues of Dicer

the large majority of small RNAs present in plant cells must be 24-nucleotides long siRNAs produced by DCL3:

- ☞ However, the plants **used for the RNA analysis were healthy, neither transfected nor infected with viruses.**
- ☞ **This means that the small RNAs extracted from these plants must represent endogenous, naturally occurring small RNAs that exist in the absence of an exogenous inducer.**
- ☞ Thus, endogenous RNAi-silencing activity may exist for siRNAs, similar to the activity of miRNA genes.

Transposable elements as an origin of endogenous siRNAs

the 24-nucleotide-long siRNAs could be mapped to genomic sites with transposons and retrotransposons that are near the centromere of the chromosome

Role of **endogenous** siRNAs

most of the DLC3-dependent, 24-nucleotide-long siRNAs originate from regions ***at or around the centromeres that are very rich in transposable elements and enriched in condensed chromatin called hetero-chromatin and are poorly transcribed by RNA polymerase II.***

There are also 24 nucleotide long siRNAs that are found in euchromatic regions of the chromosome arms.

- Interestingly, these also map to TEs.
- **They derive from these TEs as populations made from long dsRNA precursor molecules.**

Remember that many TEs form inverted repeat loci, the transcription of which can lead to self-complementary dsRNA.

Other mechanisms exist to generate dsRNA from transposons, which we will discuss in the high-level lecture.

1. transposon-derived dsRNAs is then converted into siRNAs by DLC3 (and to a lower extent, by DLC2)

2. These 24 nt siRNAs are then loaded mostly into AGO4, which belongs to a specific clade of so called "heterochromatic" AGO proteins (AGO4/6/9).
3. siRNAs loaded onto these AGO proteins anneal to nascent transcripts of the targeted DNA (that is, the DNA of the transposon) **and recruit de novo methyl-transferase to add methyl groups to the cytosines in the target DNA in an siRNA-complementary**, and thus sequence-specific manner, resulting in silencing of the TE.

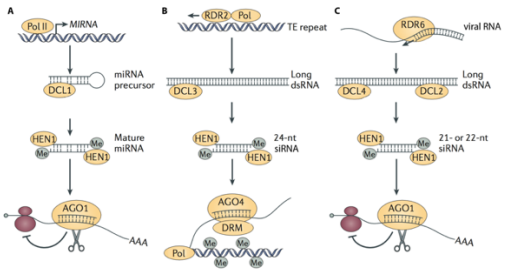


Figure 5-7 Endogenous gene-silencing pathways in plants. RNA silencing is triggered by dsRNA that can arise from various endogenous sources. All dsRNAs are cleaved by DCL proteins to produce mature small RNAs that are methylated (indicated by Me) by HEN1 to be protected from degradation. All processed, methylated small RNAs are incorporated into AGO proteins to form an RNA-induced silencing complex (RISC). RISCs can then target different mRNAs to silence gene transcription or induce transcriptional gene silencing (TGS) by DNA methylation and chromatin modification at target loci. **(A)** The miRNA pathway: Transcription of miRNA genes results in imperfectly matched hairpins that are processed by DCL1 into mature miRNA duplexes. One strand of the duplex is incorporated into AGO1 to form RISCs that target mRNAs for slicing (indicated by the scissors) or inhibit their translation (indicated by blocking ribosomes, red circles). **(B)** The siRNA pathway to silence transposable elements: RNA transcribed from transposable elements can be converted into dsRNA by RNA-dependent RNA polymerases (encoded by the cellular genome). This dsRNA serves as substrate for DCL3 to produce siRNAs that, when incorporated into AGO4, direct RISCs to specific genomic loci to induce transcriptional silencing by DNA and chromatin modifications by recruiting a methyl transferase (DRM). **(C)** The antiviral siRNA pathway: RNA produced by viruses can be converted into dsRNA by host-encoded RNA-dependent RNA polymerases. These dsRNAs are processed by DCL4 and DCL2 to produce siRNAs that are incorporated into AGO1 to form RISCs targeting mRNAs for slicing or translational inhibition. (Adapted from N. Pumplin and O. Voinnet, *Nat Rev Microbiol*, 2013)

RdDM = Transposon silencing with directed methylation

- the mobilization of TEs can cause gene disruptions and cause chromosomal breakage; **thus, cells must have evolved strategies to protect their genomic integrity by guarding their DNA from the activity of mobile elements.**
  → One of these strategies is to silence TEs by RNA-directed DNA methylation (RdDM).
- **RdDM therefore is another mechanism how siRNAs can silence genes** – <u>not by mRNA degradation on a post-transcriptional level, but by epigenetically modifying the DNA to reduce its transcription</u>.
- Gene silencing mechanisms acting on the level of transcriptional repression are collectively termed transcriptional gene silencing (TGS).
- *We will discuss RdDM and TGS in more detail in the high level lecture.*

Role of exogenous siRNAs in plants: antiviral silencing

VIGS: virus-induced gene silencing (VIGS) = antiviral silencing (virus silencing)

☞ *endogenous siRNA products of **DCL2 (22 nucleotides) and DCL4** (21 nucleotides) are very underrepresented in healthy plants*

☞ *in plants infected with viruses, these siRNA species dramatically increase*

- these **siRNAs do not map onto the plant genome, but rather onto the genome of the virus** used for infection.
- **Furthermore, plants lacking DLC4 and DCL2 function show increased susceptibility to plant viruses**; suggesting that the <u>**virus-derived siRNA products of DCL4 and DCL2 mediate antiviral defense.**</u>
- <u>Indeed, it was shown that DCL4 and DCL2 act redundantly against all RNA and DNA viruses tested so far.</u>

The mechanism by which virus infections lead to the production of siRNAs:

→ can be explained the easiest by the replication cycle of an RNA virus:

1. <u>**The 21- and 22-nucleotide-long siRNAs derive from a long dsRNA precursor that forms by replicating**</u> the genomic positive, **single-sense (+) RNA strand of the virus into a complementary negative-sense (-) RNA** (see figure 5-3).
2. This **dsRNA precursor, called the replication form (RF), is recognized by the host-plant-encoded DCL2** and DCL4 enzymes and cut **into siRNAs**
3. siRNAs can be loaded onto **AGO** proteins to mediate **gene silencing by degrading the viral mRNA (+ strand).**

☞ **DCL4 is the major enzyme in antiviral silencing, and is seconded by its surrogate, DCL2.**

☞ The major AGO proteins that mediate antiviral RNAi are **AGO1 and AGO2.** Note that AGO1 is also the main effector of most plant miRNAs.

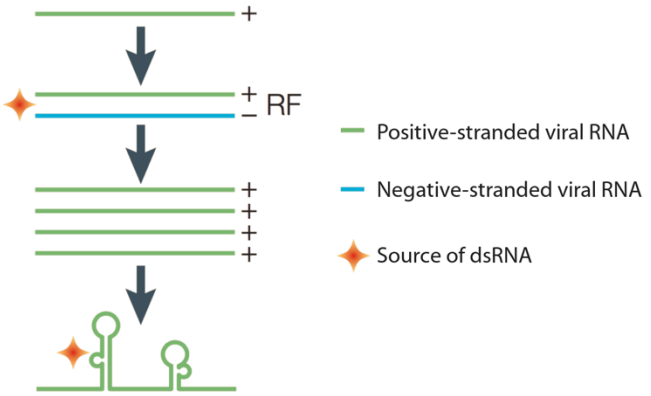


Figure 5-3 Molecules serving for RNA silencing are produced during the replication of viruses. Most plant viruses, and some animal viruses, have genomes of positive, ssRNA that are replicated within the cytoplasm of the host. Viral RNA-dependent RNA polymerases synthesize complementary negative-stranded (-) genomic RNA, from which numerous copies of positive-stranded (+) RNA are reproduced. These (+) strands are used for protein synthesis. Partial or complete annealing of (+) and (-) RNA strands provides one source of dsRNA. A second source is provided by the folding of replicated, single-stranded genomic (+) RNA, which forms secondary double-stranded structures. (Adapted from O. Voinnet, *Nat Rev Genet*, 2005)

how the RNA-silencing machinery of the host can keep up with the rapid replication pace of the virus → RDRs

- As you can see in figure 5-3, **the viral (-) strand is copied to produce multiple (+) strands**, leading **to a multifold amplification of the (+) strands.**
- **These (+) strands serve for protein synthesis.**
- Thus, the plant RNA-silencing machinery faces many RNAs to be targeted for destruction, and to do this, plants encode RNA-dependent RNA polymerases (RDR)
- The host-encoded RDRs can amplify the antiviral silencing action by generating dsRNA de novo from virus single-stranded RNA substrates.
  - Here, the cleaved RNA fragments resulting from AGO1/AGO2-mediated slicing serve as templates to synthesize new dsRNA that again can be processed by DCL4 and DCL2, loaded onto AGO1/AGO2 and mediate silencing of more viral RNAs (see figure 5-4).
  - The multiplication of antiviral Dicers in plants and other organisms can be seen as the result of an arms race between hosts and their parasites.

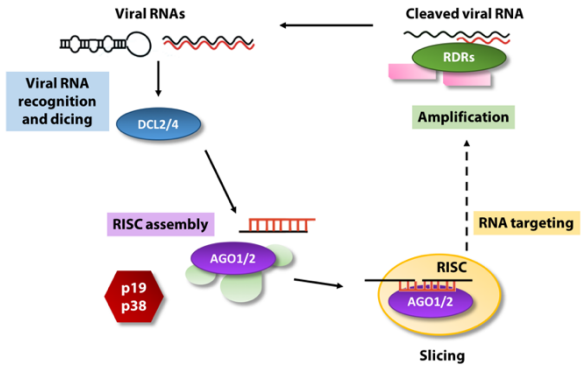


Figure 5-4 Current model of virus-induced gene silencing (VIGS) in plants. RNA silencing is initiated by the recognition of viral dsRNAs or dsRNAs with a partial hairpin, which are processed to siRNAs by DCL2 and DCL4. In the next step, AGO1/2 proteins are loaded with siRNA, thereby forming large RISCs, which probably also incorporate other unidentified proteins (green). Afterwards, the siRNA-loaded RISC targets viral RNAs by slicing or translational inhibition. Secondary siRNAs are produced in an amplification loop through the actions of host-encoded RNA-dependent RNA polymerases (RDRs) and their cofactors (pink). The viral-silencing suppressors (VSR) p19 and p38 interact with the silencing pathways (red hexagon). (Adapted from J. Burgyán and Z. Havelda, *Trends Plant Sci*, 2011)

viruses have evolved counter-defense strategies to escape virus-induced gene silencing → antiviral RNAi defense with VSR

- These viral suppressors of RNA silencing (VSR) are proteins encoded by the virus that interfere at different steps with the silencing pathway.

two strategies that are used by viruses to prevent RISC assembly:

- The p19 protein encoded by **tombusviruses** prevents RNA **silencing by binding siRNA with a high affinity, thereby sequestering it.**
  - **The protein functions as homodimer which is able to bind siRNAs in a size-dependent, sequence-independent manner**, acting like a molecular caliper by **measuring the length of siRNA duplexes and binding them**
  - p19 displays extremely high affinity for 21-bp siRNA duplexes, which are precisely the products of DCL4, the main plant antiviral Dicer. (DCL2 IST 2ND WAHL!)
  - **By binding siRNA duplexes, p19 prevents the loading of dsRNAs onto the AGO1/AGO2 protein and thus blocks antiviral silencing.**
- the p38 viral protein binds AGO1/AGO2 and neutralizes their ability to load virus-derived siRNAs.

→ The antiviral RNAi defense and coordinated production of VSR by viruses are not only found in plants but also in invertebrates including C. elegans and Drosophila.

→ Antiviral RNAi is a remarkable process, since it is entirely innate: it is not programmed by the host, but by structural (i.e., dsRNA) and nucleotide sequence features of the pathogen's RNA itself. As such, this mechanism can be virtually adapted to any virus.

Systemic transgene RNAi

We have just seen ***that antiviral RNAi includes an amplification of the RNAi response****, which can be* ***seen as a sequence-specific immune system required to keep pace with high viral replication rates,*** *because a* ***few viral RNAs can be converted into many dsRNA molecules that reinforce the silencing system****.*

RNAi (induced locally in response to long dsRNA or viruses) can move between cells and even over long distances throughout plants: How was this discovery made?

- **long dsRNA for GFP was locally delivered into only one of the leaves of tobacco plants that constitutively express a GFP transgene.**
- the GFP transgene was not only silenced in this particular part of the leaf, but the treatment also caused the GFP transgene to be silenced over time within the whole plant

- **This effect is sequence-specific**, because systemic GFP silencing is not initiated by a dsRNA carrying no sequence homology to GFP.
- This experiment **demonstrates the existence of a systemic silencing signal that gets amplified as it moves away from its sites of initiation.**
- **signal amplification occurs through the same RNA-dependent RNA polymerase (RDR6)** that also mediates the amplification of antiviral RNAi.
- the nucleic-acid component of the **signal is made of siRNA duplexes that move independently of AGO proteins between the cells** (via plasmodesmata) and through the vasculature (via the phloem) of the plant.
  - Interestingly, these routes are the same that are taken by viruses to spread within plants.

☞ ***This artificial system of non-cell autonomous silencing was found to underpin a systemic immune system that allows plant tissues to become immunized against viruses,*** ahead of the infection, in a sequence-specific manner.

☞ ***table below to remember the most important facts about the role of DCL proteins in gene silencing in plants:***

Table 5-1 Functions of the different DCL proteins in small RNA gene silencing pathways in plants.

DCL protein	Localization	Silencing mode	Size of effector sRNAs	Primary function
DCL1	Nucleus	• Processing of miRNA from stem-loop precursors	20-24 nt	• Post-transcriptional silencing (PTGS) of endogenous mRNA by slicing and/or translational inhibition
DCL2	Nucleus and cytoplasm; acts as a surrogate of DCL4 and DCL3	• Processing of siRNAs from viral dsRNA or from inverted-repeat transgenes used for experimental RNAi • Processing of siRNAs from transposon-derived dsRNA	22 nt	• Virus-induced gene silencing (VIGS) • Transposon epigenetic silencing
DCL3	Nucleus	• Processing of siRNAs from transposon-derived dsRNA	24 nt	• Epigenetic modification of transposon DNA and other endogenous loci (TGS)
DCL4	Nucleus and cytoplasm	• Processing of siRNAs from viral dsRNA or from inverted-repeat transgenes used for experimental RNAi	21 nt	• Virus-induced gene silencing (VIGS)

Summary:

We have seen that siRNAs can mediate silencing of transposable elements (and other repetitive regions within the genome) by a process called transcriptional gene silencing (TGS) via DNA modifications and chromatin remodeling. We will come back to this interesting function in the high-level lecture where we discuss the mechanisms and biological roles of this kind of silencing in more detail. Next, we have discussed how siRNAs function in antiviral silencing (virus-induced gene silencing, VIGS) and how this mechanism is exploited by plants to build an innate immune system based on RNAi amplification by specialized enzymes called RDRs. In turn, we have seen that viruses have evolved VSRs to counteract the antiviral RNAi reaction.

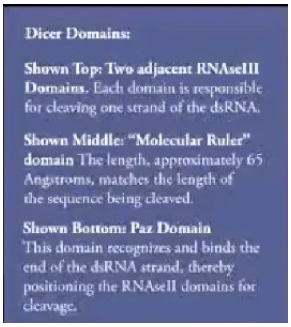
INTRODUCTION (OL) 23.11.20

RNA silencing: an historical perspective

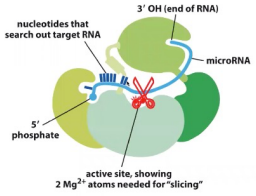
Small antisense RNA in Posttranscriptional Gene silencing

- Agrobacterium tumefaciens: a natural tool for plant transformation by T-DNA (containing tumor inducing genes) transfer trough a pore
- Scientists 1980 used this bacterium to induce genes into plant cells (this time harmeless/disarmed T-DNAs without tumor inducing genes) → introduced genes were called transgenes
- Chalcone synthase CHS = key enzyme at the beginning of pathway for petunia pruple colour as a transgene
- → wenn chalcone synthase gene induced (transgenic copies = sense RNA) = dunkel violette pflanze (selten auch loss of pigmentation wie unten)
- →wenn antisense chalcone synthase RNA induced: blocks pigmentation, since antisense-sense duplex forms and prohibits translation of chalcone synthase (white flower)
- → antisense **and** sense constructs could inhibit pigment production (CHS silencing) → they degrade both, the endogenous and the transgenic CHS post-transcriptional= process of co-supression
  - A transgenic trans-acting factor must cause the degradation of the endogenous mRNA
- GUS = bacterial compound → if GUS exprimiert in plants, they turn blue
- → 21-25nt long **sense and antisense** RNAs with sequence of the silence mRNA accumulated in the silenced tissue with dsRNA as a precurosor
- Product of induced transgene was a dsRNA in 10-15% of the cases (and 80% functional protein), since inverted repeats in the transgene were transcribed
- The discovery of RNA interference (RNAi)

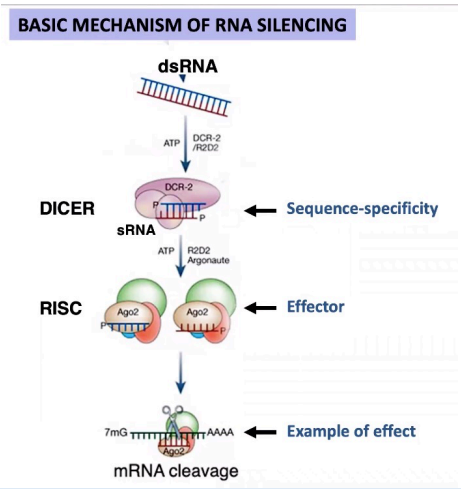
- dsRNA is transformed into small RNA by **Dicer** (also in vertebrates), which target the sequence specific sense mRNA and silence it



- Length btw RNAse domain and Paz domain bestimmt sRNA länge → so enstehen ganz viele verschiedene sRNA mit gleicher länge aus der ursprünglichen mRNA
- RISC: AGO as a core component of RISC and effector of RNA silencing
  - AGO bind sRNA at their 3' end via the PAZ domain
  - 5' end of small RNA called seed allows the micro nucleation to the target RNA
  - AGO = slicer = are endonucleases that use catalytic residues to cleave small RNA hybrids exactly at 10-11nt



  - If region at 10th or 11th nucleotide is double stranded, AGO will chop, sonst only translate



- If ago does not sense doublestrandness between the 10th and 11th position between small RNA guide strand and the targeted mRNA, it will not chop the mRNA, but will promote translational repression TR of this mRNA!
- Slicing and TR are 2 out of several outcomes of RNAi!

TUTORIAL 1 (VL) 30.11.12

1) si/miRNA strand selection

a. correctly orienting sRNA duplexes inside AGOs via "asymmetry sensing"

- 3' OH end of miRNA in PAZ domain, 5' end reads **monophosphate** of mRNA and the
- **correctly orienting sRNA d uplexes inside AGOs via** <u>assymetric sensing by R2D2 in perfectly complementary siRNA duplexes</u> (dieses Ende des siRNA duplexes braucht less energy to onwound)

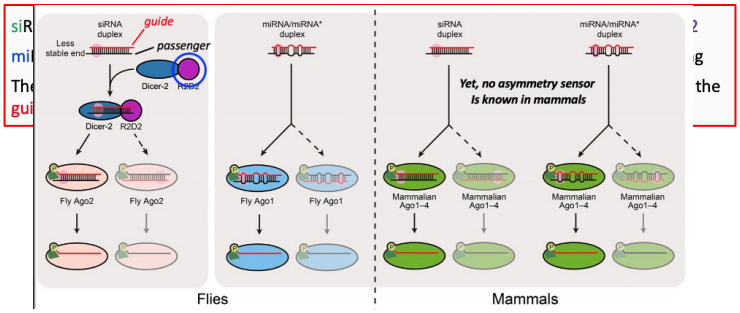
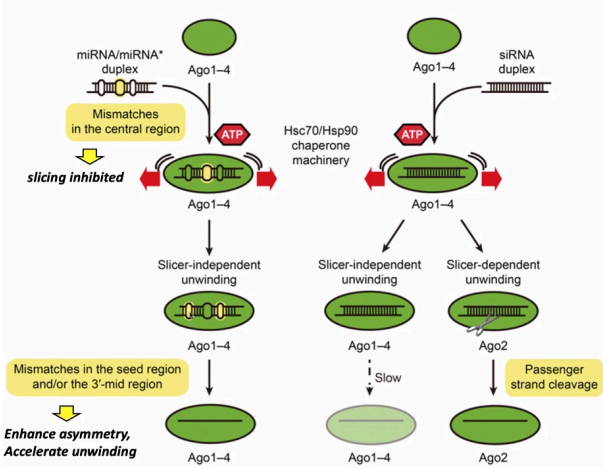


FIGURE 1. (A) miRNA/miRNA* and siRNA duplexes are functionally asymmetric. In both flies and mammals small RNA duplexes are preferentially loaded into Ago proteins with their less stable end (pink) toward the phosphate-binding pocket of the protein (light green), which results in the selection of the red strand as the guide strand. If the duplex is occasionally loaded in the opposite orientation, the black strand functions as the guide. The more asymmetric the duplex is, the more likely the red strand is exclusively selected as the guide. Besides Ago proteins themselves and the Hsc70/Hsp90 chaperone machinery **(B)**, no other factors are known to be required for asymmetric assembly of mammalian Ago1–4-RISC or fly Ago1 RISC. In contrast, in flies the Dicer-2/R2D2 heterodimer senses the asymmetry of siRNA duplexes and such binding is a prerequisite for Ago2-RISC assembly.

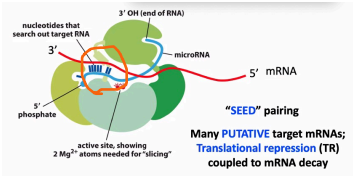
b. removing the sRNA "passenger" strand in mammals = strand selection

In mammals, central mismatches are preferred for RISC loading, whereas mismatches in the seed and 3′-mid regions promote slicer-independent unwinding.

Such mismatches are highlighted in yellow.



- Duplex loading requires ATP and the Hsc70/Hsp90 chaperone machinery, while both slicer-dependent unwinding and slicer-independent unwinding do not.
- Guide strand is loaded into AGO → important features when designing optimal siRNAs for experimental gene knockdown: Create Asymmetry (bulge) on 5 prime end at your guide strand so that red guide strand is the one loaded into AGO



AGO loading mRNA & miRNA

c. 5'-terminal nucleotide identity is also important

- 90% of loaded terminal miRNA bound to AGO2: have a 5' U and not A! (AGO1 auch U, AGO4 prefers loading with **A**-terminal small RNA)

2) Why are plant siRNA and miRNA 2'-O-methylated at their 3' ends?

All small RNAs of plants (siRNA and miRNA) are methylated at they 3' prime end!

- Extended base pairing, as in plant miRNAs and siRNAs, would imply a dissociation of the 3' end of the sRNA strand from the PAZ domain → **this would expose the 3' end to poly-U transferase** activitiea and subsequent **degradation** of the sRNA guide strand by nucleases
- **By adding a methyl group et the 3' end, HEN1 protects** perfectly or near-perfectly base paired sRNAs from degradation
- What would you predict if a mammalian mi/siRNA is provided a perfect target sequence like in plants?
  - Degradation very quickly since they have no HEN1 activity and the 3' End is not protected anymore by the PAZ domain
- in animals/metazoans the 3' end is protected in the PAZ domain, since **keine** vollständige Komplementarität von 5' bis 3'

3) Common, yet not necessarily correct assumptions of miRNA targeting in metazoans

- Most mRNA contain target sites for mutliple miRNAs, **located « mostly » in their 3'-UTRs**
- This allows an incredibly sophisticated mode of gene expression control at the post-transcriptional level.
- **Complex effects on gene expression depending on:**
  - respective cellular concentration of each miRNA, itself influenced by: cell type; cell context (stress, stimulus); miRNA isoforms (seed changes -> better/ poorer target);
  - **miRNAs act both in qualitative and quantitative manners**

Two possible modes of action of metazoan (animal) miRNAs

- mRNA cleavage : **requires a « perfect » complementarity with target (extremly rare)**
  - in other words: Rare metazoan miRNAs display perfect or near-perfect complementarity tot heir target transcripts **(like most miRNA in plants, see later in this course),** which they therefore regulate via slicing operated by the catalytic residues of the AGO2 nuclease.
- if not perfect complementary of mRNA with target mRNA: **Translational repression** (TR) coupled to prevailing decay (99% of cases)
  - in other words: Most **metazoans miRNAs** (99%) **display imperfect base-pairing** to their target sequence. **A highly conserved central mismatch prevents slicing**, leading to a mode of regulation known as translational repression (TR) through various mechanisms described in the next slide.

A. The 3'-UTR bias for animal target sites

Predict target sites of miRNAs :

Best described as an ***in silico*** '**convenience**' → based on the fact, that many **miRNA target search algorithms** initially **relied** on evolutionary conservation of these sites:

- obviously easier to detect **in 3'UTRs** im Gegensatz to ORFs → useful indeed, but **left the possibility that miRNAs target other regions** largely unexplored.
- Several experimentally validated miRNA target sites were recently found in the ORFs of mouse transcripts for pluripotency factors nanog, OCT4 and SOX2.

B. The 'Seed' notion is only valid to some extent

- *Several miRNA–mRNA interactions obey the seed rule (expects complete complementary): it helped characterize biologically important miRNA functions.*
- ***Can we be confident that seed pairing is mandatory for target recognition? NO!***
- NACHGESCHAUT SEED SEQUENCE : miRNAs regulate the gene expression by binding to the mRNA. **The seed sequence is essential for the binding of the miRNA to the mRNA.** The seed sequence or seed region is a conserved sequence which is mostly situated at positions 2-7 from the miRNA 5´-end. The seed **sequence is complementary with the binding site on the target mRNA**

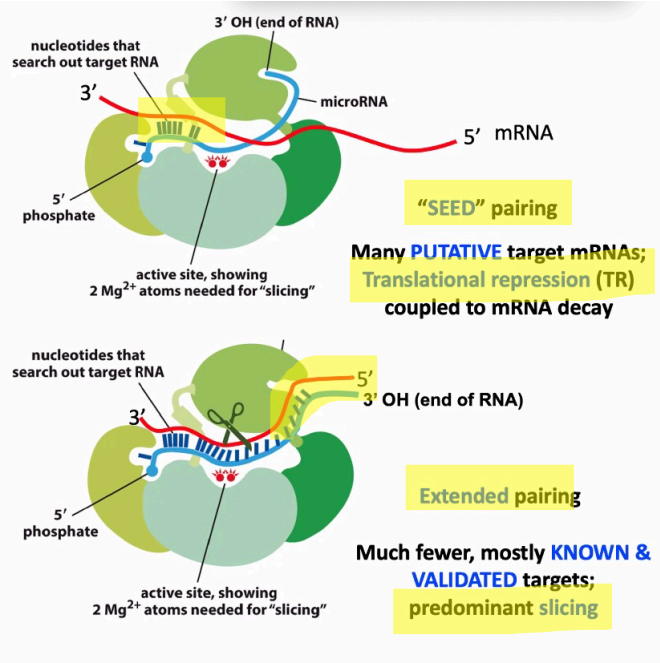
4) The almost impossible theoritical implications of a "seed"only -guided targeting in metazoans

a. many "fake" targets may play a buffering role

- Due to the limited length of the seed **(7nt),** any given metazoan **miRNA may target dozens of potential transcripts**
- With each of them undergoing approx. 1.5-2 folds regulation (!?)
- Yet most genes **are HAPLO-SUFFICIENT**
- What if all these predicted targets are NOT *bona fide* targets but **promote a general BUFFER to miRNA activity**
- Consequently, perhaps **only a few, select transcripts are genuine targets?**

- Experimentation lends some support to this notion:

microRNA target prediction programs predict many false positives

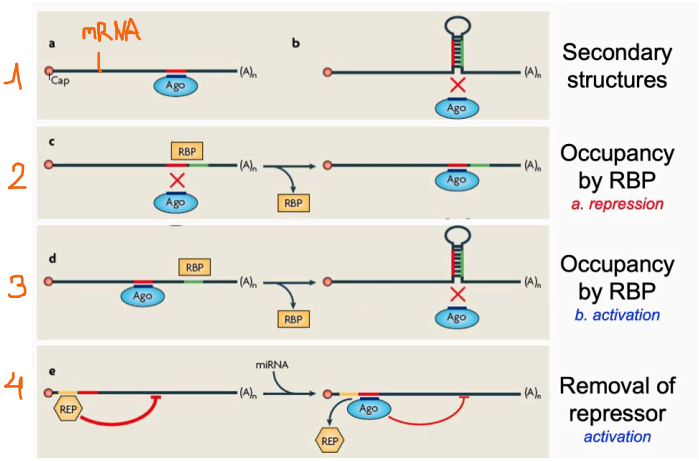


b. Whole-tissue analyses confound (verwirren) miRNA action in specific cell types

- Often, variations in target accumulation are **very modest** (2 to 2.5 folds), yet most genes are **haplo-sufficient**
- So far, miRNA expression and action have never been studied at **single-cell resolution** in any organ, let alone any whole organism, yet this is the **only way** to understand (miRNA) biology
- Tissues are a mosaic of cells and therefore have not the same amount or even diffrent miRNAs and so different miRNA interactions → we need single cell type resolution to understand miRNA biology (IN TUTORIAL 2 HOW THIS CAN BE ACHIEVED)

5) Can you think of ways by which miRNAs can actually stabilize target mRNAs or enhance their translation?

Influence of local sequence context



AUSNAHMEN : Wieso geht protein-amount up, even though miRNA induced to regulate post-transcriptionally (so regulate translation)? 4 AUSNAHME-FÄLLE :

1) Also 2nd structures play a role !
2) RBP = RNA binding protein → AGO binding is not possible if RBP is bound (but for example if in another cell type of same tissue is no RBP, the cell undergoes mRNA silencing by miRNA) → erst if RBP removed = repression of mRNA translation by miRNA
3) You can also generate activation with RBP, even though AGO is bound → since it induces the formation of the hairpin in the mRNA → mRNA can not be targeted by Ago → mRNA can be translated = activation
4) Ago miRNA interaction may be so strong, that the strong repressor of gene is pushed away/displaced, but the silencing of the mRNA is weaker than repressor, so the mRNA is more often **translated** than with active repressor REP → activation

→ Target site of mRNA to miRNA are nämlich only one part of very complex interaction! (stopped at slide 22)

Additionally :

mi-Rna's are gene encoded and read by rna-polymerase ii = so ds-rna precursor is endogenous ! (1 of 30'000 Gene is a miRNA gene)

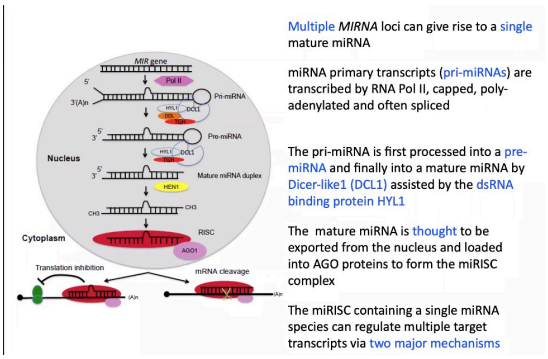
- *Transcript of miRNA, so it's mRNA, is NON CODING im Verlgeich zu normal mRNAs (selection of right strand with Dicer1 and also the generation of miRNAs)*
- *siRNA im Ggs. are from double stranded RNA which mostly are foreign*

TUTORIAL 2 (VL): RESOLVING MI-RNA ACTION IN SPACE

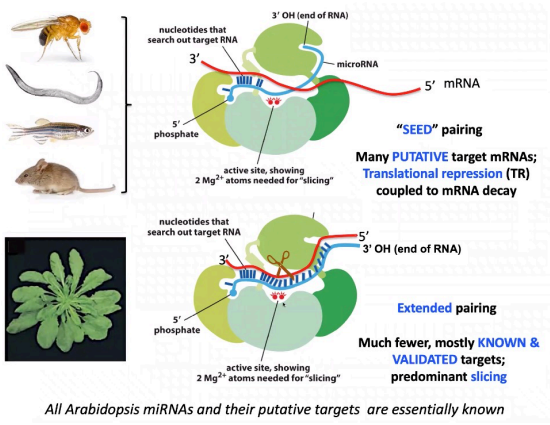
GENOME-SCALE, SINGLE-CELL-TYPE RESOLUTION OF MICRORNA ACTIVITIES WITHIN A WHOLE ORGAN

Summary Tutorial 1:

miRNA Pathway:



- have their own genes
- PolII produce mRNA → produce non cooding stem loop RNA → Piri-mRNA → sinlge stranded part are excised by DCL1 with HYL1 (double stranded binding protein) → pre-miRNA → DCL1 again to excise from from stem loop the mature duplex → Methylation by Hen1 → Methylation is protecting small RNA from degradation and polyuridilation → strand separation → AGO1 selectss strand (guide stand) → exportet outside of Nucleus by exportin1 → ago finds to the guide microRNA strand complementary mRNA by scanning and promotes cleavage at 10th or 11th nucleotide → remaining messenger RNA that are not sliced undergo translational repression



In plants compared to metazoans: extended complementary of mRNA and miRNA

- Targets of miRNA in plants are all known, bcs we can it predict it bcs of extended complementary

Whole-tissue analyses confound miRNA action in specific cell types

Often variations in target accumulation are very modest (2fold), yet most **genes are haplo-sufficient** → mosaic tissues where miRNA is active in only some cells and also the target is also available in only some cells

Start of today:

Current methods to analyze miRNAs and their action in plants and metazoans

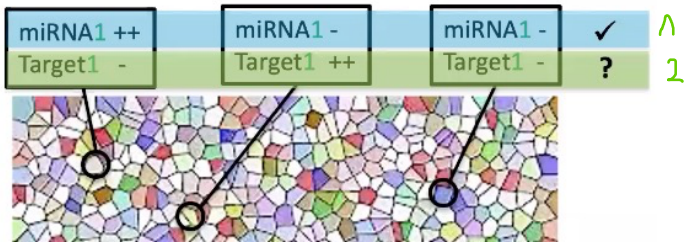
- Sensor of miRNA transcription
- Sensor of miRNA activity
- miRNA sensors, full length target fusions, miRNA promotor fusions, are useful, but only address single miRNA target combinations
- → **it is unthinkable to produce such transgenics for every single miRNA of a complex organ(ism)**

Genome-scale profiling of miRNA expression, loading and action in the whole root-tip

- All cell types in the root are organized in a concentric manner
- The different cell layers have **specific promotors**, to drive their gene expression (Endodermis specific gene expression)

Examination:

Appraoch 1: apply IP to each cell layer in the root to find all active miRNAs in the various cell layers (1)



→ *therefore* ***engineering cell specific Ago1 to capture cell specific active miRNA (generate cell specific miRNA profiles)***

- *Ago1 = main miRNA effector*
- *Just leave DCL1 active, to focus on miRNA (Dcl2,3,4 are siRNA specific)*

1) *Ergebnis: Ago1 is a cell cytoplasmic autonomous protein and and cannot move between the layers*

Ergebnisse:

- *Immunoprecipitation:* ***Ago1(bound to flag, which is expressed by the cell layer specific promotor): microRNA complex was isolated and deep sequenced in the particular cell layer, that is made up of 1 specific cell type by using Anti-FLAG antibody***
- *bewiesen dass IP based cell type specific profiling of miRNA is possible without contaimantion from surrounding cell layers leading to formation of artefactual AGO1-miRNA complexes*
- *cell specific whole active miRNAome was established → now kown in which root cell layer, which miRNA is active*
- *NO miRNA was jedoch strictly single-layer specific!*
- *Stele (stiel) is very rich in miRNA → since stele is itself composed of multiple cell layer*
- *Adapting single cell resolution: 60% increase on total known miRNAs*

→ *ZF: achieved to produce a repertoire of active miRNA in a cell layer specific manner using PI*

Approach 2: achieve cell-type specific profiling of ALL miRNA ***target transcripts in the root*** *(2) = cell specific targetome*

Reveal how miRNAs function in the root tip:

measure cell specific microRNA action (targeting targets) using polysome (ribosomes translating all mRNA) IPs:

- *RPL18 = 16s ribosome subunit component that translate all mRNA in the cell → RPL is also bound to FLAG (Flag:RPL18 complex) to immunoprecipitate the subunit of the ribosome, to isolate the mRNA that is attached to the RPL18! → we are able to produce a translatome (when we sequence all isolated mRNA, a translatome to find out how much of an mRNA is translated at this specific time point in this specific cell)*

How to find now those mRNAs that would be targeted by the miRNA?

Predict that if mRNA is a target of miRNA in a particular layer, then the abundance of the mRNA would be reduced in the layer, since AGO1 slices majority of target mRNA

- *Do one experiment with active miRNA (WT) and one with miRNA deficient mutant (hyl1 mutant= component of Ago1) → Transgenic line for each root cell layer*
- *We really have to have cell layer specificity to show this accumulation of mRNA in the WT*

miRNA mediated gene regulation focusing on the Epidermis

Compare amount of specific mRNA in hyl1 mutant and WT looking at the single cell layer "Epidermis"

- Example 1: looking at mRNA of Werwolf:
  - A lot less WER mRNA in the WT, than in hyl1 → Werwerolf is targeted by miRNA in the Epidermis

How do we analyze and display intuitively this vast amount of data to unravel how miRNAs function in the root tip?

- Digital deconvolution of each cell layer and reassembly of virtual roottips → miRoot browser (done by him)
  - Shows in which root layers which miRNA mostly abundant

→ miRNA was also abundant in cell layers of the root where it was not even made! → that means that miRNAs are non-cell autonomous (im Vergleich zu Ago1)!

→ miRNA movement to the neighboring is exquisitely regulated!

RNA-DIRECTED DNA METHYLATION IN ARABIDOPSIS (VL 3)

- DCL3 for siRNA (24nt)
  - DNA Methylation, Heterochromatin, TGS
- → most plant endo-siRNAs are made by DCL3 and map to repeats & Transposable elements
- siRNAs are produced as populations along TEs (the dsRNA derive from Transposons!)
- → the siRNAs then mediate methylation (RdDM)
- Transposon is trigger and target of the RdDM mechanism!!!

Transposons:

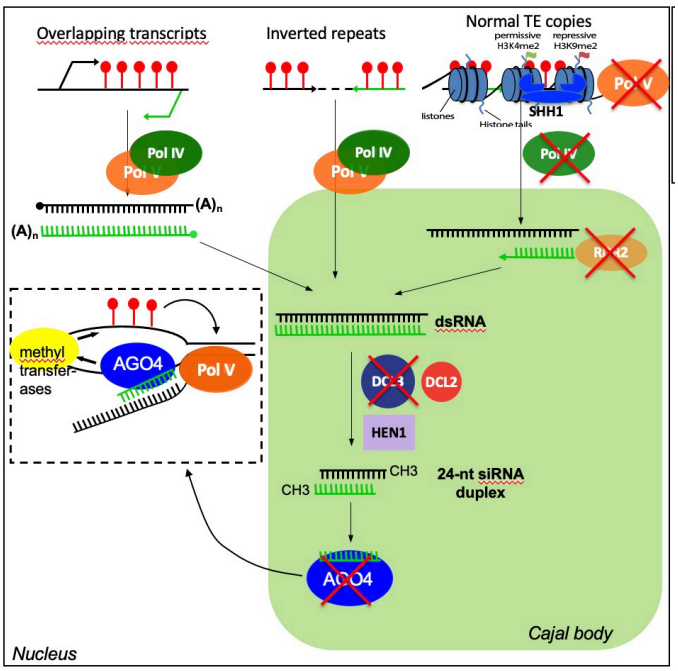
- Fragments of DNA that can insert into new chromosomal locations
- Some copy themselfes and increase in number within the genome
- Responsible for large scale chromosome rearrangements and single-gene mutagenic events
- 2 major classes of TEs
  - DNA transposons (DNA mediated TEs): cut and paste
  - Retrotransposons (RNA-mediated TEs): copy and paste (by reverse Transcription umgewandelt in DNA) → much more proliferative!
  - → The main thrat to genome integrity by far comes form retrotransposons! (50% of human genome made up of Transposons)
- Most TEs are broken = cannot transpose = remnants
- Active TEs evolved to inert into safes havens (TE rich loci) including jumping into themselfes
- Host regulates TE transcription and movement by epigenetic mechanisms
- TEs can even provide advantages to expression control!
- TEs are unusally incative, but stress conditions may activate TEs
- Active TEs increase mutation frequency → most mutations are neutral (sleten they are adaptive)
- The major threads posed by TEs are intra- and inter chromosomic rearrangements caused by the highly repetative nature of Transposons
- DNA and histone methylation cause heterochromatin formation at TE rloci, which reduces or prevents recomination and rearrangements of transposons (especially at TE rich loci)

DNA (Cytosin) methylation for Epigenetic silencing of TEs

DNA methylation is necessary to silence transposons:

- Cytosines can be methylated in all sequence contexts in pants by methyltransferase → giving rise to Epi-Alleles
- Only CG and CHG methylation (symmetric!) can be propagated/erhalten bleiben during DNA replication by MET1 (Maintenance Methylase)
  - Loss of function met1 mutants have hypomethylated DNA, same with ddm1 (DDM1: heterochromatin remodeling protein, action is coupled to MET1)
  - DDM inactivation → TE accumulate → mutations trough chromosomal rearrangements

- CHH site (asymmetric) methylation pattern can not be mainained throughout DNA replication in cell cycle
  - Requires additional information to re-methylate
  - siRNA needed to mediate de novo methylation!
    - RNA directed DNA methylation
  - Asymmetric methylation sites are maintained and initiaded by information on associated histones and RNA based mechanism = (si)RNA-directed DNA Methylation = RdDM, that directs the novo DNA methylases to these sites
  - RdDM is also required to explain how transposons and repeats become methylated for the first time once detected as foreign in the genome



Discovery of RdDM (RNA directed DNA Methylation)

- Viroids = plant pathogen, bestehend of highly complementary single stranded RNA
  - The viroid RNA does **not code for any protein (like siRNAs!)**
  - **Lines infected with exogenous vidoid,** → the transgene became methylated
  - **Since viroids are non-coding, de novo methylating was a new RNA mediated process → RNA-directed DNA methylation = RdDM**
  - **RdDM was strictly confined to the regions of homology between the transgene & inserted viroid**
  - RdDM was able to methylate cytosines in ALL SEQUENCE CONTEXTS (CG, CHG, CHH)
  - **You need therefore to maintain Transposon transcription to give always rise to new siRNAs to do de novo methylation at CHH cytosines and to mediate de novo methylation**

Summary:

RdDM needed for methylation pattern induction on symmetric AND asymmetric Cytosins and maintenance of methylation pattern at assymetric cytosins to silence transposons continuously:

1. TE invades in genome and proliferates (copy and paste)
2. TE creates difficulties: creates inverted repeats or overlapping transcripts, so that double stranded RNA is produced
3. Transcription of those loci by Pol2 → dsRNA production
4. dsRNA diced by Dicer3
5. population of heterochromatic siRNA arises with 24nt length
6. guide strand is selected by AGO4 → AGO4 looks for complementary RNA
7. when mRNA found → mRNA of transposon is sliced & Methyltransferase is recruited by AGO4 → ALL cytosins in a transposon are methylated
   a. → siRNA is derived from transposon & so complete complementary to the mRNA of the transposon and all its copies must be given
8. CHH requires constant transcription of Transposon even though its silenced, since CHH Methylation pattern cannot by maintained during replication, since the CHH methylation is asymmetric (CG & CHG can be maintained independently of RNAi pathway by Met1 and CMT3) > summary: CHH methylation is RNAi dependent
9. Now all copies of Transposons are silenced and all mRNAs of transposons degraded: Maintenance of machinery by Pol V (high affinity for methylated DNA)
   a. Pol II becomes less abundant at transposon loci, since it cannot bind to methylated DNA
10. Pol V produces non-coding RNA (unlike Pol II) = scaffold transcript so that AGO4 can go on with it's job = to anchor on transcripts of transposons
11. DNA methylation will increase K9 methylation on Histone tail → SHH1 binds to heterochromatin → SHH1 binds to Pol4
12. Pol4 (only in plants like Pol5) produces very short transcripts that are converted into long double stranded RNA by RDR2 → reinforces production of siRNA and creates independent source of double stranded RNA
13. Initial loci that were required to trigger the RdDM is not needed anymore (in case it is deleted etc.) → since Pol4/Pol5 and RDR2 is there, every single copy of the transposon can be the trigger (and target) for RdDM
14. If dumpen RdDM machinery by stress or mutation → less methylation on transposon → remnant of transposon is regulator of gene expression → danger!

Summary:

- ☞ transposon is trigger and target of RdDM at the same time!
- ☞ Transposons are silenced by the RNA mediated DNA silencing pathway, using siRNAs, that are complementary to the RNA of the transposon or the virus
- ☞ Paradoxically, epigenetic transcriptional silencing requires active transcription of the transposons (by POL4 and POL5):
- ☞ DNA GETS METHYLATED AT DNA AND AT HISTONE LEVEL!
- ☞ The RdDM machinery is a dynamic system → can be turned off and on by i.e stress

If mutation in Pol5, 4 oder in RDR2, the Transposon will reactivate

→ **DDM1 immer notwendig for the RdDM to access the heterochromatin! (lockert nämlich Heterochromatin), auch nötig für maintenance of methylation pattern of CHH sites during replication**

→ **DDM1 auch notwendig für MET1, for maintaining the methylation pattern when DNA replicates in cell cycle**

Genetic and epigenetic consequences on endogenous gene expression

How Transposons, originally seen as a threat, can create a new pattern for a gene expression and therefore be a advantage for the cell

1. Fresh TE (or virus sequence) inserts into an **endogenous gene promotor**
2. The initial genome invader and its many copies have been **inactivated by the host via epigenetic silencing by RdDM machinery**
3. Silencing and therefore no selection pressure (since not transcribed anymore) on TE promote accumulation of mutations and TE debilitation, especially at TE gene regulatory parts
   a. Only useful remnants of the invader (gene regulatory parts) are undergoing selection to retain them
4. Now in response to stress, which is known to dampen the dynamic RdDM, the mutated Transposons/Invader remnants can be used to **modulate endogene expression** and provide new expression patterns to the host

LONG DISTANCE SI-RNA MOVEMENT & EPIGENETIC INHERITANCE IN PLANTS (VL4)

Systemic transgene RNAi (RNAi moves over distances)

- We have sequence specific signals, phloem movement, cell-cell movement, amplified nucleic acids (RNA)
- **free** (AGO-**unbound**) siRNAs are the mobile signal
  - AGO is strictly cell-autonomous & does not move between cells
- **Silencing moves from phloem source tissues to phloem sink tissues and can be polarized like photo-assimilates**
  - Fully developed leaf → source leaf (guard cells of the leaf have lost plasmodesmata connection as they aged → so they did not receive the silencing signal!)
  - Young, new leaf = sink leaf
- RNAi silencing moves through the phloem and via plasmodesmata
- **All endo-siRNAs classes are also mobile**

Pathway and kinetics of RNA silencing movement resembles that of plant viruses

- Plasmodesmata are dynamically modified for the siRNA to move from one cell to other → siRNA are moving in the symplasm (and not apoplast)
- Between 3-4 days after delivery, the silencing is translocated from the leaf where the siRNA was originally made to the rest of the plant
- **Small RNA not only move through desmata like viruses do, their "infection" also follows the same kinetics** → the virus also needs 1-2 days to amplify and cover the whole plant → most plant viruses move using the same kinetics
- leaf removal following localized RNA silencing induction shows that the signal can be produced and translocated within 3-4 days of its initiation

Antiviral defense & counter defenses

- virus is feeding RNAi machinery → virus produces dsRNA as an integral part of their viral cycle (dsRNA building is mandatory)
- dsRNA gets diced by (antiviral) DICER4 (21nt) and DICER2 (22nt) → vi-siRNA are made → incorporated into AGO1 or 2 (also used for endogenous) miRNA → leads to antiviral silencing
  - bei Transposons ist DICER3 aktiv und erstellt 24nt lange siRNA's, darum gibt es in einer gesunden Pflanze auch so viele siRNA, nicht etwa weil so viele Viren bekämpft werden, sondern weil so viele Transposon gesilenced werden müssen
- the only RNA being displayed that shows complete complementary to the vi-siRNA is coming from the virus itself
- virus is like transposon trigger and target of the siRNA machinery
- pathogen evolves ways to evade the antiviral siRNA silencing machinery → VSR = proteins that prevent Dicing or AGO activity and/or loading!
  - P19 = VSR Protein → P19 forms head to tail homodimers to bind wit high affinity the antiviral siRNA (**exactly** 22 at 21nt long siRNA) made by Dicer 4 or 2
    - Man will auf keinen Fall, dass p19 auch 24nt lange siRNA angreift, denn diese sind für Transposon silencing zuständig und eine Demethylierung von Transposons wäre fatal!

- P38 = VSR Protein = binds AGO1 with high affinity to prevent loading of AGO with antiviral siRNA

- Most viruses suppress silencing but virus **could also evade silencing, since the plant is a supracellular organism**
  - But siRNAs are also able to move like virus!
  - Neighboring cells can **already be immunized as the virus enters by the siRNA itself that moves → siRNA builds a nucleotide sequence-specific immunity**
    - AGO1 bounded siRNA is restricted since AGO is cell autonomous
    -
  - **But if p19 is active, immunization cannot take place, since the siRNA cannot move, since it is bound to p19**
  - **If p38 is active, immunization of the neighboring cells can still be achieved**

- **Sources of siRNAs** are transposons, viruses and transgenes as well as other sources which are specifically activated in response to stress
- *siRNA of 21/22nt length's aim is the induction of post transcriptional gene silencing, by degradation of the RNA → achieved when using Dicer2 or 4 and loading siRNA into AGO1*
- *siRNAs of 24nt length are part of RdDM machinery, are diced by Dicer3 and loaded by AGO4 and have the aim to induce RNA directed DNA methylation → trigger direct silencing of locus they derived from by methylation as well all the copies of the transposon (scaffold RNA needs to be de novo transcribed by Pol4 for the maintenance of AGO4 mediated RNA directed DNA methylation at asymmetric cytosins of transposons)*

Some expected consequences:
- germline of plant = post-embryonic development
- some small RNA be able to invade the gametophytes in the meristem, they will induce de novo methylation → CHG and CG methylation would then be maintained in the meristem during replication (since RdDM is not active in gametophytes, but MET1 and CMT3 are active)
  - that's how **epigenetic information is given to the gametophytes! => DNA methylation in meristems**
  - **Any organ (like all new built branches!) that will emerge from shoot apical meristem that has received this specific methylation pattern, will be an epivariant**
    - That's why some branches can look totally different than others → those branches are epivariants
    - If siRNA would have moved into flower meristem, new built flowers could suddenly look totally different, since they received epigenetic informations by RdDM induced cytosin methylation → z.B komplett neue Blütenblattanordnung an einigen Blüten einer Pflanze = **epiallelism** (Bsp: L. vulgaris) => **siRNA as a source of phenotypic variation**

THE EEEEEENNNNNDDD