RNA BIOLOGY II: NON-CODING RNAS - BIOLOGY AND THERAPEUTICS

Julia Beckmann

HS2021 ETH Zürich

INHALT

MICRO RNAS

Micro RNAs are non-coding RNAs.

Discovered in 1993 in C. elegans. Mutations in defective timing of switching of developmental stages. Found the gene variant not in protein coding genes but in small RNA → very small and unusual → lin-4 and let-7.

There are about 500 miRNAs in human and mice, the majority is not conserved but some are.

LEVELS OF REGULATION OF GENE EXPRESSION/FUNCTION

- **Chromatin and Transcription/DNA**: Promoters and Enhancers. Genes are activated or repressed by transcription factors complexes
- **pre-mRNA**: Splicing and RNA transport
- **mRNA**: Translation, RBP (RNA-binding proteins), miRNAs. Focus of this lecture
- **protein**: Protein stability, activity, localization

MIRNA BIOGENESIS

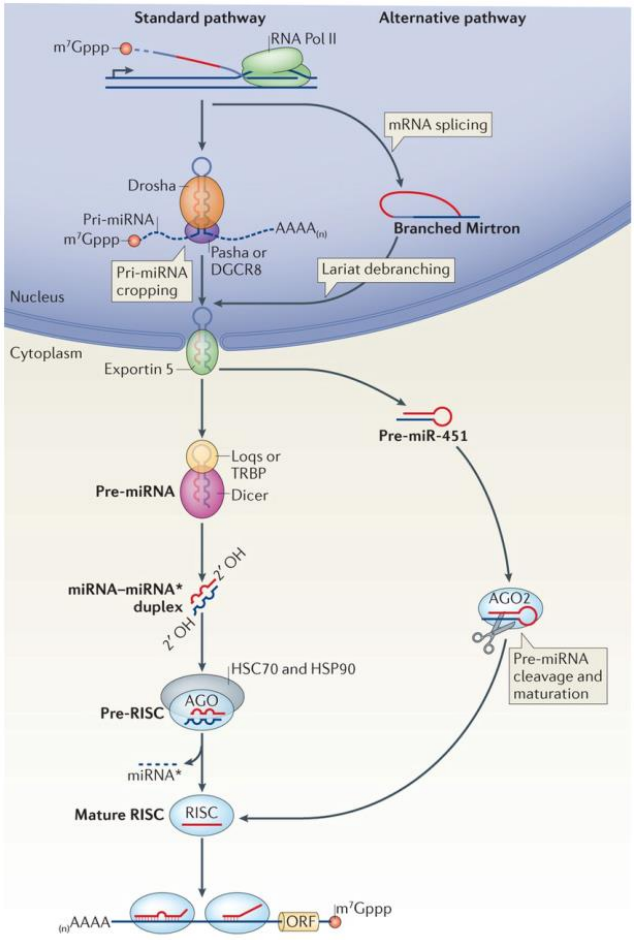
STANDARD PATHWAY = CANONICAL SIGNALING PATHWAY: miRNAs transcribed by pol II, polyadenylated. They are unique that they form a hairpin structure that is recognized by an enzyme complex called the Microprocessor. It is responsible for cleaving the hairpin at the base of the hairpin structure, getting rid of the two strands of mRNA that do not take part in the hairpin structure. This unprocessed micro RNA is called pri-miRNA. This is chopped by the Microprocessor in the hairpin-loop structure (= pri-miRNA cropping) and exported to the cytoplasm by Exportin-5 (energy requiring process). There pre-miRNA is further processed by Dicer (associated with accessory protein called TRBP, but this is not required for processing, Dicer is the important processing enzyme). Dicer cleaves of the loop on top of the hairpin-loop structure which generates a double stranded RNA → the miRNA-miRNA* duplex. It has a two nucleotide overhang at its 3'prime end at each end. The duplex is then loaded in the AGO protein where one strand is degraded, and the other is guided towards recognition elements on mRNAs. Once it finds such an element it represses or inhibits its gene expression.

The binding site is called the target recognition sequence on mRNA. The miRNA site is called the seed sequence → nucleotides at position 2-8 at 5 prime end of miRNA. There is perfect Van-der-Waals/Watson crick complementary between 2-8. There is some mismatches/wobble between the others. At the 3 prime end is again some matching which is important for stability.

We know that gene repression only occurs when the miRNA finds the target recognition site in the 3prime UTR. Because when it would bind to the ORF the ribosome usually pushes it away, meaning there is not enough time for the miRNA to gather the complex that is required for initiating the processes for repression.

miRNAs have many targets, average number of one is 150-200 targets → similar to a transcription factor, can regulate whole gene networks. miRNA are typically intergenic (between two genes) and can have their own promoter.

Majority of miRNAs exist as polycistronic clusters → transcribed as very long precursors that harbour many hairpin structures that are then processed further and give rise to individual miRNAs.



NON-CANONICAL PATHWAYS:

- **Mirtrons** → sit in exonic sequence, part of splicing can bypass the first Microprocessor, because they can just fold back and become a substrate immediately of Dicer.
- **Endogenous shRNAs** → also bypass Microprocessor. Enter pathway at Dicer stage.
- **Chimeric hairpins** → e.g. hairpins of miRNA plus tRNA. Also bypass Microprocessor, direct substrate for Dicer.
- **miR-451** → example for one that doesn't need Dicer but is dependent on Microprocessor.

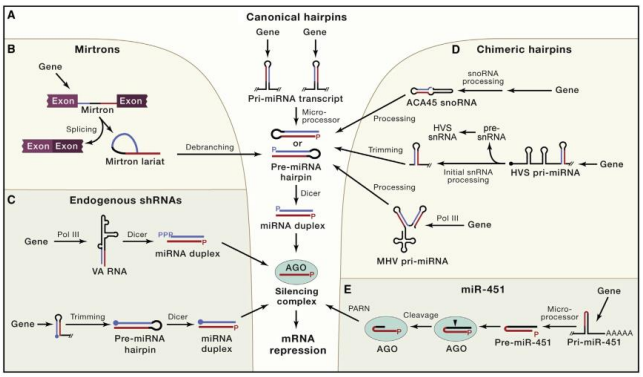
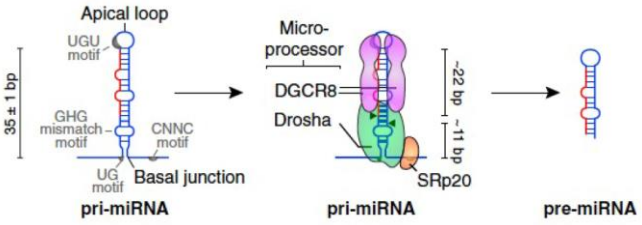


Figure 1: Comparison of Canonical and noncanonical biogenesis pathway. middle: canonical pathway. Mirtrons are second most important apart from canonicals

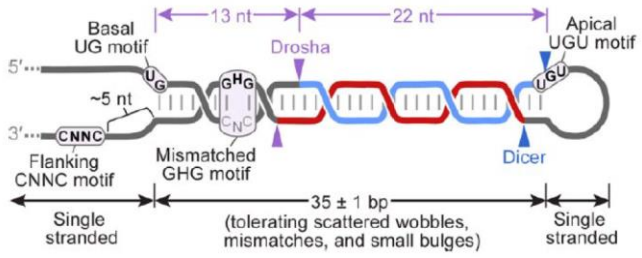
THE MICROPROCESSOR: Enzyme complex consisting of two proteins: a dimer of DGCR8 → RNA binding protein which binds to the loop and stem of the hairpin, and Drosha which is responsible for the cleavage at the loop.



Important because it is a gatekeeper of recognizing structures that need to be processed and enter the canonical RNA pathway. There needs to be a gatekeeper because there are millions of small RNAs which have some kind of hairpin structure (product of RNA degradation processes). Important that the Microprocessor recognizes authentic pri-miRNA from other small RNAs. Does this by recognizing a stem of certified base pairs (from loop to base) and an unstructured loop (apical loop) and an unstructured single stranded base. Within these structures, it recognizes certain motifs → UG motif, CNNC motif, GHG mismatch-motif, UGU motif. Microprocessor prefers perfect sequence complementary double stranded RNAs, but at 8 nucleotides from the base, it prefers a mismatch motif → GHG mismatch-motif.

All of this increases binding but it also enhances the processing of the precursors.

More detailed:



Not all precursors have all these, many have just 1 or 2.

TRANSLATIONAL REPRESSION BY MIRNAS

Can inhibit translation or lead to degradation of mRNA. Which process is dominant depends on species.

miRNA-AGO complex binds to miRNA target site in 3' UTR. Accessory protein GE182 that binds to AGO recruits or interacts with RNA binding protein → PABPC (binds to poly tails) → leads to recruitment of two de-adenylation complexes → PAN2-PAN3 complex and CCR4-NOT complex → responsible for shortening poly-A-tail → instable mRNA and increased degradation. Can also at the same time recruit decapping enzymes which further leads to instability of mRNA.

If there is *perfect sequence complementarity* between miRNA and the target (very rare in human, mice) then AGO can slice and cleave the mRNA between nucleotide 10 and 11. In plants, miRNAs have almost in the majority of cases this perfect complementarity.

Single miRNA are sufficient to initiate these degradation processes. They can have single binding sites or multiple binding sites. Statistically, they are *most likely to exert an effect if they are either towards the end of the poly-A tail or in the beginning of the 3' prime UTR*.

miRNAs can also function not as just one but with several miRNAs in close proximity. The closer the stronger the effect. In order for them to have synergism they need to be as close as 40-60 base pairs.

REGULATION OF MIRNA ACTIVITY

How is miRNA activity regulated?

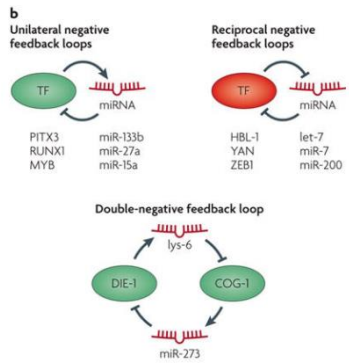
- Alternative cleavage by DGCR8 and Dicer
- Stimulation/repression of miRNA decay
- Target RNA-directed miRNA degradation
- Cooperativity
- Competitive element RNAs
- miRNA sponges
- Variation of miRNA length and sequence
- Modification of miRNA sequence (A-to-I editing)
- Post-translational modification of AGO

Strongest regulator: more miRNA → bigger effect!

ACTIVATORS AND REPRESSORS OF MIRNAS: the most important way by which miRNA activity is regulated, is by regulation of its promotors by transcription factors and transcription factor complexes (similar to any coding genes). E.g. increase expression through promotor. Can have their own promotor or if they are Mirtrons they are regulated together with the host gene (→ miRNA levels are always associated with levels of host gene).

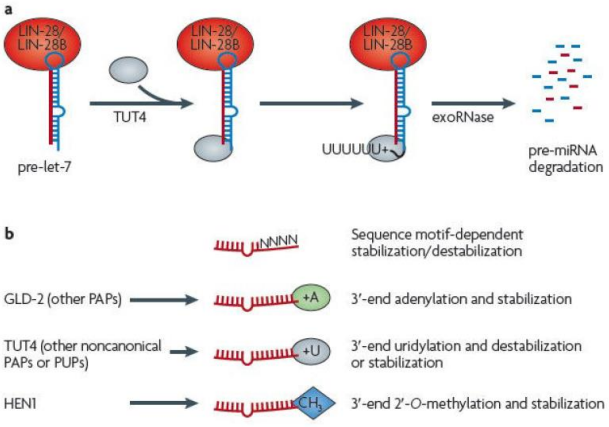
DIFFERENTIAL MIRNA REGULATION BY SPLICING: Here it is not the miRNA that is regulated but its transcript. Regulated by splicing out the target recognition site of the miRNA in the mRNA.

REGULATORS OF MIRNA PROCESSING: many features. Any step of the biogenesis pathway can be regulated! By e.g. increasing of repression of Drosha, transcription factors, Dicer etc. Knockdown of Drosha has global effect (on all miRNA). There are also examples of RBP that can bind to certain parts of pri-miRNA, can inhibit processing. Both of these has been shown to be important in cancers. Also regulation in nucleotide sequence. There are adenylases that can convert U into I → less processing, also can change recognition motif, can have effect on target.



AGO PTM REGULATING MIRNA ACTIVITY: regulating by post-transcriptional modifications (PTM) → many have been found in AGO that can affect its activity. Can have activating/enhancing activity or suppressing activity.

MODIFICATION OF THE 3' END OF MIRNAS: regulates stability. Here shown with specific RBP.



VIRAL MECHANISMS MODULATING MIRNA FUNCTION: viruses use miRNAs, e.g. in hepatitis C virus (HCV), 5' UTR of viral RNA contains two binding sites for miRNA that is highly abundant in the liver → binding of Ago2-miR-122 absolutely required to stabilize and replication of viral RNA genome → protects from detection and exonucleases. This has been exploited by using antisense miR-122 → cannot bind to viral genome anymore → leads to dramatic decline of viral DNA in blood and liver. Has been proposed as a therapeutic treatment for HCV. But less excitement because there are already drugs against it and also the highly expressed miR-122 needs to be silenced or inactivated, which can only be done with very high concentration of antisense miRNA over a long period of time. We know from mouse experiments that knockout mice eventually develop hepatocellular carcinoma. So miR-122 in the long term is needed for correct cell life.

CERNAS MODULATING MIRNA FUNCTION: competitive element RNAs. Hypothesis → if there is an RNA species that harbors a miRNA binding site this could in theory act as a sponge. It sequesters the miRNA, and this miRNA therefore cannot bind to its natural target anymore. This therefore leads to a derepression of these targets.

This mechanism is probably very rare. It might happen in certain cancers. Concept has never been proven.

One example where it is very plausible: circular RNA that harbors a lot of miRNA binding sites (70-100). Especially highly expressed in central nervous system. We know it can actually serve as a sponge!

Main concern: ceRNA expressed at physiological levels is unlikely to compete with the very large number of other miRNA targets.

RNA THERAPEUTICS

FEATURES OF SIRNA THERAPEUTICS:

- Can inhibit expression of any gene of interest, including undruggable targets
- Sequence-specific targeting of RNA
- Uses naturally occurring cellular process to silence mRNAs
- Able to inhibit highly expressed proteins with long half-lives (i.e. serum proteins)
- Can be designed to exhibit prolonged drug action
- Antidotes are affective
- Cleaved target mRNA can be measured in blood
- High thermo-stability

Initial disadvantages: toxicity (when injected, cells sense the high amount of RNA → initiate stress response), problem of delivery (solved for the liver).

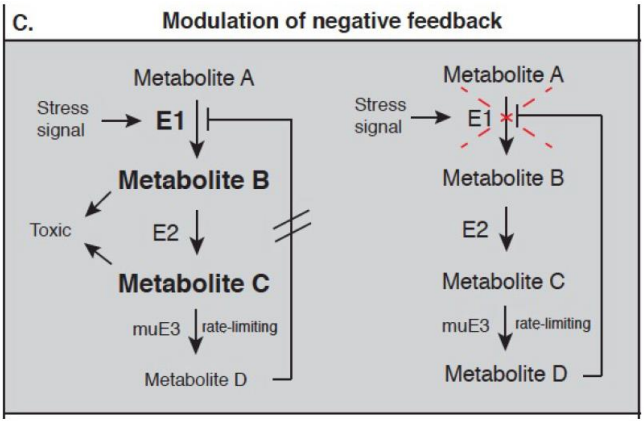
DISEASE APPLICATIONS FOR RNAI THERAPEUTICS:

Gain of function mutations → get rid of it. Example: Transthyretin Amyloidosis → heterozygous disease, patients have one allele with WT. Protein highly expressed in the liver and being excreted in the blood. Important for Osmosis and also transport protein. There is a mutation in transthyretin → people develop aggregation (amyloid) disease. These aggregates accumulate in the heart and in peripheral nerves (polyneuropathy). Most people die within 6 years after

diagnosis. Drug is now on the market which delivers via lipoprotein nanoparticles and siRNA that targets TTR and reduces it by 90% in the blood. siRNA cannot distinguish between the two heterozygous alleles and leads to silencing of both. But it is non-essential gene (function can be taken over) and therefore there is no disadvantage.

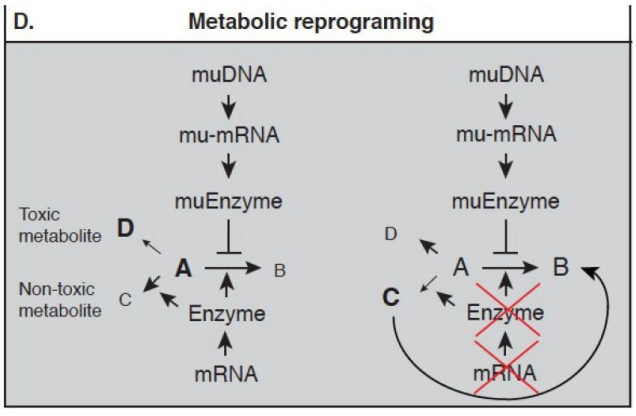
Silence a gene that is overexpressed and causes a disease. By interfering with the production of the overexpressed gene one can protect (in theory) against the disease. Example: Hypercholesterolemia → LDL removed from the blood by interacting with LDL receptors, Clathrin mediated uptake → cholesterol released → inhibits de novo synthesis, endosome recycled. This is how LDL (bad) is removed from the blood. Another important player in this is a gene called PCSK9 → binds to LDL receptor → leads to degradation in the cytosol→ prevents the LDL receptor from recycling back to the plasma membrane. There are two therapies that interfere with PCSK9. siRNA interferes on level of mRNA → no PCSK9 made → more receptors and more uptake of LDL. Antibody therapy (injected in blood) → neutralizes PCSK9 and has the same effects. Disadvantages: must be injected every two weeks, siRNA only twice a year.

Modulation of negative feedback loop.



Disease: Porphyrias → genetic defect in heat pathway. Mutation can lead to very toxic buildup. siRNA therapy targets gene and thereby prevents negative feedback loop.

Metabolic reprogramming:



By interfering with the pathway resulting in the toxic metabolite (here D) one can in theory redirect the metabolism to the pathway where there is no toxic metabolite (to C). Disease: Primary hyperoxaluria → metabolic defect in the liver responsible for Oxalate metabolism. Leads to increased Oxalate levels in the blood, crystallizes in the kidney → kidney stones and whole kidney develops microcrystals → kidney failure. Failure is that Glyoxalate cannot by metabolized to Glycine, and therefore shifts to Oxalate. This is catalyzed by GO. When GO is inhibited by siRNA the pathway shifts and diverts it away from oxalate and reverts it back to Glycolate and Glycine which is soluble and can be excreted.

RNA AND NEURODEGENERATION

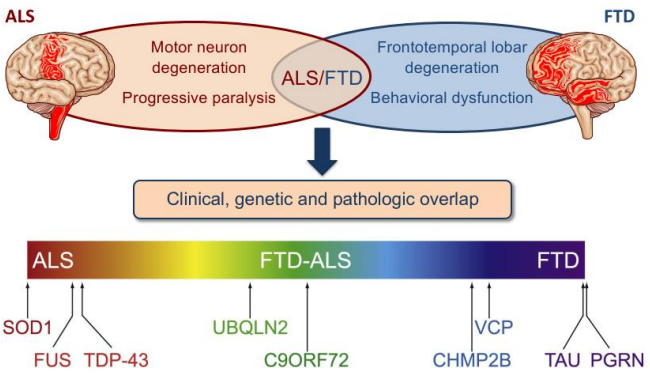
AMYOTROPHIC LATERAL SCLEROSIS (ALS)

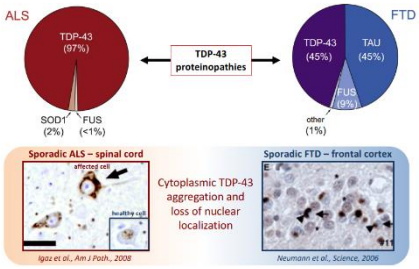
- 1 per 100'000 get it per year
- No available treatment, progressive adult-onset disease. Upper and lower motor neuron degeneration. Muscle weakness progressing to paralysis and death typically within 1-5 years from onset.
- Ubiquitin inclusions in spinal cord

FRONTOTEMPORAL DEMENTIA (FTD)

- Most common cause of dementia under 60 yrs
- 5-15 per 100'00 get it per year
- Progressive adult-onset brain disease. Frontotemporal lobar degeneration. Behavior changes, language dysfunction, emotional coldness, and apathy within 2-10 yrs. Typically no memory loss (different from AD).
- Ubiquitin protein inclusions in the frontal lobe/cortex

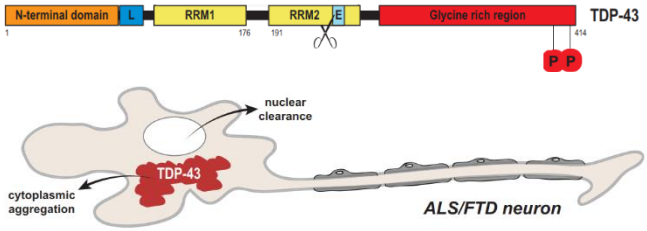
ALS AND FTD DISEASE SPECTRUM

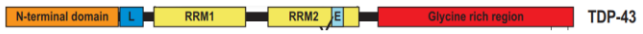




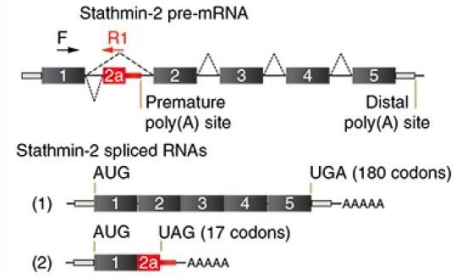
TDP-43 protein aggregates mark affected neurons in sporadic ALS and FTD. It is a multifunctional RNA-binding protein mis regulated in ALS and FTD. In a healthy neuron, TDP-43 is primarily nuclear. There is also TDP-43 in stress granules and in RNA granules (RNA transport). It acts as a splicing factor, part of alternative splicing and forms long introns leading to diff. splicings.

Loss of function:

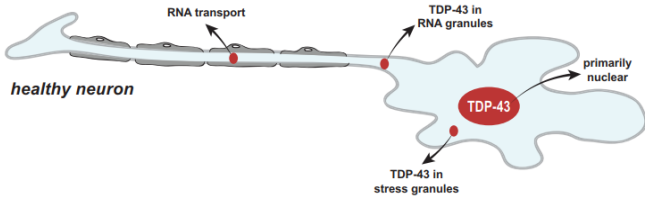


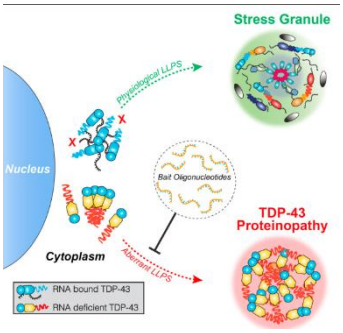


Studies has shown that **TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD**. Alternative splicing regulation and regulation of long intron containing pre-mRNA may be implicated in this.

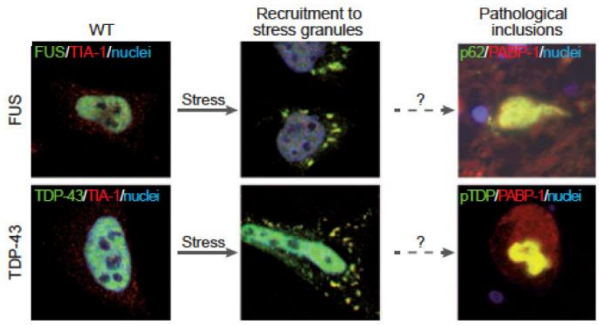


We see here that there is a premature poly(A) site, so from the Stathmin-2 pre-mRNA two different spliced RNAs can emerge. Stathmin 2 is important for neural networks. (2) is here a cryptic exon (see more later). **Without TDP-43, it is included and introduces a premature stop codon → leads to a very toxic disfunction in neurons** → direct function if TDP-43 visible.



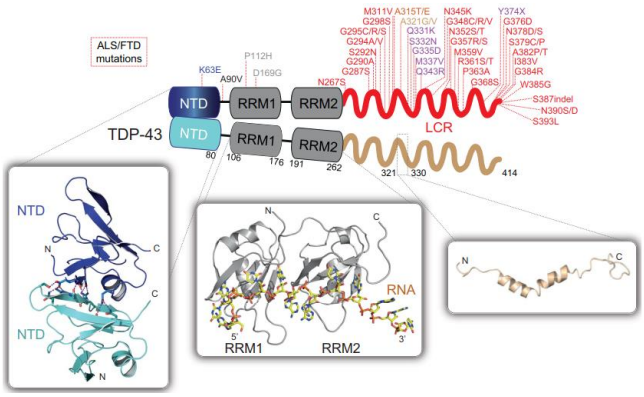


TDP-43 is also involved in RNA granules and in RNA transport. In mutant TDP-43 (M337V) and mutant TDP-43 (A315T) showed that the RNA transport to the neuromuscular junction is impaired compared to the Wild-type. The mutants no longer transport but accumulate them. This is related to stress granules which accumulate/keep the RNA granules to stop all unnecessary functions of cells that are not reacting to stress. They are complexes of RNA binding RNAs and RNAses to block transcription.



TDP-43 has a glycine rich region which allows it to be part of the stress granules. Maybe in aged neurons the disassembly mechanisms of the stress granules don't work properly and the granules remains → suppressing RNA function.

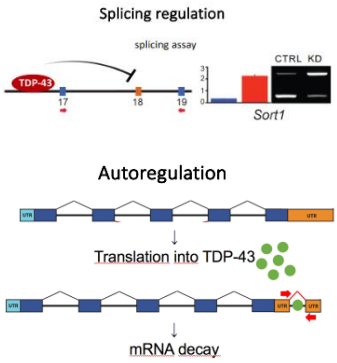
PHYSIOLOGICAL TDP-43 OLIGOMERIZATION



LCR = low complexity regions → most mutations there, because the other mutations might be lethal so the rest of TDP-43 might be essential to live.

The N-terminal domain of TDP-43 mediates its physiological polymerization (form dimers). TDP-43 oligomerization is necessary for its nuclear function.

Cryptic exon: exons that have the canonical structure of exons (sequences preceding the normal exons are the same) but they are not included in the end RNA. They either have premature stop codons or do not lead to proteins. When TDP-43 binds in their proximity, it repressed their inclusion. Without it, these exons are included, and this leads to loss of these mRNAs and proteins. This happens in patient brains.



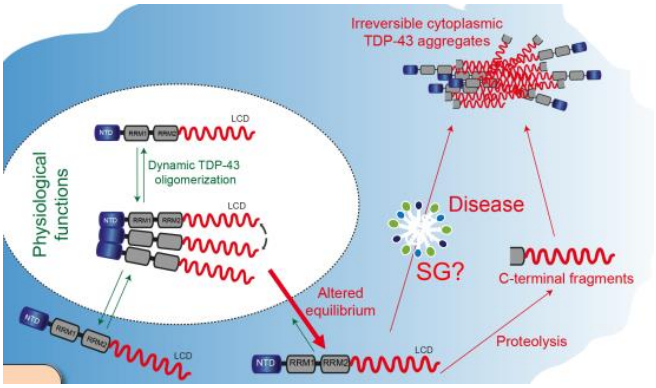
TDP-43 autoregulates by binding to its own RNA → negative feedback → TDP-43 expression leads to its own mRNA decay by binding to its 3' UTR and leading to specific degrading splicing. But it is a low affinity binding therefore a certain amount of TDP-43 is needed.

When normal TDP-43 is overexpressed (transcripts without autoregulation used) they aggregate! In mutation without oligomerization the accumulation is a lot stronger.

Assumption: in the nucleus there is a dynamic oligomerization balance → oligomers are important for splicings.

→ Dynamic polymerization of TDP-43 antagonizes its cytoplasmic aggregation!

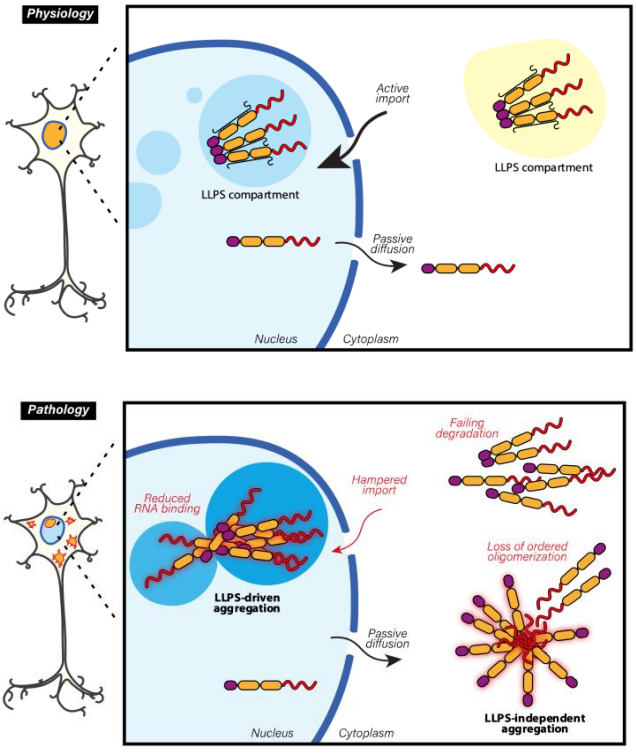
Why is oligomerization of TDP-43 necessary for RNA splicing? How does loss of TDP-43 oligomerization trigger aggregation?



Isogenic human cell lines expressing near-endogenous levels of TDP-43 variants showed that oligomerization and RNA-binding cooperatively stabilize the half-life and nuclear localization of TDP-43.

Oligomerization and RNA binding regulates the incorporation of TDP-43 into liquid-liquid phase separated bodies.

Monomeric TDP-43 binds a distinct subset of RNAs. How does loss of oligomerization lead to aggregation?



→ Distinct aggregation mechanisms driven by loss of oligomerization or RNA-binding!

Conclusion: loss of normal function and gain of TDP-43 aggregate toxicity mediates disease!

SYNAPTIC FUS RNA BINDING AND REGULATION

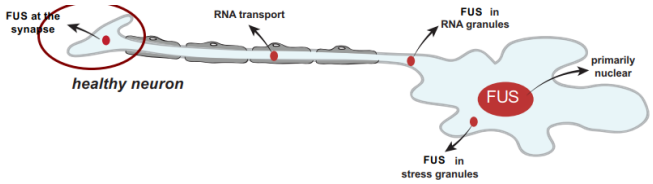
Does mislocalized, but non-aggregated protein confer early toxicity in neurons?

FUS proteinopathies is a small subset of early onset ALS/FTD. Proteinopathies seem to be mutually exclusive (either ALS or FTD), very rare but very aggressive, also early onset sporadic behavioural variant FTD.

FUS and TDP-43 show structural and functional similarities (FUS mutation is similar to TDP-43 disfunction).

WT FUS translocated to the cytoplasm in FTD-FUS.

ALS-linked mutations clustered in NLS (Nuclear localization sequence in FUS protein) results in defective nuclear localization. Degree of Mislocalization caused by ALS mutations in NLS correlates with disease onset. Does the mislocalized FUS become toxic for neurons even prior to aggregation?



What are the synaptic RNA targets of FUS? Is synaptic RNA content altered in disease? If yes, is this directly due to increased FUS in the synapse?

FUS is present at the synapses but the precise subsynaptic localization is debated. It is present in both pre- and postsynapses, but is enriched at the presynapses.

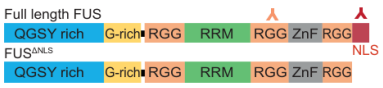
Biochemical isolation of morphologically intact synaptoneurosomes lead to identification of RNA targets of FUS at the synapses by CLIP-seq:

- LCR important for aggregation and incorporation into stress granules. FUS and TDP-43 can be included. FUS even has a Zink finger domain.
- WT FUS translocate to the cytoplasm in FTD-FUS
- ALS-linked mutations clustered in NLS result in defective nuclear localization. Mutation leading to ALS disrupt FUS binding (maybe mutation at NLS).
- Normally in total cortex we have a lot of FUS but at synaptic sites there is not that much
- They used CLIP-seq to find the binding sites of protein to the RNA via crosslinking. RNA than rimmed rather close till binding site → SDS Page, then Protein degeneration and RNA sequencing for exact position to where the protein was found.

Synaptic FUS RNA targets encode essential protein components of synapse!

Protein has both a RNA recognition motif and a zinc finger domain which both can bind RNA. FUS binds GU-rich and AC-repeat sequences at the synapse → binds stem loops. Mouse models for RNA experiment are not very accurate, here knock in mice stimulate what happens genetically in patients by removing the NLS section from FUS to be able to bind to the transporter.

Is synaptic RNA content changed in early disease stages of ALS-FUS?



$FUS^{\Delta NLS/+}$: a knockin mouse model mimicking ALS-associated FUS mutations. They lack nuclear localization signal (NLS) in one of the two mouse FUS alleles → partial mislocalization (increased synaptic localization of mutant FUS protein) without aggregation. Show mild motor impairments starting at 10 months. Hyperactivity and social disinhibition starting at 4 months. The mice show age-dependent synaptic RNA alterations. Difference really visible in SNS, showing that having so much FUS outside of the SNS really affects mice. Cytoplasmic FUS accumulation in these mice increases the stability of synaptic mRNAs.

Is stability of neuronal RNAs globally affected by FUS mislocalization?

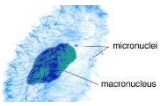
Synaptic accumulation of FUS triggers early misregulation of synaptic RNAs in ALS-FUS mice. Synaptic RNA alterations represent one of the earliest events in ALS-FUS mice. These are triggered by increased synaptic FUS, prior to aggregation.

RNA-MEDIATED GENOME REMODELING IN CILIATES

WHY STUDY CILIATES?

- Water dwelling organism, lives in water, eats bacteria
- Most extreme case of genome reduction by DNA splicing (up to 98%)
- Whole-genome-scale DNA segment reordering (similar to V(D)J)
- Epigenetic maternal inheritance of genome content and structure:
  - RNA guides DNA processing
  - RNAi-related mechanisms (Dicer, Piwi involved)
  - Histone modifications lead to DNA deletion (extreme way of chromatin silencing)
- Telomerase activity (*Oxytricha* cell has at least 120 million telomeres)
- genome is highly streamlined, everything that is not used for expression is removed
- Somatic genome

NUCLEAR DIMORPHISM IN CILIATES:



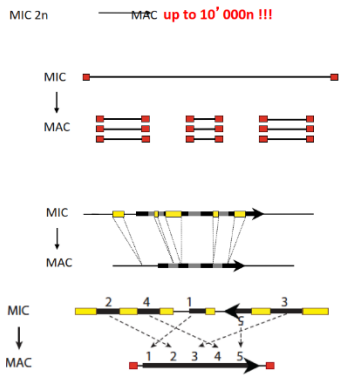
Huge nucleus to compensate for cell size. Two micronucleus and one macronucleus.

Two types of reproduction → find a partner, exchange of two cells. If they are no longer able to mate → cell division, fuse haploid product of meiosis.

Micronucleus	Macronucleus
- Germline - Transcriptionally silent - Diploid - Contains transposons and other germline-limited DNA	- Somatic - Transcriptionally active - Highly polyploid - Short chromosomes - Replaced at each sexual process

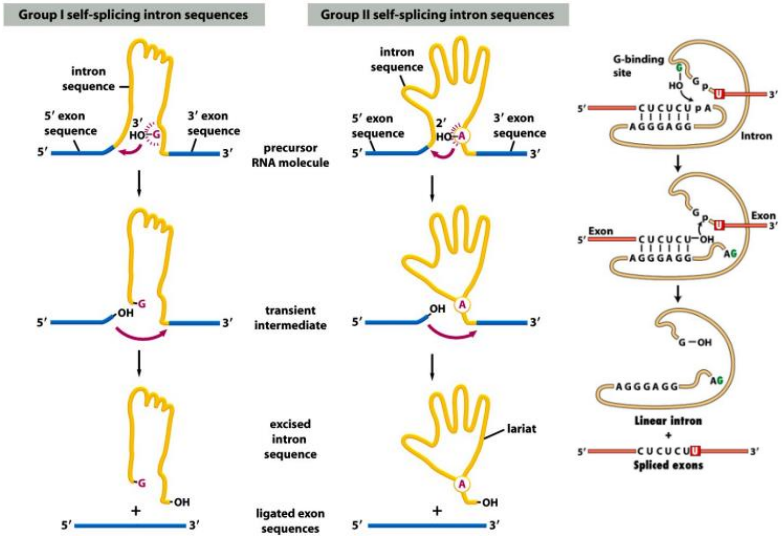
MACRONUCLEAR DEVELOPMENT:

a. Genome amplification
b. Genome reduction (up to 98%)
c. Chromosome fragmentation (Down to a 1 gene per chromosomes)
d. Precise excision of short Internal Eliminated Sequences (IES)
e. DNA unscrambling

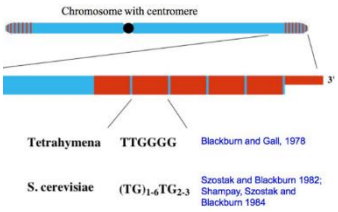


MAJOR DISCOVERIES USING CILIATES

DISCOVERY OF RIBOZYMES! Gave hypothesis for self-splicing.

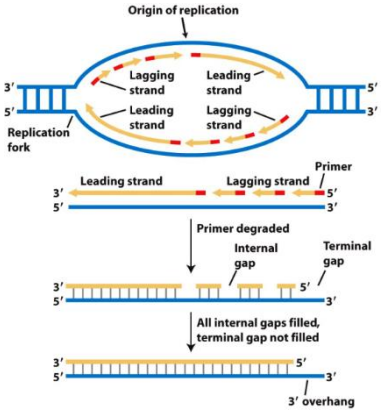


TELOMERASES: telomeric DNA contains simple tandem repeats.

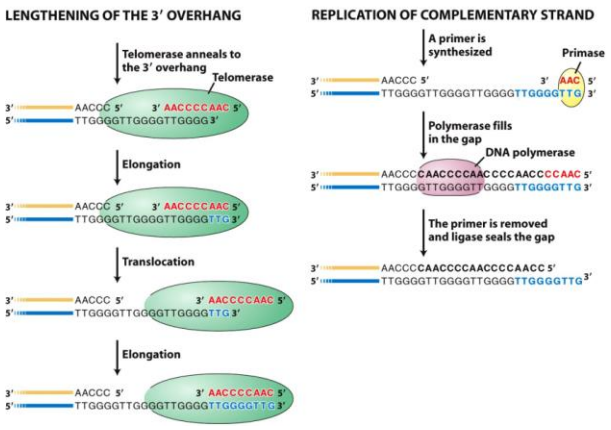


Tetrahymena thermophila contains abundant very short linear chromosomes. They end in TTGGGG repeats. How did the repeats get there?

THE REPLICATION PROBLEM AT CHROMOSOME ENDS:



→ Telomere lengthening → lengthening of the 3' overhang!



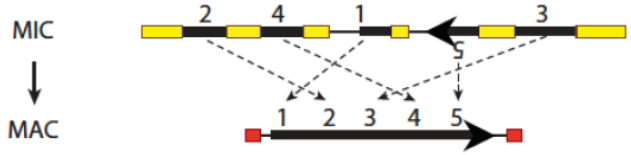
OTHER MAJOR DISCOVERIES USING CILIATES:

- First cell which showed synchronized division, which led to the first insights into the existence of mechanisms which control the cell cycle
- Identification and purification of the first cytoskeleton based motor protein such as dynein.
- Aid in the discovery of lysosomes and peroxisomes
- Early molecular identification of somatic genome rearrangement
- Discovery of the molecular structure telomeres, telomerase enzyme, the templating role of telomerase RNA and their roles in cellular senescence and chromosome healing (noble prize)
- Nobel prize-winning co-discovery of catalytic ribonucleic acid (ribozyme)
- Discovery of the function of histone acetylation
- Demonstration of the roles of posttranslational modification such as acetylation and glycylation on tubulins and discovery of the enzyme responsible for some of these modifications (glutamylation)
- Crystal structure of 40S ribosome in complex with its initiation factor elF1
- First demonstration that two of the "universal" stop codons, UAA and UAG, will code for the amino acid glutamine in some eukaryotes, leaving UGA as the only termination codon in these organisms
- Discovery of self-splicing RNA

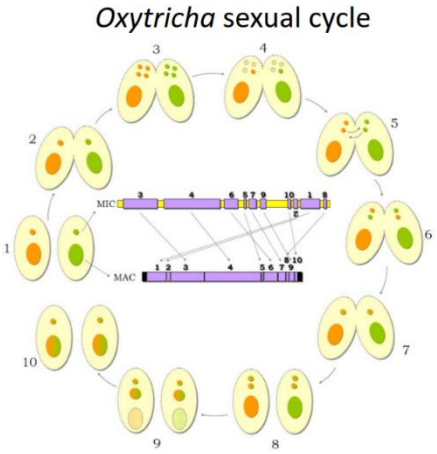
TWO DISTINCT TYPES OF GENOME REARRANGEMENT IN CILIATES:

Oligohymenophorea (Paramecium, Tetrahymena)	**Stichotrichs** (Oxytricha, Stylonychia)
- Long chromosomes - DNA excision and repair MAC Development: Long chromosomes, DNA excision and repair	- Short chromosomes (nanochromosomes) - DNA excision and repair - DNA unscrambling MAC Development: Short chromosomes (nanochromosomes). DNA excision and repair. DNA unscrampling

DNA Unscrambling:



OXYTRICHA



1. Vegetative stage, stress induces 2-10 = sexual reproduction
2. Contact between two cells (MIC begins meiosis)

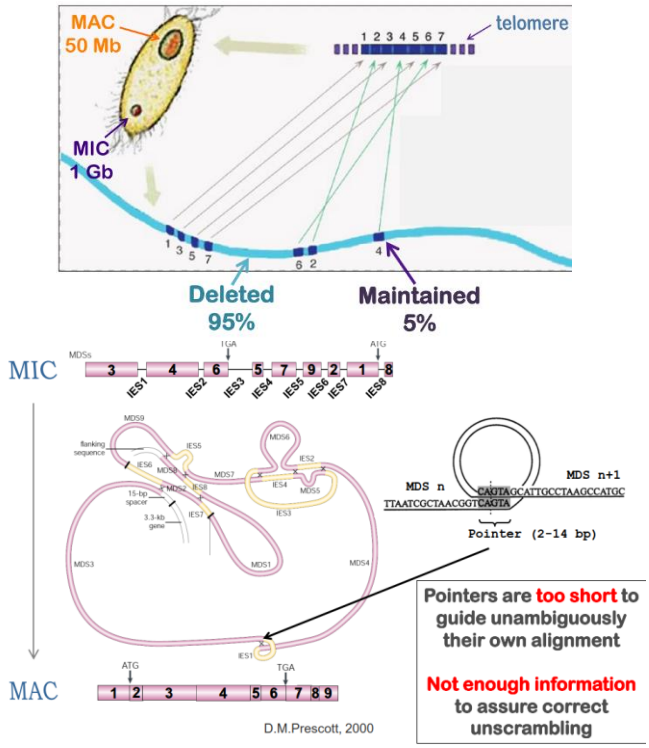
5. Mix Exchange

7. fusion of MICS - 2x haploid into diploid

8. MIC undergoes mitosis

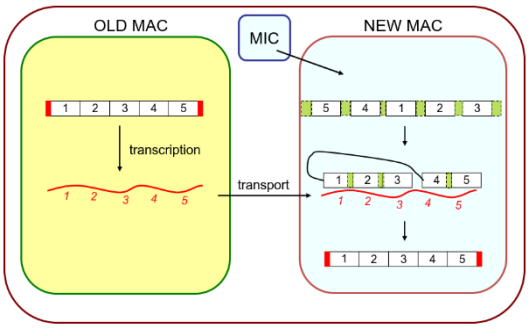
9. 1 MIC turns into new MAC

GENE UNSCRAMBLING IN OXYTRICHA:



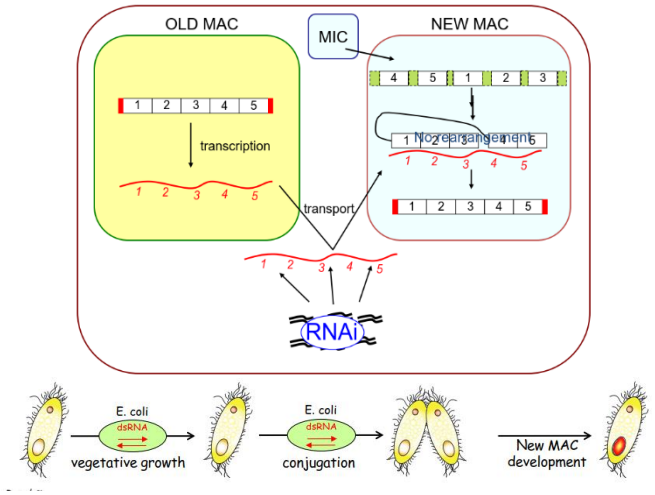
Old MAC used to produce RNA which contains the right sequence of genes, transported into the new MAC to help unscrambling (unsilencing) of the MIC/MAC.

TEMPLATE MODEL



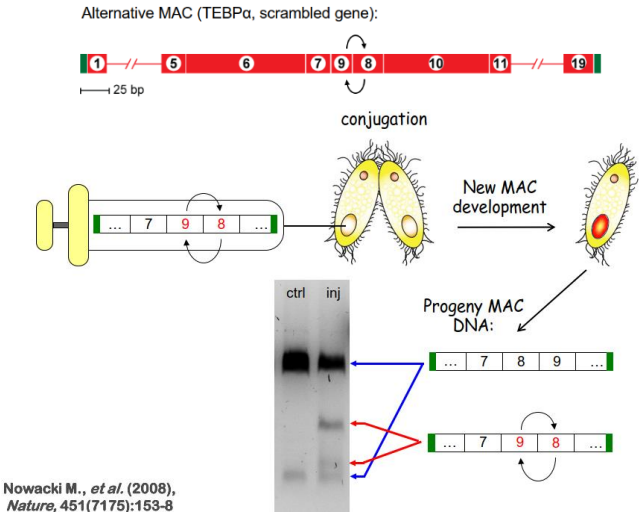
RNAI AGAINST PUTATIVE TEMPLATES

RNAi disrupts DNA rearrangement in a homology-dependent manner, leading to an excess of incorrectly processed molecules.



MICROINJECTION OF ALTERNATIVELY UNSCRAMBLED MAC CHROMOSOME

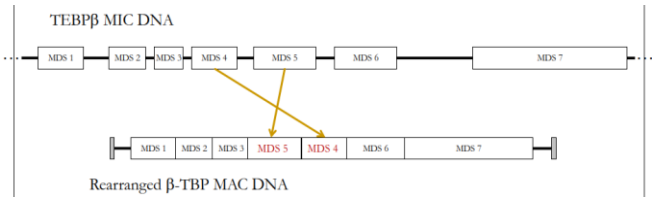
Alternative template for gene unscrambling! Features: altered MDS order (8 and 9 switched), telomeres, non-interrupted ORF, possibly functional protein, markers (single-nucleotide mismatches which allow to distinguish from the developed MAC, unique restriction sites).

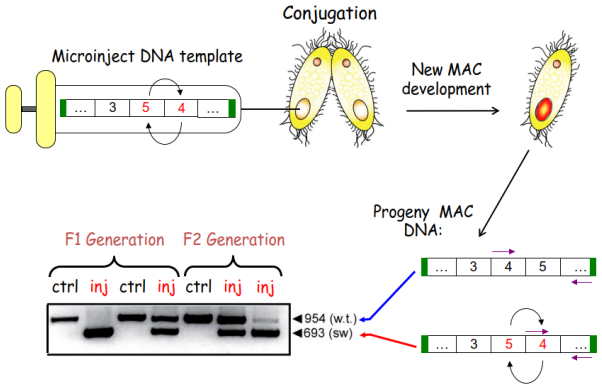


Result: Presence of alternatively unscrambled chromosome in the parental MAC leads to alternative rearrangement pattern in sexual progeny.

MICROINJECTION OF ALTERNATIVE TEBPB MAC TEMPLATE

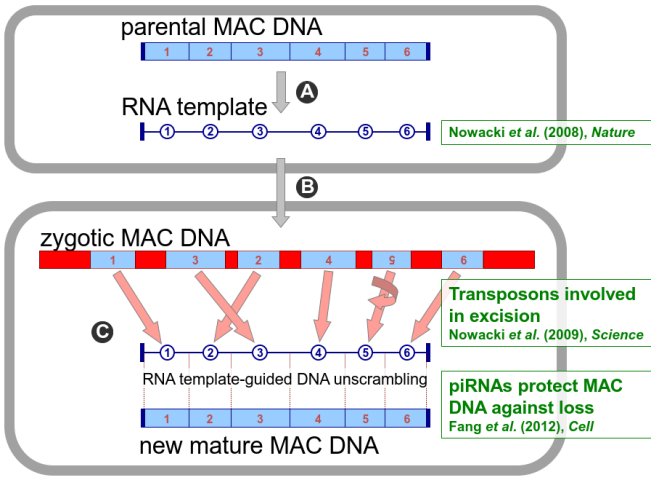
→ Switching Adjacent segments → scrambling a non-scrambled gene → Epigenetic re-programming of a DNA rearrangement pathway





Can we reprogram by alternative RNA template microinjection? Instead of Microinject DNA template we use an RNA template!

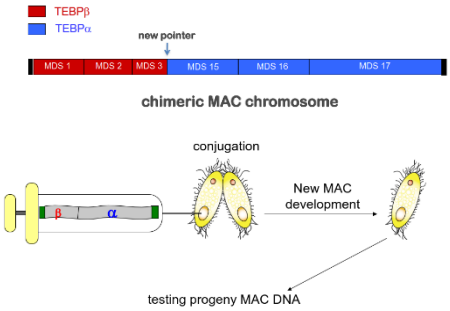
RNA TEMPLATE MODEL IN OXYTRICHA:



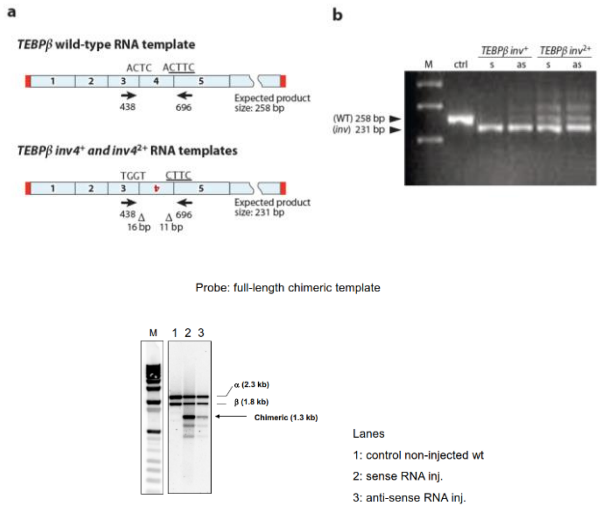
CAN RNA TEMPLATES GUIDE REARRANGEMENTS BETWEEN DISTANT DNA SEGMENTS?

- New pointer in the chimeric MAC chromosome
- Microinjection of TEBP-beta/TEBP-beta chimeric RNA template

- Screening for the chimeric construct
- Southern blot
- RNA-mediated epigenetic reprogramming of a DNA inversion in Oxytricha



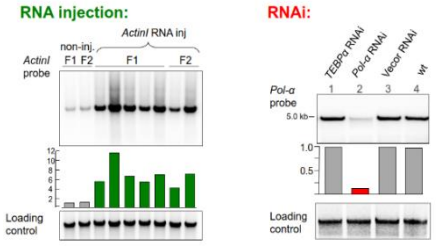
RNA-mediated epigenetic reprogramming of a DNA inversion in Oxytricha:



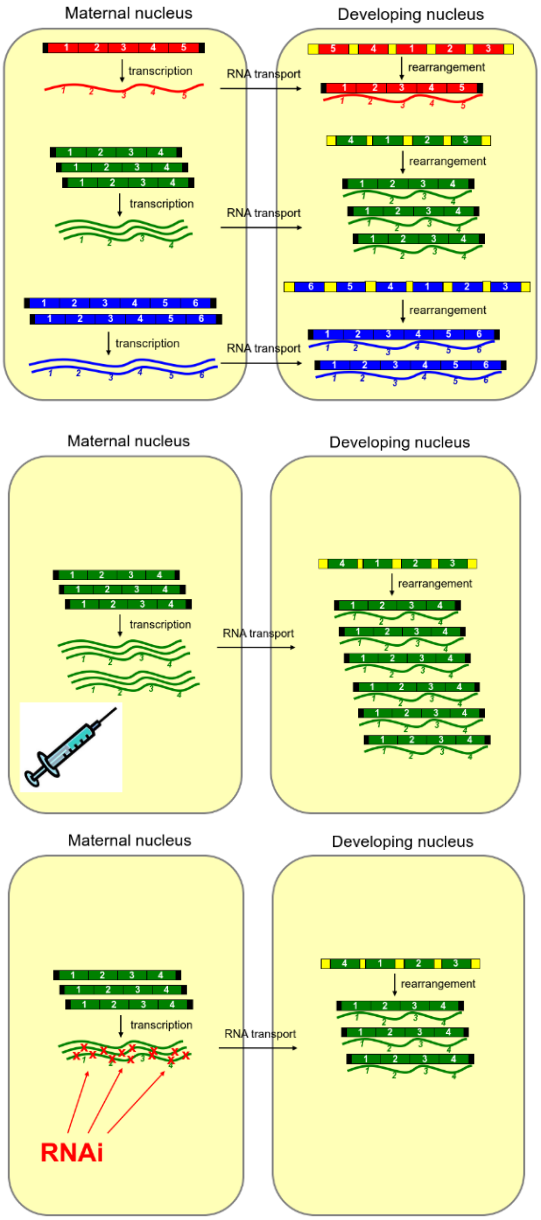
Testing epigenetic control of MAC chromosome copy number in Oxytricha trifallax: MAC chromosome are amplified to copy numbers ranging from 10^3 to 10^6 copies. Final chromosome copy number is achieved shortly after the polytene stage. Does chromosome copy number in the parental MAC have an influence on DNA amplification level in the new MAC? Does chromosome copy number in the new MAC depend on the amount of maternal RNA templates during conjugation?

Epigenetic inheritance of gene copy number in Oxytricha

Injection of RNA sequence increased amount of that sequence in gene copy number of the offspring's. But only if the sequence if already present in the genome (RNAi target guide RNA).



Epigenetic model for chromosome copy number regulation:

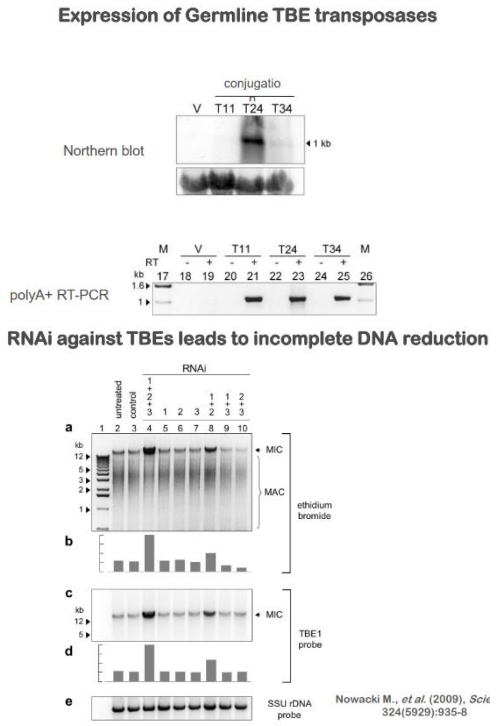


CONCLUSION

In Oxytricha, maternal RNA play an essential role during development by:

- providing templates for DNA unscrambling
- regulating chromosome amount in new macronucleus
- More chromosomes = more transcription = more RNA guides = says how many times DNA should be amplified in the new MAC.
- manipulation of guide RNA amount will lead to alteration of related gene sequence in future generations

WHAT PROTEIN MACHINERY IS INVOLVED IN MEDIATING REARRANGEMENTS?



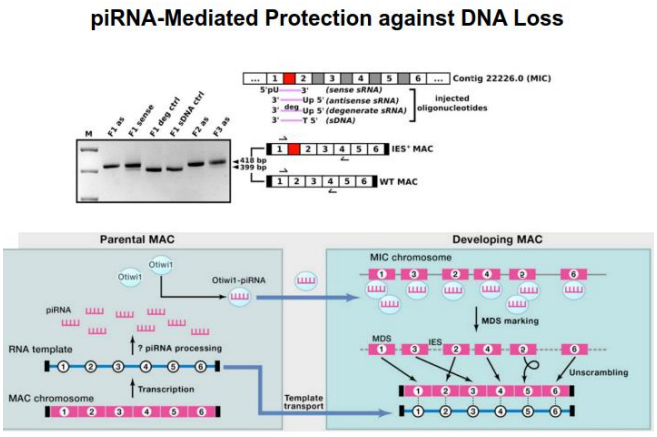
RNAi against TBEs leads to incomplete DNA reduction. Incomplete DNA rearrangement after TBE knockdown

➔ TBE transposons can persist in the germline genome as long as they clean themselves out of the somatic nucleus.

➔ Oxytricha uses transposases to completely remodel its genome.

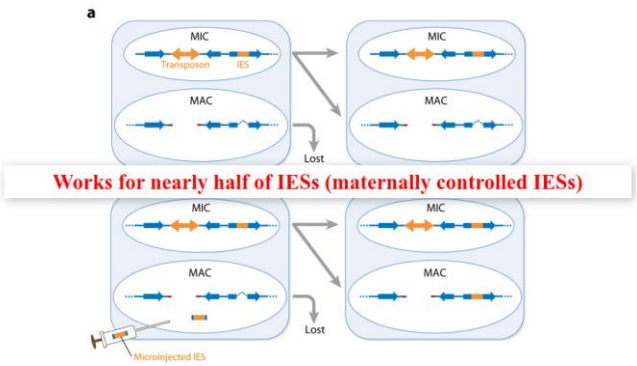
ROLE OF SMALL RNA DURING OXYTRICHA DEVELOPMENT: Piwi-interacting RNAs protect DNA against Loss during Oxytricha genome rearrangement.

Macronuclear development 20-26 h after conjugation.



EPIGENETIC INHIBITION OF DNA EXCISION

- Process is guided by RNA molecules
- Cell with germline and somatic genome
- Orange is absent in the somatic nucleus, not needed for gene expression

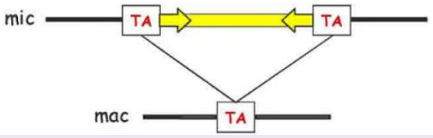


Maternally controlled IESs require Dcl2/3-produced scnRNAs for excision. Non-Maternally controlled IESs do not need scnRNAs for their excision.

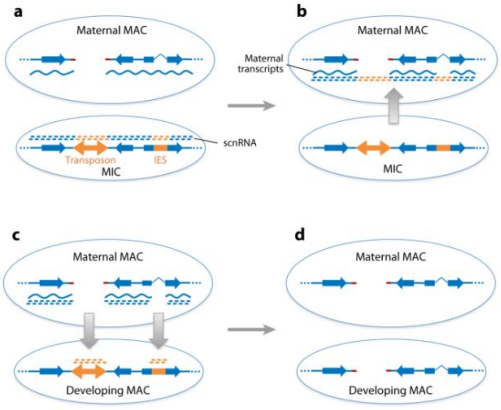
Sequence of single IEs is injected into the parental nucleus, where normally the sequence is absent. Microinjected IES is integrated in the next generation, epigenetic maintenance of DNA.

PARAMECIUM INTERNAL ELIMINATED SEQUENCES (IES)

- About 45'000 per haploid genome
- Present in genic and intergenic regions
- Short (28 bp length mode), single-copy, non-coding sequences
- 5'-TAYAGYNR-3' consensus sequence extending into IES resembles that of Tc1/Mariner transposons
- The longer the IES size, the more recently they evolved



THE SCAN RNA MODEL FOR PROGRAMMED GENOME REARRANGEMENTS

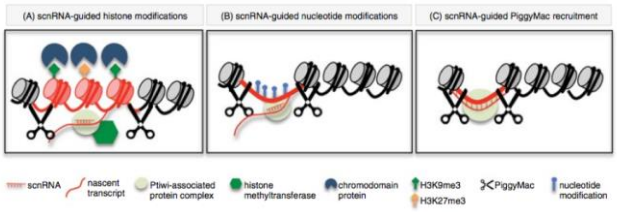


Complementary scRNA sequences to maternal scripts are being cut out.

- Mic is getting transcribed
- Transcripts are not cleaved
- Small RNA try to find complementary transcripts, orange parts cannot be found because the sequences are lost
- Complementary parts (homology with the genome) will stay but orange part is gone (process unknown)
- Piwies bring selected small RNAs.

Three hypothesis behind process.

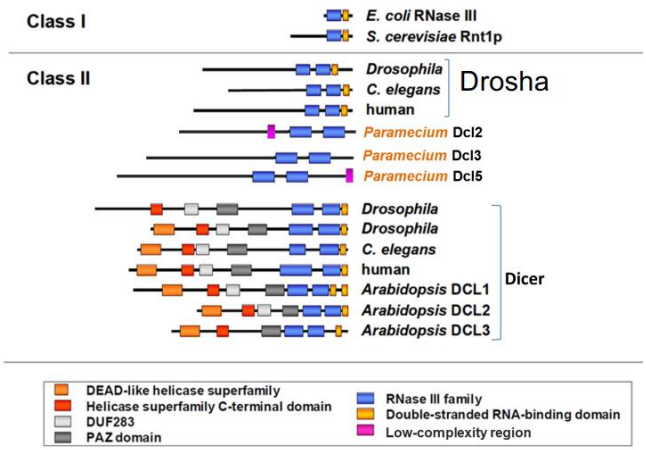
MODELS FOR SRNA-GUIDED DNA EXCISION



New class of small RNAs in Paramecium?

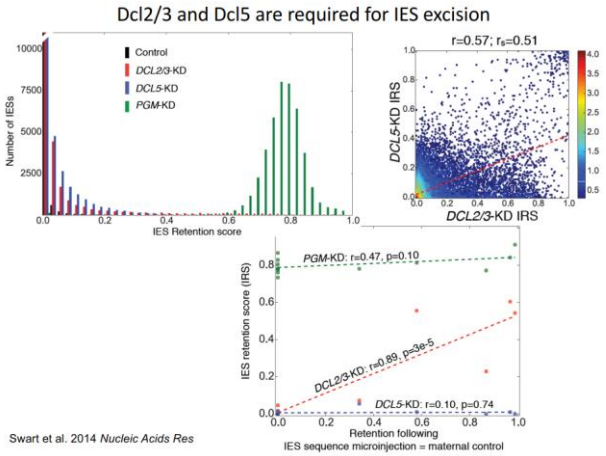
- Characterized products of different Dcl proteins
  - Do Dcl2 and Dcl3 produce the same sRNA molecules?
  - Is Dcl5 responsible for the late 26-29nt sRNAs
  - Possible role for the 26-29nt sRNAs?

DOMAIN ARCHITECTURE OF RNASE III ENZYMES



DCL2/3 AND DCL5 ARE REQUIRED FOR IES EXCISION:

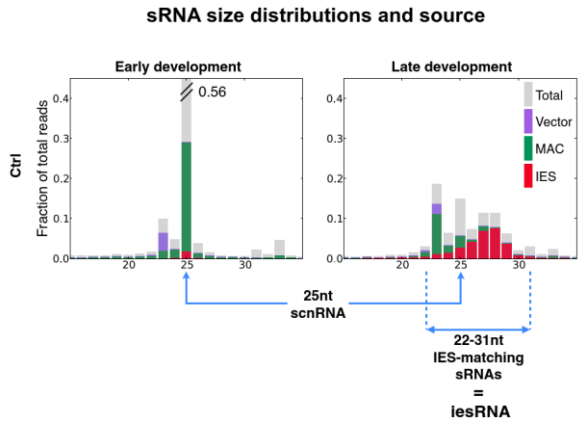
- Transposon in the wt (should have 800 copied = 1 (100%), if it would be 400 → means retention score 0.5
- If completely excised, score = 0
- Bioinformatically: silence dicers, look at transposons in the genome
- Silence dcl5, look at effect, if they are involved in the same pathway you would get the same effect
- No correlation between the proteins
- Each point = excised transposon or IES
- If correlation, should be a diagonal line

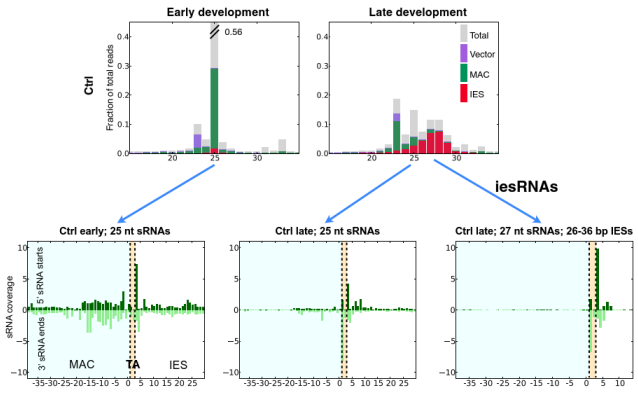


Swart et al. 2014 *Nucleic Acids Res*

SRNA SIZE DISTRIBUTIONS AND SOURCE:

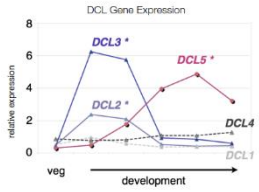
- 22-32nt IES-matching sRNAs = iesRNA
- Precisely excises small RNA to find the position where IES starts/stops
- scnRNAs overlap MAC/Ies junctions and concentrate at IES ends
- iesRNAs map exclusively to IES and concentrate at IES ends
- TA: first two nucleotides of the IES
- Each peak is a single nucleotide of a sRNA





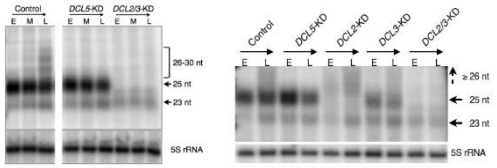
- scnRNAs overlap MAC/IES junctions and concentrate at IES ends
- iesRNAs map exclusively to IESs and concentrate at IES ends

ANALYSIS OF DCL PROTEINS CLEAVAGE PRODUCTS



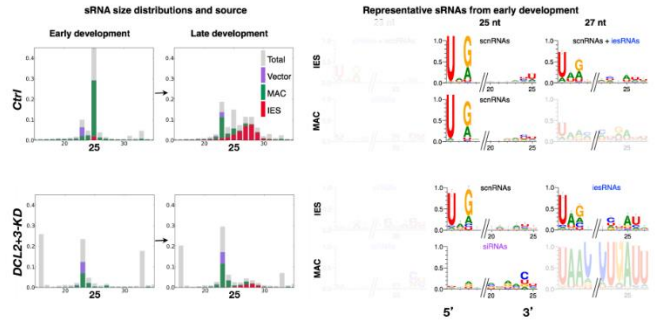
Silencing of Dcl2, Dcl3 and Dcl5 followed by sRNA sequencing and analysis.

Knockdowns of Dcls causes dramatic changes in developmental sRNAs.

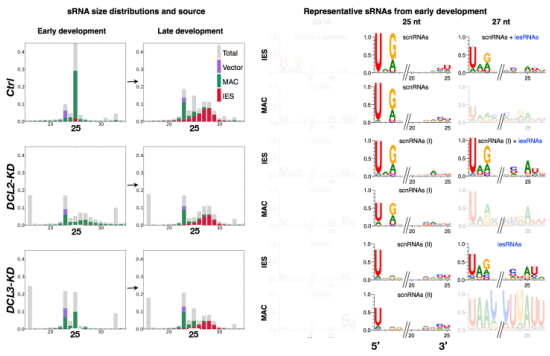


IESRNAS ARE A NEW CLASS OF DEVELOPMENTAL SRNAS

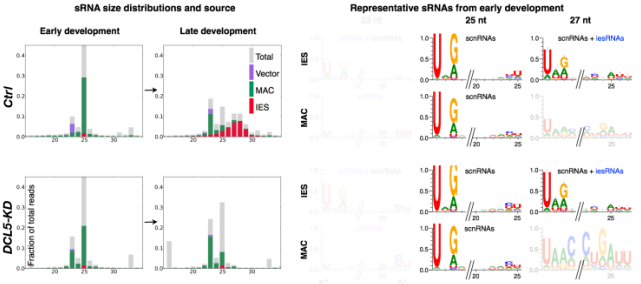
A combined *DCL2* and *DCL3* knockdown severely depletes scnRNAs

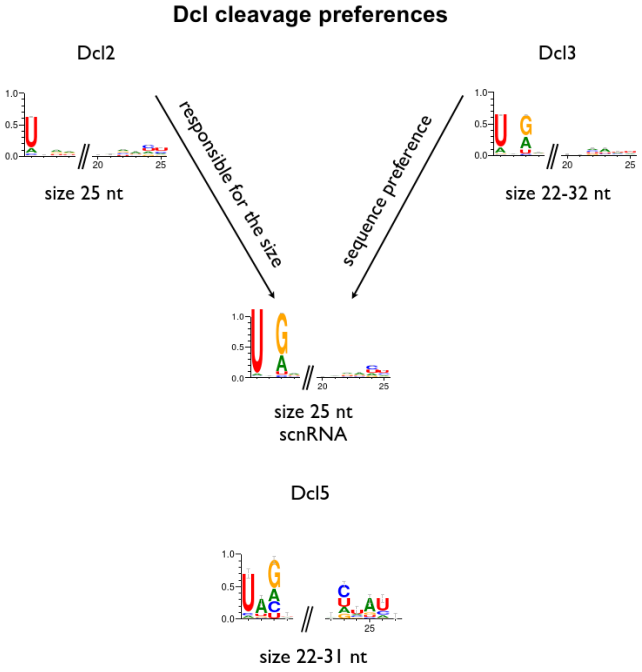


Independent *DCL2* and *DCL3* knockdowns reveal different sRNA size and length preferences



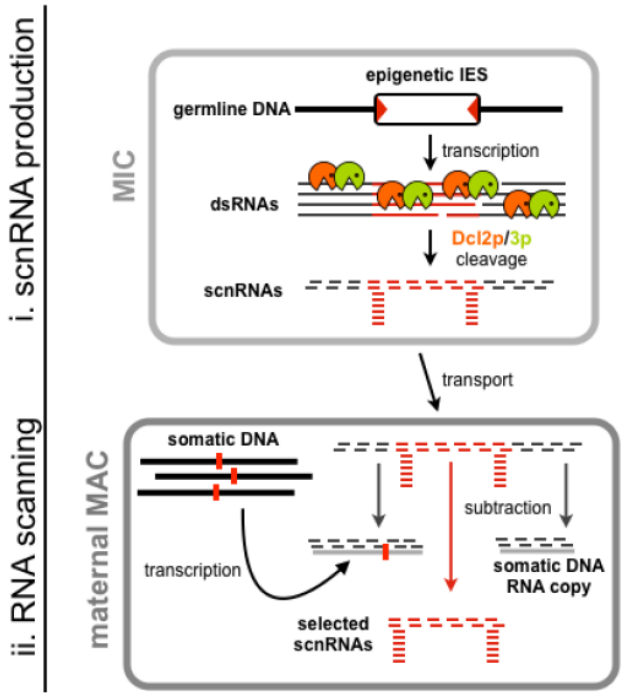
Knockdown of *DCL5* severely depletes iesRNAs

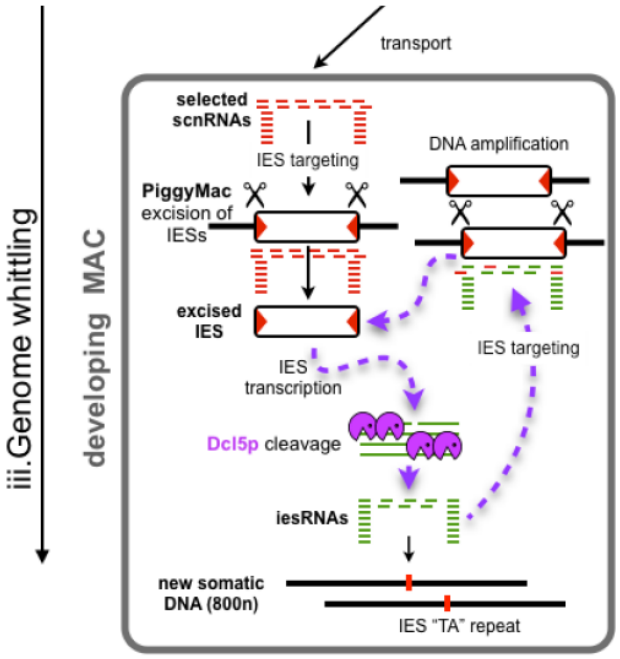




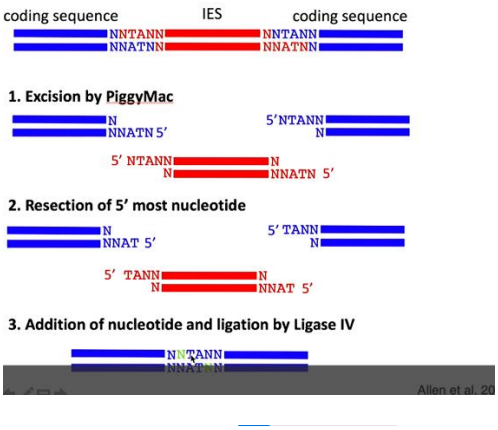
PARAMECIUM SRNAS CONCENTRATE AT IES BOUNDARIES

- iesRNAs locate 10 fold more frequently at the terminal TNG than at internal TNGs, suggesting that iesRNAs production is biased towards the ends of IES
- scnRNA mapping onto the genome: presence of sequence preference pairs sequentially with the IES ends, more abundant at the end

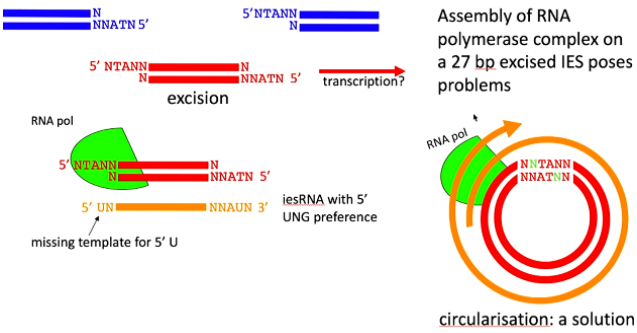




IES EXCISION - CURRENT MODEL



Are iesRNAs formed from excised IESs?



Assembly of RNA polymerase complex on a 27 bp excised IES poses problems

circularisation: a solution

THE CONCATEMER MODEL

Most IESs are too short to form circles individually. Concatenation and circcu-latirsation of short IESs would allow transcription of dsRNA Dcl5 template.

Some iesRNAs map to IES-IES junctions.

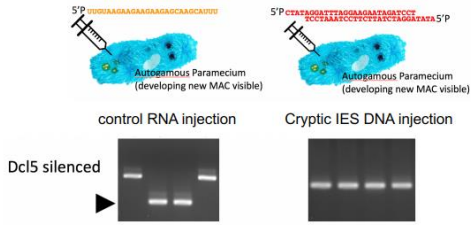
Long IES (>200 bp)

Experimental approach: 'cryptic' IES excision.

- 27 bp long
- Consensus 5'TATAG2
- Located in unknown genes

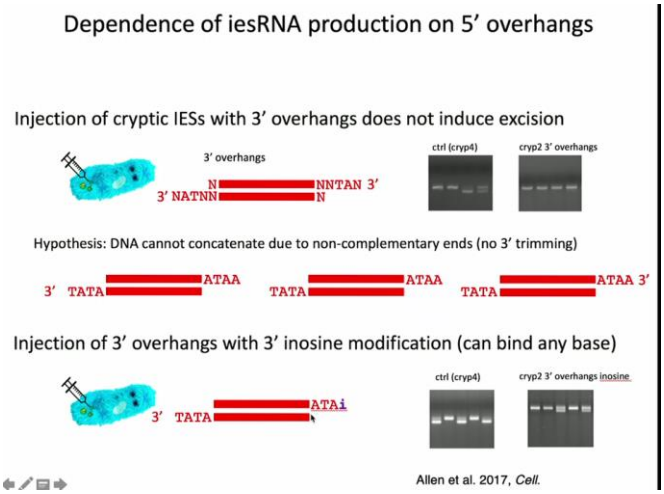
Cryptic IES excision induced by injection of dsDNA corresponding to theoretical excision product. Excision is precise and takes place at TA repeats - just like IES excision.

Is cryptic IES excision dependent on iesRNAs? Dcl5 silencing

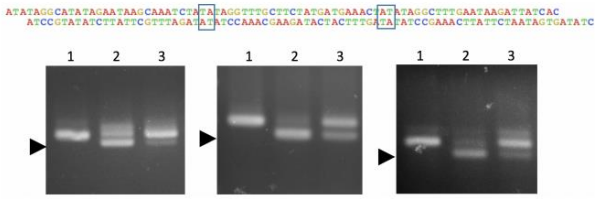


➔ Silence Dcl5 and provide the cell with iesRNA and double stranded DNA
  - Dcl5 is needed if the cell gets DAN excised
  - If the cell gets RNA, it doesn't need to excise

Dependence of iesRNA production on 5' overhangs

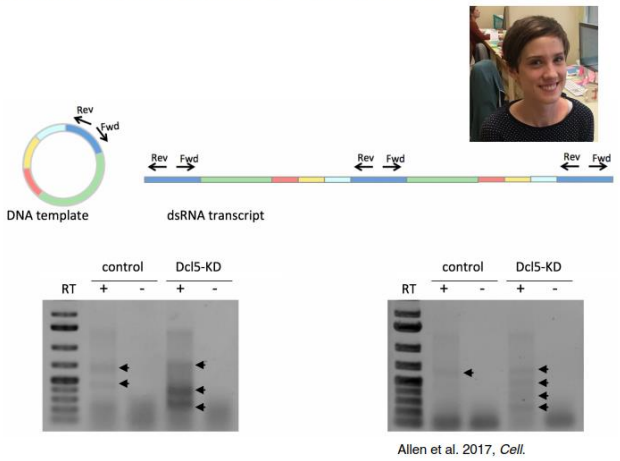


Injection of three concatenated cryptic IESs



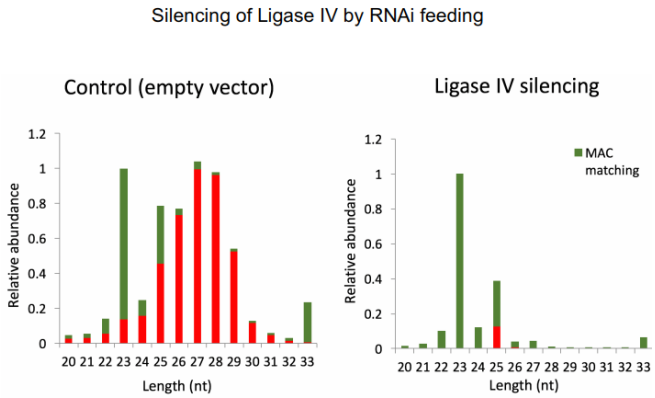
- And 3' overhang prevents edge excision but not middle

Amplification of putative iesRNA precursors (Iris Hug)



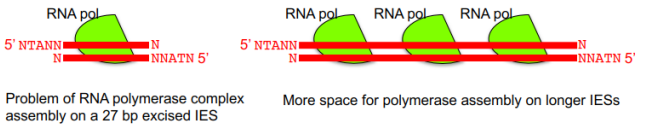
Sequencing identifies RNA corresponding to concatemers of IESs: A concatemers is a long continuous DNA molecule that contains multiple copies of the same DNA sequence linked in serues. These polymeric molecules are usually copies of an entire genome linked end to end and separated by cos sites (a protein binding nucleotide sequence that occurs once in each copy of the genome).

How are putative concatemers formed? Ligase IV?

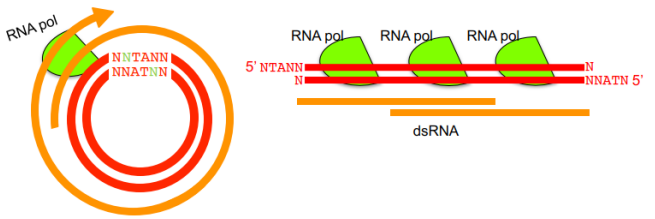


➔ Silencing is never 100 % effective

Ligase IV-KD iesRNAs and the concatemer model

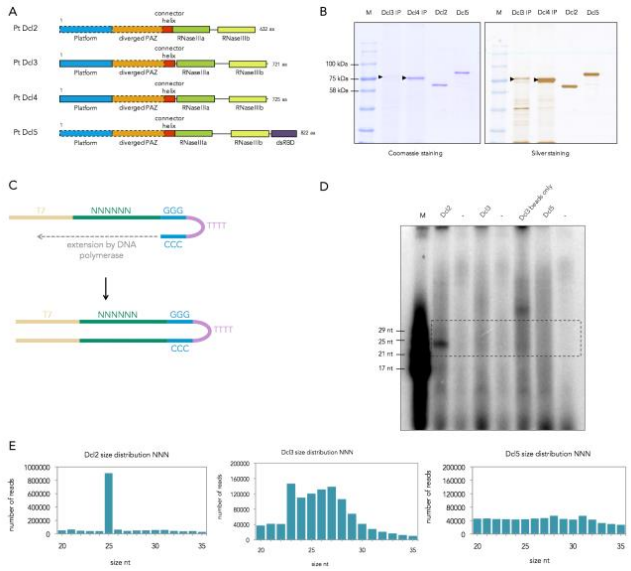


Ligase IV-KD iesRNAs and the concatemer model



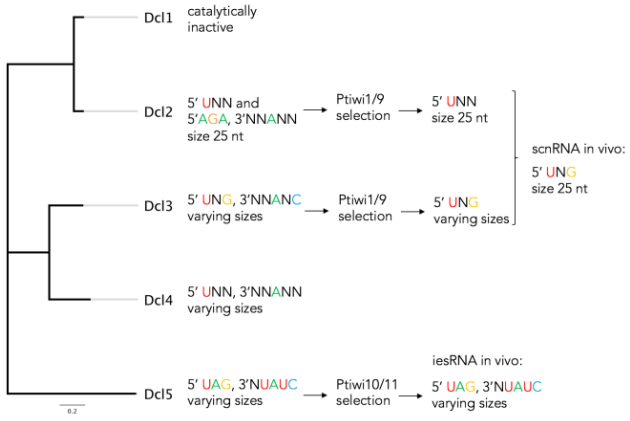
Dicer-like enzymes with sequence cleavage preferences:

- Dcl2: 25nt
- Dcl3: 22-32 nt
- Dcl5: 22-31 nt

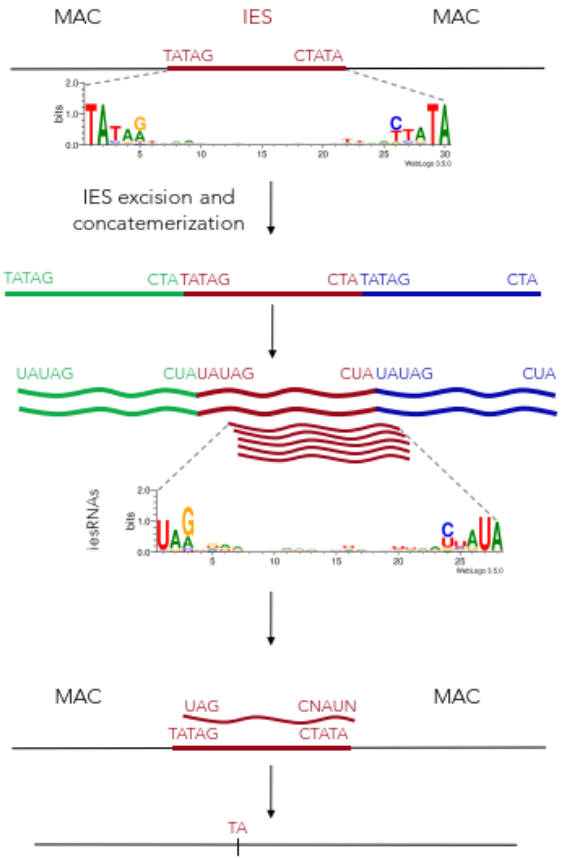


- In vitro: purify protein, overexpress Dicer, mix with longer RNA, seep sequencing on the product → see the cut RNA pieces
- Extended by polymerase, transcribes a complementary sequence, double stranded RNA hairpin, give it to Dicer and let it be cleaved
- Dcl5:
  - Random sequences are not cleaved in the same manner
  - Cleaves preferred at the border of IES
  - Enzymes know where IES are joined and cleaves there, not dependent on the length of IES

Summary of Dcl cleavage preferences



Small RNAs are cleaved by dicer and target other IES, over-lapping because if the Dcl5 can cleave exactly where the IES is, it is perfectly complementary.

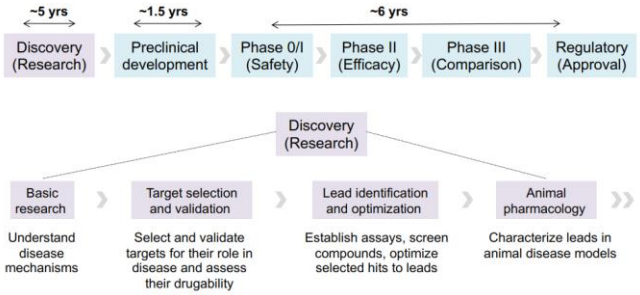


- ➔ **Dicer AND Argonaut have preferences!!!!**
- ➔ This small RNAs that are cleaved by dicer are used
- ➔ They have a function – targeting other IES
- ➔ What is the point behind overlapping?
- ➔ If the DCL5 can cleave exactly to where the IES is it can have perfect complementary products.

NUCLEIC ACID-BASED DRUGS

CHANGES IN APPROACHES TO DRUG DISCOVERY OVER TIME

- Folk medicine: Observation, experimentation
- "Classical" drug discovery: Isolation and analysis of natural products, structure elucidation, structural modification. E.g.: Morphine, Caffeine, Cocaine etc.
- Modern drug discovery. Discovery and development comprises distinct phases, many of which run in parallel. Each project is different



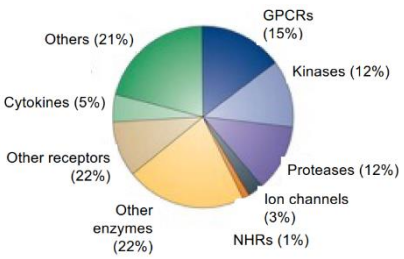
PROPERTIES OF DRUGS

Biochemical systems encountered by a drug in vivo are extremely complex: many factors contribute to its final pharmacological effect. **Pharmacodynamics (PD)** is the study of how a drug interacts with its target. **Pharmacokinetics (PK)** is how a drug distributes throughout the body in reaching its target. Both are equally important to successful drug development and in medicinal chemistry are optimized together.

Successful drug discovery and development requires initiation of many projects in order to bring one drug successfully to market. From 2009 through 2017, CDER has averaged abut 33 novel drug approvals per year. Drugs:

- Produce the desired therapeutic effect
- Show acceptable eves of toxicity or side-effects
- Bind to the target molecule with high affinity
- Show selectivity for the target over other structurally-similar biomolecules
- Show favorable absorption, distribution, metabolism and excretion in vivo
- Can be synthesized economically on large scale

TARGETS OF MEDICINAL CHEMISTRY



DRUGS TARGETING DNA

Drugs targeting DNA and RNA in the cell represent a small but significant proportion of historical and new medicines. DNAs and RNAs can be considered as receptors for drugs, but their heavily solvated structures present difficulties for the development of potent Drugs. The design of small-molecule drugs which bind sequence-selectively to DNA/RNA has not been achieved. Many natural product drugs targeting DNA/RNA have been discovered.

Generally, DNA-targeting drugs are toxic to healthy cells and therefore they are reserved for serious disease (e.g. cancer). Toxicity is particularly prominent where rapid cell division occurs (GI tract, mucosa, and hair).

Almost all small-molecule DNA- (and RNA-) targeting drugs are natural products and were discovered serendipitously. Drugs targeting the DNA structure include the intercalating agents, alkylating agents, DNA cleavers or nucleoside analogues.

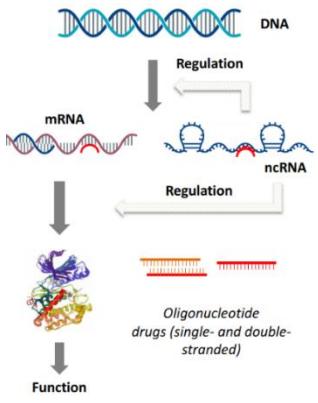
- Alkylating agents covalently modify DNA and are irreversible binders
- Reversible binders, e.g., intercalators, minor groove binders: bind the helix and disrupt replication/transcription
- Cleavage reagents covalently modify the DNA and cause cleavage

Drugs targeting indirectly DNA or RNA include inhibitors of enzymes such as topoisomerases and Polymerases, e.g. nucleoside analogues mimicking DNA building blocks.

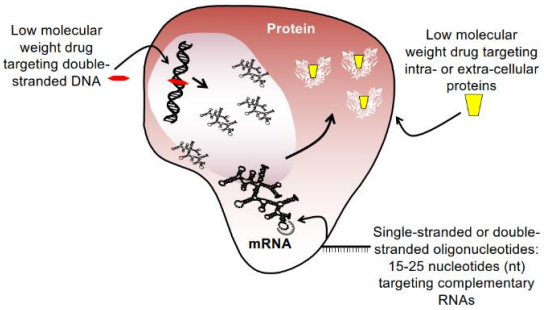
STRUCTURE OF DNA: DNA has 4 building blocks comprising purine (Adenosine, Guanosine) and pyrimidine (Cytidine, Thymidine) nucleotides. Nucleotides are linked by a negatively-charged phosphodiester backbone. Pairing through H-bonds between DNA strands is specific: A-T, G-C.

Under physiological conditions there is little tautomerism of DNA nucleobases. Hydrophobic forces between the aromatic systems of the bases are strong. The phosphate groups are exposed to water on the outside of the helix.

SMALL MOLECULE- AND POLYNUCLEOTIDE-TARGETING DRUGS:



DRUGS ACTING ON CELLULAR NUCLEIC ACIDS:



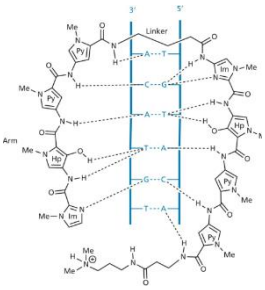
INTERCALATING AGENTS: bind between nucleobases and disrupt replication/transcription. They are hydrophobic heterocyclic ring molecules that resemble the ring structure of base pairs. . There are 3 main drug classes: acridines, actinomycins and anthracylins

- Drugs are fused ring systems which allow for charge transfer and hydrophobic interactions with the bases, vertically distorting regular helical structure
- Compounds usually prefer G-C rich regions: positive charges may increase affinity by binding electrostatically to helix phosphate groups
- Planar tricyclic systems with N^+H_3 groups are common structures in chemotherapeutics
- Dactinomycin (natural antibiotic from Streptomyces) binds DNA & prevents unwinding prior to transcription
  - It favours G-C base pairs, peptidic groups interact with the helix via H-bonding
  - The resulting Complex is very stable, duplex unwinding is prevented.
- The anthracyclines are natural antibiotics from *Streptomyces peucetius*. They bind DNA and prevent duplex unwinding prior to replication
  - The N^+H_3 function is important - without it the molecule is poorly active
  - Often cardiotoxic
- Intercalation into DNA inhibits topoisomerase II, an enzyme which temporarily cleaves the DNA during unwinding
- Mitoxantrone is an anthracine dione with similar mechanism of action to the anthracyclines
  - easier to synthesize and carries less risk of cardiotoxicity
  - intercalates into the DNA helix

RATIONAL DESIGN OF MINOR GROOVE-BINDING DRUGS:

- Proteins generally bind dsDNA in the major groove
- Some natural products bind dsDNA in the minor groove
- Aromatic rings connected by single bonds allow for torsional rotation in order to fit into the helical curve of the minor groove, particularly in A-T rich regions
- NH groups on the interior curve of the structure form H-bonds with A-T bases
- Netropsin is specific for A-T regions: its pyrrole rings contact the C-2 positions of adenines: the drug widens the groove and disturbs topoisomerase II

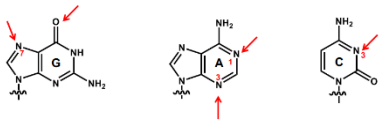
RATIONAL DESIGN OF DRUGS TARGETING TRANSCRIPTION:



- Rational specific-drug design with polyamide structures

DNA ALKYLATORS: NUCLEOPHILIC GROUPS ON THE BASES

- Nucleobases have nucleophilic heteroatoms: generally, nucleophilicity is weak and depends on the electrophile
- Nucleophilicity follows N-7 G > N-3 A > N-7 A > N-3 G > N-1 A, N-1 C

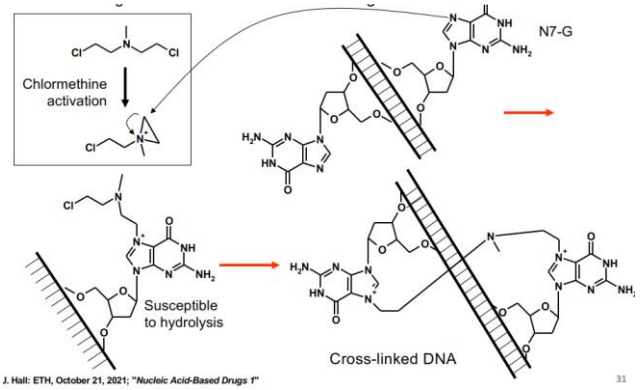


- A variety of alkylation reactions are characterized on double-stranded DNA

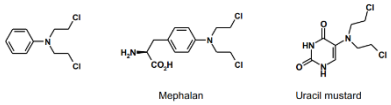
DRUGS ACTING ON DNA: ALKYLATING AGENTS: Alkylation is the transfer of an alkyl group (C_nH_{2n+1}).

- DNA alkylators react with DNA to form covalent bonds
- Mustard gas is mutagenic, vesicular and a strong carcinogen: exposure leads to white blood cell depletion

  - Nitrogen mustards for treatment of Acute Lymphoblastic Leukaemia in children began in 1970s
  - Mustards react non-selectively with the amino groups in DNA in both aberrant and normal cells, causing side effects
  - Chlormethine was the first mustard to be tested in patients
  - Chlormethine reacts with water in blood and tissues, and cross-links DNA
  - Cross-linking is more effective than mono-linking



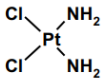
- Mephalan mimics phenylalanine: it is given orally for multiple myeloma, ovarian and breast cancers; L- and D-isomers are equi-potent



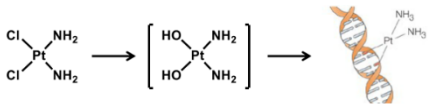
- Uracil mustard used to treat chronic lymphatic leukaemia
- Conjugation to phenylalanine and uracil was intended to improve selectivity of the mustards for tumour cells or tumour DNA

METABOLISM-ACTIVATED DNA ALKYLATING AGENTS:

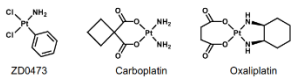
- Cisplatin is an important chemotherapeutic drug used to treat testicular and ovarian, cervical, head and neck cancers



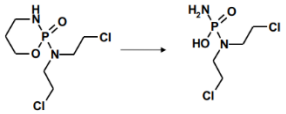
- It was discovered (1965) serendipitously during electrolysis of E. coli in a water bath with platinum electrodes
- It causes bacteria to cease to divide
- *In vivo* cisplatin is hydrolysed, and the complex reacts and cross-links G-rich DNA
- Pt-DNA adducts in the nucleus activate processes that mediate cytotoxicity incl. replication arrest, transcription inhibition, cell cycle arrest. DNA repair and apoptosis



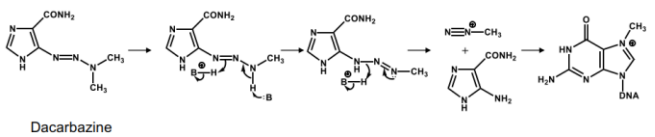
- Cisplatin causes nausea/vomiting (treated with ondansetron; 5-HT3 receptor antagonist)
- Derivates of cisplatin have been developed with improved side-effect profiles (carboplatin), oral activity (JM216) and improved tumor resistance
- Oxaliplatin is postulated to work by inhibiting ribosome biogenesis



- Cyclophosphamide is metabolically-activated in the liver

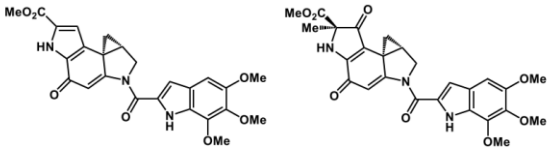


- The methydiazonium moiety of dacarbazine is produced in the liver



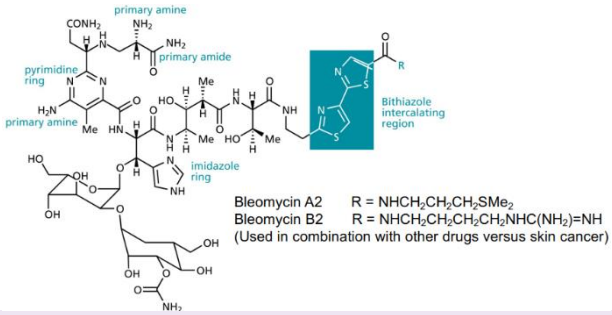
NATURAL PRODUCT DNA ALKYLATING AGENTS:

- N-3 of A reacts as a nucleophile at the cyclopropyl ring in A-T-rich regions of DNA
- The electrophilicity of the molecule is greatly reduced by the conjugated N's



DRUGS ACTING ON DNA: DNA CLEAVAGE

- Bleomycins were isolated from Streptomyces verticillus
- N-atoms complex Fe2+ ion which is oxidized and abstracts H from DNA to generate radicals: these reacts with oxygen causing DNA cleavage



SUMMARY PART 1

- Drug discovery is a long expensive process
- Many projects are initiated to bring one project successfully to market
- A drug encounters extremely complex biochemical systems in vivo and therefore many factors contribute to its pharmacological effects
- Both PK and PD are equally important to successful drug development
- Natural products are of key importance in historical and modern medicines
- Drugs targeting DNA and RNA represent a small proportion of medicines
- DNA and RNA can be considered as a receptor, but their heavily solvated structures present difficulties for the development of potent/selective drugs
- "All things are poison and nothing is without poison, only the dose permit something to be not poisonous"
- A large fraction of cancers are caused by damage to DNA which cannot be repaired: drugs targeting DNA replication prevent aberrant cell growth
- Toxicity of DNA targeting drugs occurs in tissues where rapid cell division occurs: mucosa, hair, bone marrow
- The selectivity of DNA-targeting drugs is somewhat empirical

RNA-TARGETING OLIGONUCLEOTIDE DRUGS

- Oligonucleotide drugs (~15-25 nt) hybridize to target and modulate RNA function
- Any RNA can be targeted: specificity is given
- The same drug development infrastructure is used for all oligonucleotide drugs
- Important developments since ~2000s:
  - Elucidation of their mechanisms: RNA sequestration, enzyme-directed cleavage, splice modulation
  - Discovery of siRNAs
  - Emergence of non-coding RNA biology, incl. miRNAs and long ncRNAs

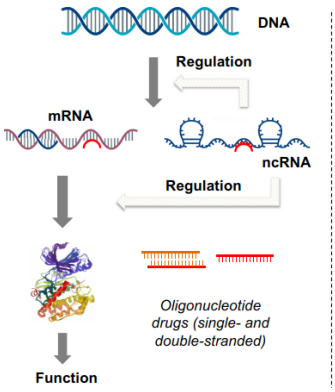
The minimum number of nucleotides in an oligonucleotide required to achieve a theoretical specificity is approximately 16:

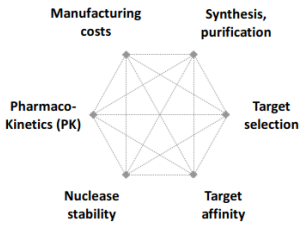
- Human genome is: 3 * 10^9 base pairs
- 4 building blocks: A, G, C, T. 4^x = 3 * 10^9, x = 16

The statistical calculation is only partially correct because:

- Particular RNA sequences occur more frequently than predicted
- The longer an oligonucleotide becomes, the higher affinity it shows for a complementary sequence, but also the higher is its tolerance for mismatches
- Binding of oligonucleotides to mismatched mRNAs gives loss of selectivity
- False pairing in the centre of a sequence can give $\Delta\Delta G^0$ of 5 kcal/mol
- In the design of RNA-targeting oligonucleotides, sequence homology to all other possible sequences in the genome is incorporated
- In rare cases close homology can be advantageous, e.g. Bcl-2 and Bcl-xL

HURDLES TO THE DEVELOPMENT OF OLIGONUCLEOTIDE DRUGS





GENERAL PROPERTIES OF ANTISENSE OLIGONUCLEOTIDES

- Non-natural oligonucleotides composed of mixed chemistries, depending on the application
- Lengths typically 16-25nt to ensure potency and selectivity
- Poor pharmacokinetic properties: high molecular weights, polyanionic
- Distribute to most tissues, particularly liver, adipose, spleen, kidney, bone marrow, intestine, macrophages
- Not orally bioavailable, poor understanding of cell uptake
- Manufactured by specialist companies

PK OF OLIGONUCLEOTIDES: IN VIVO STABILITY AND RETENTION

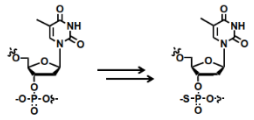
Oligonucleotide drugs composed of natural RNA or DNA are rapidly removed from the body:

- Cleaved by cellular nucleases
- Excreted in the urine (renal excretion)

Medicinal chemists modified the phosphodiester backbone of oligonucleotide drugs with the phosphorthioate functionality. This resulted in:

- Binding to plasma proteins - increased distribution, reduced renal excretion
- Increased resistance to nuclease cleavage

But:



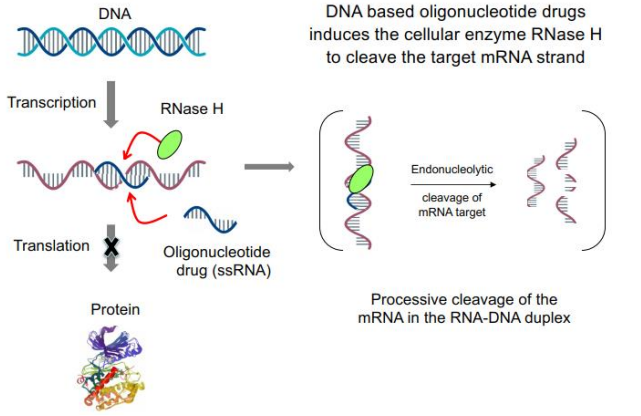
- Reduce affinity to RNA
- Introduction of diastereoisomers

MECHANISMS OF ACTION OF OLIGONUCLEOTIDE DRUGS

Oligonucleotide drugs act by a variety of mechanisms, dependent on their target RNAs and their chemistry:

- Terminating mechanisms result in degradation of the target RNA: they include induction of RNase H and RISC for which there are stringent chemistry requirements
- Non-terminating mechanisms do not cleave the RNA: they include drugs to modulate splicing or the function of non-coding RNAs
- Single-stranded antisense ASOs work by both terminating and non-terminating mechanisms; double-stranded siRNAs work by terminating mechanisms

RNase H directed cleavage of mRNA by oligonucleotides



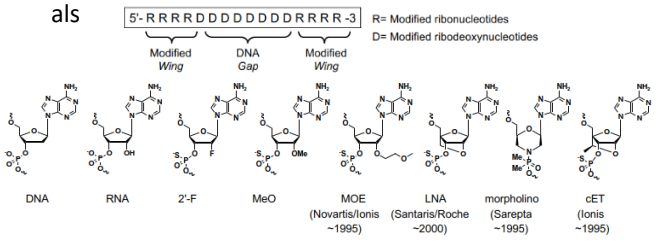
CONTRIBUTIONS OF MEDICINAL CHEMISTRY TO OLIGONUCLEOTIDE DRUGS:

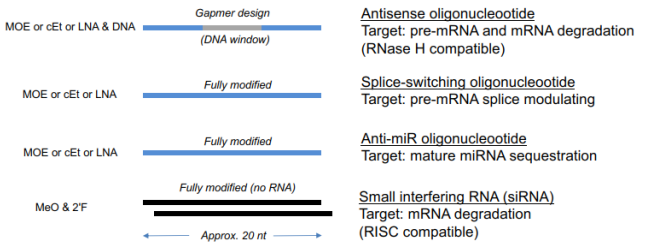
The three most limiting factors of first generation oligonucleotides were:

- Sensitivity to nucleases in vivo
- Binding affinity for RNA
- Requirements for RNase H activity

Hundreds of modifications were designed/synthesized by chemists in search of a nucleoside modification which would satisfy all criteria, including unusual structures.

- None of the nuclease-stable, high affinity modifications are able to activate RNase H: thus they were inactive for the most part
- The oligonucleotide PS-Gapmer was a (very good) compromise solution
- High affinity nucleotide modifications on the end of the sequence provided high binding affinity for RNA and excellent protection against nucleases
- The central gap of at least 5 deoxynucleotides allowed activation of RNase H
- The molecule is fully phosphorothioated
- Dozens of gapmer ASOs have been tested in clinical trials



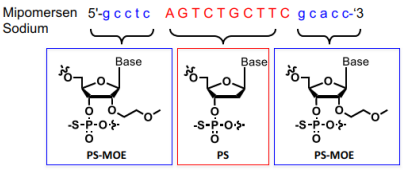


TECHNICAL FEASIBILITY: SYNTHESIS OF AN LNA

Synthesis of modified nucleotides is enormously demanding.

CLINICAL DEVELOPMENT OF MIPOMERSEN:

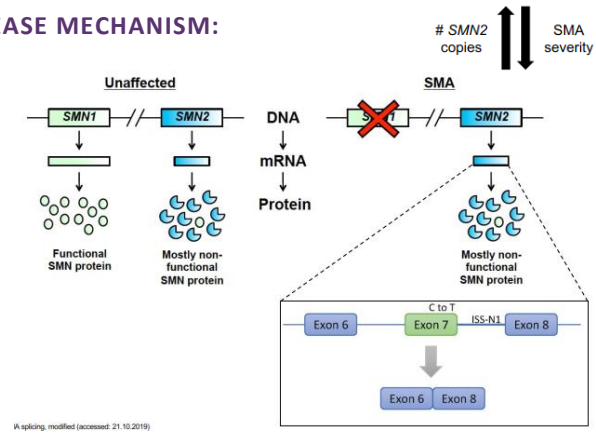
- Mipomersen: a MOE Gapmer oligonucleotide approved 2013



- Proof of concept: first approved systemically-delivered oligonucleotide drug
- Mipomersen targets Apolipoprotein B (Apo-B) mRNA, inhibits Apo-B protein, lowers circulating LDL-C
- Apo-B protein is an ideal target for an oligonucleotide drugs:
  - Synthesized in the liver
  - Genetically validated
- The dose-dependent activity of Mipomersen could by assayed in blood samples
- Showed decreased Apo-B protein and decreased LDL-cholesterol in healthy individuals and also LDL-cholesterol and triglycerides and ApoCIII, Lp(a) in patients with high cholesterol and/or familial hypercholesterolemia

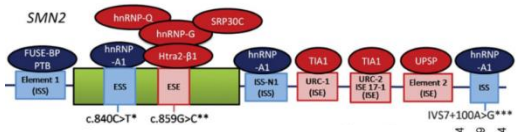
SPINAL MUSCULAR ATROPHY

DISEASE MECHANISM:

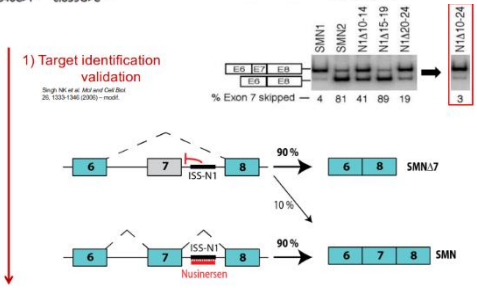


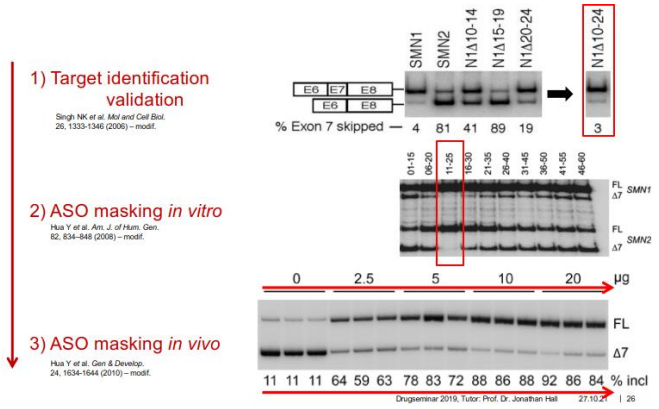
Approaches: Gene therapy acting on red cross, antisense oligonucleotides (on ISS-N1), small molecules on Exon 7.

PRE-MRNA SPLICING REGULATION OF SMN GENES:



PRECLINICAL EVALUATION OF ISS-N1 MASKING:

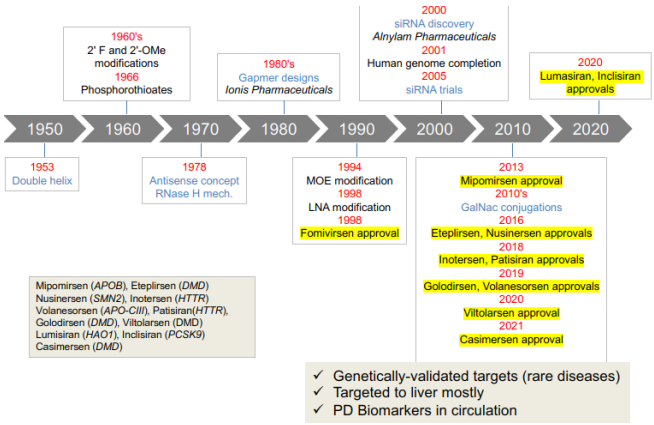




NUSINERSEN CORRECTS A SPLICING DEFECT TO TREAT SMA: it was approved to treat spinal muscular atrophy (SMA) in children.

- 5'-UCACUUUCAUAAUGCUGG-3', 18-nt
- MOE-PS chemistry (Novartis/Ionis)
- Promotes exon 7 inclusion in pre-mRNA
- Delivered by intracathecal injections
- Improved nuclease stability
- Higher affinity for target RNA

MILESTONES IN OLIGONUCLEOTIDE DRUGS



CONCLUSIONS FROM CLINICAL STUDIES OF OLIGONUCLEOTIDE DRUGS

- Oligonucleotide drugs are very safe drugs with class-specific side effects (e.g. injection irritations, thrombocytopenia)
- Dosing is typically 100-100 mg, sub-cutaneous or intravenous, weekly
- Reduced target (mRNA/protein) levels observed in blood, tumors, PBMCs, bone marrow cells, lymph nodes, muscle fibers
- Variable levels of target inhibition, sometimes very potent
- Most of the promising targets are "undruggable" → free of competition (small molecules, biologics)
- For some, indications not clear at outset of clinical development (ICAM1, CRP). Successful targets are genetically validated for the disease
- Clinical failures usually associated with unvalidated targets
- Patient numbers usually small, but drugs are usually breakthrough

SUCCESS IN RESEARCH AND DEVELOPMENT

Successful drug discovery and development required initiation of many projects in order to bring one drug successfully to market.

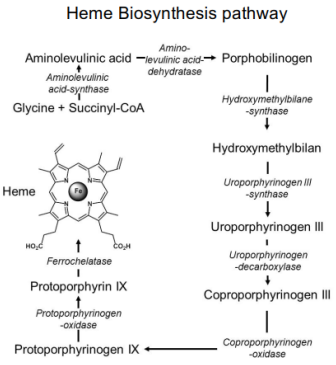
	Research			Development				
	Target selection	Lead selection	Lead optimizn.	Pre-clinical dev.	Phase I	Phase II	Phase III	Approval
Project approvals	57	36	22	14	8	4.5	2	1

# of projects
(data from accumulated experience of several leading pharmaceutical companies)

Probability of success for an untested target: 9% vs. 23% for on with an established mechanism.

NEW INDIVATIONS FOR OLIGONUCLEOTIDE DRUGS: EPP

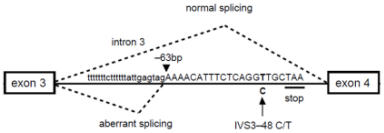
- Erythropoietic protoporphyria (EPP) is a severe skin disorder resulting from exposure to blue light
- ~80 patients in Switzerland
- Mutations in ferrochelatase (FECH) gene cause EPP



EPP GENETICS:

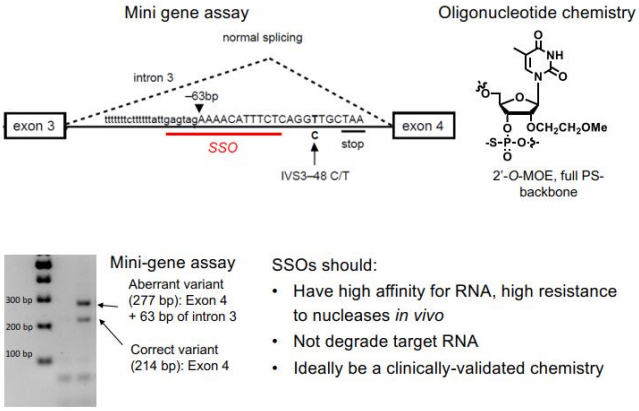
Genotype	Phenotype	FECH activity
T/T, T/C, C/C	Normal	50-100% (of T/T)
M/T	Normal	50%
M/C	EPP	25%

- Complex inheritance from both alleles of ferrochelatase gene (FECH)
- Loss of FECH functions from:
  - Mutation on first allele
  - Intronic T>C SNP on second allele
- The -48T>C variant enhances use of a cryptic splice site



Genotype	% aberrant splicing	% normal splicing
-48.T	20	80
-48.C	60	40

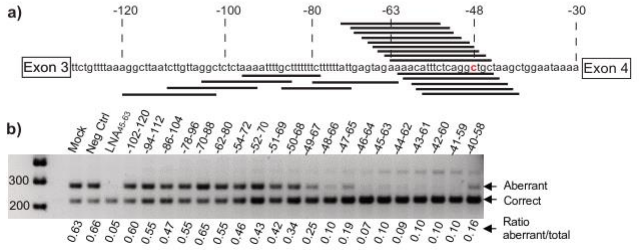
EPP: SPLICE-SWITCHING OLIGONUCLEOTIDES (SSO):



SSOs should:

- Have high affinity for RNA, high resistance to nucleases *in vivo*
- Not degrade target RNA
- Ideally be a clinically-validated chemistry

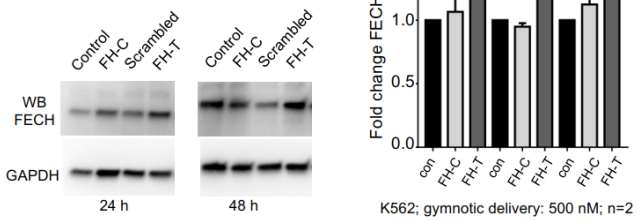
Identification of effective SSOs



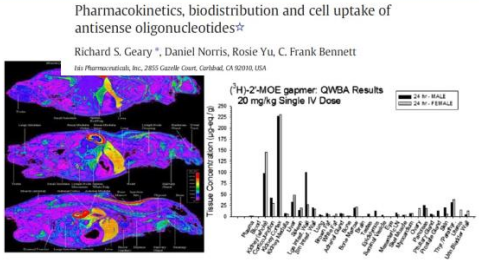
Sequence selectivity of SSOs

K562 cells with T/T genotype

SSO	Sequence 5'-3'
FH-C (ORN1)	AGCAGCCUGAGAAAUGUUUUCU (complementary to FECH-C variant)
FH-T	AGCAACCUGAGAAAUGUUUUCU (complementary to FECH-T variant)

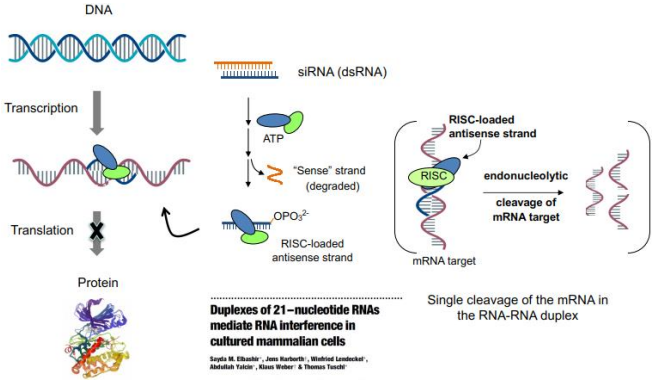


Delivery of SSOs to target tissue in vivo



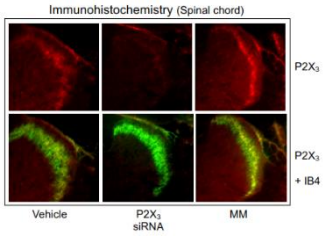
- MOE oligonucleotides show low levels of accumulation in bone marrow in mice (above) and in humans
- Oligonucleotides can be conjugated to targeting molecules, such as peptides or oligosaccharides to facilitate selective targeting of cell types

THE INVENTION OF SIRNAS

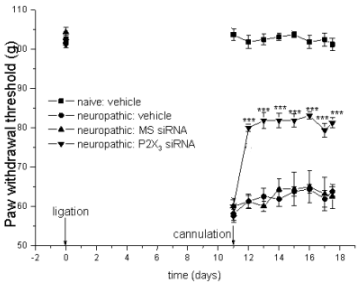


THERPAEUTIC EFFECTS OF OLIGONUCLEOTIDES FOR PAIN

Oligonucleotides administered intrathecally for 7 days on established neuropathic mechanical hyperalgesia in rats (validated models of neuropathic pain)

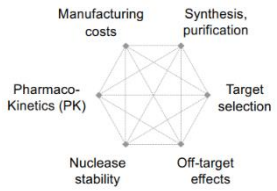


Oligonucleotides administered intrathecally for 7 days on established neuropathic mechanical hyperalgesia in rats (validated models of neuropathic pain).



HURDLES TO THE DEVELOPMENT OF SIRNA-BASED DRUGS

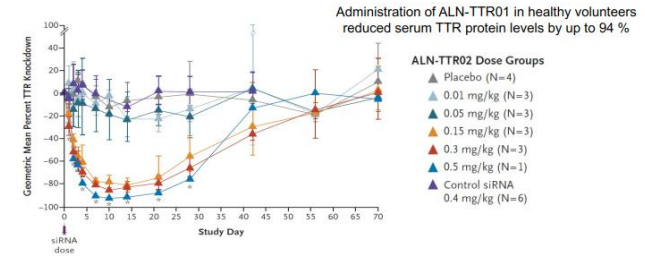
Challenges for the development of siRNA-based drugs were similar to those for the single-stranded ASOs.



Systemic delivery of siRNA drugs has been particularly taxing.

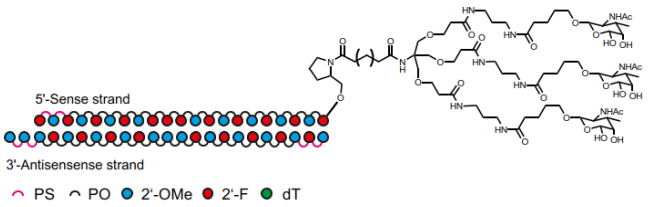
A CLINICAL PROOF OF CONCEPT FOR SIRNAS

- ALN-TTR01 is a siRNA, encapsulated in a lipid nanoparticle, in Phase III clinical trials for transthyretin amyloidosis
- ALN-TTR01 targets transthyretin (TTR) mRNA
- TTR is an ideal target for a siRNA
    - Synthesized in the liver
    - Genetically validated



NEXT GENERATION SIRNAS

The ESC chemistry of siRNAs comprises highly modified RNA strands conjugated to a GalNAc ligand.



- Givosiran targeting ALAS1 (2019)
- Inclisiran targeting PCSK9 (2021)

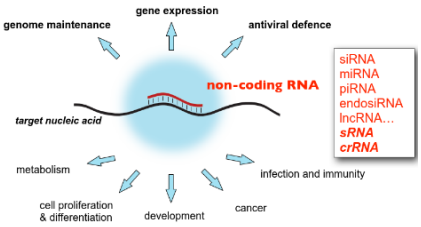
SUMMARY PART 2

- Antisense oligonucleotides are single-stranded oligonucleotide drugs which bind with high selectivity to complementary RNA targets
- Major hurdles to their developments as a class of drugs were: stability in vivo, tissue delivery, manufacturing, target selection
- Chemical modifications to the phosphodiester backbone and the ribose's were essential optimizations of their structures
- ASOs can only be delivered to a small number of tissues and organs
- Most ASOs in clinical development mediate degradation of their target RNAs
- Target selection for oligonucleotide drugs is of critical importance to commercial success
- siRNAs are double-stranded oligonucleotide drugs which mediate degradation of their target RNAs through the RISC
- ASOs can be used to target the function of non-coding RNAs
- To date there are nine approved oligonucleotide drugs on the market

CRIPR-CAS SYSTEMS: FROM BIOLOGY TO GENOME EDITING APPLICATIONS

NON-CODING RNAS

Non-coding RNAs guide genome regulation and defence



In bacteria, non-coding RNAs are involved in:

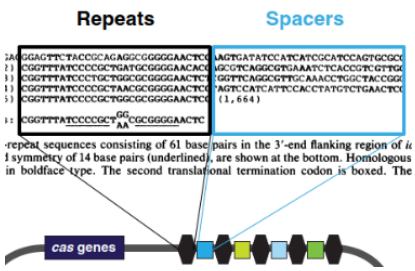
- Gene expression regulation (transcriptional and post-transcriptional)
- Genome defence against invasive nucleic acid elements

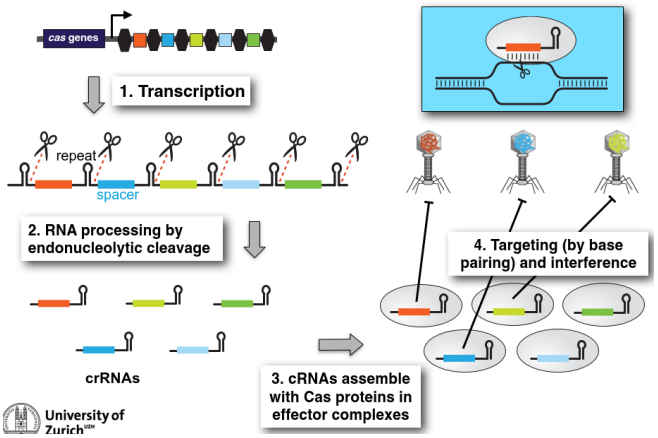
CRISPR-CAS GENOME EDITING : MECHANISMS AND APPLICATIONS

WHERE DOES CRIPSR-CAS COME FROM? HOW DO CRISPR-CAS GENOME EDITORS WORK? HOW CAN WE IMPLEMENT CRISPR-CAS EDITING? HOW CAN WE USE CRISPR-CAS FOR GENE THERAPIES?

CRISPR provides acquired resistance against viruses in prokaryotes. CRISPR-Cas systems are guided by short RNAs.

CRISPR = clustered regularly interspaced short palindromic repeats. Repeats plus nonrepetitive spacers that all have the same length.





Initially the spacer repeat array is transcribed as a single long precursor RNA and then subsequently processed by endonucleotlytic cleavage that takes places within the RNA repeat sequence, to generate individual crisper RNAs → crRNAs.

Typically done by a mechanism that involves at least one of the protein products of the so called CRISPR-associated genes (cas genes on the graphic).

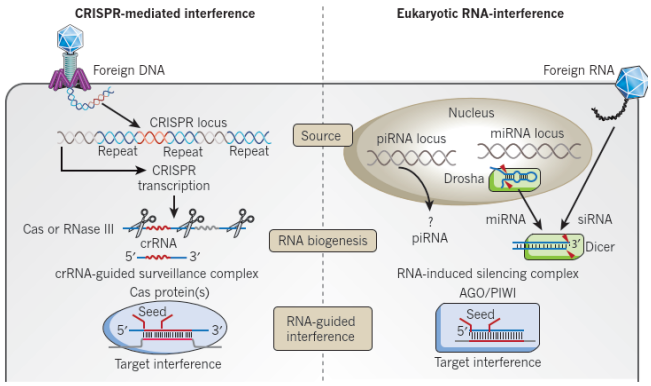
End up with a library of guide RNA that associate with proteins and function as molecular guides → use the information coming from the spacer sequence to then target the nucleic acids associated with the acids by simple base pairing. This targets the nucleic acid for destruction, typically by some kind of nuclease degradation mechanism.

There are also CRISPR-systems that target single stranded DNA or RNA (see later). The ones used in genome editing are all basically ones that target double-stranded DNA.

THE THREE PHASES OF CRIPSR INTERFERENCE:

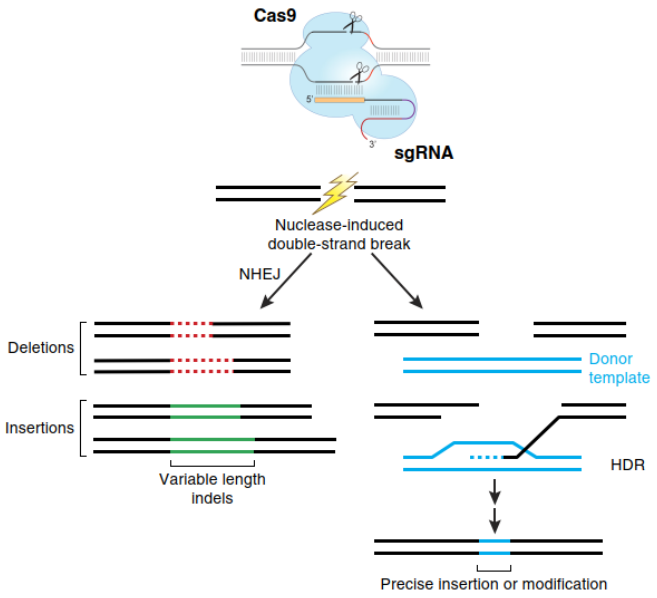
1. Adaptation
   a. Integration of new spacers into the CRISPR locus
   b. Requires Cas1 and Cas2 proteins
2. Guide RNA processing
   a. Transcription of long precursor crRNA (pre-crRNA)
   b. Endonucleolytic processing using CRISPR-associated proteins (Cas6) or housekeeping enzymes (RNase III)
3. Interference
   a. crRNAs associate with Cas proteins in effector complexes
   b. recognition of target nucleic acid - DNA (CRISPR systems type I, II, IV, V) or RNA (CRISPR system types III and VI)
   c. target destruction by single-site endonucleolytic cleavage (type II, V) or complete endo/exonuclease degradation (type I and III)

MECHANISTIC PARALLELS BETWEEN CRISPR AND RNAI:

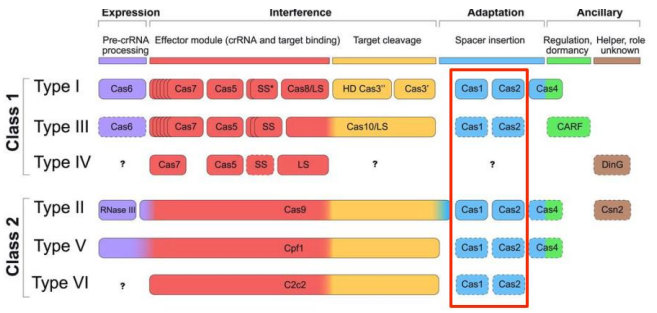


Conceptually similar mechanism. Both have source of targeting information (Spacer repeat array in CRISPR, genes or transcription units in the RNAi). Transcribed into guide RNAs and associate with protein effectors.

CRISPR-CAS - A POWERFUL GENOME ENGINEERING TECHNOLOGY:

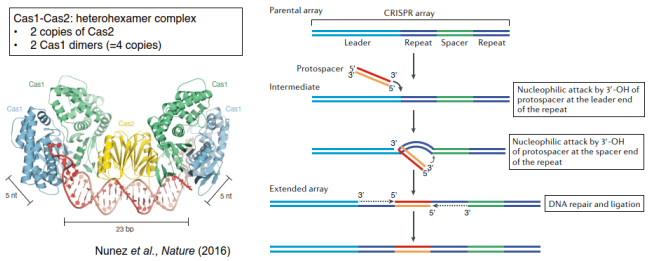


DIVERSITY OF CRISPR-CAS SYSTEMS



CRISPR systems are diverse but share components. Principle difference between Class 1 and Class 2 are that class 1 the targeting machinery is essentially a Multisubunit, multiprotein complex. In class 2 systems, the targeting machinery is a single polypeptide chain. It is the Class 2 system that have found use as genome editing tools and technologies.

What they have in common is the mechanism for spacer acquisition (adaptation). Mediated by two CRISPR associated genes that produce the proteins Cas1 and Cas2 that come together and form a complex that functions as an integrase, that inserts the DNA fragments derived from viruses into the CRISPR → CRISPR adaptation relies on the integrase activity of Cas1:

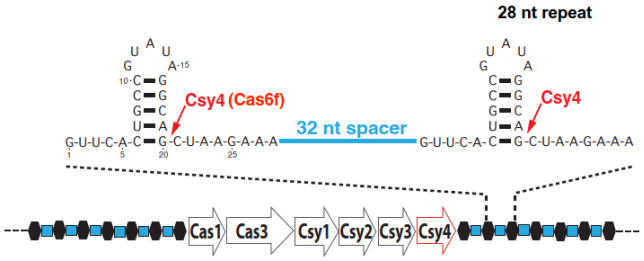


Nunez *et al.*, *Nature* (2016)

Occurs at the junction of the leader and the first repeat at the CRISPR array. The leader acts as a promoter to drive the transcription of the spacer repeat array. Mechanism that involves two transesterification reactions. Involves insertion of the fragment and then DNA repair and ligation to recreate the sequence of the repeat.

CLASS I CRISPR SYSTEMS: MULTISUBUNIT INTERFERENCE EFFECTORS

TYPE I SYSTEMS: CRRNA PROCESSING BY CAS6-FAMILY ENZYMES

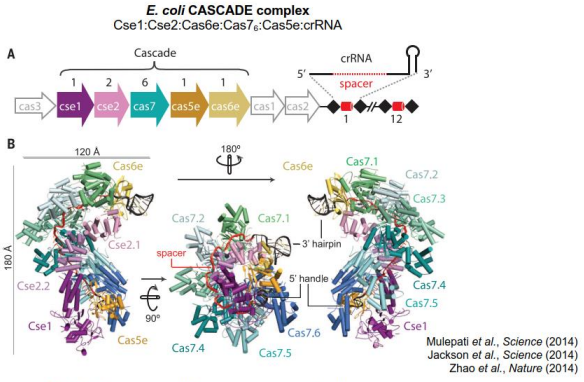


Pseudomonas aeruginosa

Palindromic → cause repeat sequences often look like palindromes → if you transcribe a palindromic sequence into an RNA product, then the product will often form these kinds of stem loops (two parts of the sequence complementary to each other → palindromic → self-pairing). This is used for the RNA processing mechanism.

In this Csy4 is a protein products that functions as an RNA processing nuclease. Its job is it to find the RNA within the repeat derived sequence and catalyse the breakage of the RNA at this location. End up with a product that contains the spacer and then a fragment of the repeat at the 5' prime end and one at the 3' prime end. These enzymes involved in the guide RNA processing have RNA recognition mechanisms. Basic principles involve recognition of the sequence or recognition of the shape. Combination of the two is commonly used by RNA-binding proteins.

TARGETING AND INTERFERENCE IN TYPE I AND TYPE III SYSTEMS:



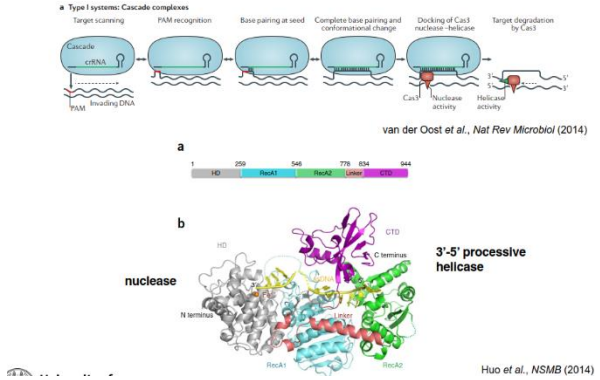
Multisubunit surveillance complexes facilitate target search

Targeting machinery is multiprotein complex that assembles onto the RNA. Example here → CASCADE complex. 5 different protein types. CRISPR RNA is embedded in the centre. Cas7 are winding along its length (unusual stoichiometry). Cas6e is the processing endonuclease, is part of the CASCADE complex and remains bound to the stem that is found at the 3' prime end that is found on the mature CRSPR-RNA.

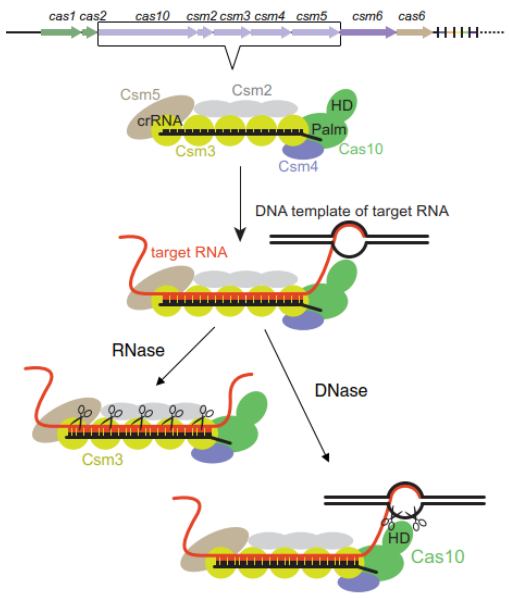
Type I and III assemble on processed RNA, facilitate recognition of the target nucleic acid.

TARGET DEGRADATION IN TYPE I SYSTEMS:

Targeting by CASCADE, followed by degradation of target DNA, by another CRISPR associated protein, Cas3. Acts as a ATP-dependent helicase nuclease enzyme, which is able to bind to one of the strands to the DNA (displaced, not base paired to the guide RNA), and is able to cleave the displaced strand. Strand then becomes a target for degradation, because of the 3'-5' processive helicase activity of the enzyme → loaded on double stranded DNA and translocate along the displaced DNA strand (ATP dependent) and as it moves, it feeds the 3'-end in the nuclease domain for processive degradation. Results in the degradation of one of the DNA strands and destruction of the target.



TYPE III SYSTEMS TARGET NASCENT RNA TRANSCRIPTS:

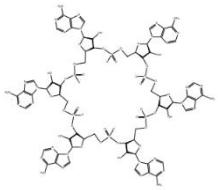


Multiprotein complex, with principal component being Cas10. Mechanism: instead of targeting DNA, they target RNA transcripts of viruses etc. Targeting complex when assembled on the guide RNA will target an RNA as being transcribed from the virus and there are two nuclease based mechanism that this systems have:

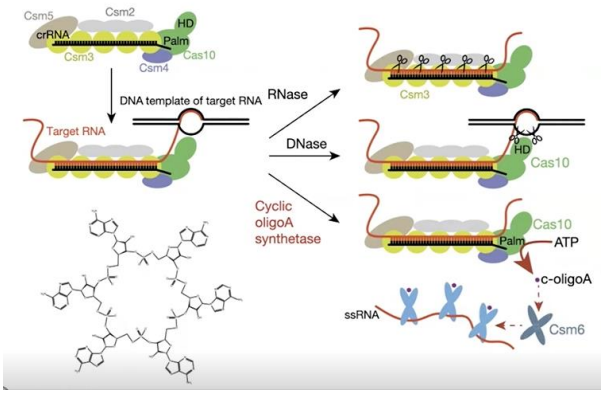
1. There is a subunit, Csm3, which degraded the RNA directly by cutting
2. There is a nuclease domain within the Cas10 protein which acts as a single stranded DNA nuclease, targets the DNA from which the target RNA is being transcribed

→ Dual nuclease mechanism that allows to target both the RNA transcripts as well as the DNA from which the transcripts are being made.

Also have other CRISPR-associated genes and proteins, e.g. Csm6. If deleted from some of CRISPR systems → targeting activity significantly reduced! Bit strange because It is not part of targeting machinery, and allows functions as ribonuclease. Contains ribonuclease domain and CARF domain (CRISPR-associated ?? fold domain). Function first unknown, but later found systems have a mechanism by which effector nucleases, such as Csm6, are allosterically controlled → based on short cyclic oligonucleotide molecules, made by Cas10 containing targeting complex by oligomerization of ATP → target RNA binding stimulates Csm1/Cas10 to produce and allosteric activator of Csm6.



How is the activity of Csm6 regulated? Type III interference machinery generates a Csm6 activator!

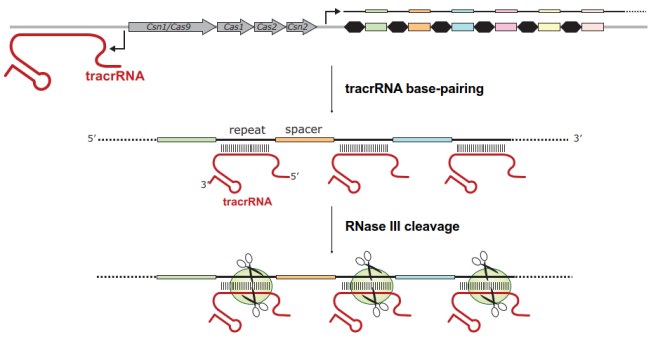


These systems have an RNase → Csm3, a DNase → Cas10, and also a third enzymatic activity → activated when complex recognizes its target and converts ATP into these cyclic oligonucleotide molecule, which then go and activate proteins like Csm6.

CLASS 2 CRISPR SYSTEMS: SINGLE PROTEIN EFFECTORS

Type II: Cas9. Type VI: Cas13

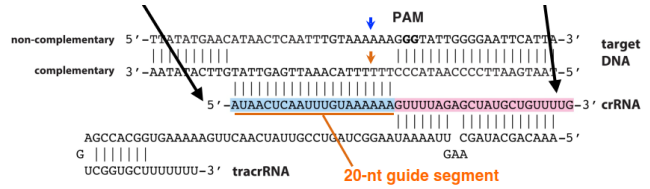
CRRNA BIOGENESIS IN TYPE II CRISPR SYSTEMS:



cRNA biogenesis requires tracrRNA, RNaseIII and Cas9
Cas9, tracrRNA and crRNA are required for interference

Not really a dedicated nuclease, but rather they contain another non-coding RNA, generated by transcription of e gene in the CRISPR-locus → tracrRNA, contains a sequence that is complementary to the sequence of the repeat in the precursor. House-keeping enzyme RNase III that specifically cleaves double stranded RNA cleaves. Cas9 is effector protein.

INTERFERENCE IN TYPE II SYSTEMS: RNA-GUIDED DNA CLEAVAGE BY CAS9:



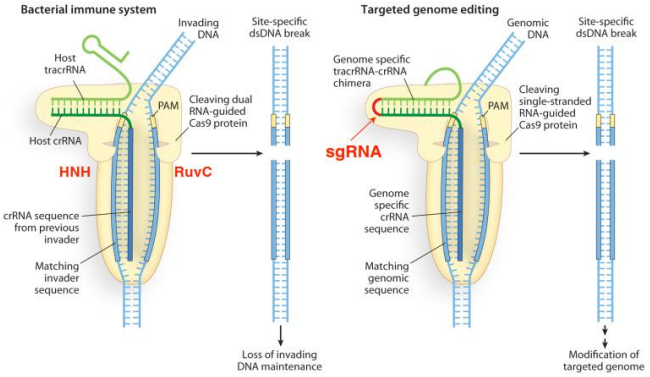
Requires crRNA and tracrRNA to be able to target. Once crRNA is processed it contains 20nt sequence that is derived from the spacer, and a sequence that is derived from the repeat (partially complementary to the tracrRNA (pink)).

The 5’ end of the mature crRNA is trimmed by an unspecific 5’-3’ exonuclease after the RNA has been bound by Cas9. The length of the guide segment (20nt) is therefore shorter than the length of the spacer in the CRISPR locus and is determined by the “footprint” of the Cas9 protein on the bound RNA.

The 3’ end of the mature crRNA and the 5’end of the processed tracrRNA are generated by RNase III cleavage - hence the 2-nt 3’ overhang typical for RNase III cleavage products.

CAS9 IS AN RNA-GUIDED DNA-CUTTING ENZYME:

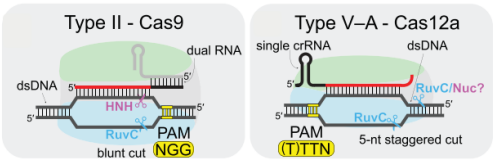
Could take RNA molecules and fuse them artificially by connecting the 3’-end with the crRNA and 5’-end with the tracrRNA → generate a single molecule, guide RNA = sgRNA, which simplified the system (one protein for cutting, one for program specificity) → basis of CRISPR-Cas technology that we know now.



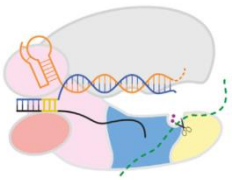
Programmable one-protein, one-RNA system

TYPE V CRISPR LOCI CONTAIN EFFECTOR NUCLEASE CAS12/CPF1:

Type V have single protein that function as an RNA guided nuclease → Cas12a. RNA guided DNA cutting enzyme, but has slightly different mechanism, rather complementary to Cas9.



	Cas9	Cas12a (Cpf1)
tracrRNA	Yes	No
crRNA biogenesis	RNase III	Autocatalytic processing
Guide region	20 nt	20 nt
Substrate	dsDNA	dsDNA
PAM	5’-**NGG**-3; 3’-terminal	5’-(**T**)**TTV**-3; 5’-terminal
Cleavage pattern	blunt or 1-nt 5’-overhang	staggered 5-nt 5’-overhang
Nuclease domains	RuvC and HNH	RuvC
Seed sequence	Yes (~10 nt)	Yes (~6 nt)
trans-nuclease	No	Yes

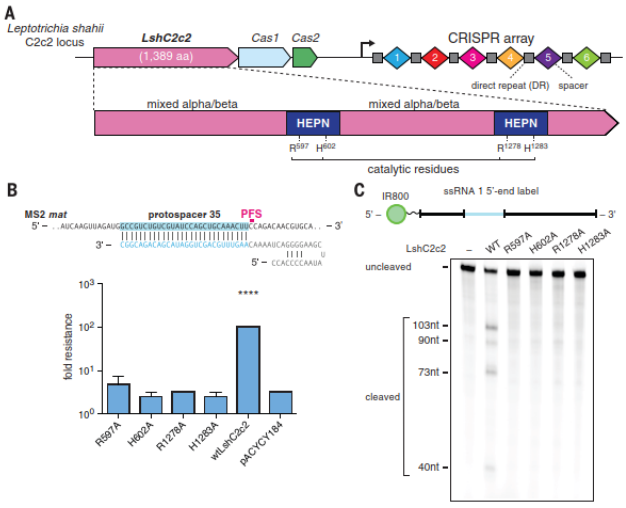


Trans-cleavage of non-target DNA

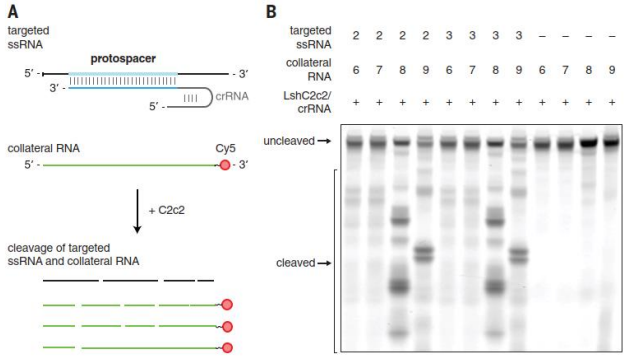
ssDNA binding activates trans-nuclease activity of Cas12a: once the enzyme is programmed with the guide RNA finds its target, remains bound to the target and remains in a chemically active site in which active site is exposed and is able to degrade other nucleic acid (does not have to part of the target DNA). Found use in applications that use cas12a as a mechanism of detecting nucleic acids.

TYPE VI:

Cas13 is an RNA-guided ribonuclease. It has collateral RNase activity. Targets single stranded RNA molecules, rather than double stranded DNA molecules.

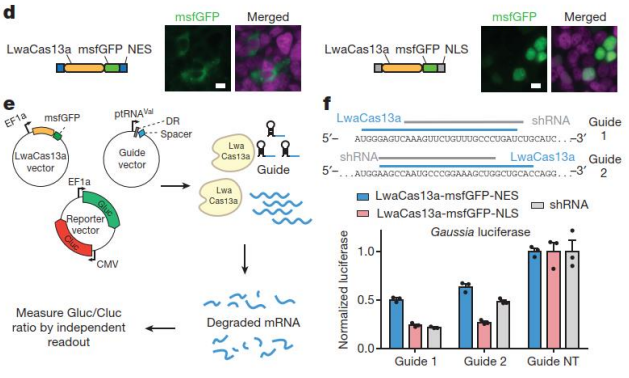


Contain two HEPN domains which form molecular domain and a ribonuclease active site which is activated when the Cas13 protein is associated with the guideRNA and base pairs to a complementary target RNA.



Also have a collateral trans nuclease activity. When Cas13 is activated by target RNA, ribonuclease is non-specific, will not only degrade base paired RNA but also other RNAs in the cell and in the organism and can therefore stop propagation of the virus in the cell.

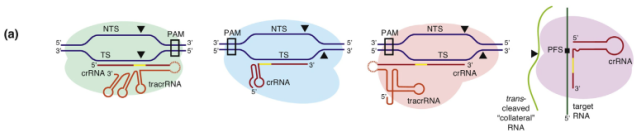
RNA targeting using Cas13: also used as a molecular tool, not exactly for genome editing but for genome expression toll modifications.



SUMMARY: CRISPR-CAS SYSTEMS

- Non-coding RNAs in bacteria are involved in genome regulation and defence
- CRISPR systems provide adaptive RNA-guided immunity
- 3 phases: adaptation, processing, interference
- crRNAs provide sequence specificity
- Cas proteins are the enzymatic machinery
- Class 1 CRISPR-Cas systems have Multisubunit effector complexes
- Class 2 systems use single-subunit effector nucleases
- Type III systems use cyclic oligoadenylate signalling
- Type II and V systems use RNA-guided DNA nucleases Cas9 and Cas12
- Type VI systems use RNA-guided RNase Cas13

SUMMARY: CLASS 2 CRISPR-ASSOCIATED NUCLEASES

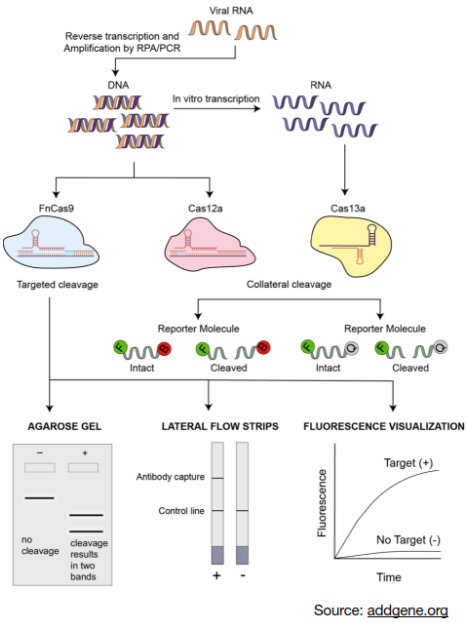


(b)

	Cas9	Cas12a	Cas12b	Cas13
pre-crRNA processing	No	Yes	No	Yes
tracrRNA	Yes	No	Yes	No
PAM/PFS	3', G-rich	5',T-rich	5',T-rich	3',non-G PFS
Nuclease domains	RuvC HNH	RuvC	RuvC	2x HEPN
Substrate	dsDNA	dsDNA	dsDNA	ssRNA
Cleavage pattern	Blunt	Staggered	Staggered	Near U or A
Cleavage cis/trans	Cis	Cis	Cis	Trans
Guide-target duplex length	20 bp	20 bp	20 bp	24 bp

CRISPR-BASED VIRAL DIAGNOSTICS: EXPLOITING COLLATERAL NUCLEASE ACTIVITY

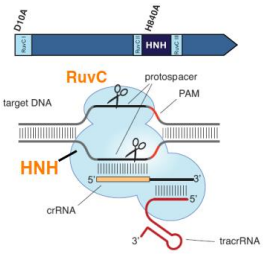
CRSIPR-associated tools also used for viral diagnostics. Couple systems to fluorescent reporter molecules and in presence of target DNA/RNA are activated and cleave reporter→ detected.



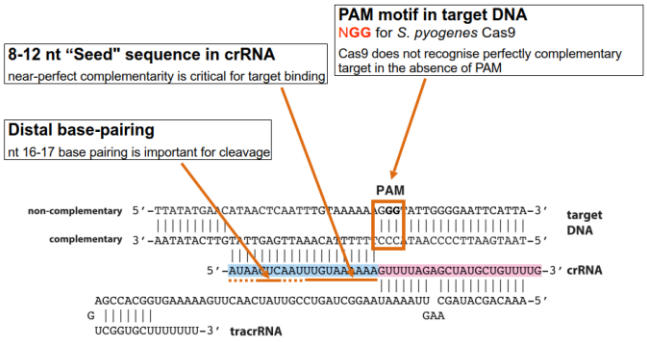
Source: addgene.org

HOW DO CRISPR-CAS GENOME EDITORS WORK?

Dual RNA-guided DNA cleavage by Cas9:



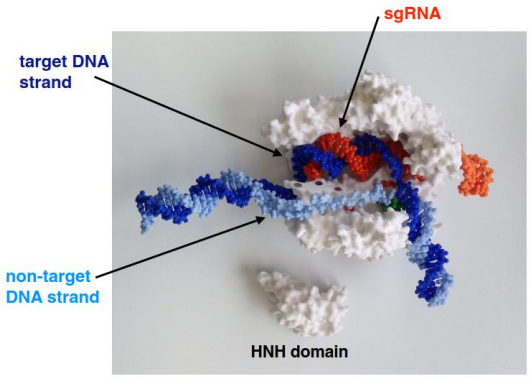
Requirements for Cas9-catalysed DNA cleavage:



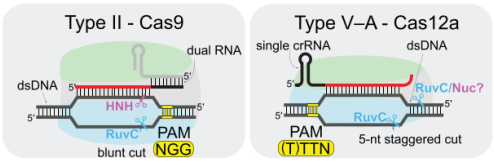
Short sequence must also be present downstream from the target site → PAM (protospacer adjacent motif) → characteristic for each of the different Cas9 proteins, has to be present, otherwise enzyme will not be able to cleave it! Even when the Target is perfectly complementary! PAM + target site is what makes it a target for the enzyme.

Additionally: complementarity between guide and target RNA. Have to have perfect or nearly perfect complementarity in the pamproximal part of the duplex → because guide-RNA has seed sequence → critical for match, very sensitive for mismatches. Also have to have base pairing in to more distal part of the sequence. Important for the cutting (less for the binding), serves as a signal, causes conformational change in the protein to activate its nuclease domains so they can cleave the DNA.

Structural studies of CRISPR-Cas genome editors:



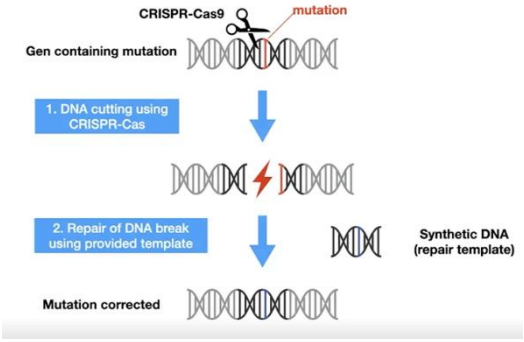
CAS12A (CPF1): AN ALTERNATIVE GENOME EDITOR NUCLEASE:



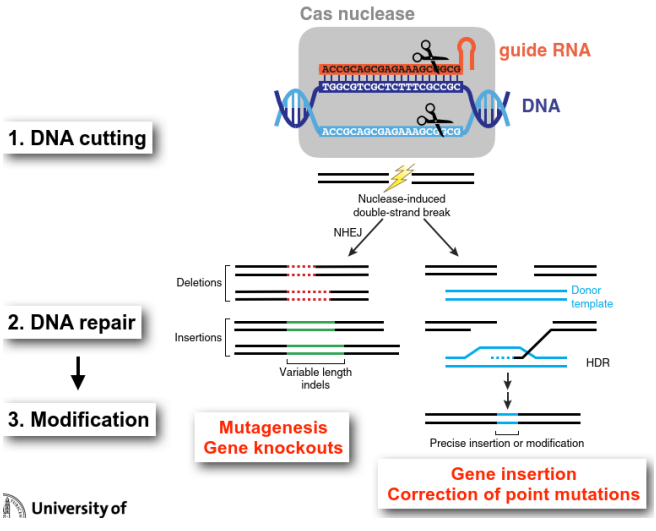
	Cas9	Cas12a (Cpf1)
tracrRNA	Yes	No
crRNA biogenesis	RNase III	*Autocatalytic processing*
Guide region	20 nt	24 nt
Substrate	dsDNA	dsDNA
PAM	5'-**NGG**-3; 3'-terminal	5'-(**T**)**TTV**-3; 5'-terminal
Cleavage pattern	blunt or 1-nt 5'-overhang	staggered 5-nt 5'-overhang
Nuclease domains	RuvC and HNH	RuvC

In Cas12 PAM is upstream, unlike Cas9, and they also target different sites (Cas9 → NGG, Cas12 → TTN).

How can we practically implement CRISPR-Cas editing? CRISPR: genome editing for the masses



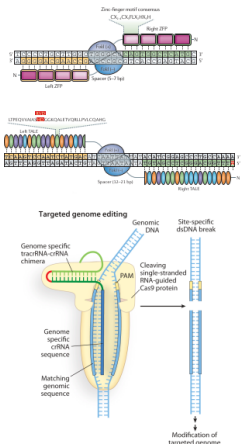
Genome editing using RNA-guided CRISPR-associated nucleases:



Cutting done by RNA guided Cas. DNA repair then leads to Modification. Different DNA repair mechanism, can results in different DNA editing outcomes. There is non-homologous end joining (left) → quick and dirty ligation mechanism. Often insertions and deletions in the junctions → can be used to engineer gene knockouts or mutations.

Second group: homology-directed repair, used to do more complicated modifications like insertions, overwriting part of sequence with another, correcting point mutations and so on. Done by introducing double strand break and at the same time providing DNA template for the repair process.

Double-strand DNA break formation stimulates DNA repair. Genome editing technologies rely on precise DNA cutting:



Zinc Finger Nucleases (ZFNs)

- developed in the 1990s, first used in 2003
- very efficient
- specificity is difficult to engineer

TALENs (Transcription Activator-Like Effector Nucleases)

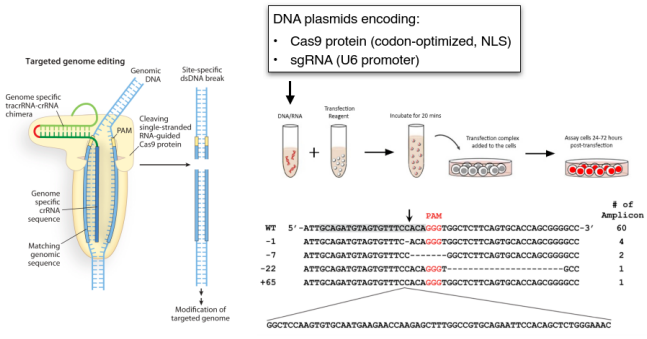
- developed in 2011
- sequence specificity is easy to engineer due to modularity
- one 34 aa repeat recognizes 1 bp
- difficult to assemble sometimes

Cas9 endonuclease

- developed in 2012-3
- specificity is programmed by changing the guide RNA sequence
- simple, fast and cheap

Initially Protein based tools → engineered protein based nucleases that combined a nonspecific nuclease in fusion with sequence specific DNA binding proteins (ZFNs and TALENs). Quite efficient but problem with design and implementation. Every time you want to target a new gene you need to design a new protein. In CRSPR one only needs to change the target sequence in the guide RNA.

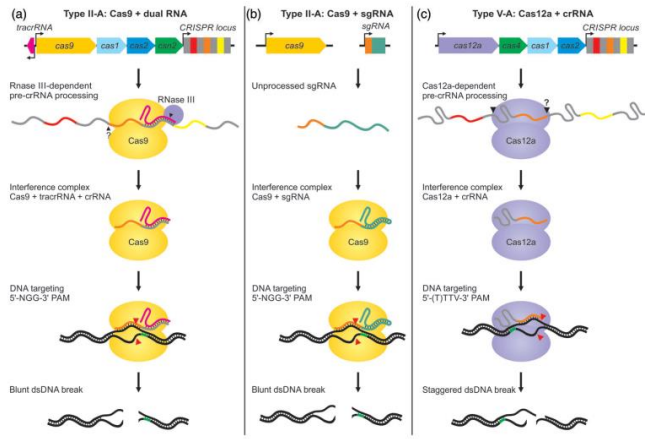
RNA-guided gene knockout using CRISPR-Cas9



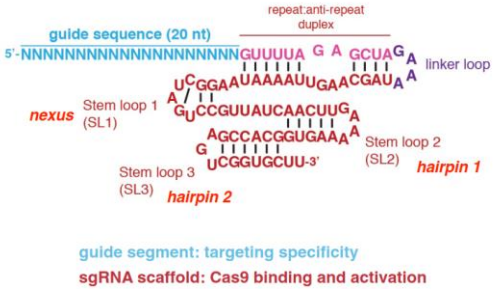
Variable indels introduced due to imperfect repair of Cas9-induced DSB by NHEJ pathway

If into cultured human cells, plasmids encoding the Cas9 proteins and a gene for expressing guide RNA are introduced we could target a genomic site and we could isolate cells and sequence clones in which the sequence was containing deletions and insertions → first indication that with using Cas9 and the guide RNA, you could make a ds-break and the non-homologous end-joining process trying to repair it (imprecisely and introducing these indels). This was an indication that the system was going to work in eukaryotic cells.

SUMMARY: CRISPR-ASSOCIATED GENOME EDITING NUCLEASES

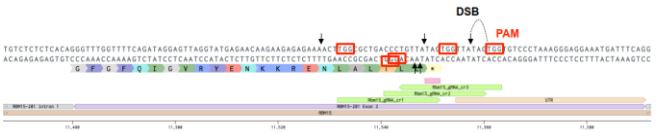


BASIC SGRNA STRUCTURE (S. PYOGENES CAS9):



First 20 nt provides the targeting information, the rest of the guide RNA structure is necessary for binding the guide RNA to the Cas9 protein and guiding it to do the cutting.

GUIDE RNA DESIGN FOR CAS9:



- Search for the presence of a PAM (NGG for Cas9) in the vicinity of the edit site
- In a random sequence, a PAM will occurs on average every 8 nt
- DSB should be located as close to the site to be edited as possible
- Locate 20-bp target site next to the PAM
- Copy target site sequence as the 5'-terminal guide sequence of the sgRNA

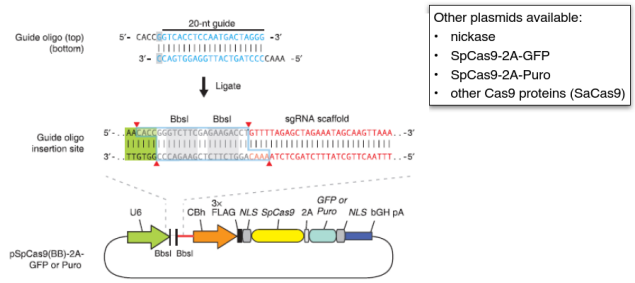
CAS9-SGRNA DELIVERY OPTIONS:

- Both Cas9 and sgRNA encoded as DNA
  - Plasmid transfection or electroporation
  - Viral delivery (lentivirus, adenovirus, AAV)
- In vivo transcribed RNAs (Cas9 mRNA, sgRNAs)
  - Transfection or electroporation of cultures cells

- Microinjection into oocytes or embryos (mice, Drosophila, zebrafish)

- In vitro assembled ribonucleoprotein complex
  - Microinjection into oocytes or embryos
  - Cell-penetrating peptide fusions
  - Transfection (liposome-mediated)
  - Electroporation
  - iTOP (induced transduction by osmosis and propane betaine)

CAS9 AND SGRNA EXPRESSION FROM A PLASMID:



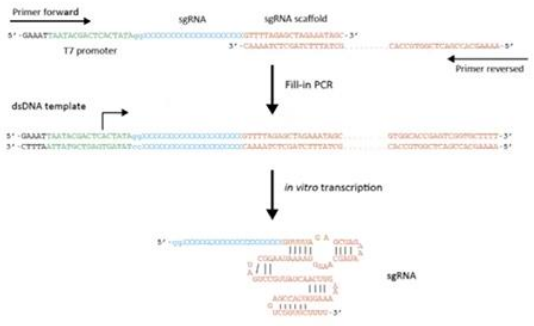
- Plasmid pX458 and derivatives (available from Addgene)
- Cut plasmid with BbsI
- Order a pair of DNA oligos
- Anneal and ligate
- Introduce to cells by transfection or electroporation

Clone guide RNA sequence in a plasmid (various are now available from companies). Plasmid expressed Cas9 and guide RNA. There are then ways to check whether the guide DNA is active, e.g. TIDE analysis → isolating DNA of the cells after the genome editing experiment, PCR amplification, trying to sequence PCR sequence (Sanger), looking where sequencing becomes scrambled → what would be expected by ds-break, indels → sequence variable, scrambled.

CAS9 PROTEIN-SGRNA COMPLEX DELIVERY (AS IN IN VITRO ASSEMBLED RIBONUCLEOPROTEIN COMPLEX):

Allows to do editing without actually delivering any DNA into cells, has two main advantages: efficiency of the on-target editing can be improved, because we deliver a fully active protein, don't have to wait for the cell to transcribe it. Second: activity of protein is transient. Protein will get degraded after time, fewer opportunity of the cell to accumulate off target mutations.

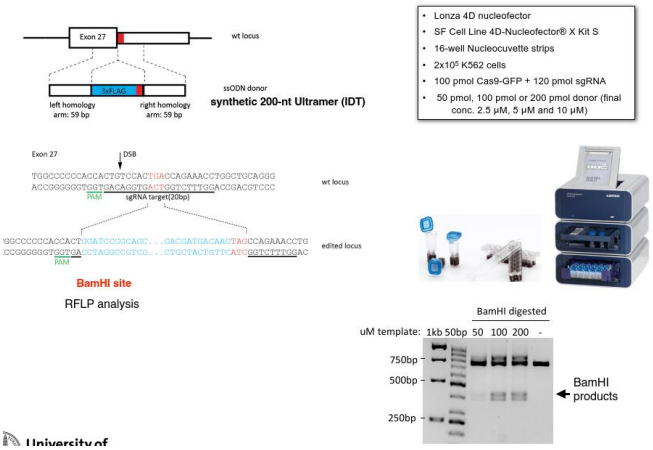
CAS9 GUIDE RNA PREPARATION :



- in vitro transcribed sgRNAs
  - To do guide RNAs
  - Fill-in PCR to prepare dsDNA transcription template
  - In vitro transcription using T7 polymerase (e.g. Roche or Promega kits) (from double stranded DNA to single-stranded guide RNA which can be isolated)
- crRNA-tracrRNA pair
  - purchase universal tracrRNA
  - order custom crRNA
  - stabilizing modifications available
  - mixing cas9 with crRNA and tracr RNA separately, advantage: crRNA is short enough to be made synthetically, and tracrRNA is basically universal, so you can buy it commercially. So just find custom crRNA and combine it with the tracrRNA.

GENOME EDITING IN CELL LINES:

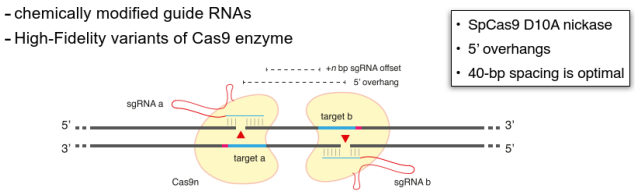
Electroporation. Cells in suspension in the cuvette, apply electric pulses to temporarily disrupt membrane and make protein RNA complex get into the cell.



IMPROVING CRISPR-CAS GENOME EDITING

- Can we use molecular insights?
- Greater specificity and minimal off-targeting
- Expanded targeting space
- Higher efficacy in specific cell types
- Better control of DNA repair outcomes
- Off-target activity of Cas9
- Specificity is not 100% perfect
- Up to ~5 mismatches can be tolerated
- Seed sequence is most sensitive to mismatches
- Cas9 also tolerates bulges and base skipping
- Improved sgRNA design and post-editing diagnostics
  - Off-target site profiling of guide RNAs as part of guide RNA design (bioinformatics tools)
  - Improved detection methods (GUIDE-Seqm Digenome-Seq, BLESS)

- Engineering the Cas9-sgRNA enzyme complex
  - Paired nickase approach
  - dCas9-FokI fusions
  - truncated sgRNAs
  - chemically modified guide RNAs
  - High-fidelity variants of Cas9 enzyme



ENGINEERED HIGH-FIDELITY CAS9 NUCLEASES

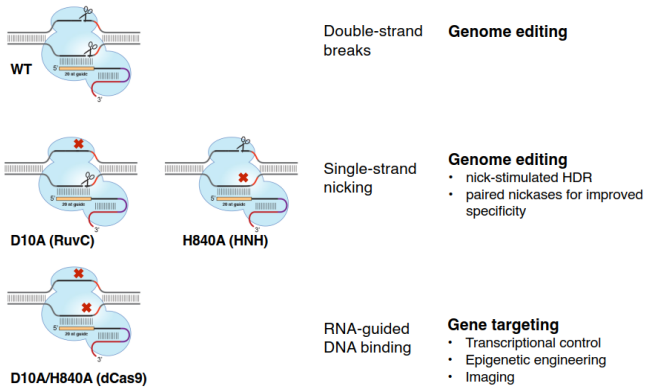
Basic idea: Make target DNA binding more dependent on RNA-DNA hybridization by disrupting molecular contacts between Cas9 and the target DNA.

Done by introducing specific amino acid substitutions in the protein (those that are in contact with the duplex), such that the system becomes more sensitive to mismatches between RNA and DNA. Enhance proof-reading mechanism of the nucleases to improve the specificity.

ENGINEERED CAS9 VARIANTS WITH ALTERNATIVE PAM SPECIFICITIES

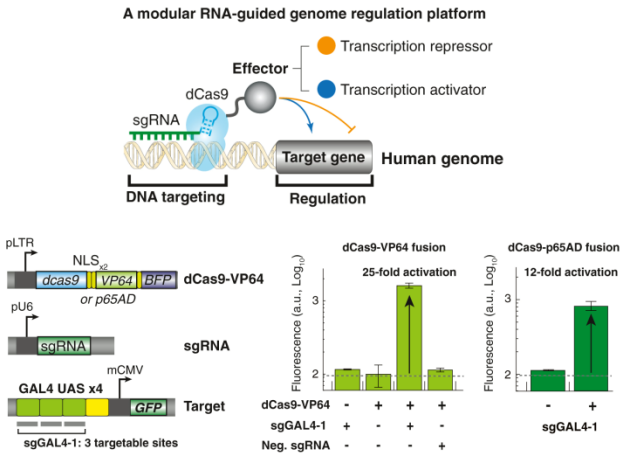
- S. pyogenes Cas9 is constrained to target sites adjacent to a cognate NGG PAM
- Use orthogonal Cas9 and Cas12a proteins from other organisms
- Engineered variants of SpCas9 that recognize alternative PAMs.
- Again done by introducing amino acid substitutions into the WT sequence of the Cas9 protein
- Greatly expanded the parts of the genome that could be targeted

CAS9 VARIANTS FOR GENOME EDITING AND GENE TARGETING:



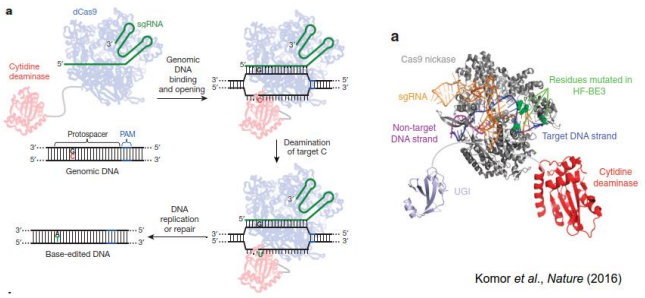
Cas9 can be mutated to make single-strand nicks, or just bind

DCAS9: RNA-GUIDED GENE EXPRESSION CONTROL WITHOUT CUTTING:



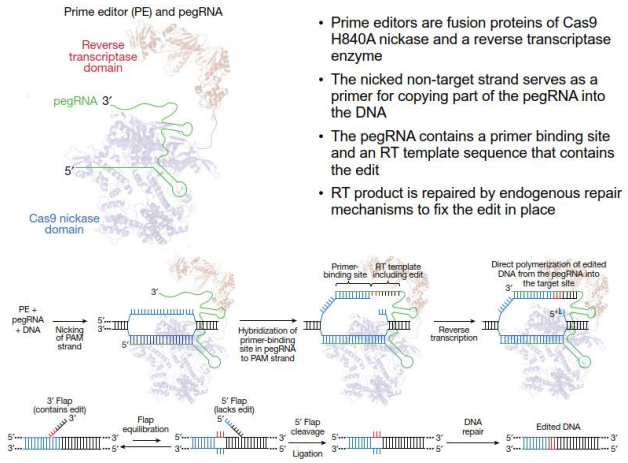
Catalytically inactive nucleases used (as in one where both nuclease domains are mutated so the protein only binds to the DNA but doesn't cut it). "Dead" Cas9 is fused with either transcription activator or repressor domains, used in a way to control the transcription of the target gene.

CRISPR BASE EDITORS: EDITING WITHOUT DNA CUTTING



- combine Cas9 as an RNA guided DNA binding platform with other enzymes than can introduce these modifications directly without relying on DNA cutting and without relying to some degree on the homologous repair mechanism
- one problem of the CRISPR genome editing is the efficiency of the homology directed approach is lather low → these are developed to circumvent that
- one is called Base editors are fusion proteins of Cas9 D10A nickase and adenine or cytidine deaminase enzyme
- Deamination of C produces U, which results in a G-A transition mutation after DNA replication. Deamination of A produces inosine (I), which results in a T-C transition mutation.
- The edited base must be located on the non-target strand upstream of the PAM.
- Make single nucleotide substitutions in the DNA

PRIME EDITING: INSERT EDIT USING REVERSE TRANSCRIPTASE



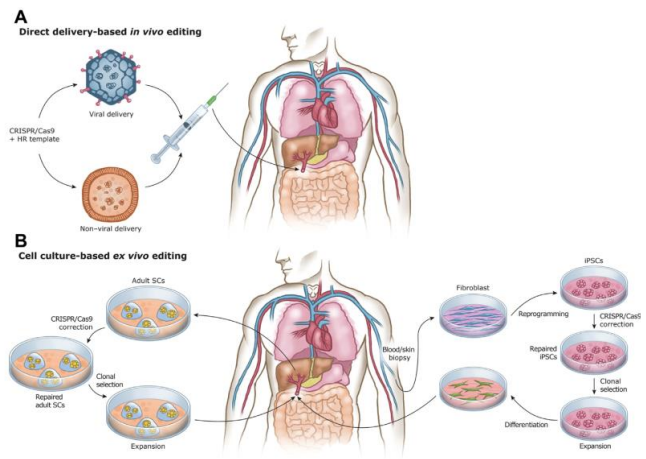
Combining a Cas9 molecule together with reverse transcriptase and a specially modified form of the guide RNA. Basic process: cas9 protein will be programmed with the guide RNA to bind to a specific site in the genome and in this case its not catalytically inactive Cas9, but it is a Cas9 nickase, so it is able to cleave the displaced DNA strand. So it generates a free three prime end in the displaced DNA strand which can be extended by the reverse transcriptase. So you can use this in combination with a special guide RNA in which the 3' end is extended so it can act as a template for the RT. Modifications in the guide RNA are then also introduced in the genome.

HOW CAN WE USE CRISPR-CAS FOR GENETIC THERAPIES?

This chapter was discussed in the last 5 mins of the lecture so its just to give an idea (don't think this needs to be learned in detail).

Basic idea: gene correction instead of gene augmentation. E.g. in sickle cell disease directly overwriting the gene defect that causes the abnormality in the cells.

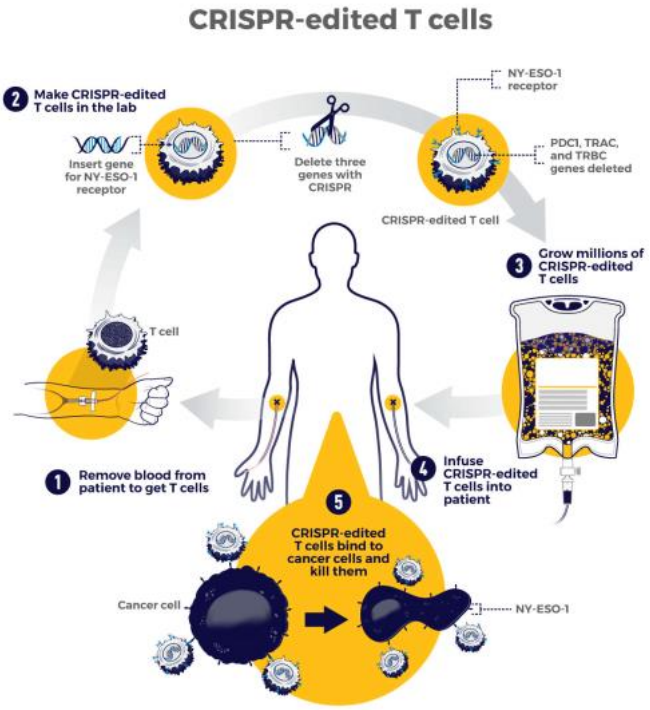
IN-VIVO VS. EX-VIVO GENOME EDITING:



EX VIVO APPROACHES FOR THERAPEUTIC GENE EDITING

- CRISPR Therapeutics & Vertex CTX001 Phase ½ trials
- CRISPR-Cas9 used to edit autologous HSCs to increase HbF production
- Disruption of beta-globin gene reactivates fetal gamma-globin expression in adult erythroblasts

EX VIVO EDITING OF T-CELLS FOR IMMUNOTHERAPIES



KEY GOALS FOR THERAPEUTICS APPLICATIONS OF CRISPR-CAS EDITING

Improving specificity and avoiding off-target modifications

- **Improved sgRNA design and post-editing diagnostics**
  - off-target site profiling of guide RNAs as part of guide RNA design (bioinformatics tools)
  - *improved detection methods* (GUIDE-Seq, Digenome-Seq, BLESS, SITE-Seq)
- **Engineering the Cas9-sgRNA enzyme complex**
  - Paired nickase approach, dCas9-FokI fusions, truncated sgRNAs
  - *High-Fidelity Cas9 variants*

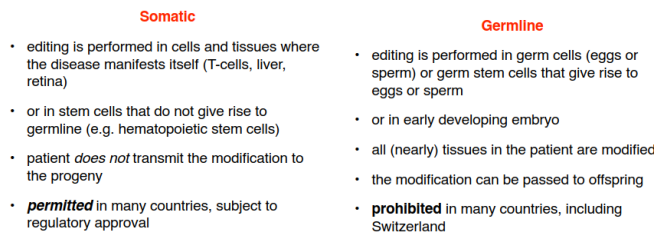
Improved cellular/tissue-specific delivery of CRISPR components

- **Viral vectors**
  - AAVS, lentiviral vector with tissue-specific tropism
- **lipid nanoparticles**
  - reduced cytotoxicity, improved efficacy, cell/tissue-specific delivery
- **RNP nucleofection**
  - improve efficiency, scale-up for ex vivo approaches

Improved control of DNA repair outcomes

- **Maximize HDR and minimize NHEJ indels**
  - cell cycle control
  - homology template design and delivery
  - alternative nucleases (e.g. Cas12a/Cpf1)
- **Base editing**
  - DNA repair/modification without cleavage

SOMATIC VS. GERMLINE GENOME EDITING:

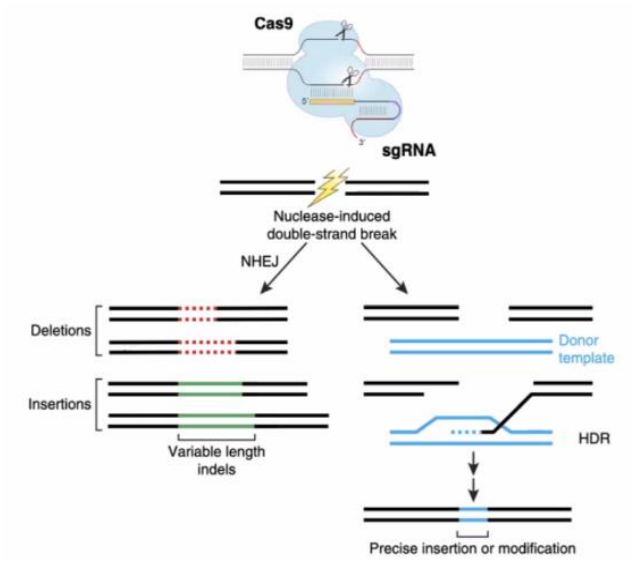


Human genome editing: global consensus needed.

SUMMARY: CRISPR-CAS GENOME EDITING

- CRISPR-Cas genome editors are programmable RNA-guided DNA endonucleases
- DNA cleavage requires PAM in the target DNA and a near-perfect match to the seed sequence of the crRNA
- CRISPR-Cas nucleases undergo conformational rearrangements during their catalytic mechanism
- Structure-guided engineering of genome editors for improved specificity, efficiency and versatility
- RNP approach for genome editing can improve efficiency and reduce off-targeting
- CRISPR-Cas genome editing has many potential applications in basic research, biotechnology and molecular medicine
- Somatic cell editing for therapeutic purposes is under way in a number of clinical trials around the world
- Applications of CRISPR-Cas genome editing in the human germline are controversial and banned by law in many jurisdictions

CRISPR-CAS GENOME EDITING II



Basic usage of CRISPR-Cas is for genome editing: through making DNA double strand breaks.Repair mechanism of DNA double strand breaks:

1. Non-homologous End Joining (NHEJ), Alternative End Joining (Alt-EJ), Microhomology-mediated End Joining (MMEJ) → error-prone; may get insertions or deletions (indels).
2. Adding donor templates (such as ss or dsDNA) to be integrated into the ds-break → Homology-directed repair (HDR) → more precise DNA repair
3. Unresolved DNA break = cells die. Some cell types don't repair breaks, they just arrest (p53 response)

Mammalian cells are better at NHEJ, alt-EJ, MMEJ compared to HDR, but yeast cells are better at HDR. Genome editing method depends on the type of cell and vary a lot (guide RNA design in mammalian cells and prokaryotes is very different).

CRISPR:

- tracrRNA: to recognize the genome
- WC base pairing

GUIDE RNAS:

- PAM: Protospacer adjacent motif
  - requirement that is in the genome
  - function: to be recognized and crack open genome (then detection of protospacer sequence possible)
  - exist because in bacteria: guideRNAs are stored in CRISPR arrays, that contain several guideRNAs but not PAM)
    - recognition but no cutting itself
    - but cutting if recognition of sequences on viruses
    - Arms race question: tuning of acquisition and rate of loss of ability of viruses possibly not be detected (?)
  - different Cas have different PAMS parts of Cas grab bases
    - directed evolution to develop systems with different PAM motifs/ no PAM
    - SpRY: mostly "PAMless" → Cas cuts everything containing sequence, also plasmid
      - huge targeting options
      - specific
  - if no PAM in target gene → either use PAMless methods, and/or make less specific
  - specificity: WT Cas9 has a protospacer of 20 bp
    - $1:4^{20}$
    - Seed sequence/region: 8-9 bp (at 3' end) → important
    - non-seed region (upstream from seed, 5' end) → less important, mismatches in this region is not too bad
  - off-target effect:
    - guideRNA engineering: create synthetic guideRNA to improve it (insert a loop)
    - shift guide RNA by a couple of bases
    - in vitro revolution: **engineer the CAS** (protein) to remove or add specificity → HI-5 versions → sometimes mechanisms are unknown, but it works

SYSTEMS BEFORE CRISPR

- Zinc fingers (ZFNs):
  - Relative modular DNA binding proteins
  - Hard to program
  - Sangamo: first ones to develop method that worked and is safe
- TALENs
  - Modular (more than ZFNs) DNA binding proteins

CRISPR IN THERAPY

SICKLE CELL DISEASE:

- HSC in bone marrow → develop to RBC
- RBC turn-over: 20 days, also no nucleus
- bone marrow transplant (filter out HSC from blood) → **ex vivo editing** (some flexibility)
  - (different from in vivo: used in cystic fibrosis)
  - HBB (gene) → HBS (point mutation) causes SCD
    - revert point mutation back to WT with HDR → why not common? HDR really hard in stem cells
    - make different protein together with HBG1/2 (for hemoglobin production)

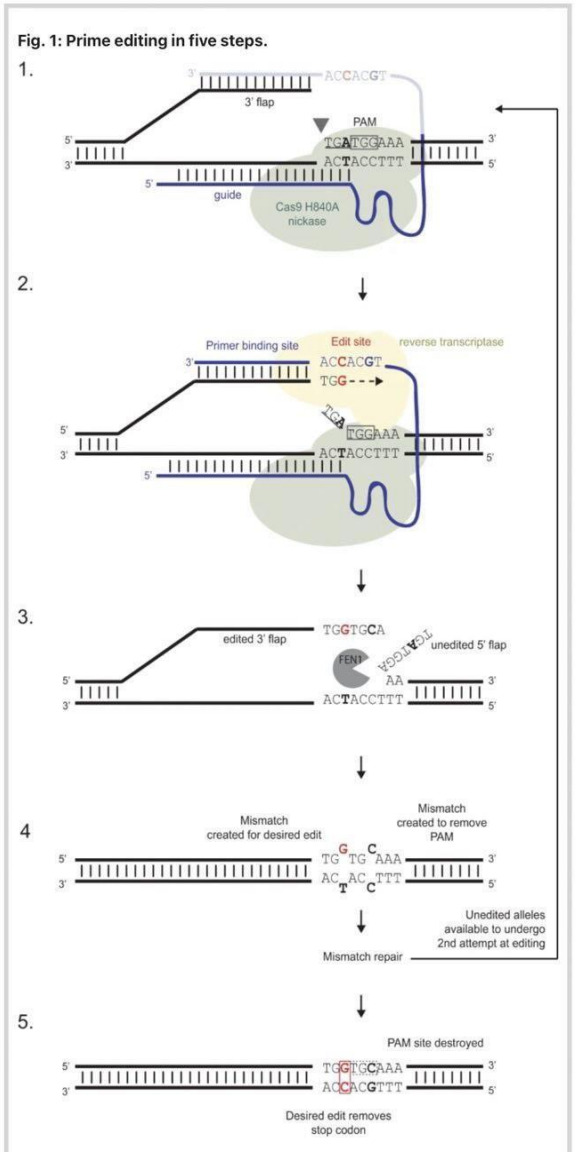
    - HPFH: have sickle mutation but are healthy → turn on fetal hemoglobin back on (that usually decreases in adults)
    - make HPFH synthetically with KO
    - make indels
    - because fetal Hemoglobin is silent: how?
    - repressor Bcl11A: try to KO that gene (actually KO enhancer within gene that is specific for erythrocytes) to turn fetal hemoglobin back on
- HDR is hard, with indel therapies: double negatives (remove repressors, works with dominant-negative diseases (KO of the mutated allele, WT enough for function)

TOOLS: USING CAS9 HYBRID PROTEINS:

- Base editing:
    - Cas9 is targeting component and brings deaminase to the right place for base editing (deaminase is combined with Cas9)
    - C → U → T
    - A → I → G
      → Adenine base editors: took RNA deaminase and made it a DNA deaminase, but it still has some RNA targets and makes RNA damage, effect still unknown (seems okay), → it is not Cas directed, so damage can happen anywhere
    - large, clunky → work best for ex vivo engineering
- Prime editing:
    - instead of HDR donor, all DNA information is incorporated into PEgRNA
    - with reverse transcriptase
        - problematic: off-target profile very unknown
        - e.g. only works in cells with mismatch repair pathways deficiencies
    - Cas9 nicks DNA (only 1 strand is cut): reverse transcriptase copies off of PEgRNA and adds to the 3' end of the cut DNA sequence
    - advantage: it does not induce a ds break, does not depend on cell cycle → many cells in body are postmitotic so it still applicable, just make sure there are not any off-target effects
- → Prime editing is way more complicated than base editing, takes time to make it more robust

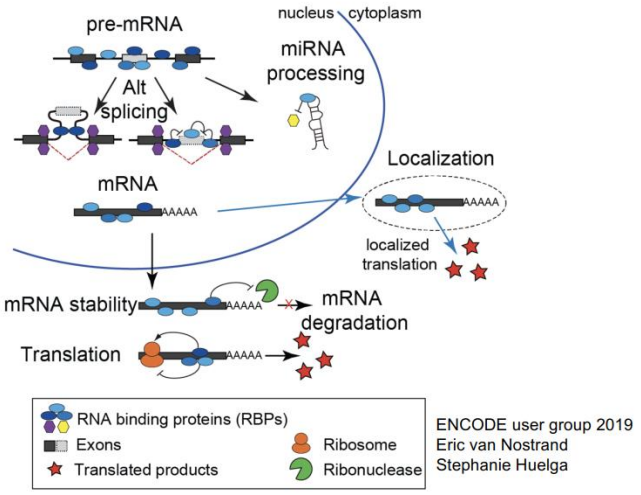
(Pyogenes Cas9 very robust and works very well, other Cas9: not robust)

- Immunogenicity against Cas9:
    - edited cells are not recognized by immune system
    - but you only get 1 shot: if editing unsuccessful, you probably cannot repeat the process since there will be an immune response against Cas9



RNA PROCESSING CODES

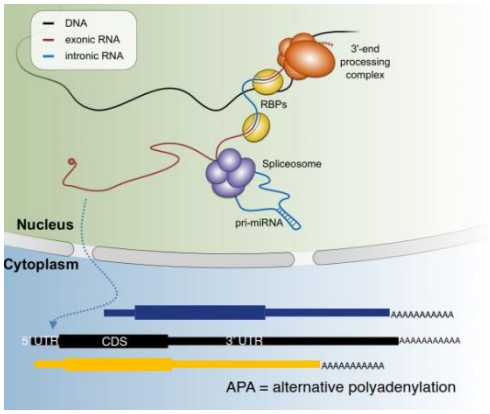
RNA PROCESSING STEPS IN EUKARYOTIC GENE EXPRESSION



RNA PROCESSING "CODE"

- A code = system of rules for converting one type of information into another form
- Focusing on RNA: each step of RNA processing is carried out by some complex, which has to be recruited to the specific RNA
- The recruitment depends on signals (Sequence/structure/etc.)
- The set of signals and rules for combining them to achieve various processing outcomes form a code
- How do we decipher these process-specific codes?
- How do we derive quantitative models linking the signals to the processing outcome?

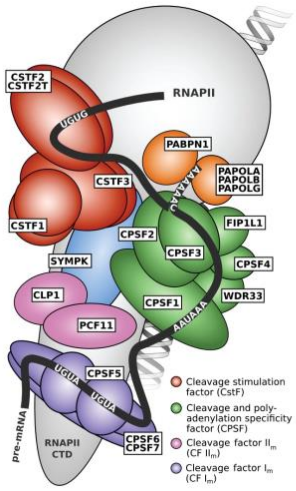
EXAMPLE: ALTERNATIVE POLYADENYLATION

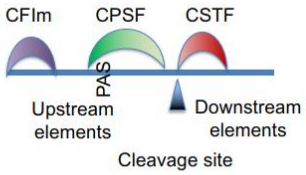


mRNA coming out of the mRNA polymerase start to get associated with complexes (RBPs - RNA-binding proteins), which give rise to alternative isoforms, which can even change the protein coding potential of the gene, example if an alternative exon is included or not in the result. Most of the isoform variation comes not from alternative splicing but from the use of alternative promoters and alternative 3' end processing sites. So, the start and the end are to most varying across isoforms.

The most common form of alternative polyadenylation is one that occurs in the same terminal exon in the gene but at different sites → see on the graph above, depending on which polyA site is chosen you get a short or a logn 3' prime UTR. This has implications for what they can bind to and associate with, and the transcripts therefore have different outcomes and functions.

CORE 3'END PROCESSING FACTORS:

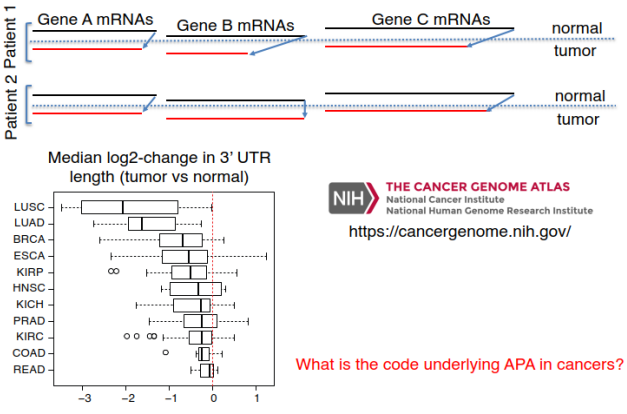




Gruber et al. WIRE 2014

(Not important to remember all now). In the region where the 3' prime cleavage happens, there is a sequence composition bias. There are three important elements, plus a lot of smaller ones. Upstream are sequence elements that recruit complexes. CFIm complex is a tetrameric complex that binds UGA motifs. CPSF does the cleavage of the RNA. One of the proteins in the complex recognizes the AAUAAA nucleotide pattern. These elements tend to occur at most of the canonical processing sites, as well as this CSTF binding sequence element which is downstream and binds to a region that tis G rich.

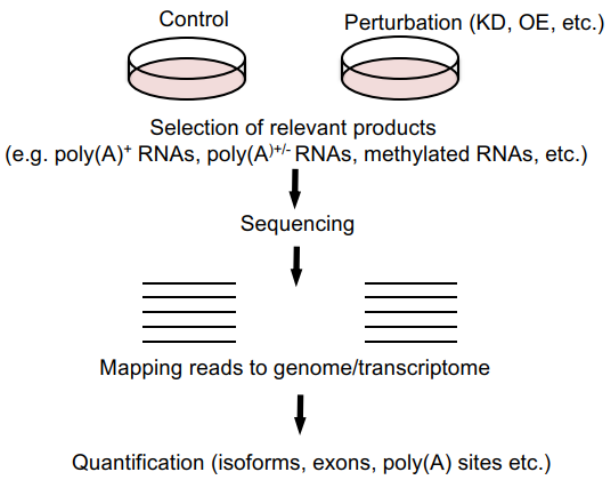
APA IN CANCERS:



What is the code underlying APA in cancers?

3' UTR is typically shorter in cancer. What determines this? It is a systematic thing so it should be something that affects all genes. But we do not know what it is yet.

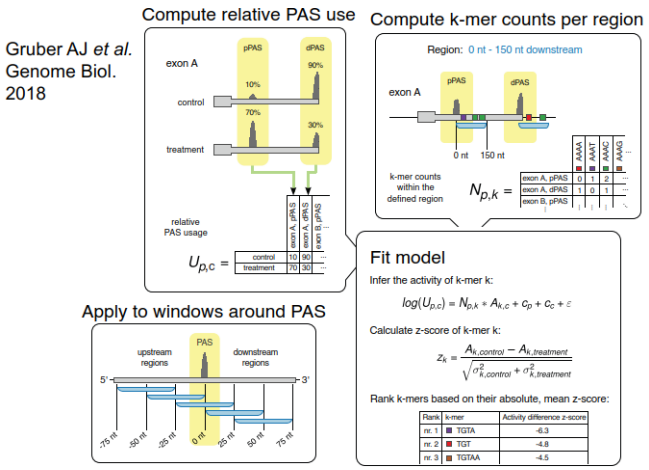
MEASURING THE OUTCOME OF RNA PROCESSING



Which products have changed abundance and by how much?

LINKING SIGNALS TO OUTCOMES WITH MATHEMATICAL MODELS

KAPAC: K-MER ACTIVITY IN POLY(A) SITE CHOICE



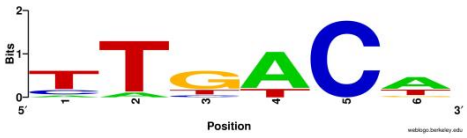
Alternative polyadenylation. If we look at a region around it, we expect some motifs contribute to the process, and those should be at all poly-A sites that either have enhanced or decreased usage in some condition. We model usage of poly-A site as a linear function (combination of all the motifs that occur in the vicinity of the site multiplied by some activity that each of them have).

Example: PTB1 may drive 3' UTRs shortening in glioblastoma. Which factors are responsible for that? Took at mRNA and looked which poly-A sites are important. What motifs occur in the vicinity of the poly-A site? Which are most strongly associated?

LIMITATIONS OF THE MODEL:

Sequence specificity of regulators represented in terms of k-mers (variables) is very restrictive. Most proteins don't just bind one sequence, it also binds variants of that sequence. Ways to deal with that:

- Generalizing the sequence specificity of RBPs as positional weight matrixes. Summarizes what a TF or RNA binding protein likes to bind. . Position-specific weight matrix (PWM)



Yaxis = information score. Go position by position and sum up probability of observing the nucleotide at that position times the logarithms of that probability divided by the probability that we would expect in a random sequence (every nucleotide same chance of occurring)

Constructing a PWM from known binding sites:

Sites\Position	1	2	3	4	5	6
1	T	T	G	A	C	T
2	T	T	G	A	C	A
3	T	T	G	A	C	A
4	T	T	T	A	C	A
5	T	T	T	A	C	A
6	C	T	G	A	C	G
7	T	T	G	T	C	A
8	T	T	G	A	C	T
9	C	A	C	A	C	T
10	A	T	G	T	C	A

$$w_b^i = \frac{n_b^i + \alpha_b^i}{\sum_b (n_b^i + \alpha_b^i)}$$

n_b^i count of nucleotide b at position i

α_b^i pseudo-count of nucleotide b at position i (here 0.25)

$$P(S[i..i+l-1]|PWM) = \prod_j w^j_{s[i+j]}$$

Nucl.\Pos.	1	2	3	4	5	6
A	1	1	0	8	0	6
C	2	0	1	0	10	0
G	0	0	7	0	0	1
T	7	9	2	2	0	3

Nucl.\Pos.	1	2	3	4	5	6
A	0.11	0.11	0.02	0.75	0.02	0.57
C	0.20	0.02	0.11	0.02	0.93	0.02
G	0.02	0.02	0.66	0.02	0.02	0.11
T	0.66	0.84	0.20	0.20	0.02	0.30

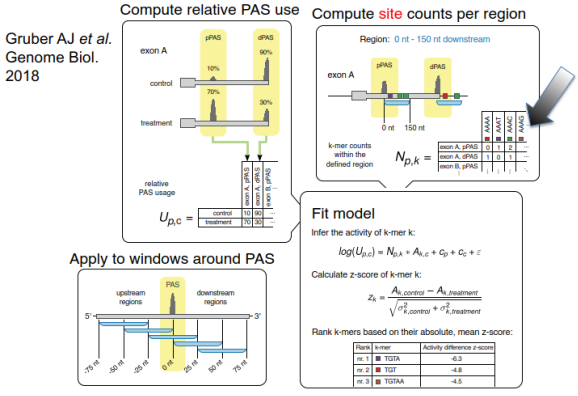
Information score (y-axis) = $\sum_{\alpha\in\{A,C,G,T\}} w_\alpha log_2\left(\frac{w_\alpha}{b_\alpha}\right)$

Height of individual letters $h_\alpha = w_\alpha \sum_{\alpha\in\{A,C,G,T\}} w_\alpha log_2\left(\frac{w_\alpha}{b_\alpha}\right)$

Posterior probability of a sequence corresponding to a PWM match

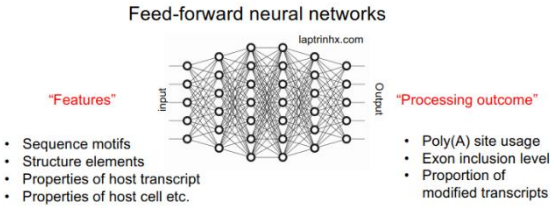
$$P(PWM|S) = \frac{P(S|PWM)\pi_{PWM}}{P(S|PWM)\pi_{PWM} + P(S|bg)\pi_{bg}}$$

PWM matches can be used in motif-activity models

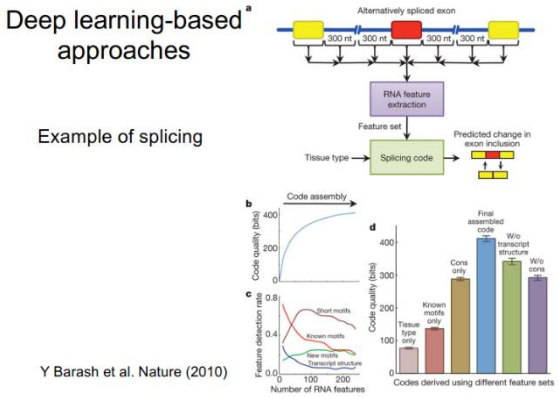


Linear models - no interactions between regulators. Ways to deal with that:

- Deep learning-based approaches. Can calculate features we want to look at. These are all long vector of features,



and we have one of these for every single example of a polyA site, so we get a big matrix. This we can multiply by weight (?) and carry the calculations out over multiple layers. As we get from one layer to the next, the information from the nodes feeds into the next ones, which gives us the possibility of considering the interactions. Once we do this, we want to predict an outcome, example the usage of the polyA site in a given condition (or some other kind of quantitative measure). Example:



Potential pitfalls:

- Models have large numbers of parameters
- Unclear what model architecture to choose
  - User-defined vs. inferred features
- Models are not always easy to interpret mechanistically

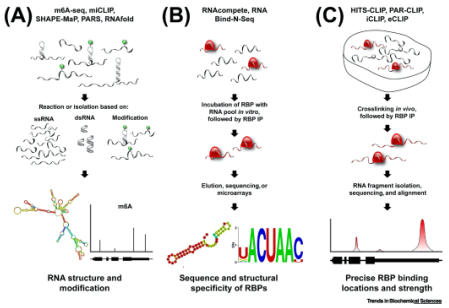
Similar approaches can be applied to:

- RNA localization - compare RNAs from different subcellular compartments
- RNA methylation - compare modified with unmodified RNAs
- RNA degradation - compare RNAs with high and low decay rate
- Etc.

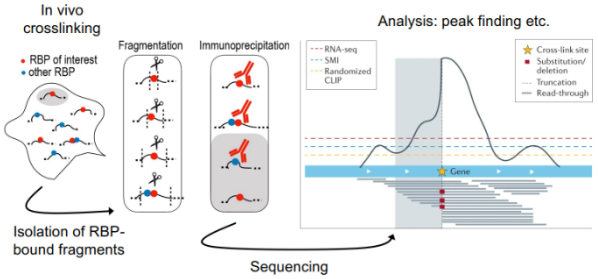
Effects could be indirect (causation vs correlation). Ways to address this:

- Identifying of direct RBP-RNA interaction sites in vivo. There are low-throughput methods still used but here we talk primarily about high-throughput methods:

Global Approaches in Studying RNA-Binding Protein Interaction Networks



The CLIP method to capture sites bound *in vivo*

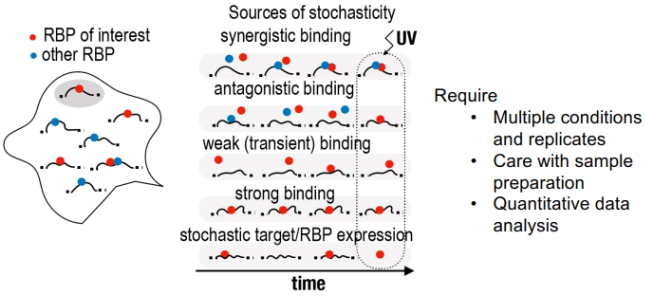


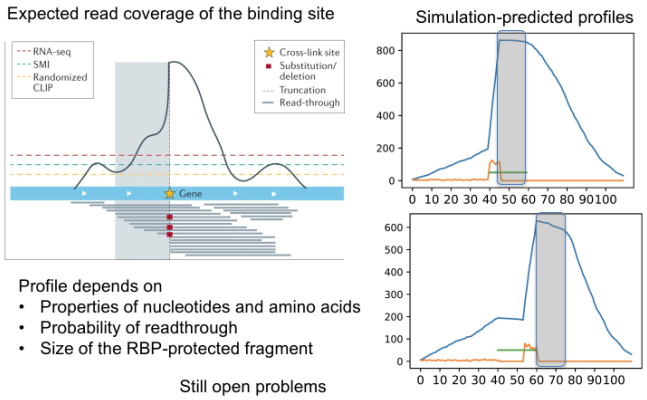
Idea behind Clip: cells on dish, UV light → crosslinks between proteins and RNAs. If we are able to chop down the RNA to the small bits that are protected by the RBP → take and antibody specific to the RBP, fish out fragments bound by protein crosslinked by it → get reads with sequences. Map them to genome and get this map on the right (above). Potential pitfalls:

- Incomplete data (false negatives)
- Biased data (e.g. due to RNA fragmentation, adaptor ligation)
- Dependence on RNA expression levels (false positives)

How to identify binding sites? There is discussions about multiple modes of binding. One mode of binding would be very strong, and give a nice sharp peak, but it is also apparent that there are reads that spread over regions. Is this because the experiment didn't work well or is it because of the nature of binding? → different modes! Interaction of RBPs with other proteins also important, could be in a complex and therefore not show on the correct molecular weight.

Sources of "noise" in CLIP data





How do we distinguish signal from noise? There are different approaches. When we do the CLIP, we sequence the fragments to which the RNA binding protein was crosslinked. After you digest the protein away, you have to do the cDNA → reverse transcriptase. Observation: at the site where the protein was crosslinked, the protein digestion doesn't eliminate everything, there is kind of a stub left. Makes it so when the RT tries to go through the site, it either falls of (80%) or it goes through and introduces some kind of error.

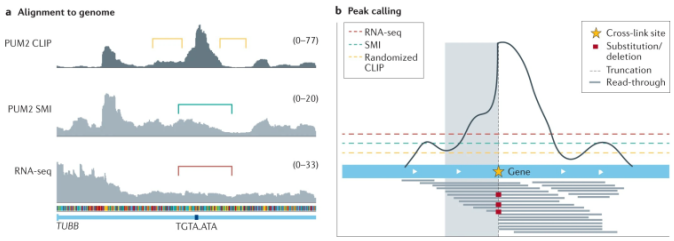
Ways to try to narrow down the binding site of the RBP simply based on these fall offs and the mutations → one method: circularizing the DNA, transcribed until the RT transcribed, falls of, then the RNA is circularized, adapter caught in the middle used to sequence fragment (??).

iCLIP or enhanceCLIP → instead of doing immunoprecipitation, make the whole prep run on the gel, no Ip no specific selection of protein and cut on specific size → generating a background sample.

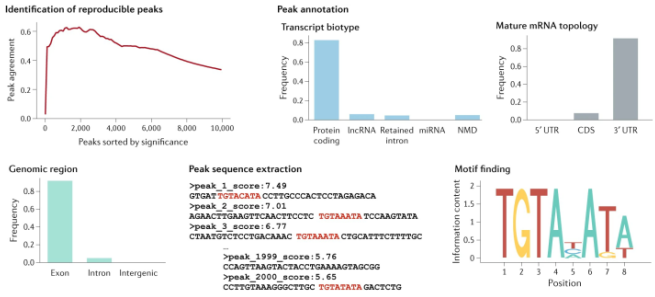
Looking at two types of regions: mRNA seq data, look at mRNA that are very abundant and maybe geenrate coverage anyway. Also look at neighboring regions and determine if they may be specific binding sites or not.

Delineating binding sites based of read profiles

Key issue: separating signal from noise



From peaks to PWMs



What motif is enriched in the peaks?

Identifying of direct RBP-RNA interaction sites in vitro:

Measurements of interaction affinity

$$\frac{dC}{dt} = k_{on}(R_T - C)\,(P_T - C) - k_{off}C$$

In equilibrium we have $\frac{dC}{dt} = 0$, meaning that $k_{on}(R_T - C)\,(P_T - C) = k_{off}C$

This gives us a quadratic equation in C,

$$C^2 - (R_T + P_T + K_D)C + 4R_TP_T = 0,$$

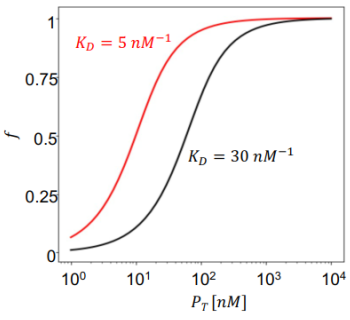
with $K_D = \frac{k_{off}}{k_{on}}$. This equation has one valid solution, which we can use to get the fraction of RNAs/motifs bound,

$$f = \frac{C}{R_T}$$

RNA sequence that interact with protein: can write law of mass action and then derive some secondary equation that allows us to get the fraction of bound RNA, or if we know that and the concentration of protein we can get the KD, the dissociation constant. Can do this at different concentrations and derive this curve:

Inferring PWMs from affinity measurements

Conversely, knowing R_T, P_T and $f(P_T)$, we can get K_D.



If we can get the KD we can relate it to the free energy of interaction between the RNA and the protein. Can take this and compose it in some way → assume a model that can assume that every position in the RNA contributes with some energy to the energy of interaction. How much depends on the nucleotides on that position.

Take the consensus binding sequence of a protein and make all the single point mutants, and say that the model is: has as many parameters as the length (each position contributes independently) times the nucleotides. Can derive a weight matrix from KD measurements of enough variants. So if I know roughly what the sequence looks like, its relatively easy to generate through all the single point mutants → have enough data to infer these parameters. But can still be quite tedious to do.

Inferring PWMs from affinity measurements

The binding affinity described by the K_D can be related to the free energy of interaction: $\Delta G = -RT\ log(K_D)$

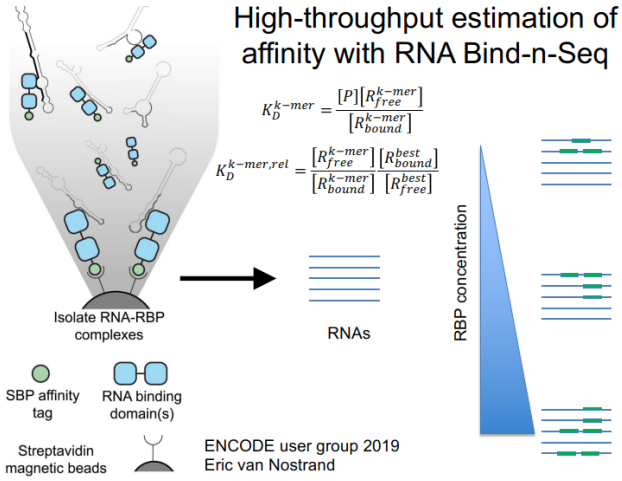
ΔG = Gibbs free energy
R = gas constant
T = temperature

Assuming that each base in the binding site contributes independently to the energy of interaction, $E = \sum_{i=1}^{n} \epsilon_i^{s[i]}$

we derive a procedure for inferring a PWM describing the sequence specificity of the RBP: find the set energy contributions ϵ_i^{α} of each of the 4 bases α at each of the n positions in the binding site that minimize the error in predicting the measured K_D for a set of distinct binding sites.

The entries of the PWM will be given by $w_i^{\alpha} = \frac{e^{-\epsilon_i^{\alpha}/RT}}{\sum_{\alpha' \in \{A,C,G,T\}} e^{-\epsilon_i^{\alpha'}/RT}}$

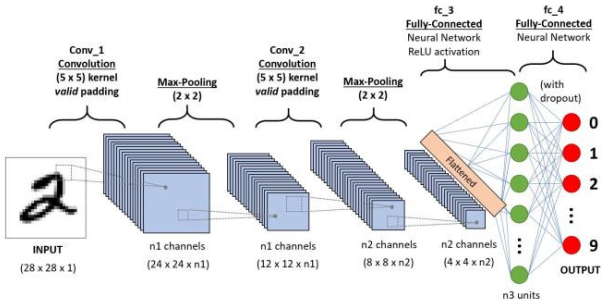
High-throughput estimation of affinity with RNA Bind-n-Seq



RBP experiments: take RNA binding domain of proteins, immobilize them on beads and take a pool of random RNAs, sequence the pool, determiner relative frequency of oligonucleotides, wash them over the RBP → look which sequences are affinity selected by the RBPs. Gives enrichment and affinities of kmers. Don't really try to model all the possible ways of the protein to bind along the sequence, but they assume there is one way to bind the kmer per sequence. Based on the enrichment they derive relative affinities. Measure free kmer versus bound kmer. If we know protein concentration, we can derive KD. If we don't know protein concentration, you can look relatively speaking across all the kmers, what is the range of KDs → can calculate relative KD compared to the one that binds the strongest.

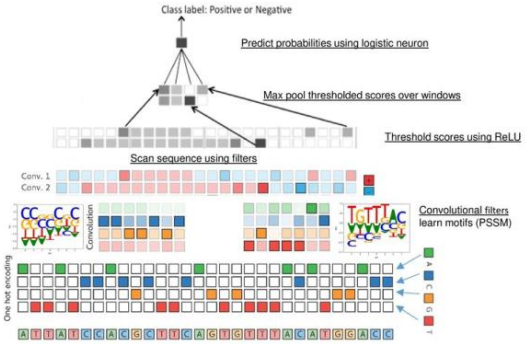
Inferring regulatory motifs with deep learning

Convolutional neural networks

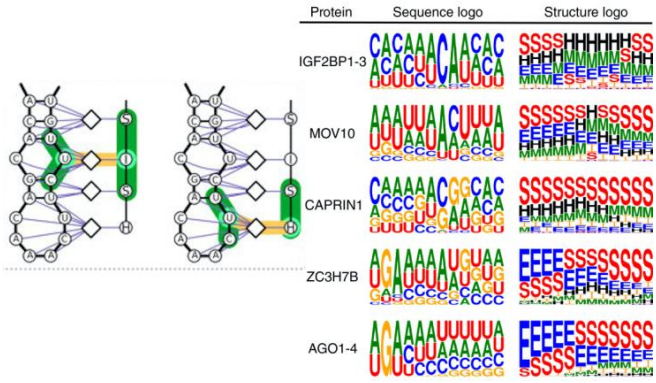


Idea: have sequences you are trying to analyze. Looking for filters (short patterns that occur across the sequence) → a way to find enriched patterns that occur in the sequence.

Inferring regulatory motifs with deep learning



Representing sequence-structure motifs



From CLIP date: if we know for what kind of binding site we look for → look for enriched sequence motifs. Can also predict with computational methods the RNA structure to which the binding site binds to. E.g. look at neighboring interactions etc. and can give probability of one nucleotide interacting with another nucleotide in the sequence, and also give a predicted structure. Look at all the instances where out specific binding site motif occurs and look at what kind of structure is predicted. H - hairpin loop, S - stem, etc.

POTENTIAL PITFALLS:

- High-throughput measurements of affinity have limited accuracy
  - Measurements usually done with protein domains not full-length protein
  - Non-specific interactions - “sticky” domains, interactions outside of the region of the RNA of interest
  - Sample preparations artifacts - biased adaptor ligation, amplification
- Model-driven approaches for analyzing the data are still missing
  - Current approach - k-mer-based
- “black box” aspect of deep learning models
  - Does the architecture allow the inference of relevant features?
  - Is the encoding efficient?

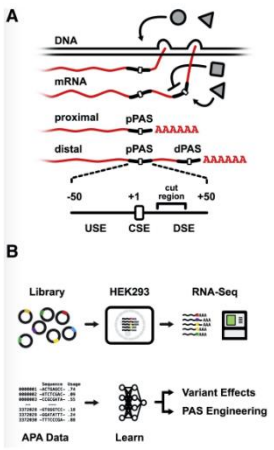
LINKING SIGNALS TO OUTCOMES WITH DEEP LEARNING MODELS

MASSIVELY PARALLEL REPORTER ASSAYS

- Synthesis of a large number of sequence variants
- Evaluation of the variant effect in cells
- Development of a model linking the sequence to the outcome
  - mRNA expression
  - Polyadenylation
  - Splicing
  - Translation
- Advantages: Very large data sets (10^4-10^6) data points, well-control experiment
- Disadvantages: one condition/cell type. Focused on a specific type of sequence/region

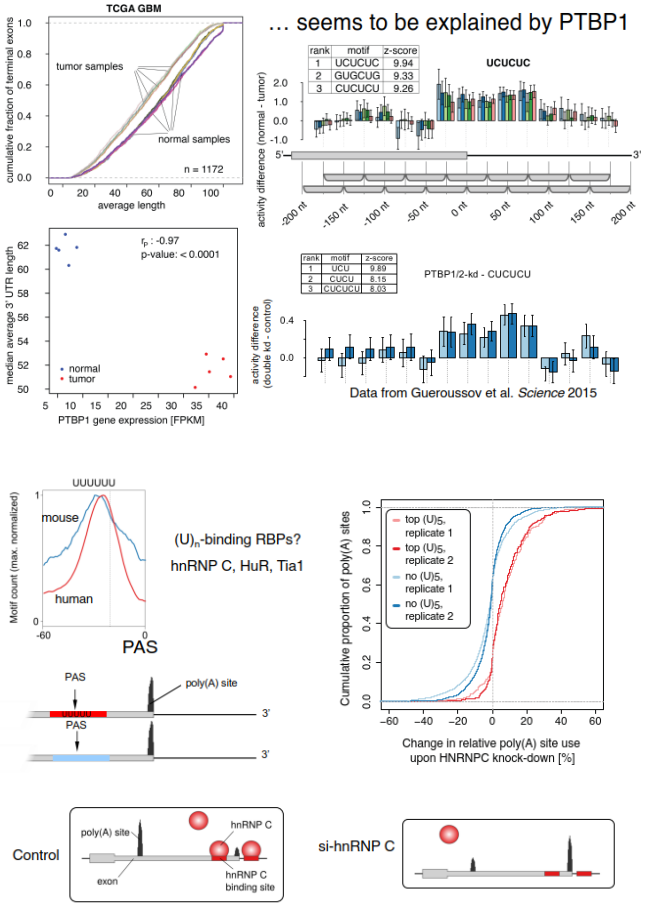
An MPRA approach to model polyadenylation:

Taken the sequence around the polyA site of a particular gene, took the sequence (around 100 nt around cleavage site) and generated thousands of constructs in which one particular subregion has been randomized → basically generate lots of oligos and replace region in construct with randoms. Transfect these in the cells and sequence the outcome of the process. Where did the cut happen? Can also measure which variant has been processed more, which less. Get a table with sequences and associated outcomes → can apply machine learning.



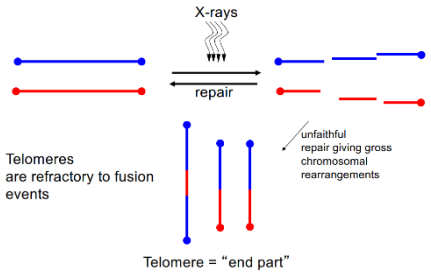
VALIDATING A CODE BY COMBINING DATA TYPES

Motif interference and targeted perturbation:



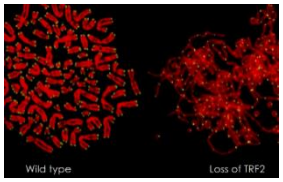
TELOMERES AND TELOMERASE

Chromosome ends are different:

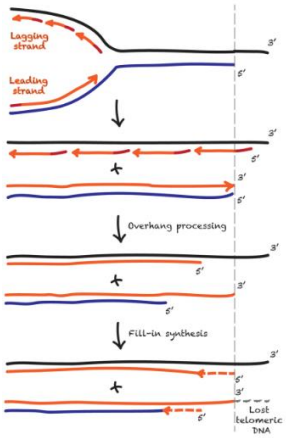


Telomere function to protect from end fusions:

DNA is replicated in a semi-conservative manner, there must be a problem replicating the ends → Telomeres are difficult to replicate!



The Telomere end replication problem:



Is there an Enzyme that can solve the end replication problem? Try to establish reaction conditions that reproduce the suspected biochemical reaction in vitro:

Ingredients: telomeric DNA oligonucleotide substrage, alpha-^{32}P-dGTP, dTTP, dATP, Mg^{2+}, extract

→ incubate at 30°/37° C, analyze reaction products on a polyacrylamide sequencing gel → Discoery of telomerase activity in extracts from Tetrahymena thermophila. Telomerase consists of proteins and RNA.

Telomeric DNA length: 5'000-15'000 in humans; ~50-200 nucleotide 3' overhang in humans, ~300 bp in S. cerevisiae; 3-5 nucleotide 3' overhang

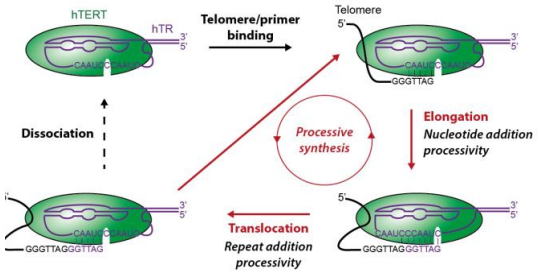
SOLVING THE END REPLICATION PROBLEM:

- Most eukaryotes: telomerase
- Drosophila: HeTa and TART non-LTR retrotransposons
- S. cerevisiae that survived telomerase deletion, some human cancers (sarcomas): ALT-pathway (recombination-based mechanism)
- S.pombe: strains with circular chromosomes can be selected upon deletion of telomerase

A mutant with a defect in telomere elongation leads to senescence in yeast.

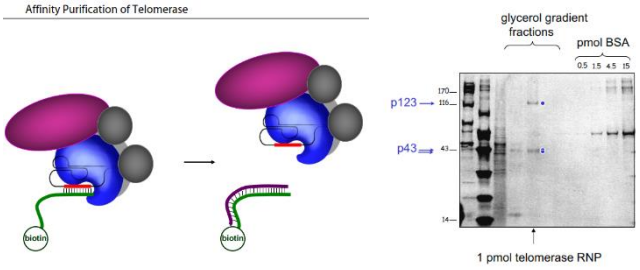
TELOMERASE REACTION CYCLE

- hTERT: human telomerase reverse transcriptase
- hTR: human telomerase RNA

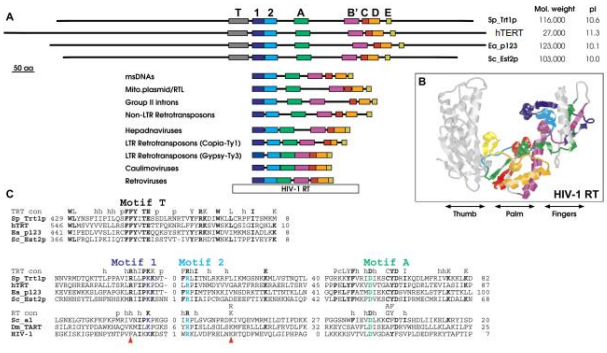


EUPLOTES AEDICULATUS: TELOMERASE RNA

Purified telomerase from Euplotes aediculatus:

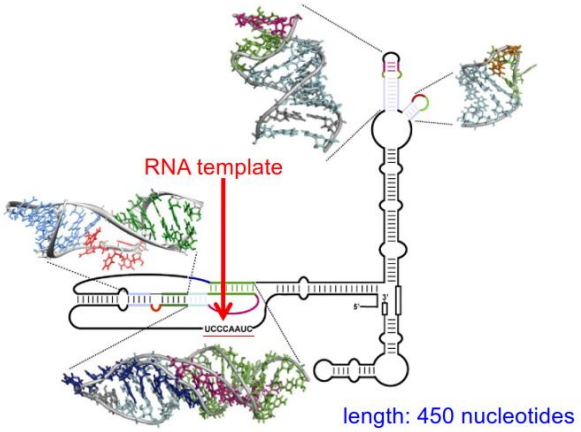


RT-motifs in Telomerase Reverse Transcriptases (TERT)

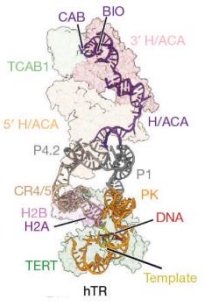


TELOMERASE STRUCTURE

Human Telomerase RNA Secondary Structure

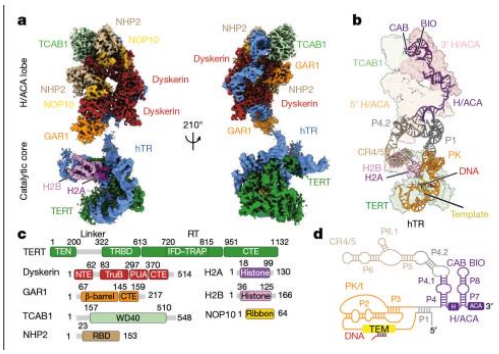


A structure of human telomerase was obtained using cryo-electron microscopy. The overall structure is bilobal and consist of an extended RNA that acts as a scaffold for various proteins.



There is a stem and a loop, and the loop is base-paired with another region in the RNA → pseudo-knot. This structure and model is derived from sequence comparison, by looking at what kind of structure can be conserved between different sequences. This RNA has quickly derived in evolution and has no resemblance to telomeres from ciliates or yeast for example. Much quicker evolutionary clock than, for example, ribosomal RNA.

A structure for the human telomerase holoenzyme was published (cryoEM), and we can see the two views at the bottom.



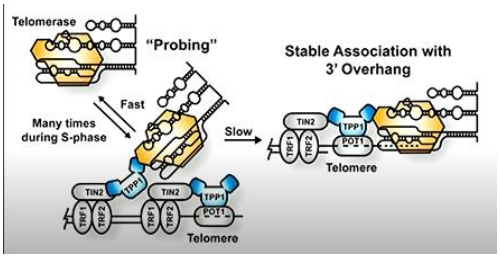
We see the telomerase RNA and several protein subunits and the TERT (telomerase reverse transcriptase) present on one lobe of this two lobe structures. The other proteins are important for the stability and assembly of the telomerase. They also find histone H2A and H2B associated, and it is not clear yet what the significance of the finding is.

TCAB1: required for Cajal body association and telomerase recruitment to telomeres. Important for telomerase assembly in Cajal bodies. H/ACA proteins: hTR stability and telomerase maturation (3' end formation).

TELOMERASE RECRUITMENT TO TELOMERES?

→ Est1 protein in S. cerevisiae

→ TPP1 in human cells: recruits telomerase and stimulates the processivity of the telomerase enzyme (i.e. the propensity to add multiple telomeric repeats prior to dissociation).



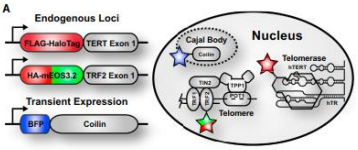
Telomerase forms frequent, short "probing" interactions with telomeres, mediated by the TPP1-TERT interaction, sampling each telomere thousands of times during S-phase. Conversion of a "probing" interaction into a long-lasting static interaction (and elongation), which involved engagement of the 3' overhang of the chromosome, is slow and infrequent. This telomerase-monitoring mechanism allows telomerase to take advantage of small windows of time during which the telomeric single-stranded overhang is available for binding.

Hypotheses: telomerase associates with TPP1, which is a telomere binding protein present at the telomere through a protein bridge with **TRF1** and **TRF2** (important for capping the telomeres, which protects them from DNA repair activity, prevents chromosome end to end fusion).

How they got this data:

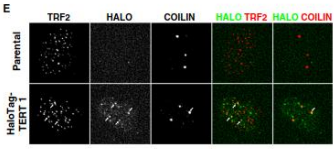
Methods to Assess Telomerase Recruitment

- Chromatin-IP
- Live cell imaging:



HaloTag: enzyme that can attach covalently a fluorescent dye (cells are incubated with fluorescent HaloTag ligand which crosses the cell membrane and becomes linked to the HaloTag).
→ **single molecule imaging !!!**

EOStag: fluorescent protein (EosFP emits a strong fluorescence (516 nm)).



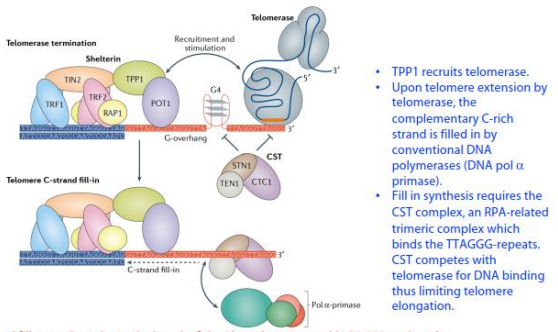
HaloTag is a self-labeling protein tag. It is a 297 residue peptide (33kDa) derived from a bacterial enzyme, designed to covalently bind to a synthetic ligand. The bacterial enzyme can be fused to various proteins of interest.

Analyses of movies (done with HaloTag):

- ~250 telomerase RNPs/ cancer cell (HeLa)
- Thousands of transient telomerase-telomere interactions per S phase: <1 sec
- Few long-lasting interactions: ~0.5 - 8 min (average 3.7 min)
- Thought to correspond to telomere elongation events (telomerase may add ~15-30 nucleotides/min)
- S. cerevisiae: telomerase extends preferentially short telomerase

Termination of telomere elongation and fill-in synthesis:

Termination of Telomere Elongation and Fill-in Synthesis



CST-assisted fill-in synthesis limits the length of the 3' overhang to roughly 50-200 nucleotides

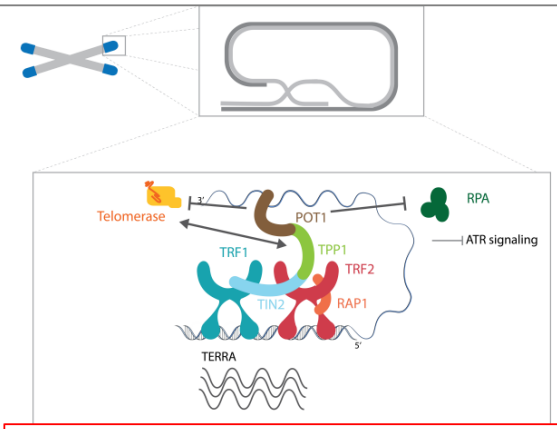
From: NatRevMolBiol **22**, 283 (2021)

CST-assisted fill-in synthesis limits the length of the 3' overhang to roughly 50-200 nucleotides.

TPP1 recruits telomerase. Upon telomere extension by telomerase, the complementary C-rich strand is filled in by conventional DNA polymerases (DNA pol alpha primase)

Fill in synthesis requires the CST complex, an RPA-related trimeric complex which binds the TTAGGG-repeats. CST competes with telomerase for DNA binding thus limiting telomere elongation.

Introduction Telomeres

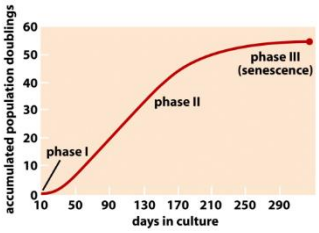


Telomeric proteome: >100 different proteins? Identity and Functions?
How does the telomeric proteome change during aging and in disease?

Telomeres can form these T-loop structures → 3'prime overhang is tucked into the double stranded part of the telomere and makes this displacement loop. Are important, way how telomere hides its end, and makes it different from a double-strand break. Structure requires TRF2. Once structure opens up and it isn't fully stable (e.g. must open up for semi-conservative replication, structure would inhibit replication fork), there is a helicase to open it up, if it opens up it allows DNA checkpoint signaling → I think maybe the same kind of signaling as in double strand breaks → repair response (ATM). Telomeres also prevent repair by homologous recombination and NHEJ. For this you need the Shelterin complex.

WHAT ROLES DO TELOMERES PLAY IN CANCER?

Normal human somatic cells undergo cellular senescence. Adult human body: 10^{13}-10^{14} cells. Life span: 10^{16} cell divisions.



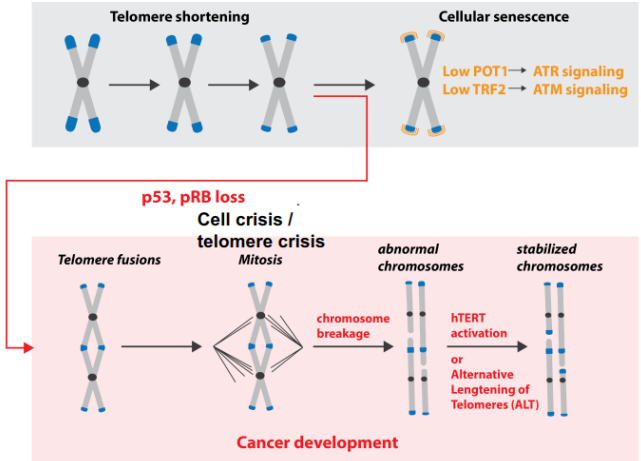
In contrast to primary cells, many cancer cells are immortal in culture (immortal cell lineage). Example: HeLa cells, derived in 1951 from a cervical carcinoma of Henrietta Lacks.

The Regulation of Telomerase and Telomeres shortening regulate cellular lifespan.

Most normal human somatic cells do not express telomerase, simply because the hTERT (catalytic subunit) transcription is repressed, the other subunits are constantly expressed. But: Telomerase is expressed in the germ line, during early embryogenesis, and in stem cells in the adult.

It is expressed in the germline and during embryogenesis (important), and in the stem cells in the adult to some extent (not very high). It is important for our offspring to be born with telomeres that are long enough.

CELLULAR CLOCK: TELOMERES DURING TUMORIGENESIS



Telomere shortening because of end replication problem and the nucleolytic processing of chromosome ends. We lose about 100-200 nt with every round of DNA replication. When the telomeres are short, POT1 and TRF2 are less abundant, T-loop may unfold, may not successfully keep of RPA from telomeres, obtains checkpoint activation with ATR and ATM, which then induces a permanent cell cylce rest → cellular senescence. Are not dead, they just don't divide anymore. Can stay in the body for days to years without doing any harm. This is very important. Like this a cell may hit this senescence barrier before a life-threatening tumor will form.

TWO CRITICAL TELOMERIC STATES DURING TUMORIGENESIS:

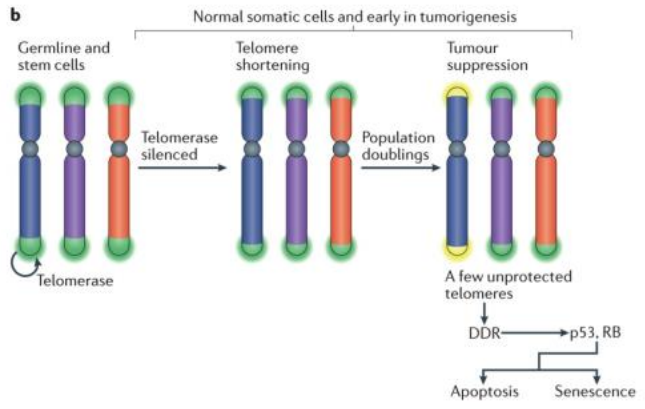
Barrier is overcome, very often checkpoint signaling is somehow perturbed. Often p53 tumor suppressor and pRB tumor suppressor function are loss → leads to ignorance of checkpoint cascade. Because this kinases phosphorylate p53, and then p53 becomes stable, which induces transcription of genes than induce cell cycle arrest → if p53 is lost, this signaling is lost. After a couple of divisions, TRF2 and POT1 levels are so low, that now the chromosome become mistaken as a double strand break, repair machinery starts to fuse chromosome -_> cells continue to divide but telomeres are fused together → get these multicentric chromosomes.

Cells then enter cell crisis which is characterized by cell division cycles (active cycling) but extreme genome instability due to the end-to-end fusions. Most cells will die at this stage. Only one in 10'000'000 can overcome this cell crisis. Those activate hTERT transcription or activate alternative lengthening of telomeres pathway. So the stabilization of telomeres is essential for immortality of cancer cells → must overcome senescence and cell crisis! So the critical states are:

1. Cellular senescence: Permanent cell cycle arrest with a G1 DNA content.
2. Cell crisis: active cell cycling, chromosome end-to-end fusions by MMEJ (alt-NJEH), chromosome fragmentation and mis-segregation, frequent cell death (by autophagy).

Senescent cells: accumulation of DNA damage markers at short telomeres. Permanent cell cycle arrest. Telomeres shorten in cells towards their progression to senescence. DNA damage signaling from short telomeres prevents cell cycle progression.

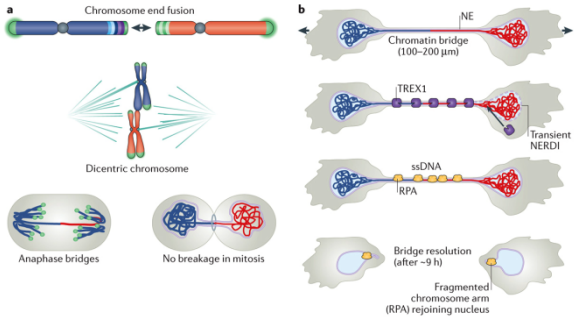
TELOMERE SHORTENING AS A BARRIER TO TUMORIGENESIS:



During development, TERT is silenced. As a results, telomeres shorten gradually. After numerous population doublings, a few telomeres become too short (yellow) and lose their protective function. The kinases ATM and ATR are activated at the unprotected chromosome ends and this DNA damage response (DDR) induces senescence or apoptosis. This limits the proliferative capacity of incipient cancer cells, thus functioning as a tumor suppressor pathway. Cells lacking p53 and RB function can avoid this replicative arrest.

Example of Human pancreatic carcinoma cell: dicentric chromosomes cannot be correctly segregated at Anaphase.

Breakage of Dicentric Chromosomes in Cell Crisis



Dicentric chromosomes formed by telomere fusion rarely break during mitosis. After cell division, TREX1 3′ nuclease resolves the bridges between daughter cells.

Thought that anaphase bridges, do not immediately break because the pulling forces are not strong enough to break a chromosome. Only later, after cell division, you still have the fused chromatin bridges, and eventually the nuclei becomes somewhat leaky and a nuclease called TREX1 then degrades these bridges, which leads to the resolution of the anaphase bridge. But in any case, there is heavy genomic instability during cell crisis.

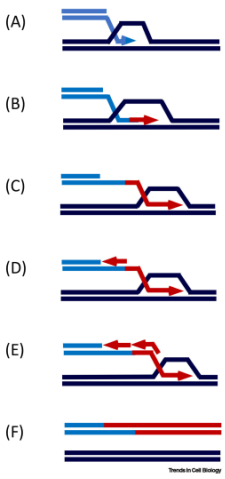
HOW CAN TELOMERE CRISIS BE OVERCOME? → Telomerase is reactivated in cancer cells

Telomerase RT promoter = TERT promoter = most frequently mutated genome region that is non-coding in cancer. The mutation creates a new binding motif for Ets transcription factors and ternary complex factors (TCFS) near the transcription start → leads to activation of TERT gene. Also found mutations in promoter, amplification of TERT gene etc. → selective pressure to turn on this gene in tumors!

THE ALT (ALTERNATIVE LENGTHENING OF TELOMERES) PATHWAY AND TERRA

Most cancers rely on telomerase for maintaining telomeres but roughly 10% rely on ALT (mostly of mesenchymal origin; some brain tumors).

ALT: Homologous Recombination (HR) Promoting Telomere Synthesis



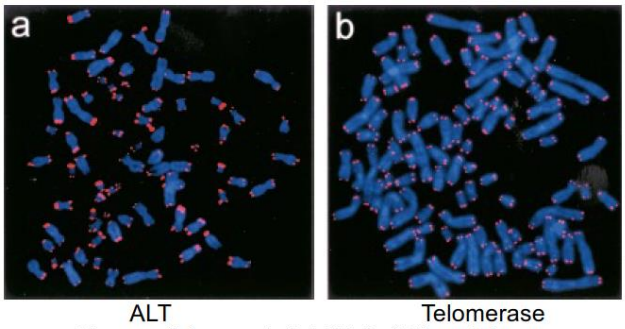
Break-induced replication (BIR). BIR begins by stand invasion (A) to create a D-loop that can migrate as new DNA synthesis (red) is initiated at the invading 3' end (B). Unlike normal DNA replication, BIR creates a long single-stranded intermediate (C) that becomes copied to form double-stranded DNA by delayed lagging-strand synthesis (E), such that all **the newly copied DNA is conservatively inherited** (F).

ALT pathway is less well understood. Idea: shortened Telomere can invade another adjacent telomere and 3' prime end can be extended, and complementary strand can be synthesized. BIR: Not a semiconservative replication but a conservative replication because both strands are newly synthesized.

SOME FEATURES OF ALT:

- Changes in telomeric chromatin: mutations in ATRX/DAXX (deposits H3.3 at telomeres) and the histone H3.3 variant
- Replication stress at telomeres
- TERRA upregulation and TERRA R-loops
- Heterogenous telomere length
- Extrachromosomal telomere repeats
- ALT-associated PML bodies (APBs). Membranes less nuclear structures. Contain promyelocytic leukemia protein (PML), telomeric DNA, proteins involved in DNA repair, recombination, and replication. Required for ALT.

Telomerase-Independent Telomere Maintenance: The ALT-Pathway

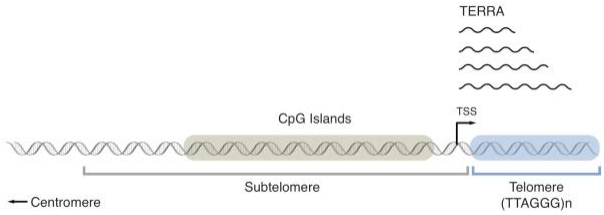


Telomere-specific fluorescence *in situ* hybridization (FISH) on metaphase chromosomes of ALT and telomerase-positive cells, illustrating the highly heterogeneous telomere lengths within individual ALT cells (red, telomere-specific probe; blue, DAPI-stained metaphase chromosomes).

ALT occurs in a smaller fraction of sarcomas (cancer of connective or supportive tissue; bone, cartilage, muscle etc.), many glioblastomas

Telomerase positive cell. Telomeric signals are relatively similar between different chromosome ends, at least compared to ALT cancer cells. Telomere length highly variable in ALT.

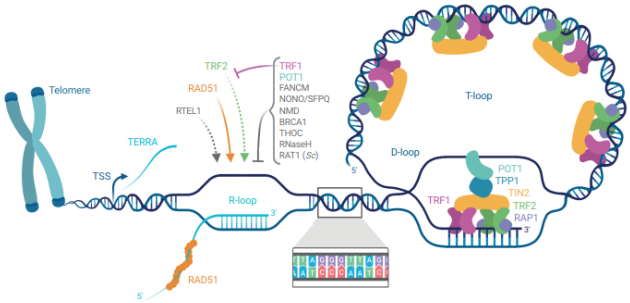
TERRA:



- Conserved in all eukaryotes
- Transcribed at a large number of chromosome ends
- 100-9000nt long
- Nuclear localization, partially co-localizing with telomeres

Strongly upregulated in ALT cancer cells. Is transcribed from sub telomeric regions into the telomeric tract at a large number of chromosome ends. So there are some promoters in the sub telomeric region, transcription starts and an RNA is being made that contains a lot of UUAGG repeats (so contains sub telomeric sequences and repeats). Length is highly heterogeneous. Many colocalize with the telomerase.

TERRA Associates with Telomeres Post Transcription Forming R-Loops in a RAD51-dependent Manner (Feretzaki et al. and Lingner, Nature 2020)



Invades the double-stranded part of the telomere and base-pairs with the C-rich strand of the telomere. Generally these loops are thought to occur during transcription from RNAs that are not successfully removed from chromatin, but it turns out that TERRA can be expressed from a plasmid in a cell, it goes to the telomere, and makes a loop-structure post-transcription. Showed that the RAD51 DNA recombinase that is required for this → RAD51 is well known for an enzyme that binds single-stranded DNA to promote homologous recombination between DNA molecules. It searches for homologous regions by strand invasion and base-pairing.

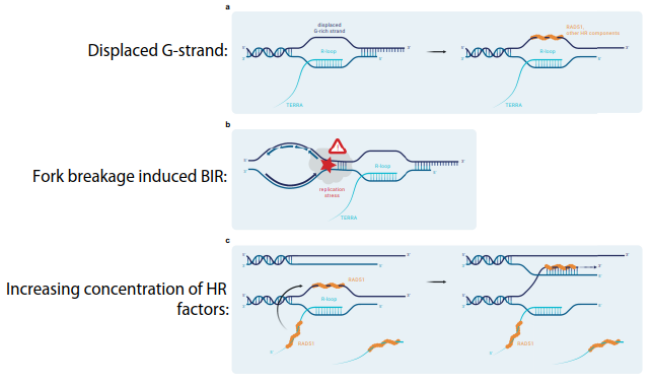
TERRA FUNCTIONS:

- Establishment of telomeric heterochromatin
- Participation in the telomeric DDR
- Regulation of telomerase

- Interference with telomere DNA replication through R-loops → Telomere fragility
- Stimulation of telomere recombination in ALT cancer cells through R-loops

How Do TERRA R-Loops Stimulate Telomere Recombination in ALT?

Models:



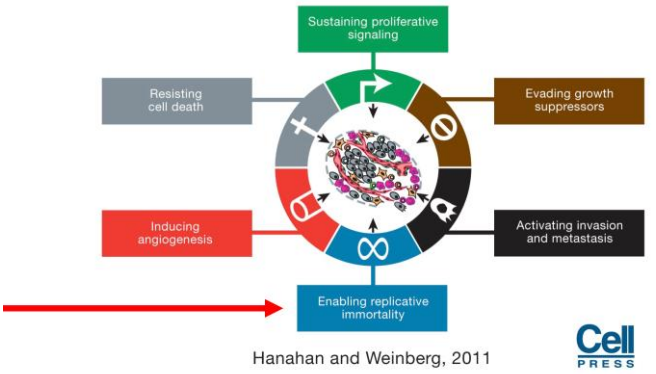
Displaced G-Strand: have R-loop structure. G-rich strand of the DNA is single strand/displaced, then more available for binding factors involved in homologous recombination that act on telomeres in the ALT pathway. Also know that TERRA in normal cells is not present in S-phase, but it is present in S-phase in ALT cells. This interferes with replication, and when there is replication stress this leads to telomere damage which leads to subsequent repair by homologous recombination. Also could be that since TERRA can associate with telomeres and form these R-loop structures by RAD51, it also brings it to telomeres, increasing its concentration near there, which also stimulates homologous recombination. Which of these model or what kind of combination is true is not known. Just know that TERRA contributes, if it isn't there, recombination is strongly reduced, and the ALT pathway is inhibited.

WHAT HAPPENS TO NORMAL CELLS IN WHICH TELOMERASE IS REACTIVATED?

Requirements for immortalization and transformation of human cells

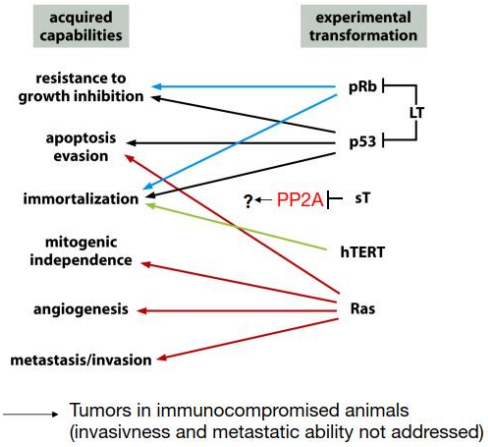
- In tissue culture, normal human cells undergo cellular senescence
- Human cells (fibroblasts, endothelial cells, epithelial cells) can be immortalized by ectopic expression of hTERT if cultured under appropriate conditions!
- hTERT-immortalized cells remain karyotypically and functionally normal
- hTERT-immortalized cells do not spontaneously acquire the characteristics of cancer cells

Acquired Capabilities of Cancer



Hanahan and Weinberg, 2011

Tumorigenic phenotype through defined genetic alterations in human fibroblasts and kidney epithelial cells:



KEY CONCEPTS

- intact telomeres protect chromosomes from end-to-end fusions (suppression of NHEJ and MMEJ)
- TFR2 suppresses NHEJ and ATM; TRF2 also mediates formation of t-loops which prevent ATM activation
- POT1 suppresses ATR
- Cellular senescence: permanent cell cycle arrest with a G1 DNA content; active ATM and ATR signaling
- Telomere crisis: frequent telomere end-to-end fusions mediated by MMEJ; chromosome missegregation and breakage events; active cell cycling accompanied by frequent cell death
- Telomerase: cellular reverse transcriptase which uses an internal RNA template; it counteracts telomere shortening caused by the end replication problem and the nucleolytic processing of chromosome ends. TERT is tightly regulated in humans (but not in mice). hTERT is re-expressed in 90% of cancer (often due to TERT-promoter mutations).
- ALT pathway: recombination-based pathway to maintain chromosome ends. TERRA stimulates homologous recombination through formation of R-loops

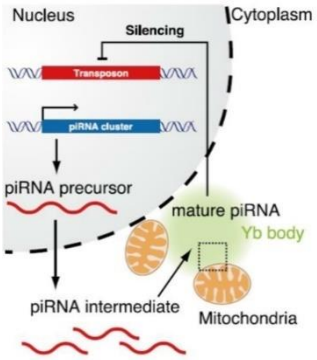
GENOME DEFENSE BY GERMLINE SMALL RNAS

piRNAs: a specialized class of germline-specific small RNAs. They are found only in animals gonads (testis, ovary), and function to repress transposable elements. Bound to Piwi proteins which is why they are called Piwi-interacting RNAs. Function both in nucleus and cytosol.

PIRNA PATHWAY

"All you need to know in one slide!" → Summary of first pathway. Transcription of piRNA precursor, biogenesis of piRNAs and transposon silencing.

- Transcription of piRNA genes or piRNA clusters in the nucleus. (Precursor → long)
- Biogenesis in the cytoplasm in perinuclear granules called nuage or Yb bodies. Non membrane bound structures, RNA protein aggregates, usually surrounded by mitochondria.



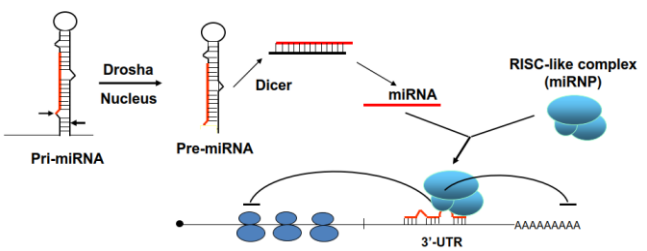
- Silencing in both nucleus and cytoplasm

Transposable elements in human genome about 50%. Move in the genome and could be integrated in new loci, if the gene is needed this could be lethal! Not all of them are active. Some are. The active ones need to be suppressed. piRNA help with this. Loaded in the protein and can silence the transposable elements.

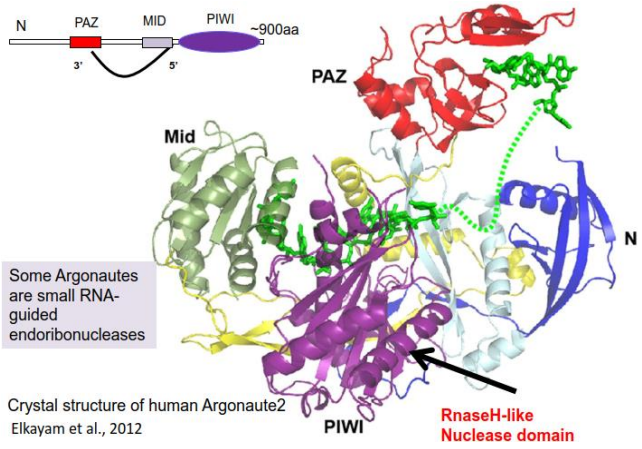
MICRORNA'S

~22-nt regulators of gene expression. Small non-coding RNAs

- C. elegans small temporal RNAs (let7, lin4): early developmental decisions
- Conserved from plants to humans
- Hundreds of different miRNAs identified by cloning & bioinformatics (~1000 in mammals?)
- Advantages of miRNA regulators: specificity, rapidity of action, dosage sensitivity, potential reversibility

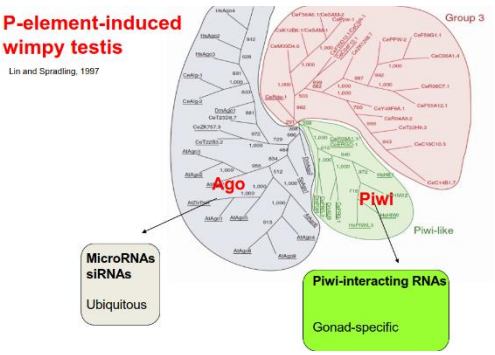


Small RNAs are bound by Argonaute proteins. Real functional unit is the smallRNA-Argonaute unit together. Complex uses sequence-complementarity to target to bind and mediate its function. Can either promote deadenylation or repress translation, both lead to downregulation of gene expression. Argonaute proteins are found in every cell. Bind the small RNA by recognizing the 5' prime end in the MID-domain, and the 3' prime end in the PAZ domain (domains are on AGO). There is a groove in the protein where the target RNA can sit and base pair to the small RNA. This is how Argonaute which are guided by the small RNA would bind the target. Then there is an important activity contained in some of the argonaute proteins = RnaseH-like nuclease domain. This domain is an endonuclease, but its endonuclease activity is activated only when the smallRNA target-RNA complementarity is perfect → Piwi domain can cleave target RNA at a precise location. In between 10th and 11th nucleotide.



Some Argonautes are small RNA-guided endoribonucleases. piRNAs are bound by Piwi proteins. Ago an Piwi proteins (an argonaute protein) bind RNAs of distinct sizes.

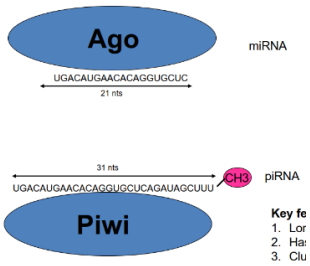
Argonaute Family:



PIWI-INTERACTING RNAS (PIRNAS)

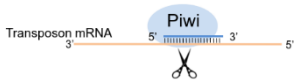
- Discovery of piwi-interacting RNAs: 25-27 nt RNAS
  - Cloned small RNAs from different tissues, found a lot of microRNAs and also some slightly longer RNAs, which were complementary to transposable elements
- Repeat-associated small interfering RNAs (rasiRNAs)

KEY FEATURES OF PIRNAS:

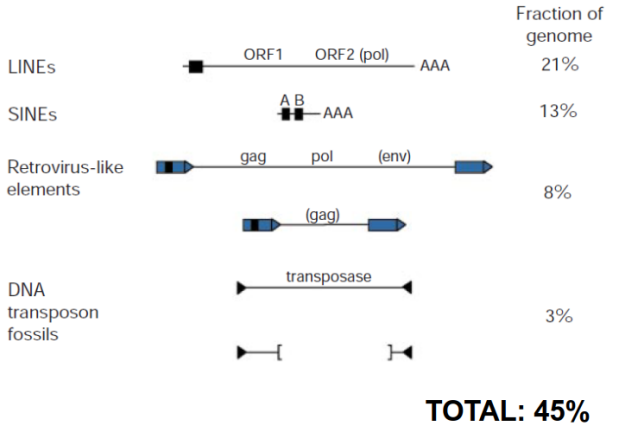


- Longer than miRNAs, 24-31 nt
- Has a 5' U-bias
- Clustered genomic origins
- Dicer-independent processing
- Poorly conserved
- Enriched in transposon sequences

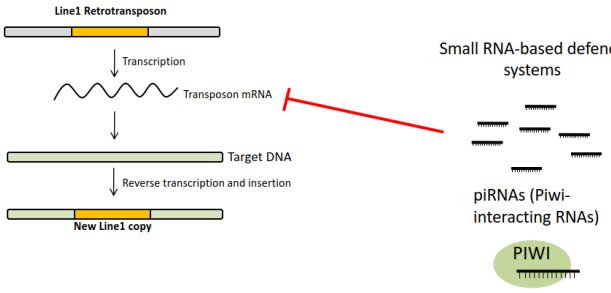
Deep sequencing reveals the complexity of piRNAs! Millions of individual piRNA sequences, while there are only a few hundred of microRNAs.



piRNAs are enriched in transposon sequences. Hypothesis: Slicer activity of piRNAs would target transposons to render them inactive. Half of the human genome is made of repeats:



PIRNAS: AN IMMUNE SYSTEM AGAINST TRANSPOSONS:



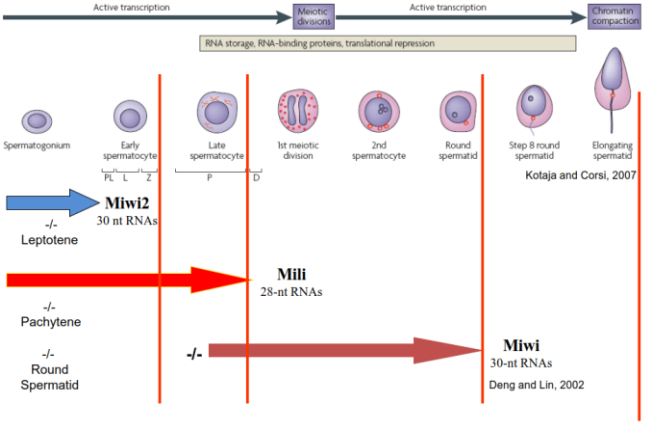
REQUIRED FOR FERTILITY IN ANIMALS: MICE, FISH, FLIES AND WORMS.

Use of small RNAs to control transposable elements is not special to animals. Plants also do this but are not called piRNAs, called siRNAs, use different biogenesis and pathway. Transposons evolve with species and so piRNAs also have to evolve which is why they are not conserved over species, and why they have a different biogenesis and AGO protein to other small RNAs. Almost a personalized defense system because transposon could even be different between mouse strains. There are millions of piRNAs, you don't know where the threat will come from, so you are ready for a lot of different stuff. If there would be millions of miRNAs, the minimal sequence to bind a target is 7 nt (seed sequence, nt nr 2-8). With millions of miRNAs, lots of off targeting would happen → not good!

Sequencing of mouse embryonic testicular piRNAs enriched in piRNAs targeting repetitive elements. In early timepoints of life, piRNA is very low abundance, gets very high in the adult.

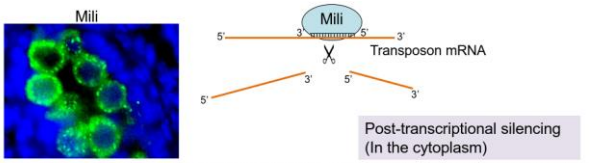
Mouse piRNAs target two active retrotransposons in the male embryonic germline: LINE1 (L1) and IAPEYI. In a Knockout of piwi proteins, these become active in the genome. LINE1 elements derepressed in piRNA mutant mice.

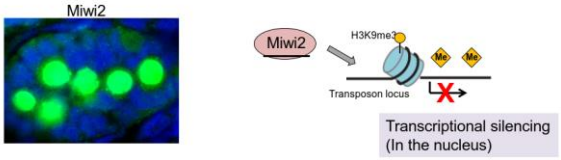
Mobile DNA: Transposon movement causes genome instability. Mouse Piwi KO's display spermatogenic arrest. In mice piRNA is not required for female fertility (but is for males).



Transposon de-repression is accompanied by failure of spermatogenesis. LINE1 transposon-encoded protein L1ORF1p is detected in piRNA mutant testicular germ cells.

piRNAs guide cytoplasmic and nuclear Piwi proteins (Piwi domain on argonaute) to transposon RNA and genomic loci.

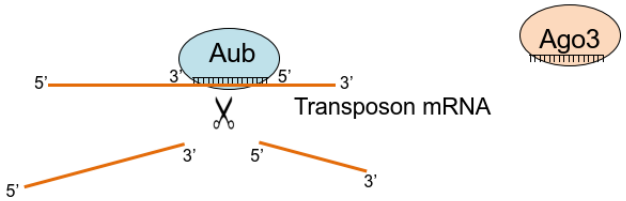




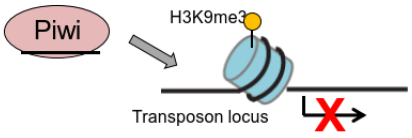
Piwi-Protein binds nascent RNA → which is freshly transcribed RNA that is still attached to the transcribing locus through the RNA polymerase. Can then recruit either histone modifying enzyme or DNA methyl transferases (assembles complexes, and then mediates transcriptional silencing). You need transcription to silence it!

Cytosolic Piwi protein is an endonuclease (cleaves). The nuclear Piwi protein should not cleave the nascent RNA, because it would lead to it falling of (the assembled silencing complex). So every nuclear Piwi-protein has mutations that it makes it catalytically dead → no cleavage.

TRANSPOSON SILENCING IN FLY GERMLINE



Post-transcriptional silencing
(In the cytoplasm)

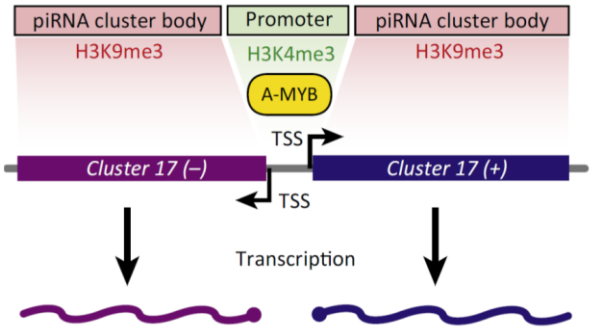


Transcriptional silencing
(In the nucleus)

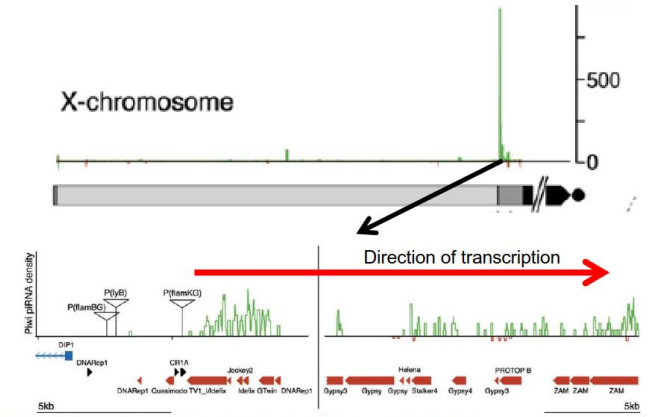
PIRNA GENES OR PIRNA CLUSTERS, TRANSCRIPTION OF PIRNAS CLUSTERS, LONG SINGLE-STRANDED RNAS ARE PRECURSORS OF PIRNAS

piRNAs originate from genomic loci called piRNA clusters. piRNA clusters are transcribed by RNA pol II transcripts are capped, spliced and polyadenylated.

A testis-specific transcription factor A-MYB drives piRNA clusters in mouse meiotic germ cells.



FLY FLAMENCO PIRNA CLUSTER HAS ALL INFORMATION NECESSARY FOR SILENCING GYPSY

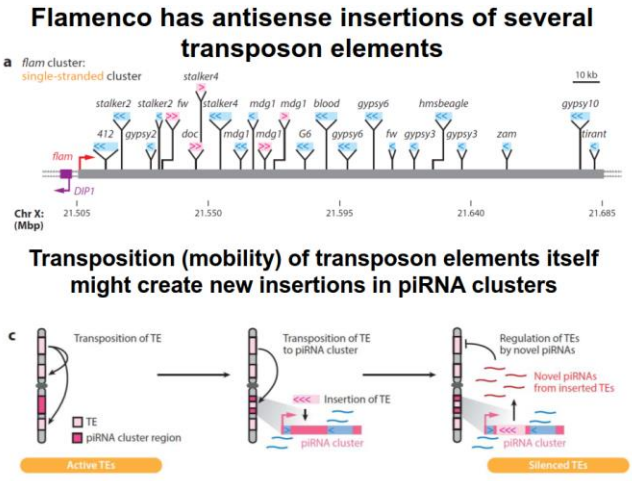


Flamenco has antisense insertions of several transposon elements. Transcription creates a long single-stranded precursor with sequences antisense to several transposon elements.

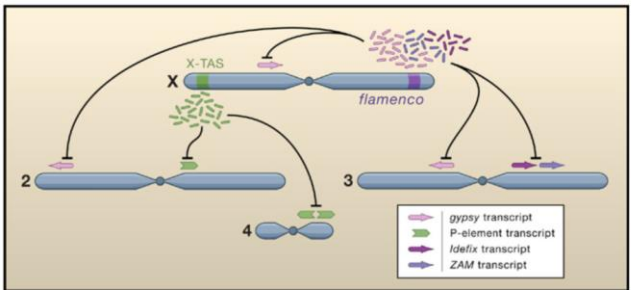
Brennecke et al

Most interesting: compare piRNA cluster to be a repository of information. All the invasion events that this species has suffered throughout its evolution, every transposon, a record of that is kept in this cluster. Have immunity against all these transposon elements → concept of innate immunity of piRNAs.

Storing of these cluster is very analog to the way CRISPR works. Cannot trace any links, but logic and circuitry is exactly the same!



Flamenco piRNAs silence Gypsy and other transposons all over the genome

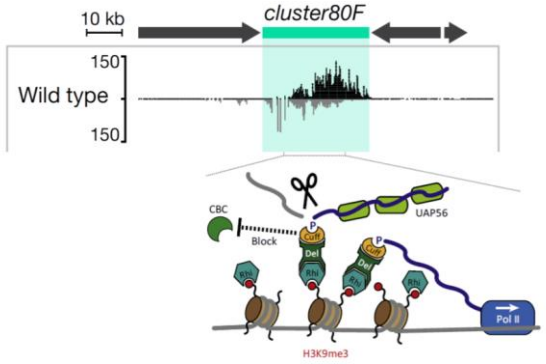


GYPSY: A TRANSPOSON THAT FORMS RETROVIRAL-LIKE PARTICLES.

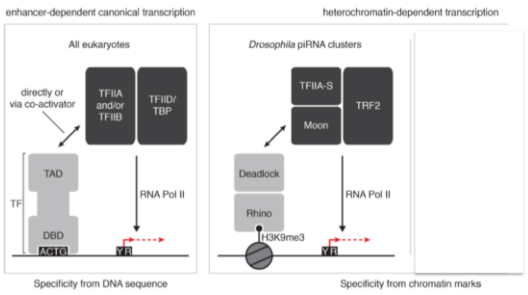
Extreme example in Drosophila. Flamenco cluster silencing gypsy transposon. Happens in somatic follicle cells → drosophila egg chamber. Inside the egg chamber are all the germ cells. Follicle cells are considered somatic cells, but they still possess the piRNA pathway, and it is where the gypsy transposons are made (active when flamenco cluster is inactive), and packaged into virus like particles, and can invade the germ line and get into the egg that is made in the chamber. Transposon self-propagates itself by mobility.

DUAL STRAND PIRNA CLUSTERS IN DROSOPHILA ARE BURIED IN THE HETEROCHROMATIN

piRNA pathway shows a lot of adaptation. Normally we expect a region that is transcribed is supposed to be in an active euchromatic region. But in the case of drosophila system, most of the piRNA clusters (which are transcribed) are embedded in the heterochromatin. How can it be transcribed there? Heterochromatic regions are recognizing through marks: H3K9me3 (histone marks). Through a protein that can recognize it -> Rhino. Can recruit other factors that lead to transcription of that locus! Use of a H3K9me3 reader Rhino to recruit RNA pol II. Both strands are transcribed, piRNAs map to both strands. (**IMPORTANT**)



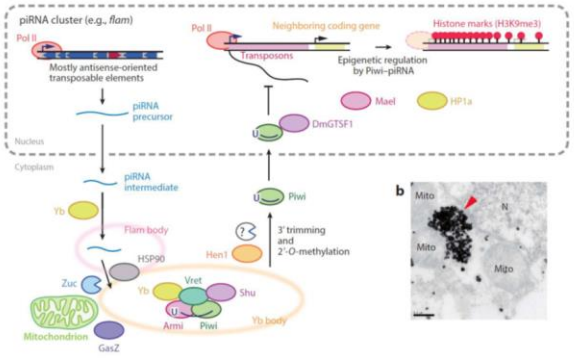
Dual strand piRNA clusters in Drosophila. Use of H3K9me3 reader Rhino to recruit RNA pol II. Rhino-Deadlock-Cutoff-Moonshiner. Use a specialized machinery to access piRNA cluster buried in heterochromatin.



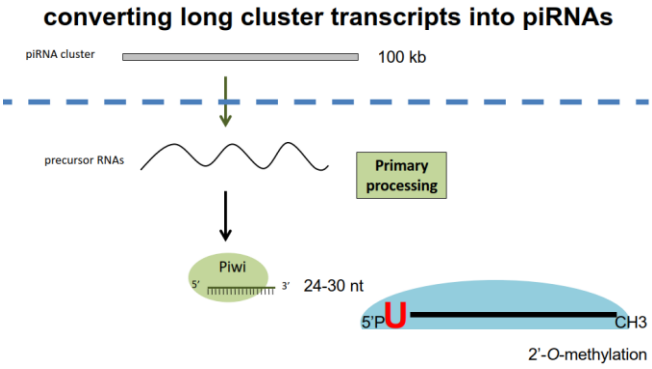
SITES OF PIRNA BIOGENESIS

piRNA biogenesis is a cytosolic process. It takes place in perinuclear granules called Nuage, Yb body, inter-mitochondrial cement.

Primary piRNA biogenesis factors in Drosophila are localized in Yb bodies. After transcription, how does the precursor end up in the granule and get processed? Long standing question, still not a very good answer to it.

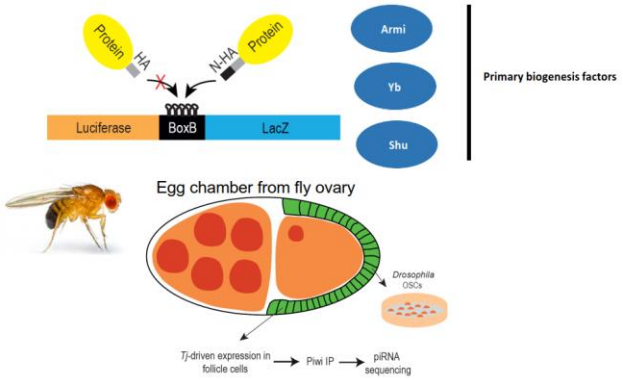


PRIMARY PIRNA BIOGENESIS: HOW TO MAKE PIRNAS DE NOVO FROM A PRECURSOR? (SLICER-INDEPENDENT PATHWAY)



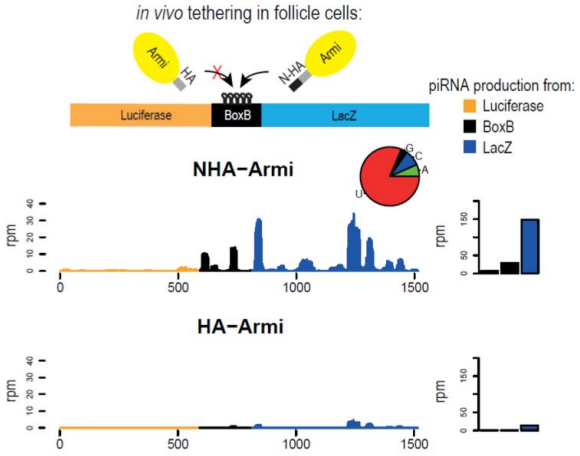
How are piRNA precursors specifically recognized?

Hypothesis: perhaps some signal on the precursor that is recognized by a processing factor. Artificial tethering of piRNA factors (to a reporter, i.e. LacZ, can ensure it this RNA becomes a piRNA) to trigger piRNA biogenesis?



Because we do not know how it works but we assume that real precursors have some kind of binding sites to recruit these factors, so we mimic that with the BoxB system → simply a stem-loop structure that is specifically bound by N-peptide, and the interaction will recruit these factors.

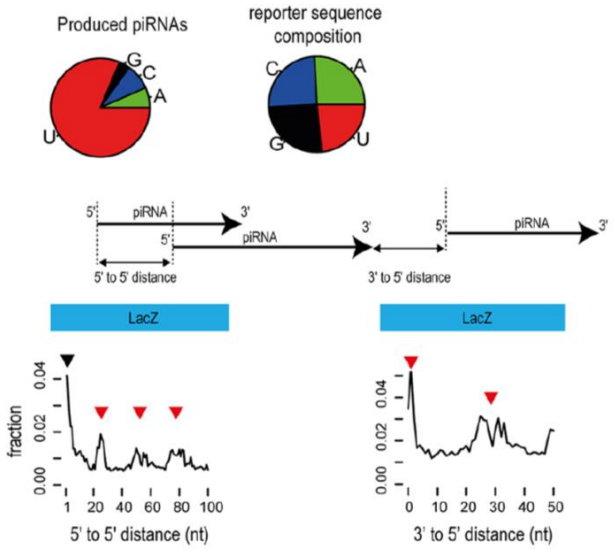
Binding of Armi triggers piRNA production from a transcript



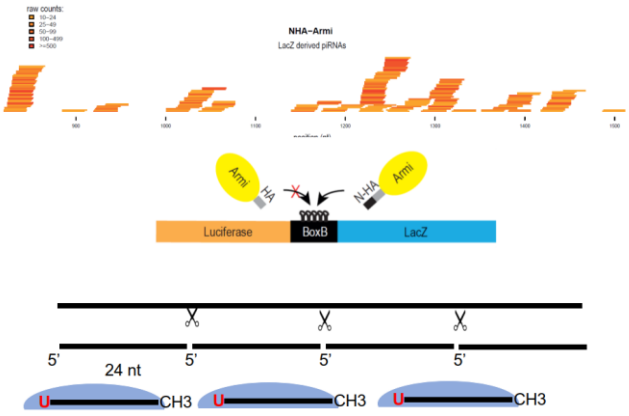
What we see here is, that if we express the reporter together with the proteins in the fly germline and extract the small RNAs and sequence it, found that if you tether biogenesis factor Armi you have piRNA production from the LacZ region.

So: we have this RNA reporter with a sequence for luciferase (orange), a central BoxB cassete (black), and has LacZ downstream. When you tether Armi to the central region (BoxB), you get piRNA made only from the downstream region! Very specific! And you see that if you do it with an HA-tag, so it lacks the N-peptide so it cannot bind to the BoxB, there are no piRNAs made. Suggest that there is a very specific event, after recruitment of the protein, and only the piRNA biogenesis takes place in a 5' prime to 3' prime direction.

PHASED PRODUCTION OF PIRNAS:

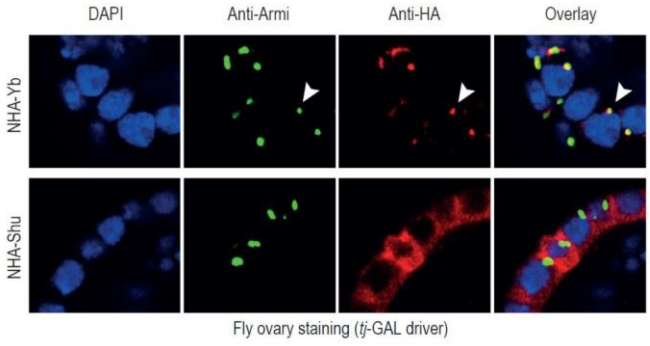


Is there a feature that we can study when looking at these piRNAs made from LacZ region? Looked at distance between one piRNA and the other and calculated this for all the piRNAs produced. A certain pattern emerged, in where the distance between piRNAs is about 24 nt. Fragment is cleaved by some enzyme into 24 nt fragments, to make a series of phased piRNAs.

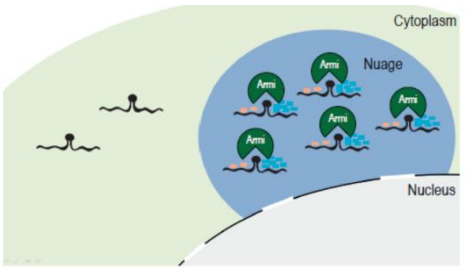


So this is how piRNAs are made: somehow they are brought into the nuage, and then there is an endonuclease that can cleave to chop up the RNA into different fragments. Also tethered other factors, but not all have robust piRNA production. Yb tethering also triggers piRNA biogenesis, but not that of Shutdown

Localization of factors to nuage determines processing

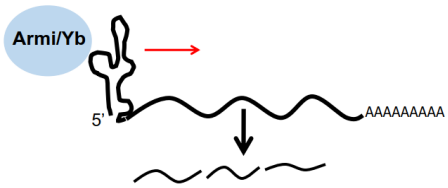


Conclusion: those proteins that have inherent ability to target themselves to the nuages, promote piRNA biogenesis! Tells you, that for piRNA production, you need the precursor specifically identified, and transported to the nuage, and in the nuage they are processed. Recruitment of Armi or Yb to a transcript identifies it as a piRNA precursor by targeting the transcript to the nuage.



What is the specific signal on the precursors, that will specify this? → no idea!

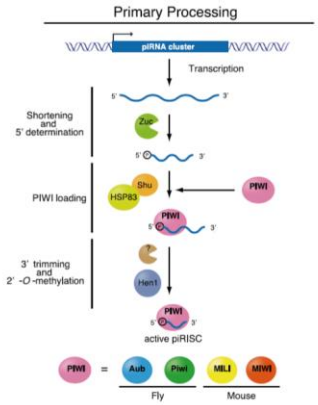
Speculations: Binding of Armi or Yb on a transcript identifies it as a precursor for primary processing,

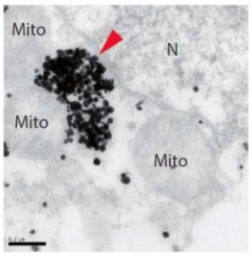


We know that if we take a fragment of the piRNA cluster (took that from the flamenco, the 5' prime most region) and stitch it together with a reporter, than this reporter becomes piRNA. So there is some information in the 5'prime end. We don't think it's a specific sequence (because the different clusters don't share a sequence in the 5' prime), but maybe some kind of feature. It is enigma how they are able to recruit the same proteins to this if there is no conservation across. Speculate that there is some sequence at the 5' end that is specifying, because of experiment where they took sequence from flamenco cluster, fused to GFP and expressed it in germ-cells, and saw that GFP now became piRNAs. So there is something in this 5' that is needed for specifying, but it's not clear what. There are at least 150 different piRNA clusters in the mouse, but no sequence in the 5' end is identical.

piRNA biogenesis proceeds only from 5' → 3'. Don't know how it moves but can speculate. Armi, which is a very essential factor → is an ATP-dependent 5'→3' RNA helicase. Can move along the RNA in the right direction, could somehow help the machinery to move in steps such that 24 nt RNAs are made.

piRNA precursor is fragmented by Zuc cleavage, loaded into Piwi, 3' processed and methylated



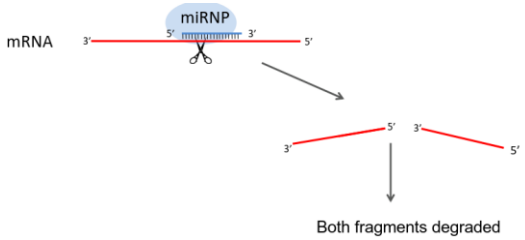


MitoPLD (Zucchini = an essential endonuclease, called Zucchini in drosophila, MitoPLD in mice) = a homodimer. Forms an endonuclease that is tethered to mitochondrial surface. Dimerization on mitochondrial surface generates an active endonuclease. Zuc creates 5' end of piRNAs. Cuts RNA in 24 nt → that's why proximity to mitochondria

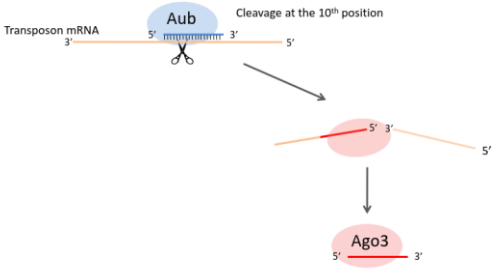
ADAPTIVE IMMUNITY BY SECONDARY PIRNA PROCESSING

Slicing by a Piwi protein initiates piRNA biogenesis from the target RNA.

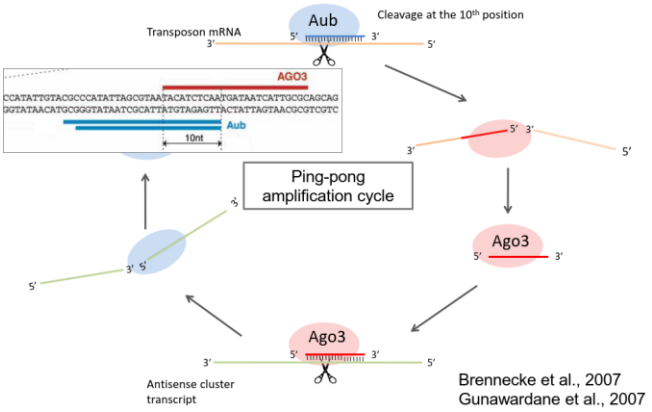
Argonaute slicing: target RNA cleavage guided by miRNAs or siRNAs lead to destruction of both RNA fragments. Perfect complementarity (21 nt complemenatirty) → slicing → degradation.



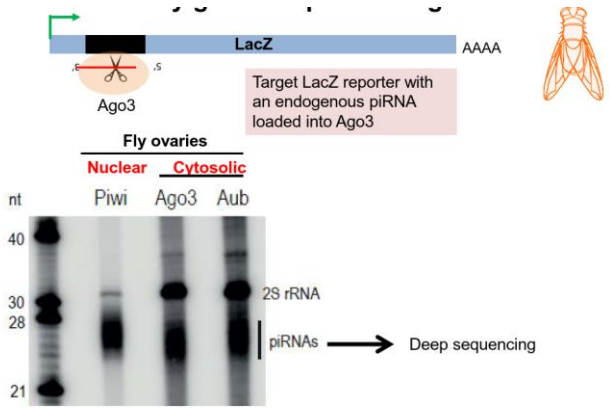
In piRNA pathway: slightly different: Piwi protein called Aubergine in Drosophila. When it cleaves a target, it cleaves it in two, one fragment with 5' end gets loaded in a new piwi protein (then matures into new Piwi protein). So the two fragments are never completely degraded! Piwi slicing leads to one of the fragments becoming substrate for a new piRNA (secondary piRNA)



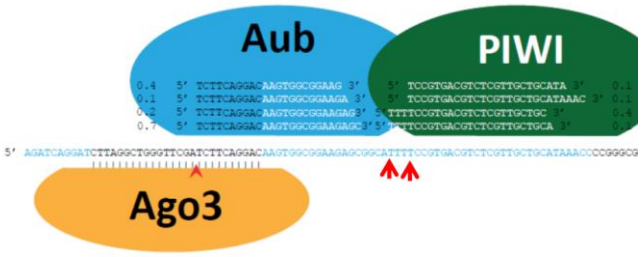
Ping-pong cycle: piRNA amplification (adaptive response): Ago3 binds complementary sequence, again cutting to generate two fragment. 5' fragment binds Aubergine and leads to production of original piRNA!



Looking at how this works at a molecular level: A transgenic fly reporter to study germline piRNA biogenesis

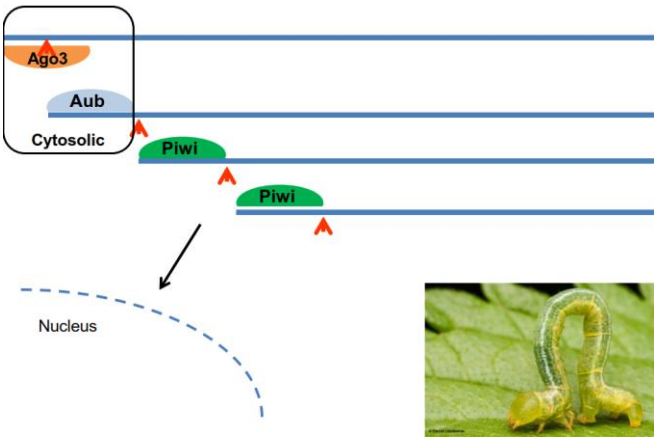


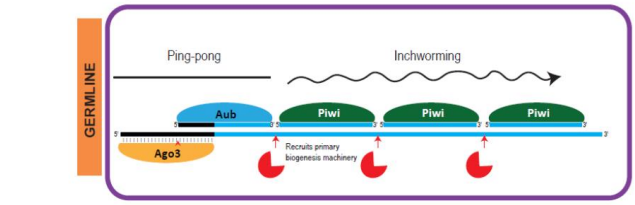
Make LacZ a transposon, by inserting a binding site for one of the abundant piRNAs loaded into Ago3 → so initiate the cleavage by Ago3. Ago3 slicing on LacZ reporter generates 5' end of a new piRNA loaded into Aub. Cleavage downstream of Aub piRNA generate 5' end of a piRNA loaded into PIWI



Surprise: every RNA sequence downstream made into piRNA go into the nuclear PIWI. So you have a cytosolic Piwi initiating cleavage on a target, the first fragment made goes into second cytosolic Piwi protein → ping-pong cycle, and then downstream cleavages introduced by Zucchini which get into the nuclear Piwi protein. Cytoplasmic slicing triggers generation of cytosolic piRNAs and nuclear piRNAs

INCHWORM MODEL FOR PIRNA BIOGENESIS:

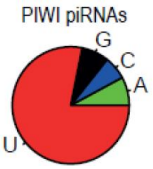


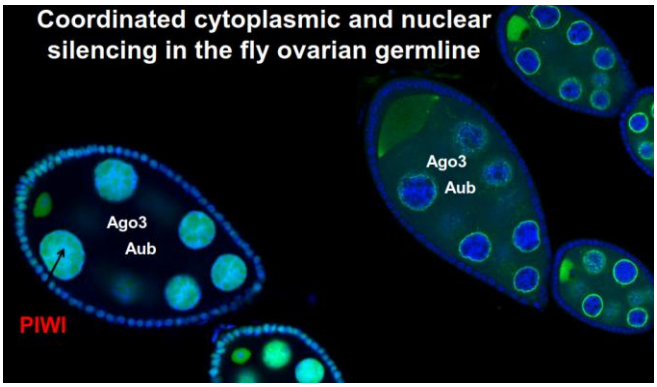


Biogenesis of primary piRNAs in the fly ovaries

Mohn et al., 2015
Han et al., 2015

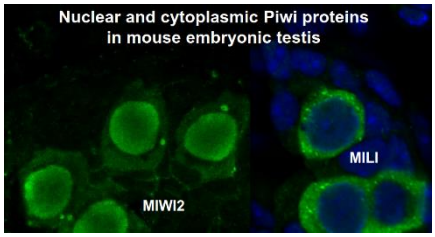
Homolka, Pandey et al., 2015



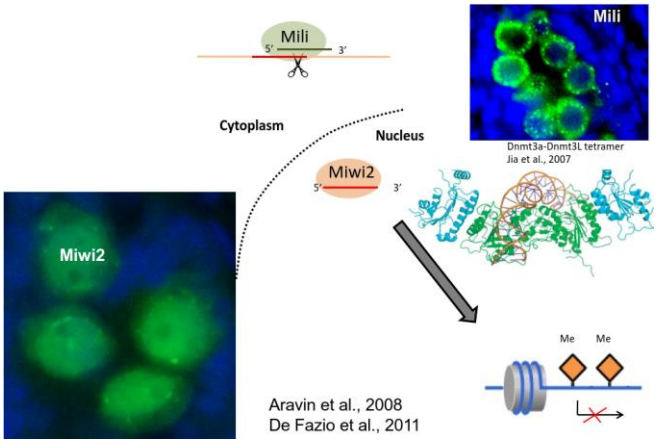


Communication → Ago3 and Aub in the cytoplasm, PIWI is in the nucleus. If you want to communicate to the nuclear compartment that there is a transposable element active, then the first identification is done at the mRNA of that active locus in the cytoplasm, and then that information is sent by taking a fragment of that and sending it into the nucleus.

Similar things happening in the mouse. Don't have to remember the names here (I think).

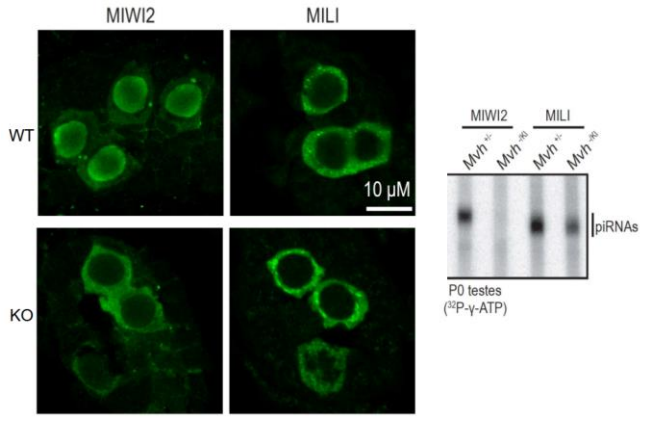


CYTOPLASMIC SLICING BY MILI IS ESSENTIAL FOR LOADING NUCLEAR MIWI2 WITH PIRNAS:



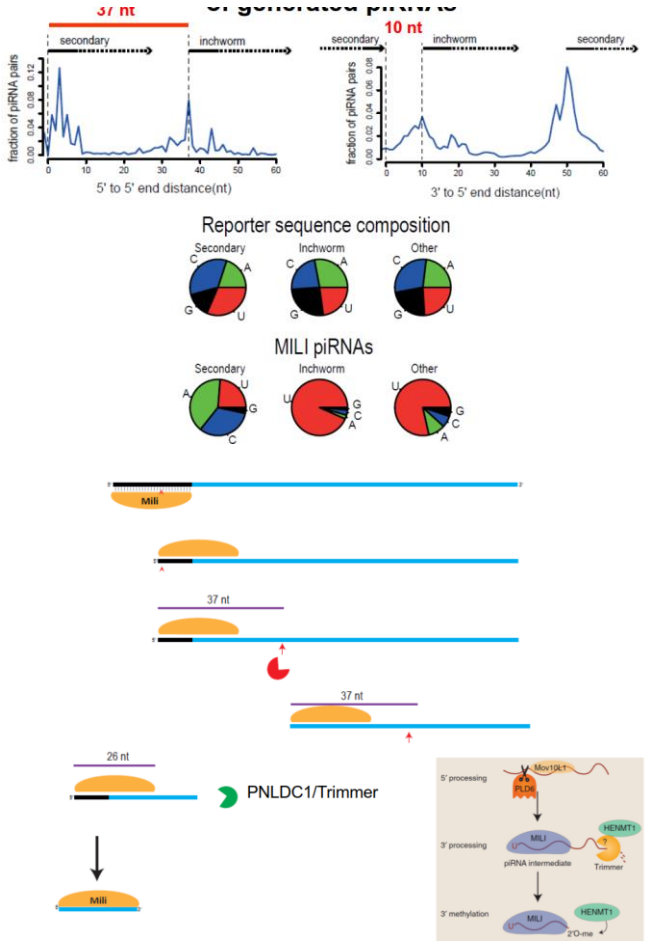
Mili activity is important for biogenesis.

In a piRNA pathway mutant, unloaded MIWI2 is not imported into the nucleus

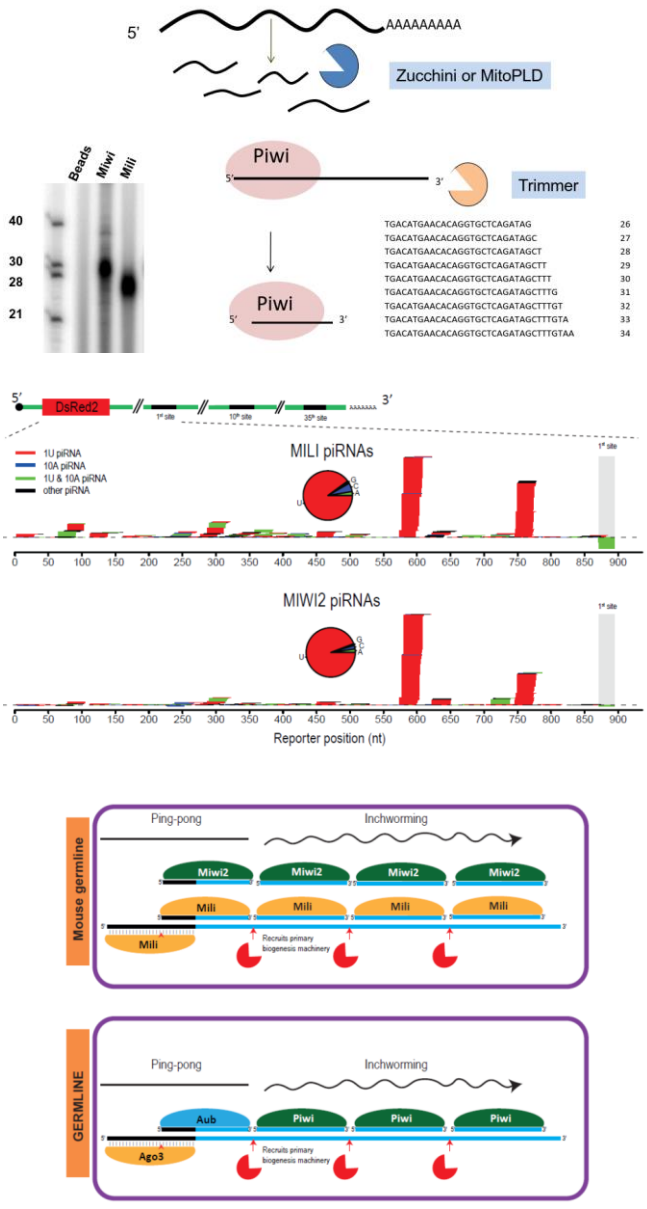


A reporter to study Mili slicer-triggered piRNA biogenesis. Targeted by cytosolic piwi protein Mili. Once cleavage happens, what happens to the LacZ reporter → immunoprecipitation of Mili and Miwi2 from PO testes of mice. Found Mili slicing triggers a series of piRNAs with 5'-3' directionality. Exactly like in Drosophila, cleaved in different fragments, cleavage of a target that initiates biogenesis. Unlike the case of drosophila, all the fragments can be loaded into both MILI and MIWI2 → Slicer dependent biogenesis!

SPACING BETWEEN PIRNAS AND 1U-BIAS OF GENERATED PIRNAS:



Have initial cleavage, and then downstream cleavage happens approximately 37 nt away (as opposed to drosophila piRNAs which are 24 nt). Fragment is trimmed at 3' end by an enzyme called Trimmer (to 26 nt?). It is a 3'-5' exonuclease, and then makes the mature piRNA. Why? No idea, different biogenesis machinery, just the way it is. Once it is loaded it will make the different piRNAs. Footprint of the PIWI protein determines final length of the piRNA



This is what we found in the mouse system: cytosolic Piwi protein cleaving the target, and the fragments are loaded into both nuclear and cytosolic Piwis. In Drosophila (bottom) its different.

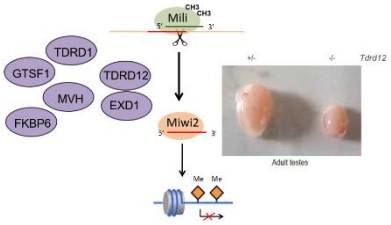
piRNAs with sense-orientation to transposons are triggers. Slicing-triggered piRNA biogenesis allows piRNAs to be generated even when a transposon diverges in sequence to escape regulation, as only one hit is sufficient.

Don't know how different pathways are initiated.

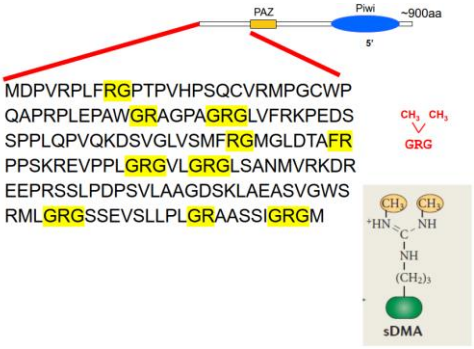
Mouse piRNA biogenesis takes place in perinuclear granules called nuage or intermitochondrial cement. MIWI2 piRNA biogenesis require a number of additional factors not required for MILI piRNAs

PIRNA BIOGENESIS FACTORS IN MICE REQUIRED ONLY FOR MIWI2 PIRNAS:

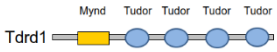
Many factors involved and each is required. If it is knocked out, piRNA biogenesis completely collapses.



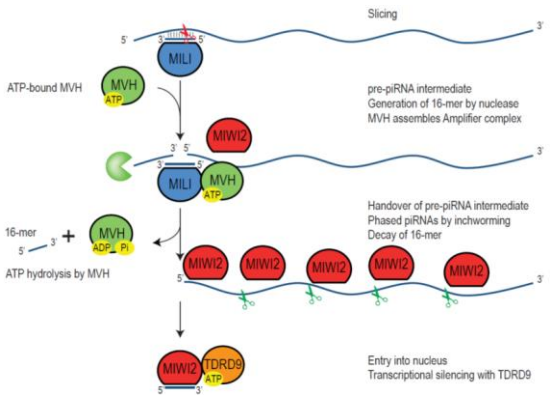
PIWI PROTEINS ARE POST-TRANSCRIPTIONALLY MODIFIED BY SYMMETRICAL DIMETHYL ARGININES (SDMAS):



These are recognized by some of these required factors → Tudor domain-containing (TDRD)proteins recognize sDMAs. Can assemble these large complexes that are required for the process.



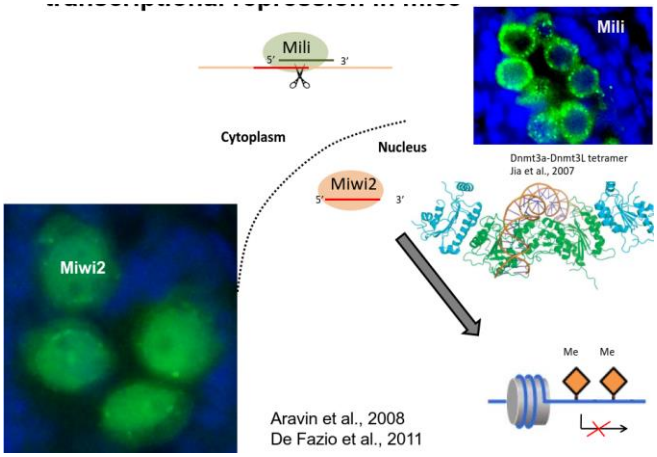
MVH FACILITATES HANDOVER OF THE SLICER-CLEAVAGE FRAGMENT TO THE PIRNA MACHINERY:



We know the cutting is done by zucchini. We know that initiation in the case of slicer-dependent is a slicer Piwi. But how this pathway happens, how the machinery works, how the machine moves in the right direction and how it is orchestrated on the mitochondria is not clear.

NUCLEAR REPRESSION MECHANISMS

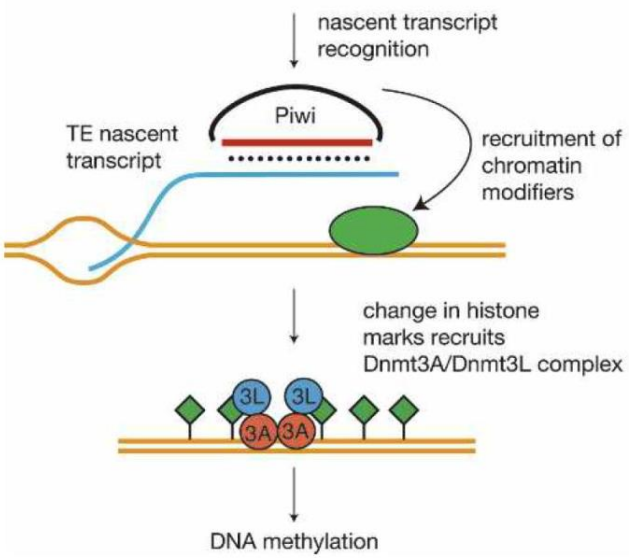
Linking cytoplasmic slicing to nuclear transcriptional repression in mice:



Nuclear components recruits the histone modifications. No DNA methylation in drosophila, in mammalian they recruit histone methyltransferase and DNA methyltransferases to silence the transposable elements (through nascent RNA).

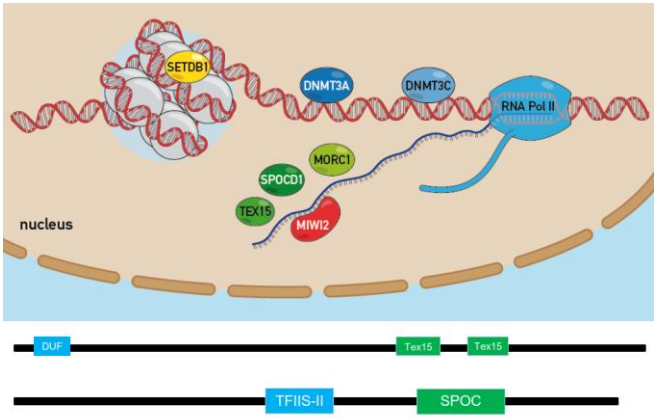
In a piRNA pathway mutant, unloaded MIWI2 is not imported into the nucleus. Absence of nuclear MIWI2 results in loss of DNA methylation on transposon genomic loci

NASCENT RNA FROM TRANSPOSON GENOMIC LOCI SERVE AS ANCHORS FOR MIWI2:



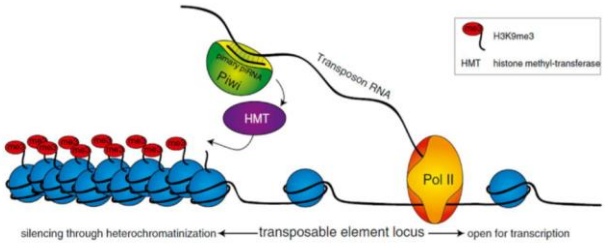
How does it work? Very complex and not understood, entirely machinery is not new. In the case of mice there is even a new DNA methyltransferase: Transcriptional Slicing in mice: the DNA methyltransferase DNMT3C protects male germ cells from transposon activity.

SUMMARY OF CURRENT MODEL IN MICE:

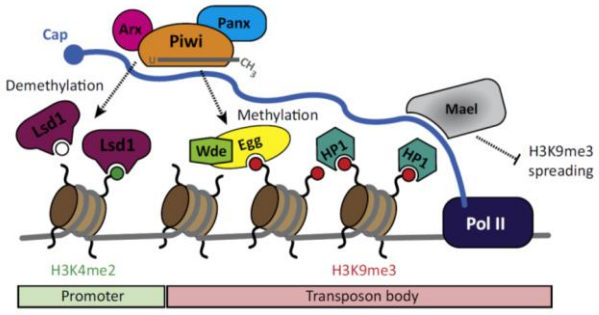


MIWI2 = nuclear PIWI protein that binds nascent RNA, and these nuclear factors that are required for silencing. Knocked out = no DNA methylation on transposable elements, all Transposons are active, and the mouse is infertile. How they work together, who is recruiting, who is interacting is all unresolved.

TRANSCRIPTIONAL SILENCING IN FLIES: Nuclear PIWI recruits histone methyltransferase



Role for Panoramix (not looked at in detail):

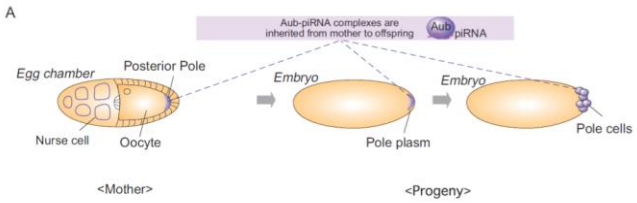


END EFFECT THE SAME: PIWI proteins are guided by piRNAs in the nucleus of germ cells in the organism, they target the genomic locus which produces the transposable elements, and then recruit many proteins which finally end up recruiting histone methyltransferases and DNA methylases which silence. While at the same time in the cytosol, other PIWI-proteins, which also have these piRNAs target the RNA from these locus to silence!

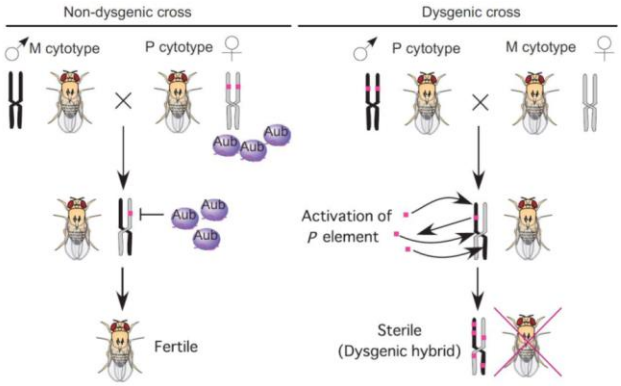
PIRNAS CAN BE MATERNALLY CONTRIBUTED IN THE EGG

In some systems like in Drosophila eggs, mothers can deposit piRNAs which are inherited through the egg into the next generation → absolutely important! Provides protection against transposons brought by the paternal genome.

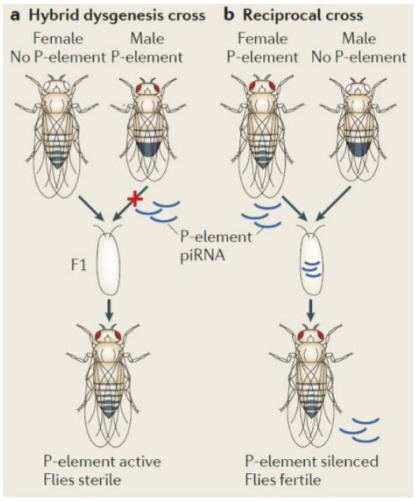
Fly mothers deposit piRNAs/Piwi complexes into egg. Accumulate in pole plasm that forms the future germline formation in the progeny.



Hybrid dysgenesis: Sterility occurs from crosses between different strains of the same Drosophila species



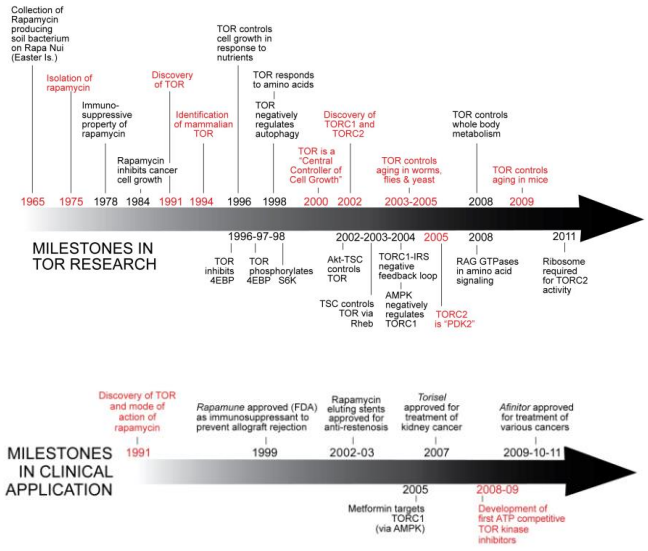
Female has the transposon, so has piRNAs active against it. Female deposits piRNAs into the egg. Progeny is fertile. But if the transposon is coming from the father and is unable to transmit piRNA to the sperm. Progeny has active transposons, destroys germline → infertile = dysgenic hybrid.



LECTURE ENDED HERE (PACHYTENE RNAS NOT INCLUDED IN LECTURE BUT ARE IN SLIDES)

TOR SIGNALING in translation, pyrimidine synthesis and ribosome biogenesis.

TIMELINE OF THE TOR FIELD

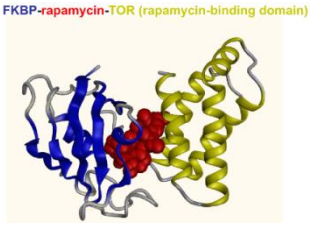


Discovered in Rapa Nui (easter islands) in search of exotic microbes that might make novel secondary metabolites that could be developed into drugs, e.g. anti-fungal (like penicillin). Found a new bacteria: produced Rapamycin, which has anti-fungal activity. But when experimenting on it, found it had the unwanted side-effect of suppressing the immune system. Therefore forgotten as an anti-fungal. But then rediscovered as an immunosuppressor.

RAPAMYCIN

TOR = TARGET OF RAPAMYCIN. Applications in three major therapeutic areas:

- Transplantation (Immunosuppression)
- Cancer
- Cardiovascular disease



Started understanding Rapamycin and its mechanisms of action by working in yeast. Assumption: whatever the target of Rapamycin might be, it probably is highly conserved from yeast to human. And it was first discovered as an anti-fungal, and since yeast is a fungus, we know Rapamycin has activity in yeast. Rapamycin resistant yeast mutants: fpr1 (recessive, FKBP) (common), TOR1 (dominant), TOR2 (dominant) (very rare). Why are there many recessive mutations in one gene that is recessive and cause resistance against rapamycin, and rare ones that are dominant? → related to the mechanism of action of rapamycin. FKBP product of FPR1 gene. Rapamycin alone can bind TOR, but is has no effect. What it needs to has an effect is to form a complex with the FKBP. So FXBP is absolutely essential for drug action, but not essential for cell viability it is a dispensable protein. A simple loss-of-function mutation gives complete rapamycin resistance → common and recessive mutations.

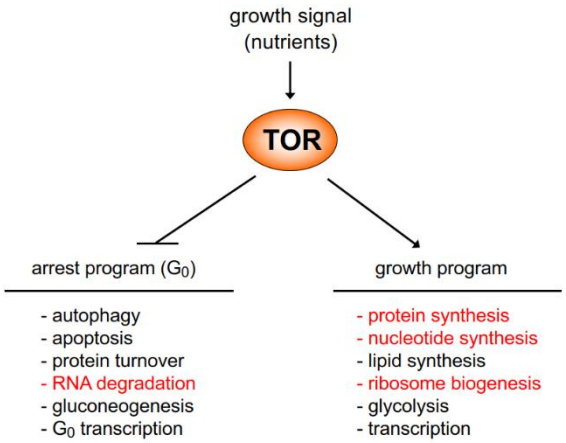
TOR is absolutely essential for cell viability! So it tolerates very little mutation. Every mutation found falls on the same codon → specifies a key residue in the alpha-helix that is a key contact site between rapamycin and TOR → prevents binding without otherwise affecting TOR activity. This is why it is so rare and dominant! Can cause rapamycin resistance even in presence of WT copy of TOR gene.

TOR

TOR is a conserved PI kinase-related protein kinase. It controls cell size. Decreased TOR signaling = smaller. Increased TOR signaling = bigger. TOR controls cell and organism size → *tor* mutants are smaller. TOR1 and TOR2 are very similar. Almost identical. mTOR = human. Shown to be highly conserved from yeast to human and all organisms in between. Yeast unique because it has two TOR gene, all other eukaryotic organism have single TOR gene.

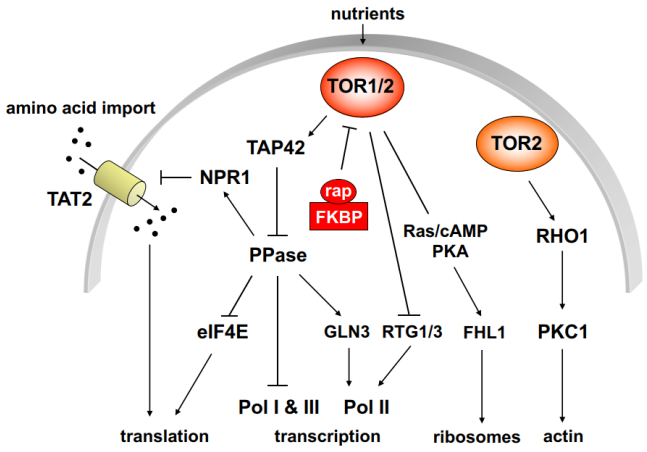
TOR controls cell growth in many contexts: Liver - fasting/refeeding, muscle – exercise (bigger muscle cells), brain - differentiation, in "nature" - making of a queen bee. Tor mutant fly is smaller but perfectional proportioned. Don't know if problem is number of cells or size of cells. Looked at wing and could measure number and circumference of cells. Saw that there is normal number of cells but are smaller →

TOR IS A CENTRAL CONTROLLER OF CELL GROWTH



How does TOR does this? → two groups, the anablic processes which TOR activates, and the catabolic processes which TOR inhibits. Balances these in response to nutrients! TOR what controls our metabolism in response to whatever our nutrient input is, so we grow in a manner that is appropriate to our nutrient status. A lot of these processes are RNA related. TOR controls this through effector signaling pathways which intersect with key proteins involved in these proteins.

CONTROL OF CELL GROWTH BY TWO TOR BRANCHES:



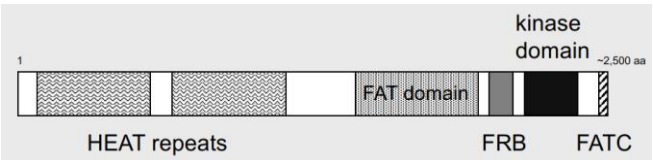
Yeast. TOR Part of two signaling pathways. One is called the TOR2 unique pathway. Controls organization of the cytoskeleton, which determines shape of cell and where new mass is laid down on a cell. Mediates Spatial control of cell growth. Other pathway: TOR shared pathway (both TORs can signal here). Controls all processes which lead to mass accumulation when nutrients are available. Temporal control of cell growth. Spatial and temporal control of cell growth have to be integrated for cell growth to occur in the right place and right tame.

FKBP-rapamycin binds and inhibits TOR but only in the TOR shared pathway branch. For some reason it cannot inhibit in the TOR2 unique pathway.

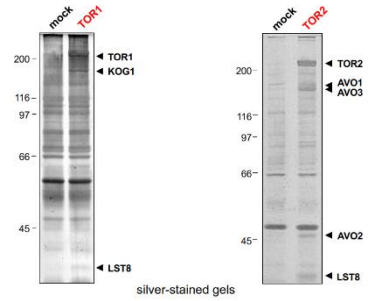
UNSOLVED TOR QUESTIONS:

1. What are the upstream regulators and direct downstream effectors of TOR?
2. How is signaling specificity determined, i.e., why is TOR2 more versatile than TOR1?
3. How does rapamycin selectively inhibit TOR only in the TOR-shaped signaling branch?

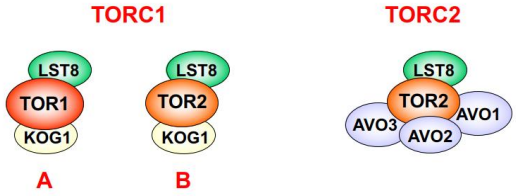
ARCHITECTURE OF TOR PROTEINS:



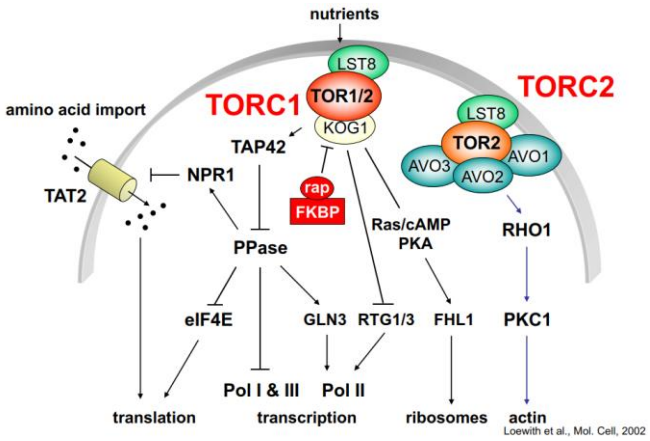
Find TOR bindings TOR by purifying TOR proteins directly out of yeast cells. Gentle enough that if there are TOR interacting proteins, we could co-purify them with TOR and identify them. But before that they wanted to have some confidence that TOR is actually interacting with proteins. So looked as sequence of the TOR protein based on the gene sequence. Very large protein (2'500 aa), made up of several domains. Nobody knew what the function of these domains were. But speculated that they were protein-protein interaction domains. Kinase domain → attached to large landing pad for other proteins to bind and somehow regulate function of this kinase domain. So this made it look like there are actually TOR-binding proteins. TORs contain several protein-protein interaction domains. But another experiment was done to be sure, which was to isolate TOR1 and TOR2 and run them over a sizing column → biochemical trick to get rough idea of the size of a protein. Predicted to have size of 300 kDA based on sequence. In the Column they behaved like they had a size of 2000 kDa! → almost 10x bigger than predicted. So → maybe part of a larger complex! → made a purification of TOR (cation exchange, affinity column) → couldn't purify TOR! Several proteins co-purify with TOR1 and TOR2.



TOR1 AND TOR2 ARE IN LARGE COMPLEXES. TWO TOR COMPLEXES:



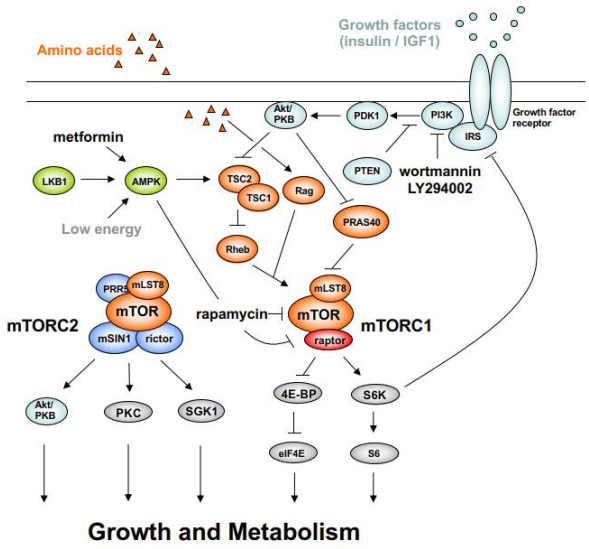
CONTROL OF CELL GROWTH BY TWO TOR COMPLEXES:



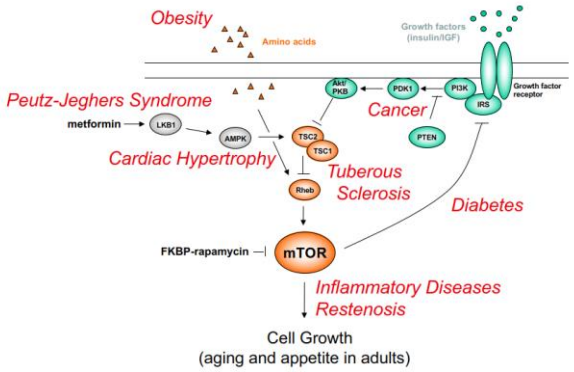
Mammalian orthologs:

yeast protein	mammalian protein	% identity
TOR	mTOR	~40
KOG1	raptor	~30
LST8	mLST8 (GβL)	~30
AVO1	mSIN1	~20
AVO2	?	
AVO3	rictor (mAVO3)	~25

MTOR SIGNALING: TWO BRANCHES AND TWO COMPLEXES CONSERVED:

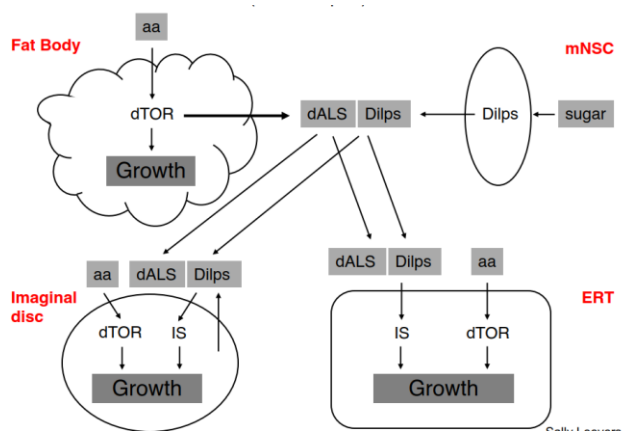


Disorders resulting from dysregulation of mTOR signaling

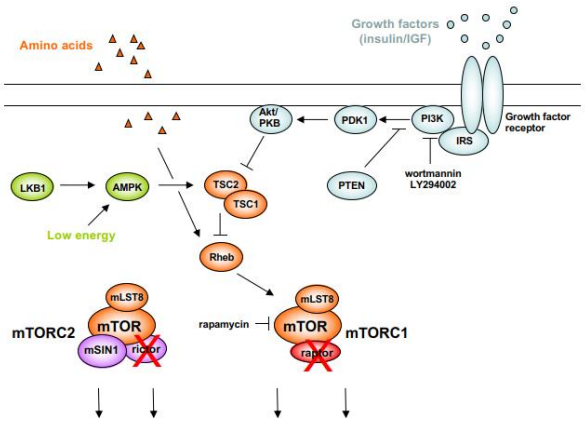


TOR signaling is linked to a large number of disease. They all are characterized by inappropriate cell growth! Rapamycin fed late in life extends lifespan in genetically heterogenous mice. TOR shown previously to control lifespan in yeast, worms and flies and now in mammals. Control of lifespan is via control of protein synthesis:

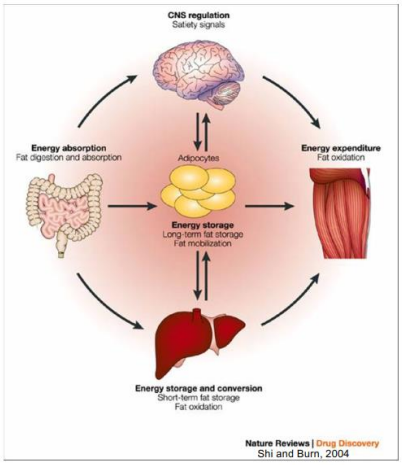
SYSTEMIC GROWTH CONTROL BY TOR: SIGNALING BETWEEN ORGANS:



MTORC1- AND MTORC2-SPECIFIC KNOCKOUTS



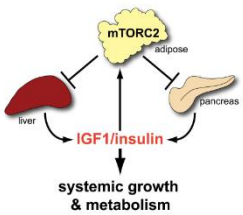
Mutant Mice! Role in metabolic tissues → the ones that are the most nutrient and insulin responsive and mediate energy homeostasis in the body. Which is also the three inputs which control TOR. Role of mTORCs in signaling between organs:



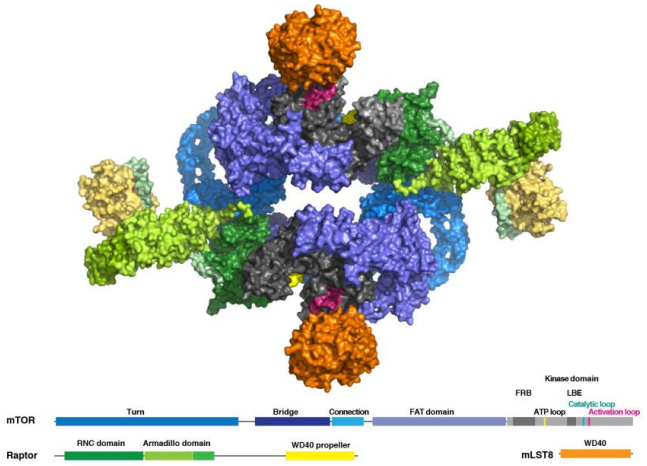
Nutrient-sensing, energy-balancing metabolic tissues

ADIPOSE MTORC1 CONTROLS WHOLE BODY METABOLISM:

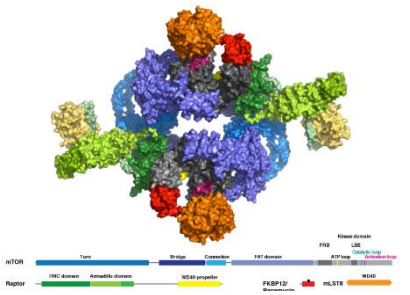
Adipose-specific mTORC1 Kos are lean and resistant to diet-induced obesity and have better metabolic parameters. Leanness is probably due to higher energy expenditure resulting from mitochondrial uncoupling. mTORC1 controls adipose metabolism, and thereby whole body energy homeostasis. Implications for drug therapy, aging? Adipose mTORC2 controls whole body size:

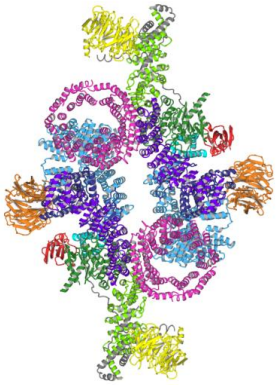


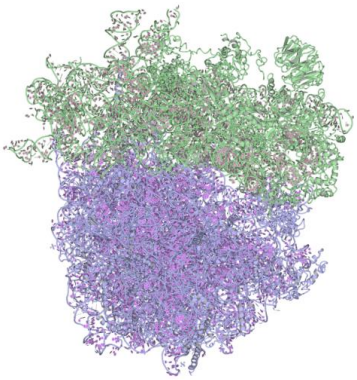
Overall architecture of mTORC1 at 5.9 Ampere - kinase face:



Overall architecture of mTORC1 in complex with FKBO12-Rapamycin:

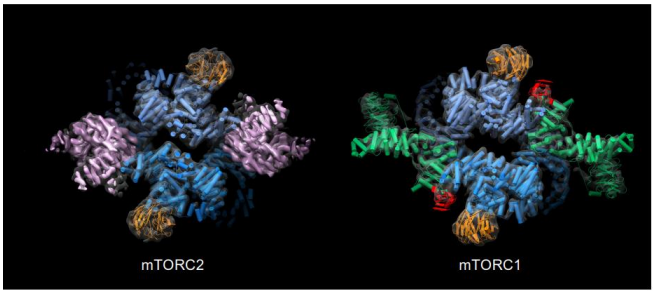






mTORC1 80S Eukaryotic ribosome

Structures of mTORC2 and mTORC1

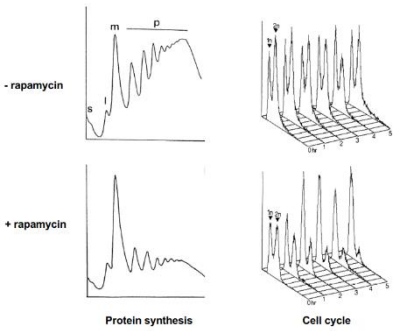


Same structure but different subunits. Binding site for FKB-rapamycin in TORC2 is masked by a subunit!

Mammalian TOR complex 2 controls the actin cytoskeleton and is rapamycin insensitive.

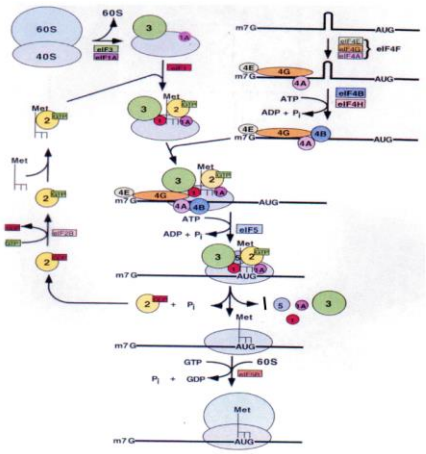
I. MTORC1 ACTIVATES TRANSLATION

TOR controls translation initiation and thereby cell cycle

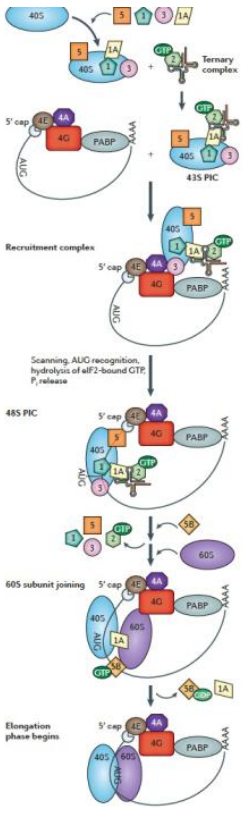


Observation that implicated TOR in protein synthesis. If you treat with rapamycin, you lose all these polysomes → translating ribosomes. Monosomes accumulate in place. Can block protein synthesis and accumulate monosomes tells you that not only protein synthesis is inhibited, but a stage of translation initiation.

Overview of translation initiation (eukaryotic)



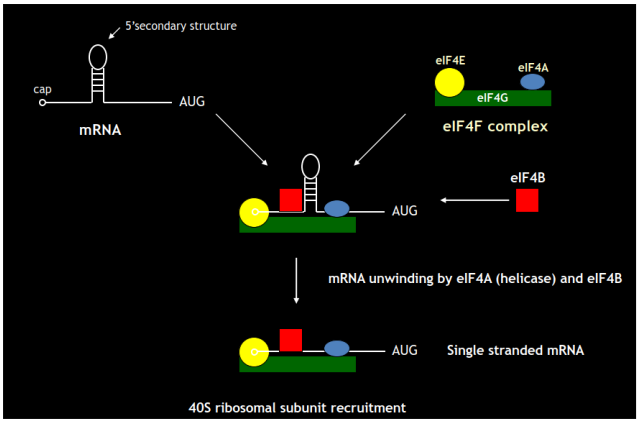
Overview of translation initiation (eukaryotic)



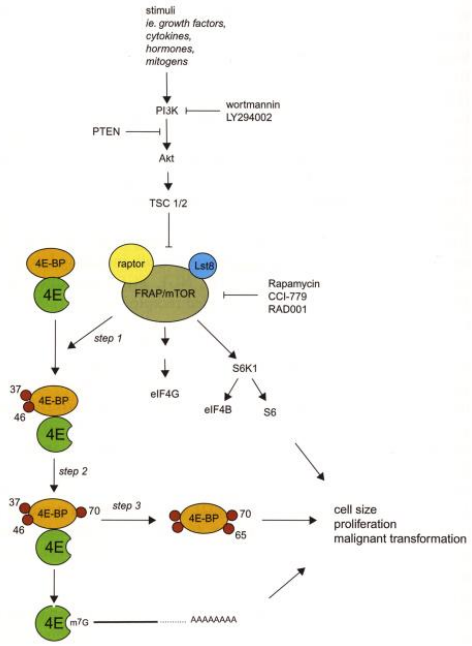
Kong and Lasko, 2012

Highly structured mRNA 5'UTR: how does ribosome bind?

Ribosome recruitment to the mRNA: mRNA unwinding:



mTORC1 activates translation initiation by phosphorylating 4E-BP and S6K



4E-BP bind eIF-4E and blocks formation of pre-initiation com-

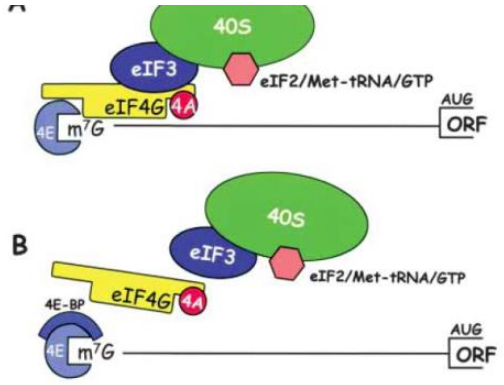


Figure 4. (*A*) Assembly of the mammalian ribosome initiation complex at the 5′ end of an mRNA. eIF4E, as part of the eIF4F complex, binds the m^{7}G-cap structure. eIF4G binds eIF3, which, in turn, recruits the 40S ribosomal subunit along with its associated ternary complex (eIF2/Met-tRNA/GTP). Not shown are other initiation factors that participate in ribosome recruitment. (*B*) 4E-BPs binds the dorsal convex surface of eIF4E to prevent its interaction with eIF4G, thereby abrogating ribosome binding.

plex:

mTORC1 controls translation initiation by phosphorylating 4E-BP and S6K: eIF3 serves as a scaffold to coordinate events leading to initiation:

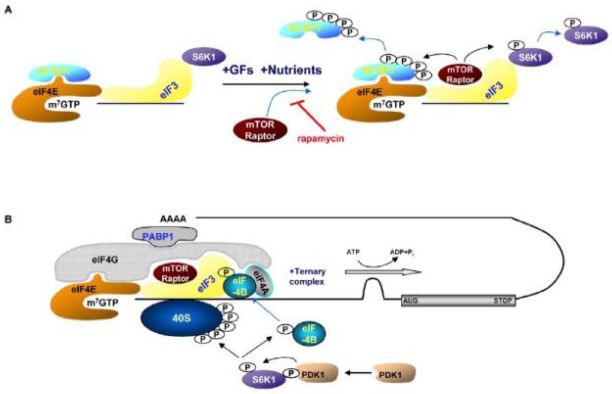
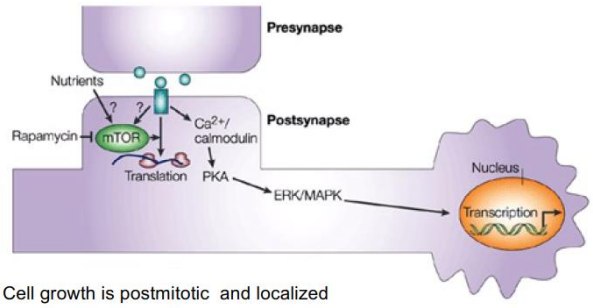


Figure 7. Model of Dynamic Interaction between the eIF3 Complex and mTOR/Raptor, S6K1, and eIF4B

Under basal conditions, S6K1 is associated with the eIF3 complex. Upon mitogen stimulation, an activation complex is formed in which mTOR/raptor is recruited to the eIF3 complex and phosphorylatea S6K1 and 4E-BP1. Phosphorylation of S6K1 at T389 leads to its dissociation from the eIF3 complex (A). T389-phosphorylated S6K1 then binds to PDK1, which phosphorylates S6K1 at T229. The fully activated S6K1 then phosphorylates eIF4B and S6 (B). Phosphorylation of eIF4B at S422 promotes its association with the translation preinitiation complex. Concurrently, mTOR/raptor phosphorylates 4E-BP1 (A), which dissociates from the cap complex, allowing recruitment of the eIF4G scaffold to the cap bound eIF4E as part of the eIF4F translation initiation complex (B). This sequence of events leads to preinitiation-complex assembly and initiation of protein translation.

Does this apply to every mRNA or only those with pioneer initiation?

A role for TOR in learning and memory (LTP): localized protein synthesis:



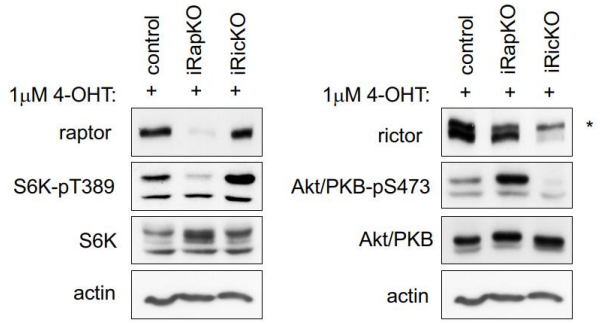
Cell growth is postmitotic and localized

II. MTORC1 ACTIVATES NUCLEOTIDE SYNTHESIS

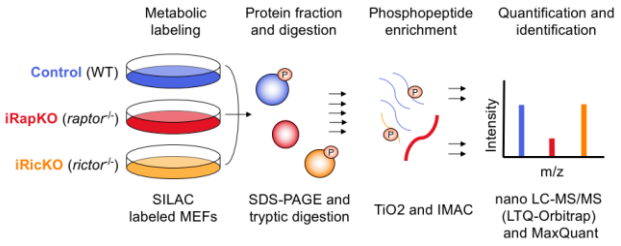
Despite central role of TOR in cell physiology, relatively few downstream effectors are known.

Cell growth and metabolism

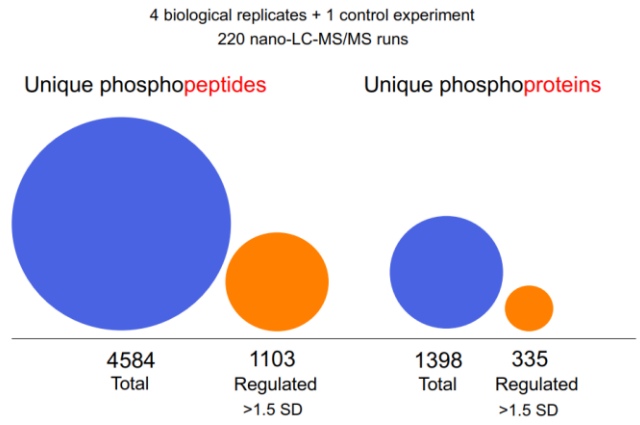
Inducible KO of raptor (mTORC1) and rictor (mTORC2)



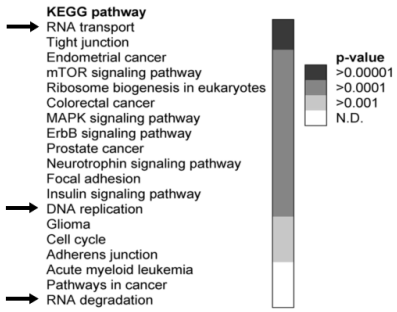
Quantitative mTOR phosphoproteome:



Statistics of the mTOR phosphoproteome

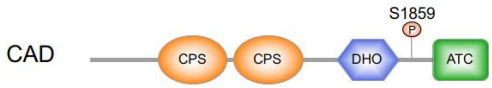


Cellular processes represented in mTOR phosphoproteome:



All terms related to Nucleotide metabolism → unexpected

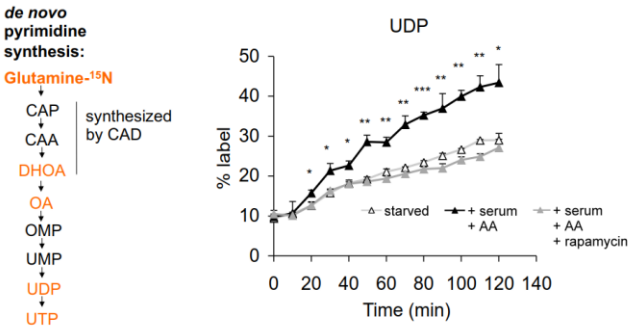
MTORC1 PHOSPHORYLATED CAS (SER1859):



CAD protein encodes three distinct enzymatic activities that mediate the three initial steps in de novo pyrimidine synthesis. The pyrimidine synthesis pathway is an essential, conserved metabolic pathway. mTORC1 phosphorylated CAD S1859 in vivo. Not a direct substrate, but a substrate of S6-kinase → S6K is required for CAD-S1859 phosphorylation. Downstream of TORc1 not phosphorylated by the complex itself. S6K directly phosphorylated CAD-S1859

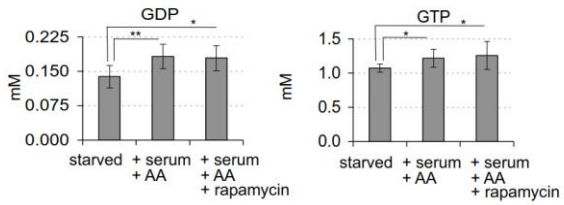
DOES MTORC1 REGULATE DE NOVO PYRIMIDINE SYNTHESIS?

Rapamycin inhibits growth factor stimulated de novo pyrimidine synthesis



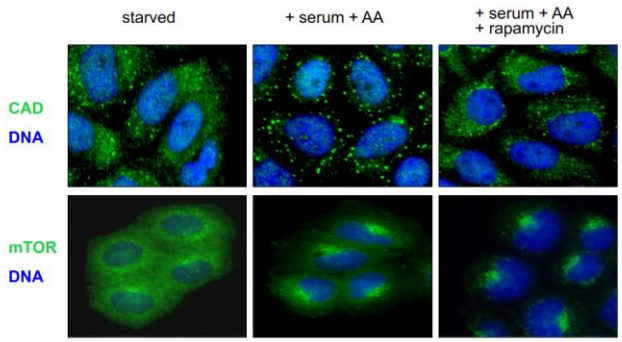
Note: similar results obtained with DHOA, OA and UTP.

Rapamycin does not inhibit growth factor stimulated de novo purine synthesis



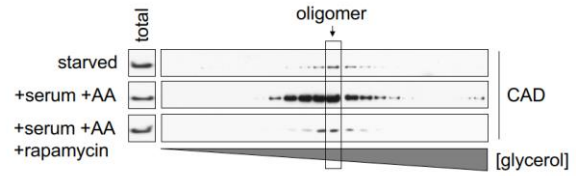
What is the molecular mechanism by which mTORC1 activated de novo pyrimidine synthesis - role of CAD phosphorylation?

mTORC1 promotes CAD 'puncta' formation



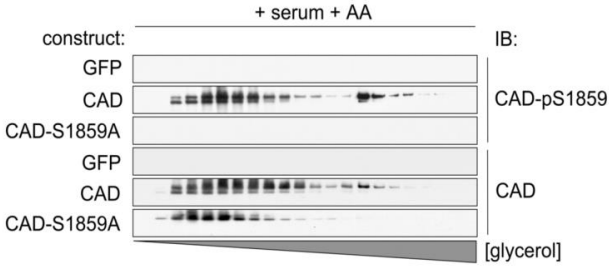
Puncta are cytoplasmic CAD clusters

mTORC1 promotes CAD oligomerization



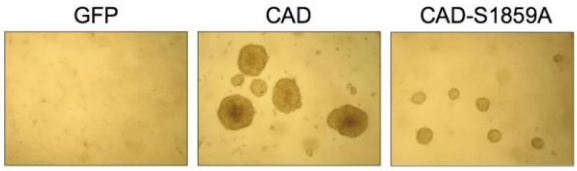
CAD is essential for cell proliferation

CAD-S1859A phosphorylation promotes oligomerization



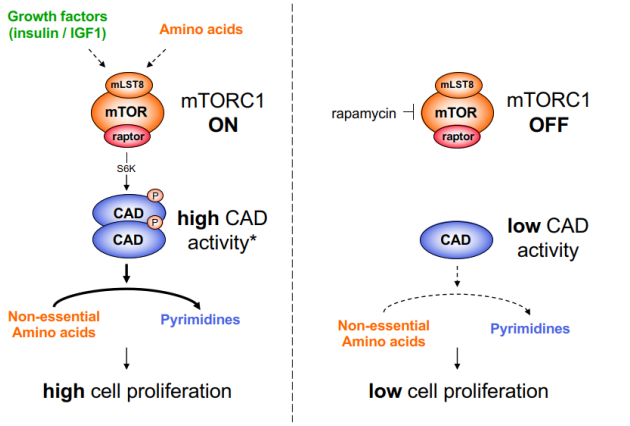
G9c cells reconstituted with wt or mutant CAD

CAD-S1859A is defective for cell proliferation



CAD reconstituted G9c cells Growth in soft agar, 8 days, no uridine

mTORC1 phosphorylation of CAD activated de novo pyrimidine synthesis and cell proliferation



*steric channeling of substrates by oligomer

Robitaille et al., Science,

TRNA BIOLOGY

STRUCTURAL FEATURES

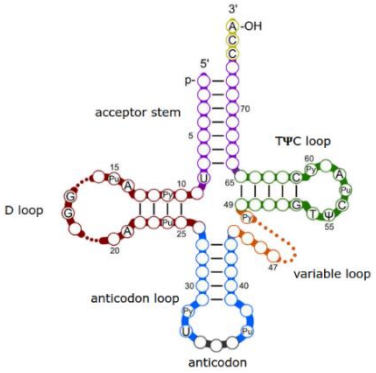
OVERVIEW

tRNA = transfer RNA, physically link the genetic code (codon) to an amino acid

- Proposed by Francis Crick as necessary adapter molecules
- First described as soluble RNA (sRNA): activated with aa (ATP-dependent)
- tRNA are amongst the oldest biomolecules that are present (RNA world)
- first function in evolution likely unrelated to translation
- rather a tag for amplification in the RNA world

SECONDARY STRUCTURE

Secondary structure is a clover leaf



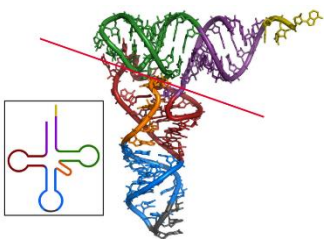
- 5’ terminal phosphate
- Acceptor stem: 7bp long, some non-Watson-Crick base pairs → special, structure is kept flexible
- D arm: 3-4 bp long stem; contains dihydrouridine (D)
- Anticodon arm: 5 bp long stem; contains the anticodon
- Variable arm: 3-21 nt long; may contain a stem of up to 7 bp
- TΨC arm: thymidine, pseudouridine, cytosine
- CCA tail: generally enzymatically added, caries the amino acid
- Some positions consist of identical nucleotides in all tRNAs

TERTIARY STRUCTURE

The tertiary structure is an L-shape

Folding order: D => anticodon => TpsiC => acceptor

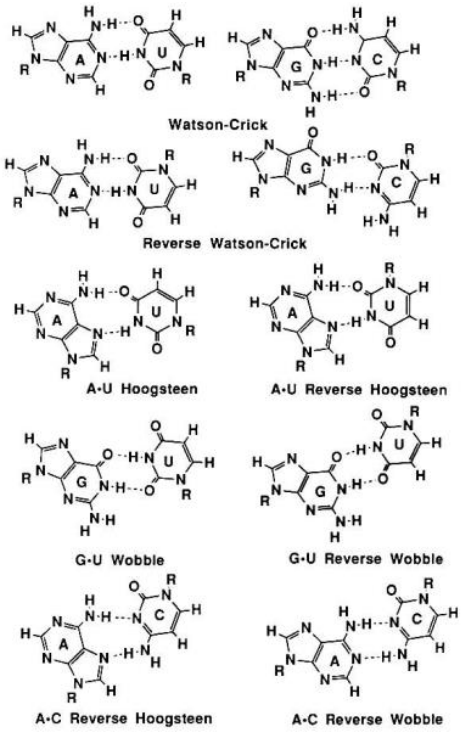
tRNAs consist of two largely independent domains.



TRNA FEATURES

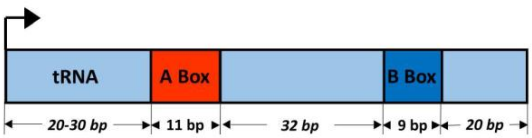
- Many nucleotides are chemically modified (12-15 on average)
- Most nucleotides are engaged in base-pairing interactions
- Few nucleotides are exposed to the solvent
- Tertiary interactions between functional elements = stabilization
- Most 3D interactions are not Watson-Crick base-pairs
- Non-canonical base pairs
  - Most tertiary interactions are not Watson-Crick pairing
- 3D interactions often involve modified bases

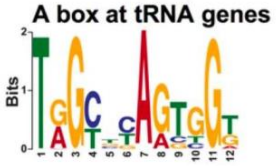
BASE PAIR INTERACTIONS - OVERVIEW

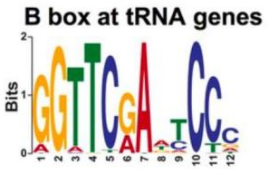


TRANSCRIPTION

- tRNA genes carry their own promoter
- A- and B-boxes facilitate transcription factor binding

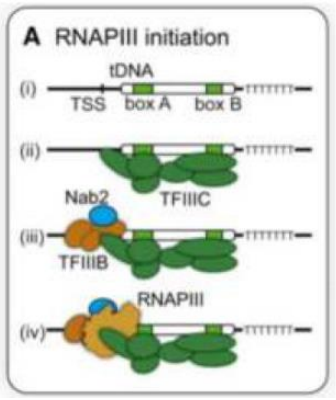






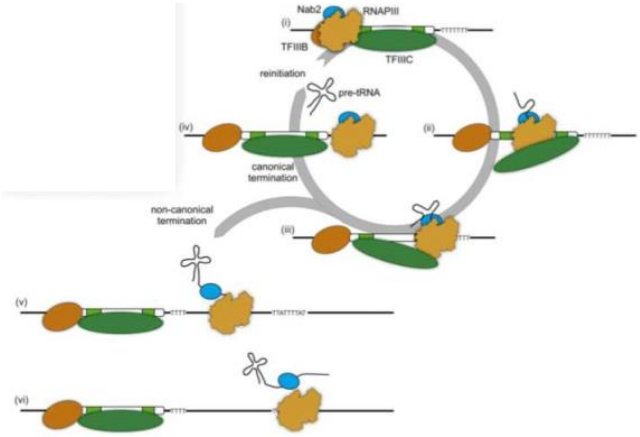
TRNA TRANSCRIPTION - INITIATION

- Binding of TFIIIC (six-subunit complex) binds A and B box, recruits TFIIIB
- Binding of TFIIIB (facilitated by Nab2, nuclear polyA binding protein) → facilitates binding of RNAPIII
- Recruitment of RNAPIII → transcription



TRNA TRANSCRIPTION - ELONGATION AND RECYCLING

- TFIIIC stays associated with RNAPIII (slowdown during Box A or B transit)
- Termination signal: PolyA tract (Nab2 stays bound to RNAPIII, important for efficient tRNA initiation)
- Re-initiation/Recycling following (strong) termination

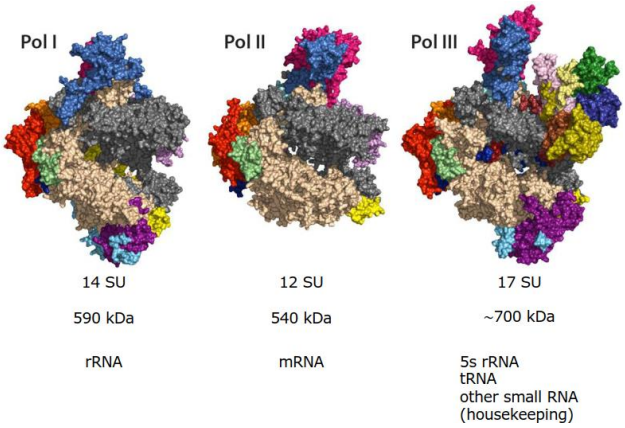


Alternative termination via Nab2: tRNA gets longer and longer, also termination initiation via Nab2, but here Nab2 gets released from Polymerase III)

TRNA TRANSCRIPTION

- PolI: Factors for processivity
- PolIII: factors for efficient initiation and termination (tRNAs are small so need this efficiency! PolII could never, would take to long, which is why PolIII has been optimized for this task!)

How does it doe this? PolIII carries transcription-factor like elements for promoter melting → reduces need for other transcription factors and mediators. Don't need to wait until certain factors bind because functionality is already there with PolIII.

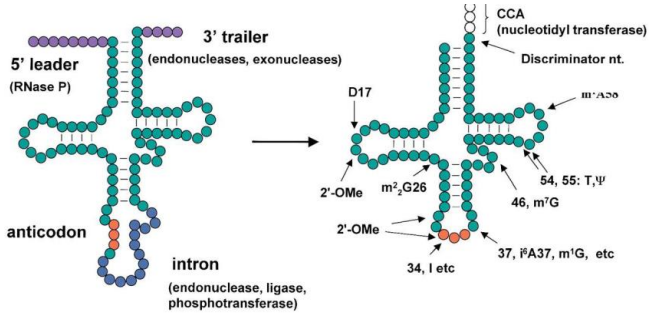


REGULATION

- Maf1 is the critical transcriptional repressor
- Maf1 is phosphorylated by TORC1 via Sch9, PKA and CK2
- Maf1 prevents binding of RNAPIII to TFIIIB (via Brf1)
- Maf1 shuttles between cytoplasm and nucleus
- During stress Maf1 translocates to the nucleus to bind PolIII
- Maf1 integrates many different signals
- Maf1 shuttling is mediated by its phosphorylation levels

In reality, life is obviously a bit more complex. tRNA transcription depends on many factors (chromatin, context, etc.). Not all tRNA genes are expressed all the time.

PROCESSING AND TURNOVER



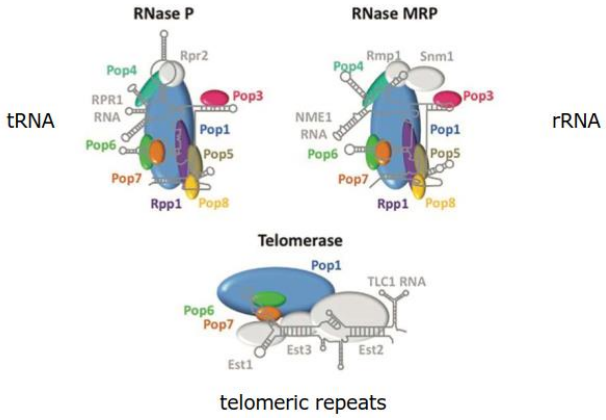
- Pre-tRNA transcripts contain 5'- and 3' leader sequences
- Most pre-tRNA do not contain a CCA tail (exception e.g. E. coli)
- Pre-tRNA can contain introns
- Pre-tRNA contain no chemical modifications

PROCESSING OF 5' AND 3' ENDS

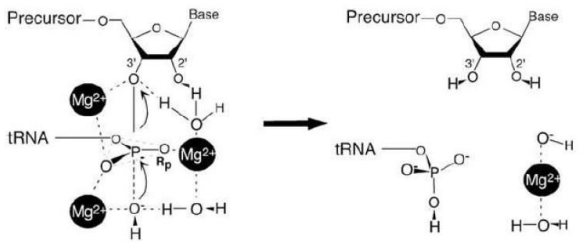
- End processing is critical for tRNA (not so much for many other RNA)
- 5' end processing is very conserved
- 3' end processing differs between species

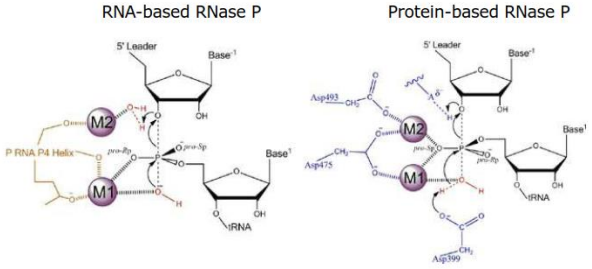
PROCESSING THE 5' END - RNASE P

- Bacteria: M1 RNA and C5 protein → RNA is a ribozyme
- Archaea: 4-5 proteins (all dispensable)
- Eukaryotes: 9-10 proteins, several shared with RNase MRP (rRNA)
- Exception: Human mitochondrial R in tRNANase P: protein, does not use RNA. Enzyme that does not use RNA. Amongst the first ribozymes transcribed
- Few subunits are required in eukaryotes, but not really for the tRNA processing
- Most subunits are dispensable (critical role in other complexes)



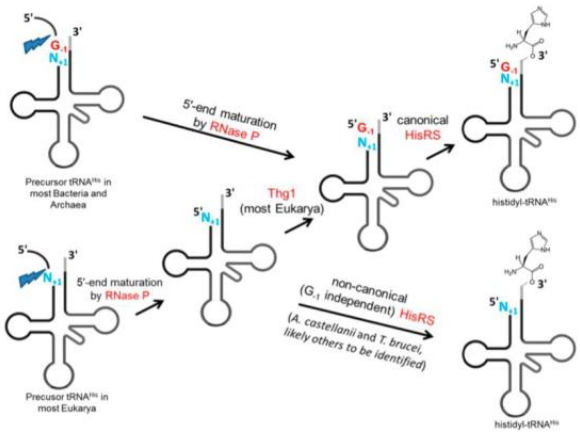
- Mg^{2+} ions are coordinated by nonbridging phosphodiesters
- First Mg^{2+} stabilizes the attacking OH^--nucleophile
- Second Mg^{2+} positions the H_2O as leaving group
- S_N2-reaction (bond broken and formed synchronously)
- End product is a 5'-phosphate on the tRNA





- Aspartate catalyzed the activation of OH^- via Mg^{2+}
- Stabilization of leaving group by an active-site acid (residue unclear)
- Mg^{2+} increase electrophilicity and stabilize the transition state

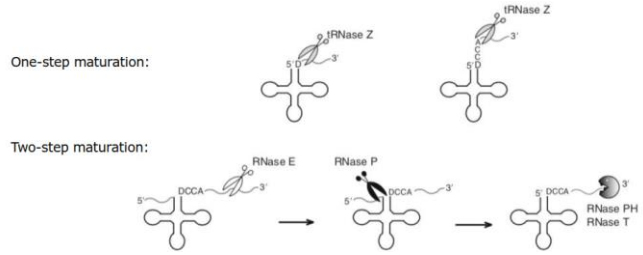
PROCESSING THE 5' END - G-ADDITION IN TRNA^{HIS}



- Enzymatic feature of 5'prime end of tRNA in tRNA his
- Most bacteria and archaea have the G encoded, Rnase P just needs to cleave it, but in most eukaryotes don't have the G and needs to be added
- Either miscleavage or missing G_{-1} in tRNA^{His}
- Aminoacyl-tRNA synthetases cannot charge these tRNA
- Thg1-enzymes in eukaryotes add an additional G in tRNA^{His}
- 3'-5' RNA synthesis! Unusual!

PROCESSING THE 3' END

- CCA sequence is essential for charging amino acids
- Some species encode CCA (e.g. E.coli)
- 3' end trimming in a one-step or two-step reaction
- Differs between species



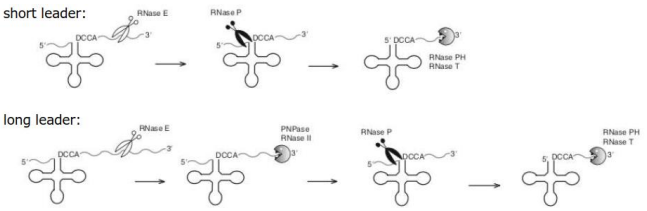
PROCESSING THE 3'END - ONE STEP MATURATION

- One-step maturation is always catalyzed by tRNase Z
- tRNase Z recognizes a discriminator nucleotide (usually cuts 3' of it)
- 5'-leader prevents 3' trimming
- Homologues in all three domains of life
- Short form (280-360aa) in all kingdoms
- Long form (750-930aa) only in eukaryotes

PROCESSING THE 3' END - TWO STEP MATURATION

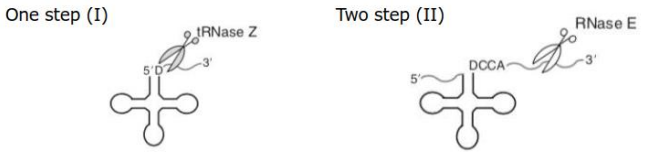
- 1. Endonucleolytic cleavage by RNase E (downstream of CCA) (3'prime)
- 2. Endonucleolytic cleavage by RNase P (5'prime)

- 3. Trimming by exonuclease(of 3'prime)
- Four exonuclease are possible: RNase II, D, T & PH
- RNase T & PH are most efficient



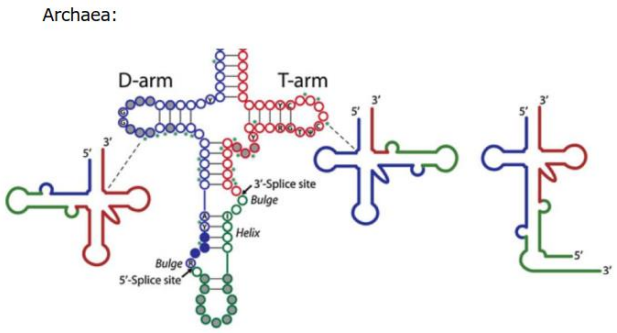
- Short leaders are removed in a three-step process
- Long leaders in E.coli can require four steps!
- Processing of the final nucleotides requires 5' processing (base pairing)
- Long leader pathway also used for cistrons

RELATIONSHIP OF 3' MATURATION PATHWAYS



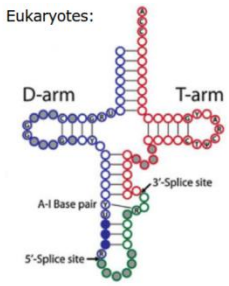
- Multicistronic transcripts: either (I) directly of further downstream (II)
- If 3'-trailer is long and structured → (I) is more likely
- CCA transcripts → (II) is more likely
- Some organisms use pathway (I) and pathway (II) as backup
- If only one pathway exists, it becomes essential

TRNA SPLICING - ARCHAEA



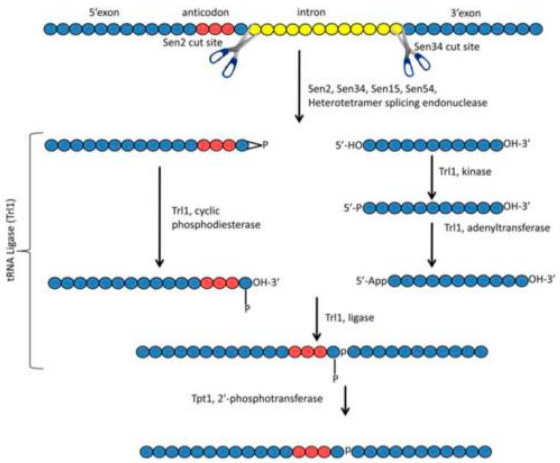
- Up to 70% of tRNAs contain introns
- Bulge-helix-bulge recognition motif - position is variable
- Split tRNA exist
- Introns : 16-44 nt long

TRNA SPLICING - EUKARYOTES



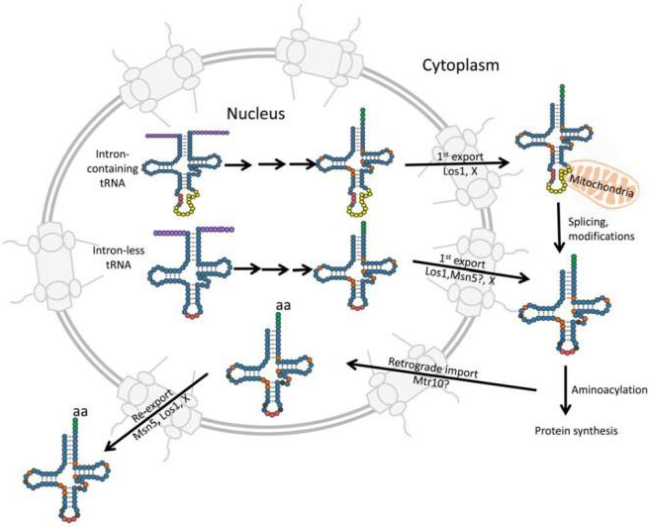
- in eukaryotes: 6% (humans) - 20% (yeast) contain introns
- Don't really know why tRNAs have introns. Have been knocked out an replaced and cells didn't really show a phenotype
- No clear motif, but position of intron is highly conserved
- Introns: 12-102 nt long (can be quite different!)

TRNA SPLICING - YEAST

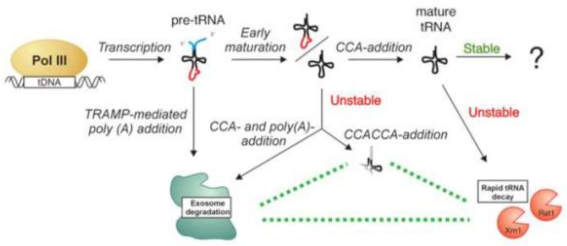


The yeast process is well understood

TRNA PROCESSING - CELLULAR LOCATION (YEAST)



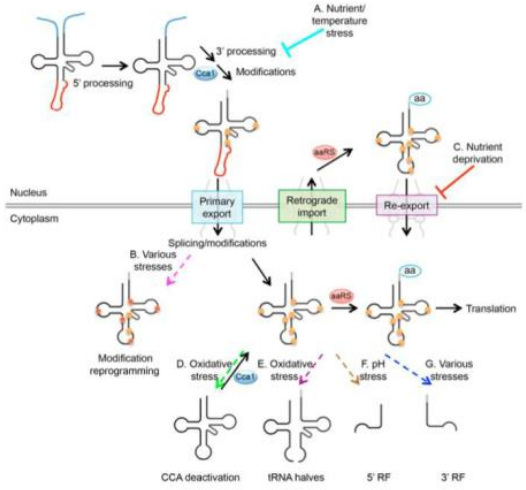
DEGRADATION OF NON-FUNCTIONAL TRNA



- Degradation by Exosome (nucl.) or Rapid tRNA Decay Pathway (cyto.)
- tRNA are marked by polyA or CCACCA tails
- Degradation of mature tRNA is unclear (high $t_{1/2}$)

TRNA TURNOVER

- PolIII transcription and processing need to be coordinated
- Maf1 (PolIII) repressor coordinates the balance
- Early processing and Exosome degradation (3'-5') compete with each other
- Exosome interacts with PolIII
- Hypomodified tRNA 5'-3' degradation Xrn1 (cytosol) & Rat1 (nucleus) - RTD
- RTD is blocked by PolIII inhibition and decreased tRNA levels
- Immature tRNA block translation via Gcn4 (Gcn2 independent)
- tRNA surveillance is used in regulation of tRNA fate:

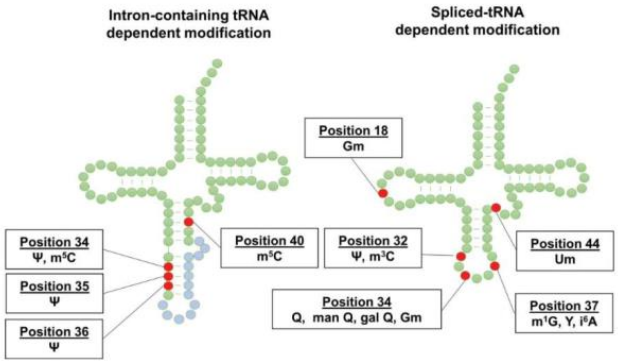


CHEMICAL MODIFICATIONS

WHY HAS MODIFIED RNA TAKEN CENTER STAGE?

- ➔ Chemical modifications ensure stability and efficiency of mRNA vaccines
- ➔ Chemical modifications determine self vs. non-self recognition
- ➔ Today we know 161 different RNA modifications
- ➔ Virtually every cellular RNA is modified (including mRNA)

TRNA SPLICING AND CHEMICAL MODIFICATIONS



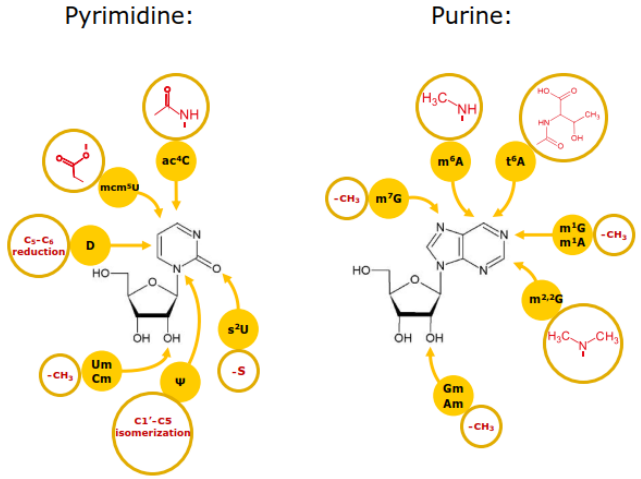
- Splicing provides a checkpoint for other processing steps
- tRNA modifications are linked to splicing

CHEMICAL MODIFICATIONS OF TRNA

- tRNAs molecules carry on average 12-15 chemical modifications

Why do we need chemically modified RNA? → the four canonical nucleosides are chemically limiting. They are all quite similar, especially compared to amino acids. This si why modifications exist.

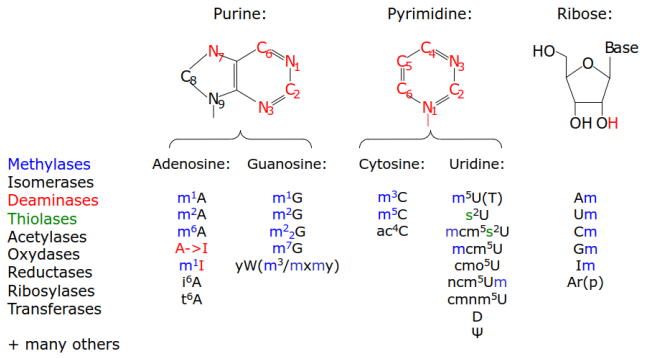
WHAT TYPES OF MODIFICATIONS EXIST?



- most are just methyl groups but there are also very complex ones
- RNA modifications increase the chemical space of RNA

Are tRNA modifications conserved? tRNA modifications are found in all domains of life.

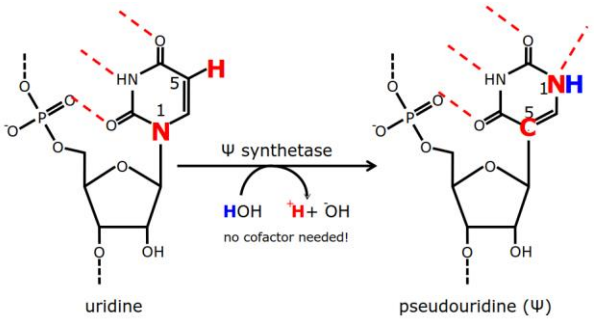
WHICH TYPE OF ENZYMES INTRODUCE MODIFICATIONS?



- Between 1-10% of genes encode for RNA modifications
- First nucleoside discovered is pseudouridine. Now one of the two nucleosides in the vaccines. Is the most abundant one!

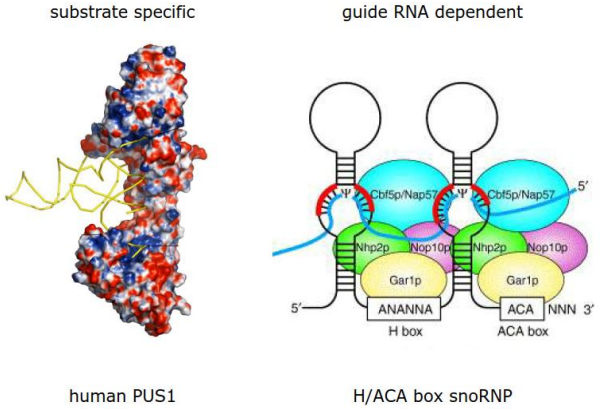
PSEUDOURIDINE

How is Pseudouridine generated?



- Pseudouridine is formed through an isomerization reaction
- Pseudouridine can form one additional H-bond

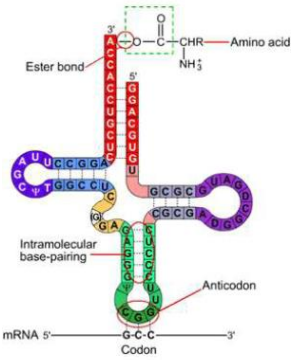
HOW IS SPECIFICITY DETERMINED?

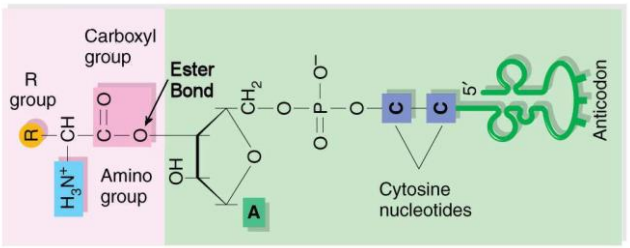


- Enzymes recognize specific substrates or use a guide RNA

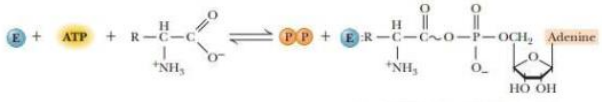
AMINOACYL-TRNA SYNTHETASES

- tRNA carry the amino-acid that matches the codon
- Aminoacyl-tRNA synthetases perform the reaction
- The link is a carboxylate ester between tRNA and amino acid
- Amino acids can be transferred to the 2'OH or 3'OH of the ribose

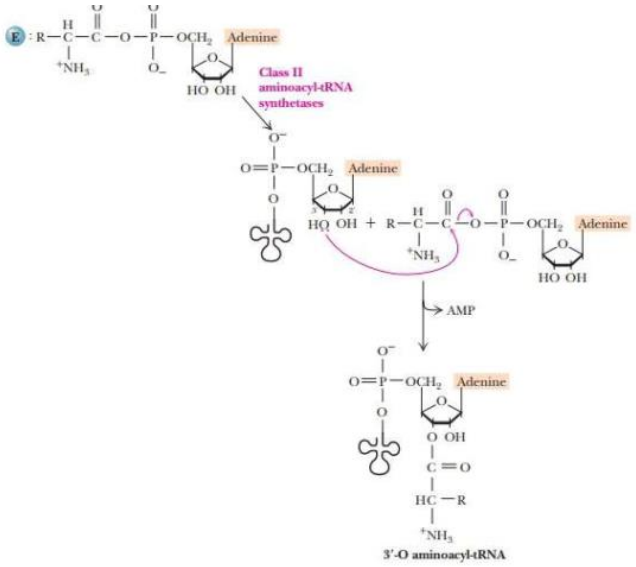




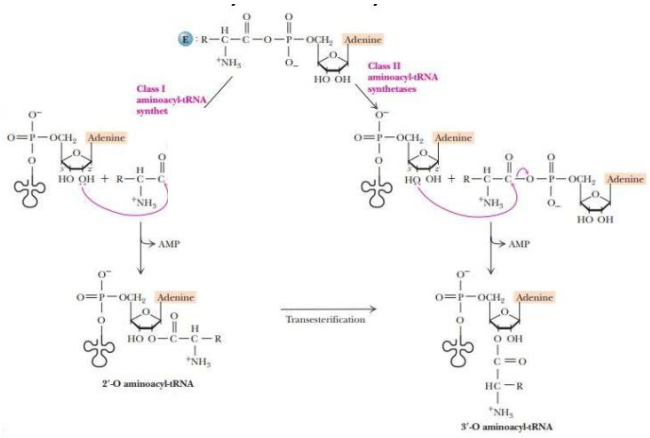
- Amino acid is first adenylated to aminoacyl-AMP



- Second step forms aminoacyl-tRNA



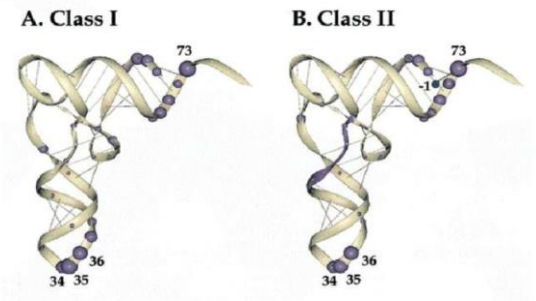
- Class I or Class II enzymes attach to 2' or 3' OH of the ribose
- Transesterification moves the amino acid to 3'OH



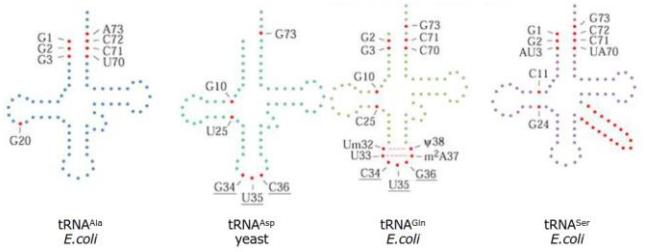
ACCURAY OF AMINOACYLATION

- Synthetases need to link the correct amino acid to right anticodon
- Error rate is 1:10'000 (expected is 1:1000)
- Double-sieve strategy: two checkpoints: Aminoacylation (only specific amino acids that fit in there will actually be coupled to the tRNA, but may not always be the right ones) and editing site (checks if tRNA is charged with the right AA, if not → hydrolyses) → decreases error rate
- Editing site hydrolyses wrong amino acids efficiently
- But sometimes mis-acylation may be beneficial: sometimes cell senses severe stress, can be beneficial, almost like mutation you introduce without altering genome → if cell has nothing to loose may have advantages

IDENTITY ELEMENTS OF TRNA

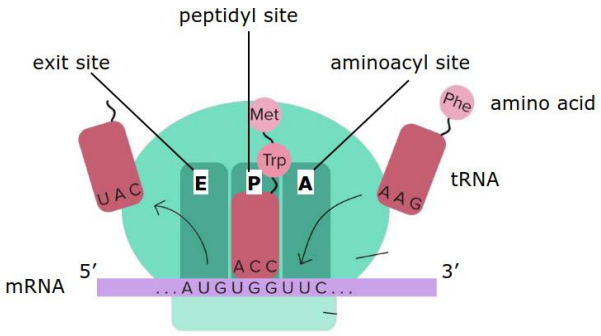


- Identity elements differ between class I and class II
- Mostly anticodon and 5' and 3' ends of tRNA

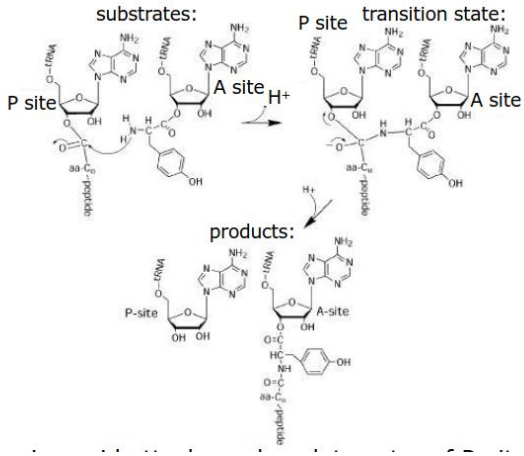


- Identity elements differ more between tRNA than between classes
- Interactions between tRNA and aaRS differ massively

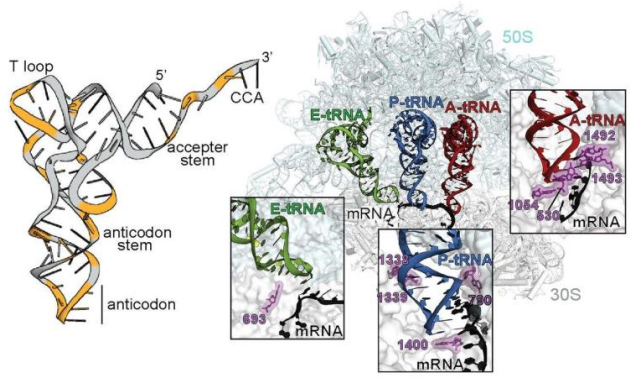
TRNA IN DECODING



- A site: Binding of charged tRNA → reading
- P site: Nascent chain is attached to the P-site tRNA
- E site: Release site of empty tRNA



- A-site amino acid attacks carboxylate ester of P-site chain
- Formation of a transition state
- Hydrolysis of P-site tRNA → Transfer of peptide to A-site tRNAs

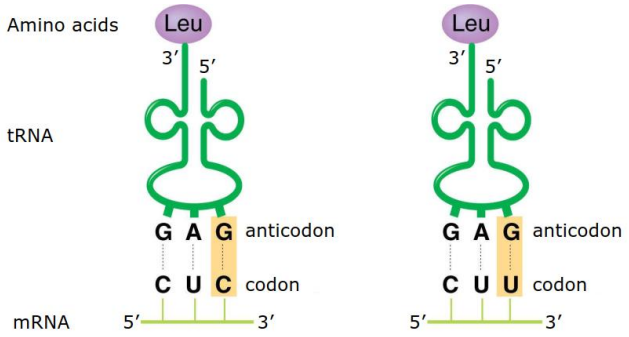


- Formation of mini-helix between anticodon and codon in A- and P-site
- Transit through ribosome leads to tRNA bending
- Release at E-site (mRNA kink)

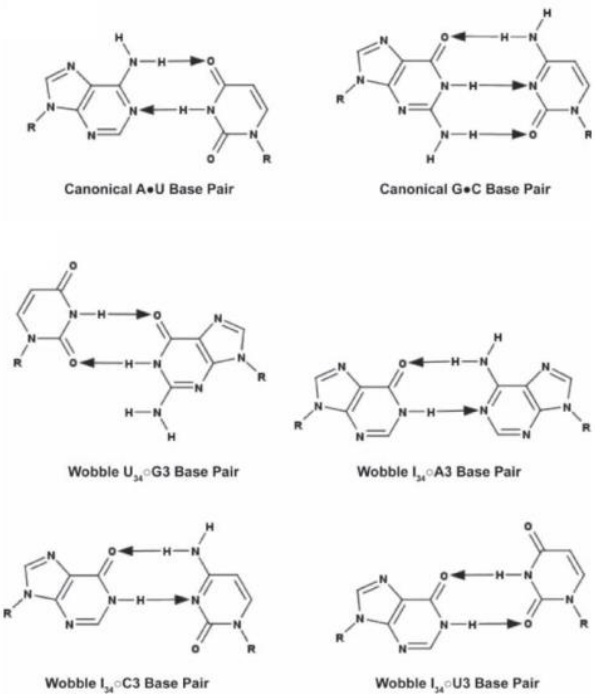
TRNA - MRNA INTERACTION

- Anticodon and codon form an antiparallel minihelix

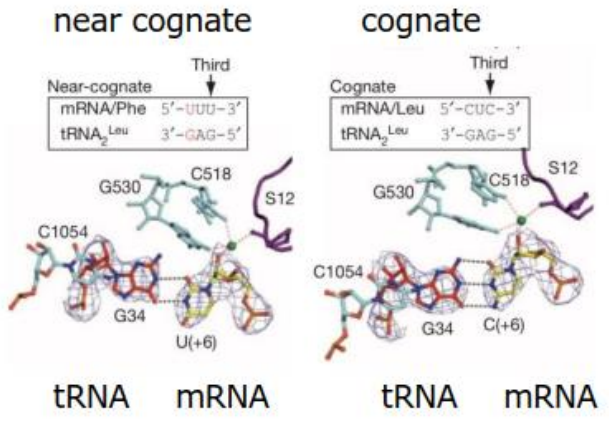
- Codon-anticodon interaction can be inaccurate → wobble



TYPICALLY OCCURRING WOBBLE PAIRS:



WOBBLE DECODING

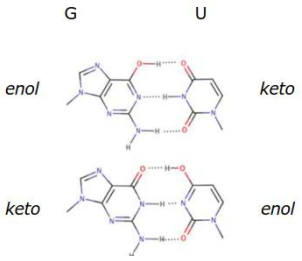


- Mispairing in the 1st and 2nd position of codon-anticodon disrupt binding
- 3rd position forms base-pairing interaction

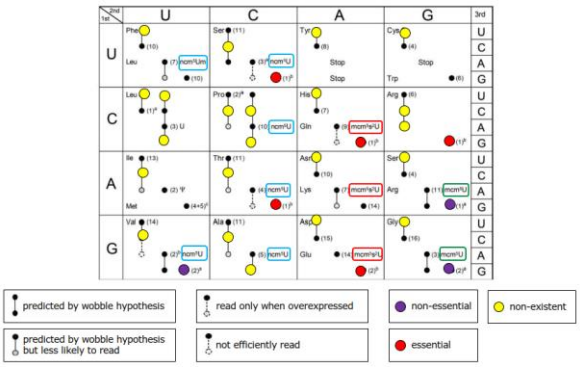
- Wobble pair adopts Watson-Crick geometry!

WOBBLE DECODING - TAUTOMERISM

- Many nucleotides occur in their keto or enol form
- Enol-keto tautomerism affects codon-anticodon interaction
- Enol- and keto forms have been observed in crystal structures

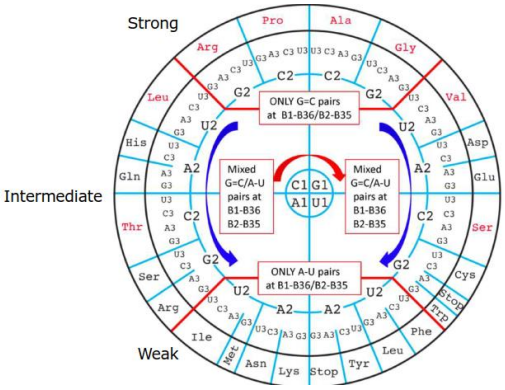


WOBBLE INTERACTIONS - YEAST EXAMPLE



- Not all tRNA exist → wobble decoding is essential

AN ENERGY-BASED CODON REPRESENTATION



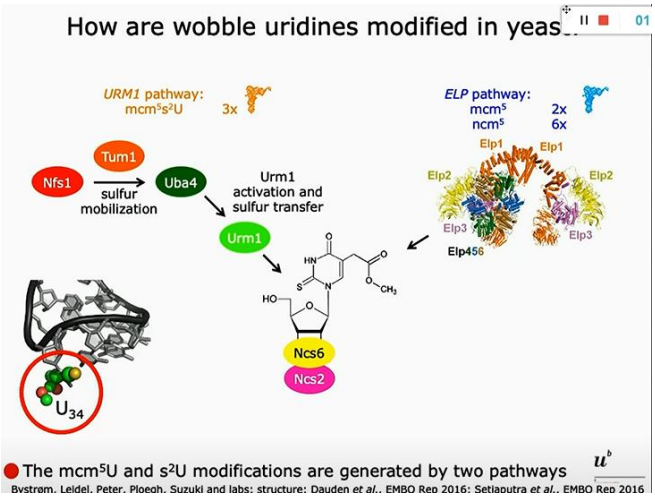
- Codons can be sorted based on the strength of codon-anticodon interaction
- Strong codons → strong interactions: Wobble pairing does not matter
- Weak codons → weak interactions: modifications can tune interactions
- Modifications are more frequent/effective in weak codons
- Modifications allowed the introduction of new amino acids
- Interactions strength limits synthetic biology approaches

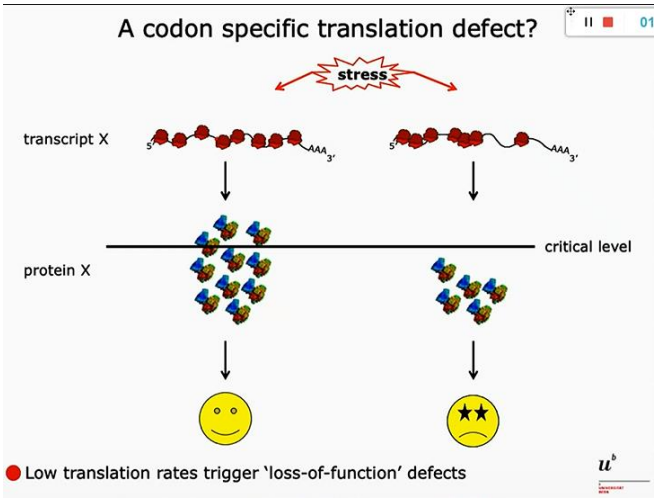
WHAT DOES TRANSLATION NEED TO ACHIEVE?

Ribosome needs to find the tRNA where the anticodon matches the codon in the A site. Synthesis of protein and reading of the mRNA are not two independent processes as soon as translation happens the nascent polypeptide already forms outside of the ribosome. Translation dynamics and protein folding are tightly coordinates. This happen within a complex tRNA pool. Very difficult to quantify → tRNAs are the dark matter of translations. People are fighting over what role tRNAs have in translation. Big open questions.

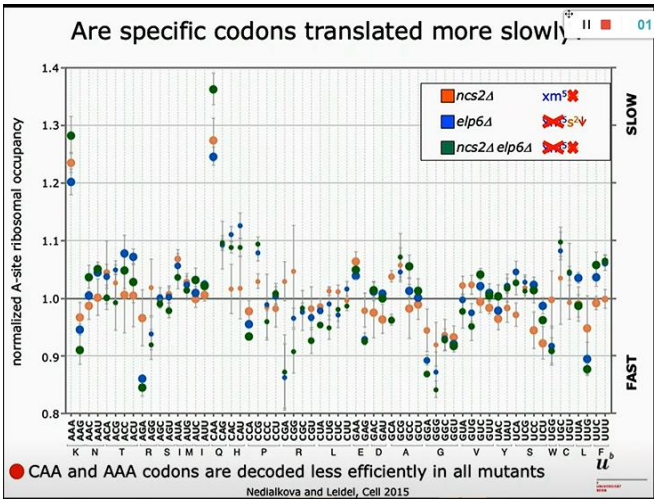
WOBBLE URIDINE IN THE LIMELIGHT:

Modification of Wobble uridines are (almost) invariably modified (in eukaryotes). Barely uridine in that position that is not modified.





How to characterize such translation defects? Find out what happens in cells that don't translate well. Ribosome profiling allows for a quantitative analysis of translation → know which ribosome have been translated, which codon have been read. RNA seq at the same time in parallel → how often do we find a specific codon in the A site in the mutant and compare it to wildtype. On the line: as frequent as in WT, further up: found more often than in WT, interpretation: ribosome is slower in decoding than the WT.



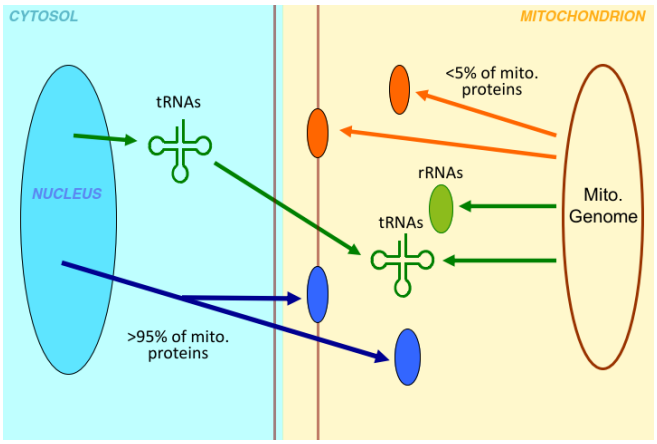
CAA and AAA codons are decoded less efficiently in all mutants! → exciting → two out of three codons that are ncm5 modified.

Why do the mutants die? Slow down but why do they die? Look at mRNA and translation levels. Gene expression of mutants is dramatically altered. Gene expression points to protein homeostasis defects. Proteins aggregate in the cells (hypothesis).

MITOCHONDRIAL TRNA IMPORT

- Mitochondrial tRNA import is widespread (essentially all plants, many protozoa, a lot of fungi, few invertebrates)
- No tRNA import in vertebrates
  - No convincing reports of tRNA import in mitochondria in vertebrates
  - Including humans
- Number of imported tRNAs are variable
- Applied aspect: lot of diseases → mitochondrial diseases caused by mutations in mitochondrial DNA. 40 % affect mitochondrial tRNAs, non-functional or not sufficiently functional tRNAs → could be important in treating diseases!

ORIGIN OF MITOCHONDRIAL LOCALIZED GENE PRODUCTS



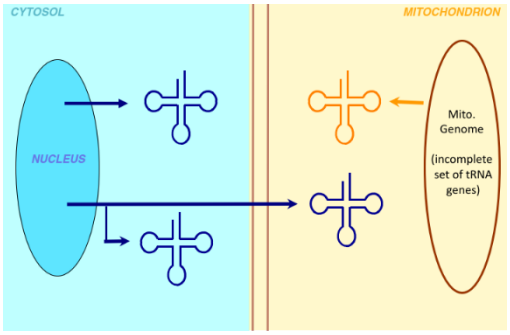
All mitochondria have a genome, but only a small number of proteins are encoded on these (essential for mitochondrial function → mitochondria need to have their own translation system!!) → all components of respiratory chain. This means the large majority of mitochondrial proteins actually comes from outside the mitochondrion.

tRNAs for mitochondrial translation in human mitochondria are all encoded in the mitochondrial genome. In plants, protozoa and fungi is different, some tRNAs are imported because set encoded of mitochondrial genome is incomplete → mitochondrial tRNA import!

tRNA has not the same flexibility as proteins when imported through a membrane → unusual process!

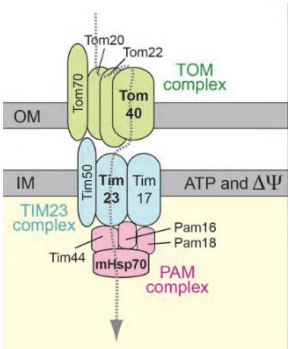
TRNAS IN EUKARYOTES

Divide tRNAs in different classes: old textbooks → nuclear encoded tRNAs (code for tRNAs in cytosolic translation) and mitochondrial encoded tRNAs (work in mitochondrial translation). In plants maybe also Chloroplast-encoded. Organisms which import tRNA → third class → small fraction of cytosolic tRNAs imported in mitochondria → dualy localized tRNA, most part works in cytosolic, small fraction works in mitochondria. Not found so far a nuclear-encoded tRNA which a 100% is imported in mitochondria → probs doesn't exist. Cytosolic → eukaryotic type, mitochondrial encoded → bacterial type

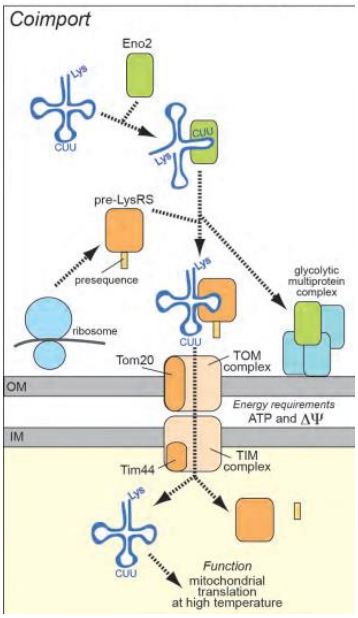


MITOCHONDRIAL PROTEIN IMPORT

- N-terminal presequences provide targeting signals
- Protein is unfolded during import
- Import across the IM requires the membrane potential and ATP
- mHsp70 functions as import motor
- Tom40, Tom20, Tom22, Tom70
- Tom20 and Tom70 are protein import receptors -> factors at outer membrane that recognize sequences



MITOCHONDRIAL TRNA IMPORT IN S. CEREVISIAE



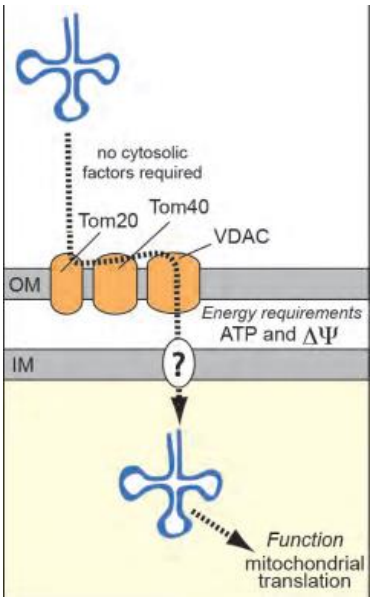
Only a single tRNA is imported → lysine tRNA. Function of the tRNA inside mitochondria is not well understood. Corresponding mitochondrial lysine tRNA is present, so maybe redundant?

Lysine tRNA binds to Enolase 2, induces conformational change in tRNA which is then handed over to other protein: precursor of the mitochondrial lysine tRNA synthetase (imported into mitochondria). Unusual: mitochondrial lysine tRNA synthetase → but binds to cytosolic lysine tRNA! Synthetase would not be able to charge the llysine tRNA inside mitochondria, because it would charge the mitochondrial lysine tRNA.

Imported via protein import system: TOM and TIM complex. Direct involvement of Tom20 and Tim44. Model implies that protein takes tRNA along through the pore.

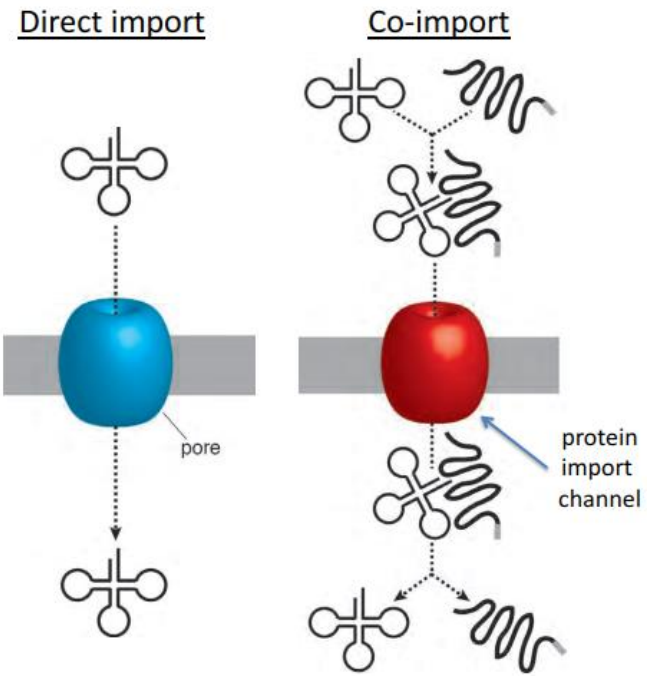
Problem with the model: import of protein required unfolding of the transported protein, so difficult to imagine how interaction with tRNA can be maintained when the protein is unfolded during the import! Also not clear why this tRNA is imported because there is mitochondrial lysine tRNA → maybe needed at high temperature.

MITOCHODNRIAL TRNA IMPORT IN PLANTS

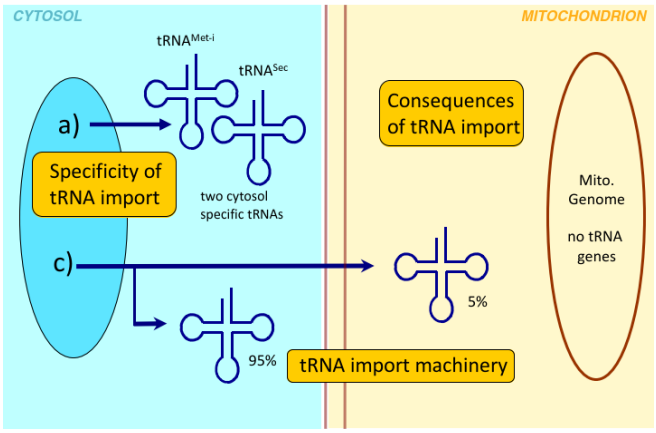


Direct import. Much less is known. Has been shown that Tom20 and Tom40 is required. But doesn't really cross membrane through Tom40 pore or other components of the protein import system. Suggested: pore is VDAC, which is a beta barrel protein, transport ATP to the intermembrane space of cytosol. No cytosolic factors required for tRNA import. In Inner membrane process is unknown but is known it needs membrane potential and ATP. More than one tRNA is imported and lack the corresponding mitochondrial tRNA.

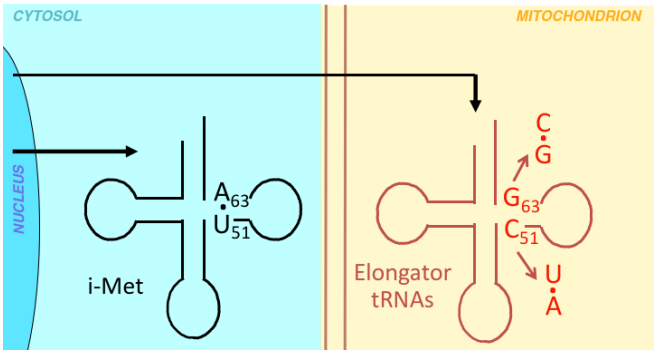
TWO MODELS FOR MITOCHONDRIAL TRNA IMPORT



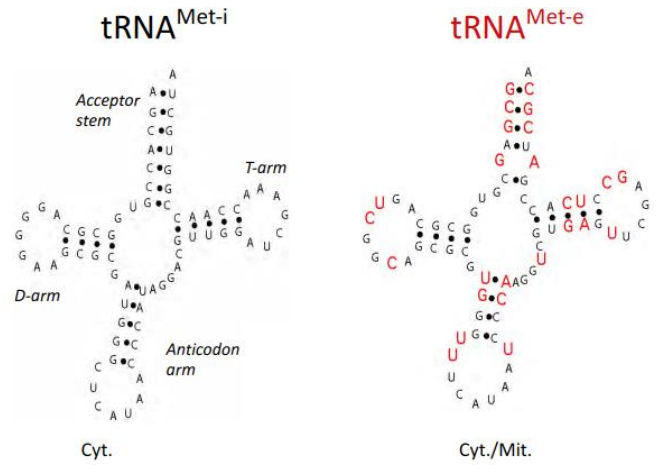
TRNAS IN T. BRUCEI



No tRNA genes are encoded on the mitochondrial genome! All are imported! Need very efficient import system. Nucleotide pair 63:51 is the main localization determinant in T. brucei

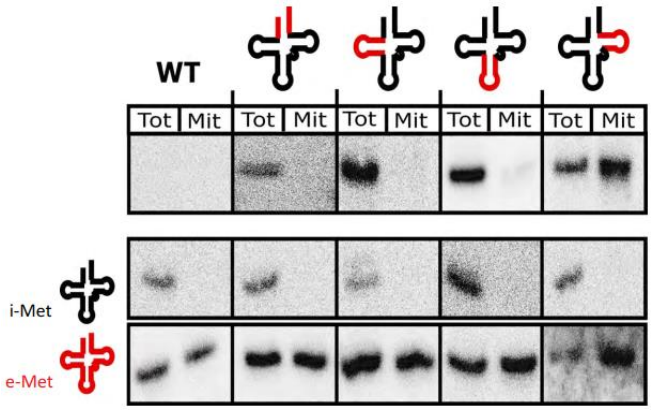


TRYPANOSOMAL TRNA^{MET}

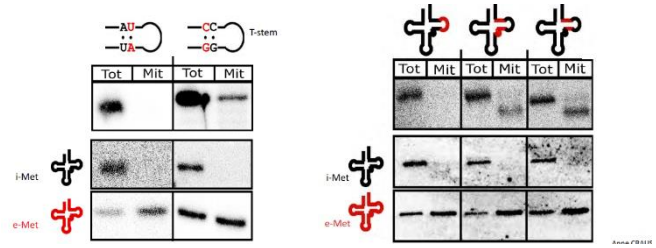


Initiator Met-I and elongator Met-e. Left is not imported into mitochondria, but the right one is. But they are very similar, only differences are the nucleotides indicated in red. Whatever is responsible for the fact that the right is imported must reside there, because it is the only difference! Which exact nucleotide is actually responsible for the import?

IN VIVO IMPORT: VARIANT TRNASMET-I

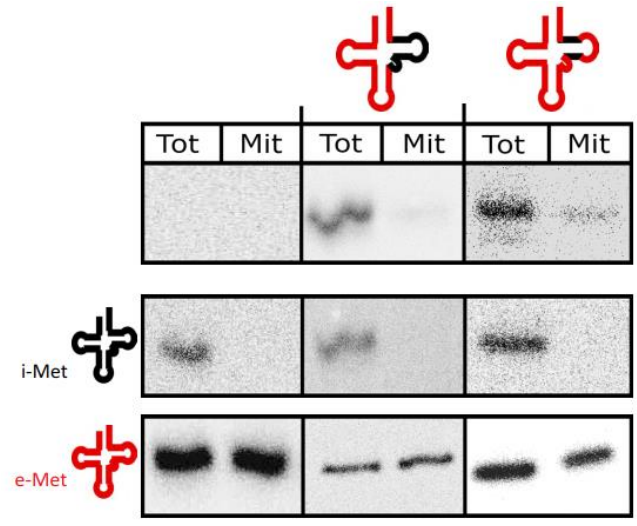


Expressed variants of these tRNAs ins trypanosomes. Analyzed localization of tRNAs by looking at total and mitochondrial fractions. Total is essentially the same as cytosolic, because most tRNAs are cytosolic. Tried to induce import of i-Met by transplanting regions from the elongated e-Met, successful in the right part. More detailed: just the stem is needed to induce import!

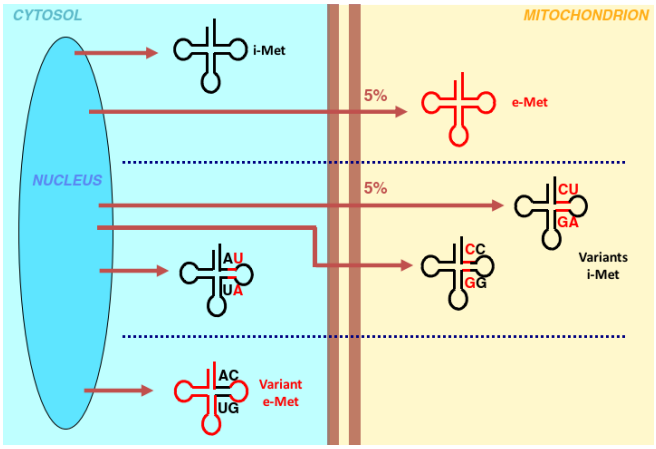


Could narrow it down to a single nucleotide-pair!

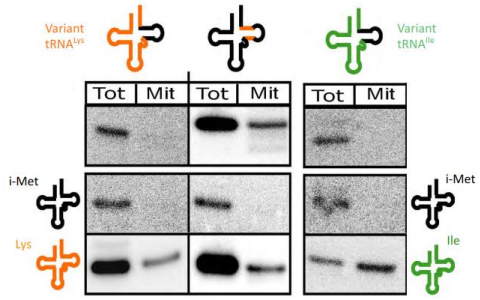
IN VIVO IMPORT: VARIANT TRNASMET-E: What do we need to transplant to prevent import?



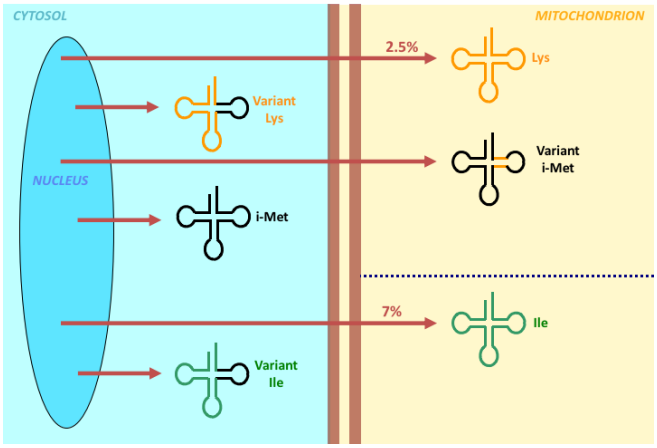
SUMMARY: LOCALIZATION OF TRNAMET VARIANTS



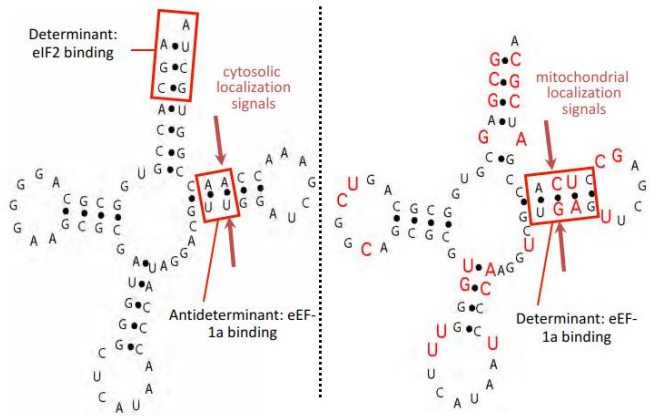
IN VIVO IMPORT: VARIANT TRNA^{ILE} AND TRNA^{LYS}



SUMMARY: LOCALIZATION OF TRNA VARIANTS

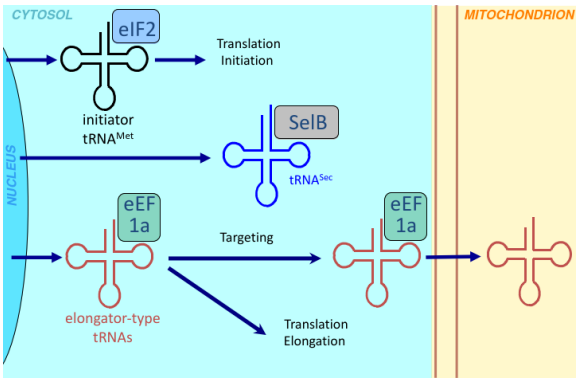


MET-I VS. MET-E



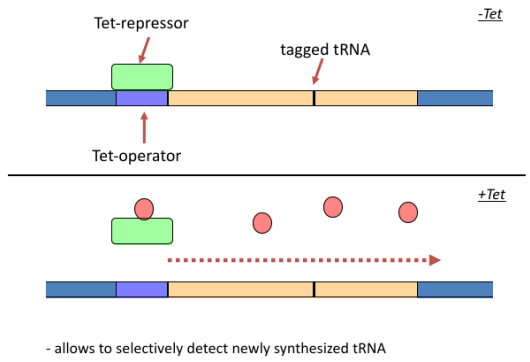
CORRELATION:

- mitochondrial tRNA import with binding to eEF-1a
- cytosolic localization with absence of eEF-1a binding
- is eEF-1a required for the targeting step?

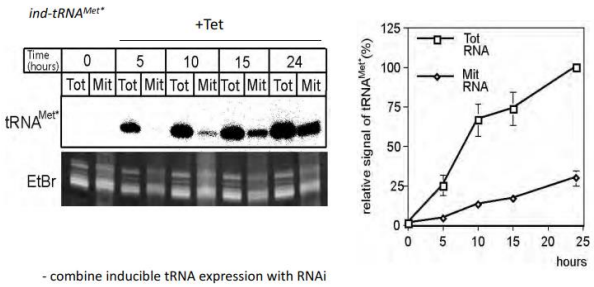


- tRNA import machinery
  - involvement of elongation factor 1a
  - inducible tRNA expression system

TET-INDUCIBLE TRNA EXPRESSION SYSTEM

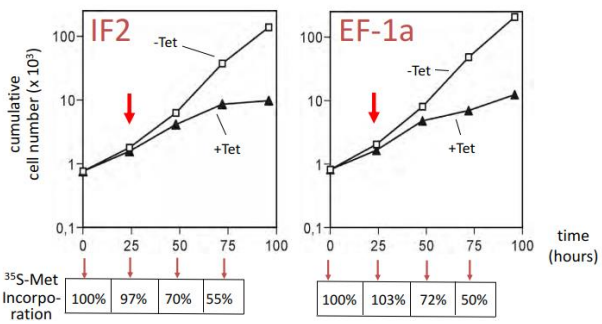


- allows to selectively detect newly synthesized tRNA



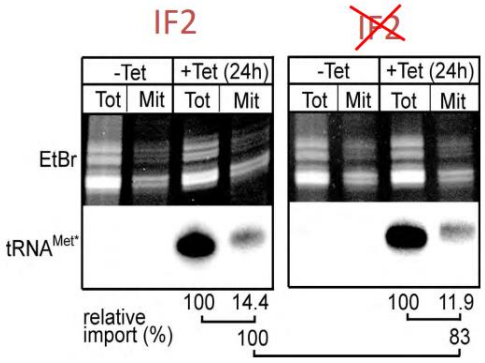
- combine inducible tRNA expression with RNAi

Tet-inducible tRNA expression combined with RNAi: effects on growth and translation.



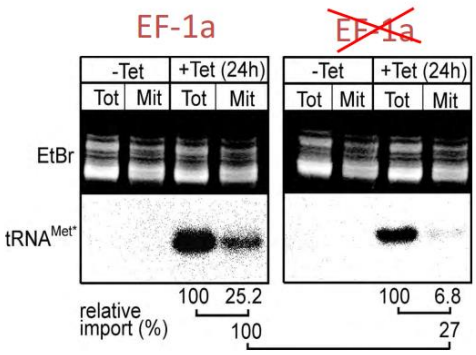
Two cell-lines which upon addition of tetracycline express this tRNA → inducible tRNA expression, a tRNA that is imported. Show the expected effect in protein translation.

EFFECT OF IF2-RNAI ON TRNA IMPORT:



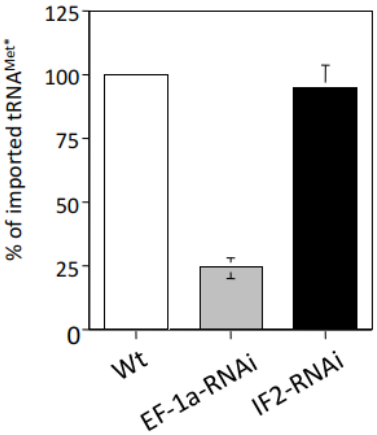
Normal IF2: addition of tetracycline only leads to expression of tRNA. Only expressed and imported in the presence of tetracycline. Initiation factor 2 is depleted → expression and import still works, difference is not significant.

EFFECT OF EF-1A-RNAI ON TRNA IMPORT:

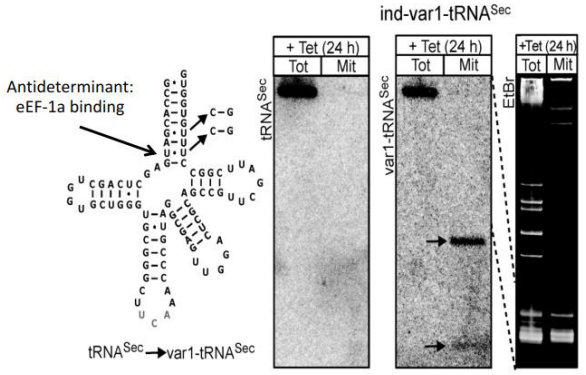


Here we see a significant difference in tRNA import!

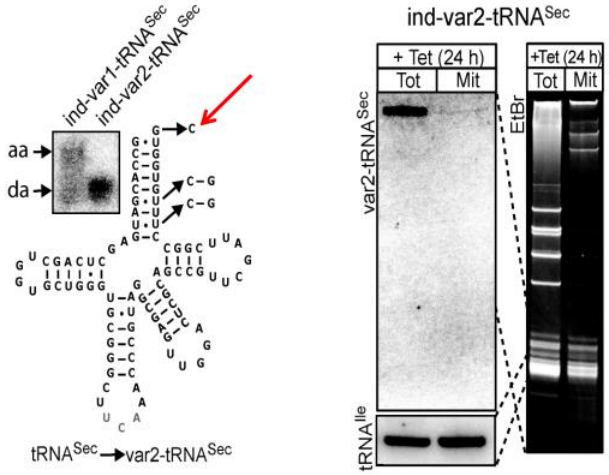
EFFECT OF IF2/EF-1A-RNAI ON TRNA IMPORT:



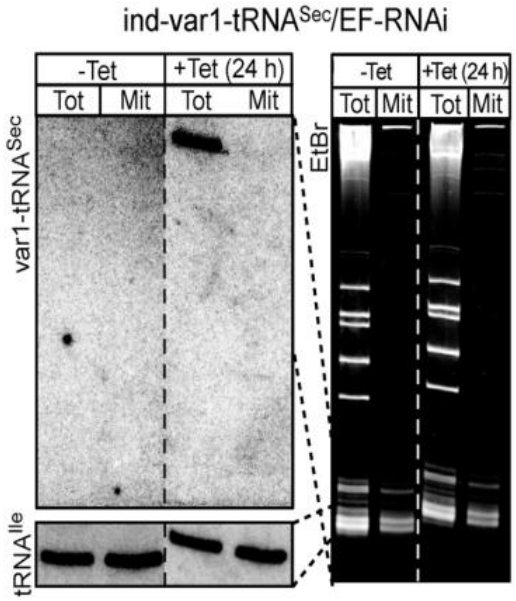
TRNA IMPORT MACHINERY - SELENOCYSTEINE TRNA



Variant tRNA^{Sec} deficient for aminoacylation is not imported:



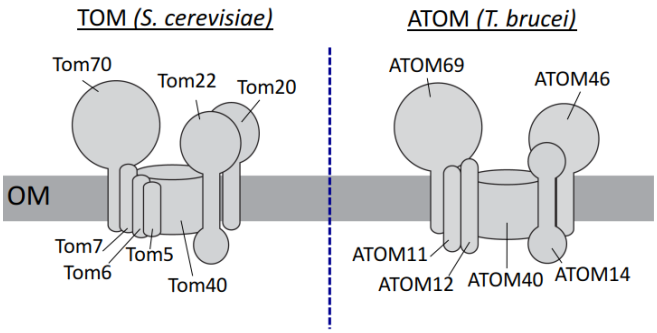
Ablation of EF1a prevents import of the variant tRNA^{Sec}



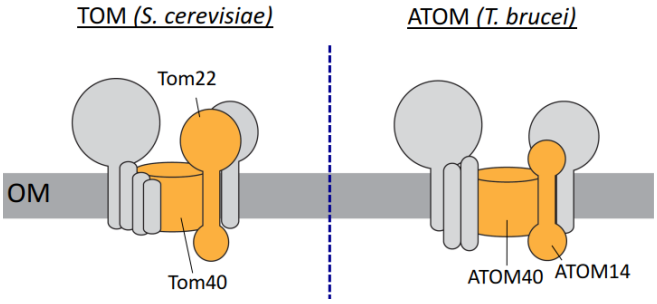
Is eEF-1a required for the targeting step? → model says yes, requirement for tRNA to end up in mitochondria, but must be involved in a targeting step, because protein itself is not imported into mitochondria, rather works in bringing tRNA in vicinity of mitochondria.

IS THERE A CONNECTION BETWEEN MITOCHONDRIAL PROTEIN IMPORT AND TRNA IMPORT?

MITOCHONDRIAL OUTER MEMBRANE PROTEIN TRANSLOCASES :



Only two translocase subunits are conserved between yeast and trypanosomes:

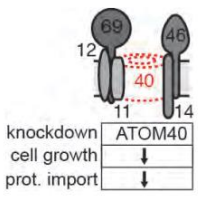


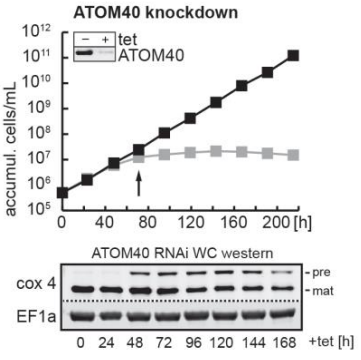
Both translocases contain two protein import receptors. Tom70, Tom20 and ATOM69, ATOM46. Sequence wise not identical at all, so they evolved independently. ATOM = atypical TOM.

EXPERIMENTAL APPROACH:

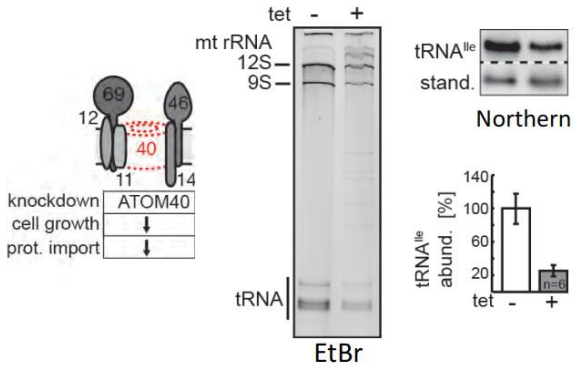
- use inducible RNAi cell lines ablated for subunits of the OM protein translocase (ATOM subunits)
- Determine mitochondrial tRNA level at the time of the growth arrest
- Compare with mitochondrial protein import phenotype

EXAMPLE: ATMO40

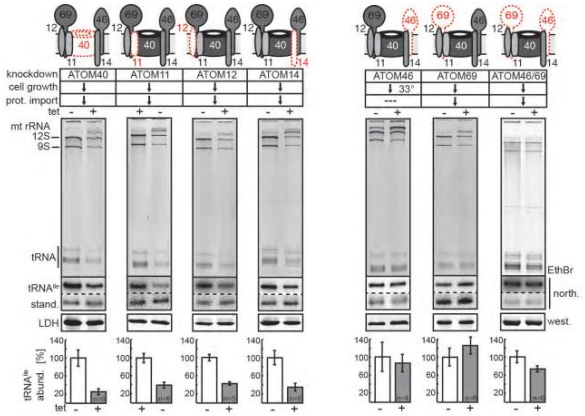




ATOM 40 builds channel.



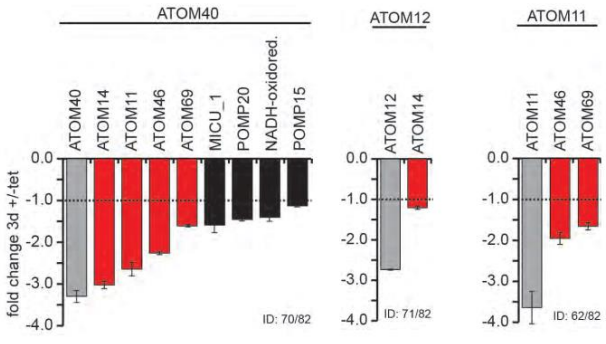
ATOM core subunits are required for mitochondrial tRNA import.



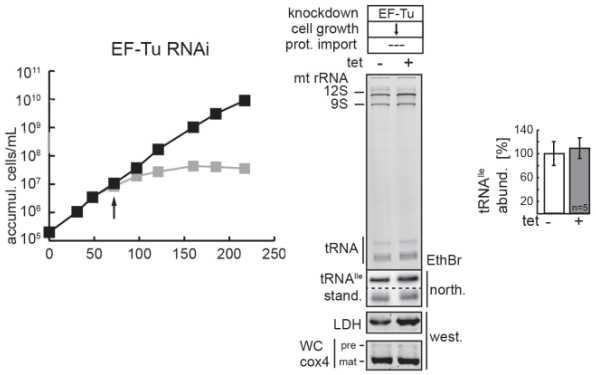
INDIRECT EFFECTS?

Does ablation of protein import prevent import of an unknown tRNA import factor in the outer membrane? Does ablation of protein import lead to degradation of mitochondrial tRNAs because it inhibits mitochondrial translation?

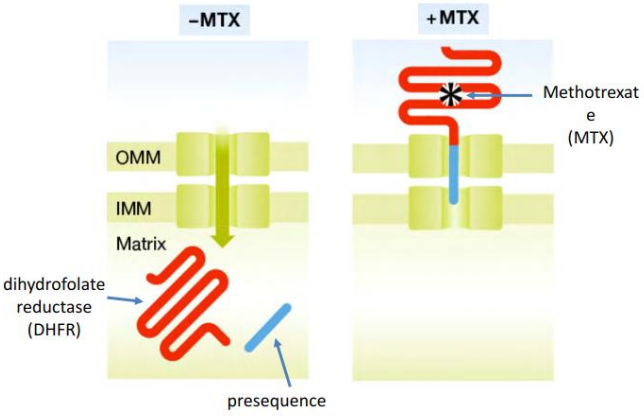
OM proteins reduced in abundance in ATOM subunit-lacking cells:

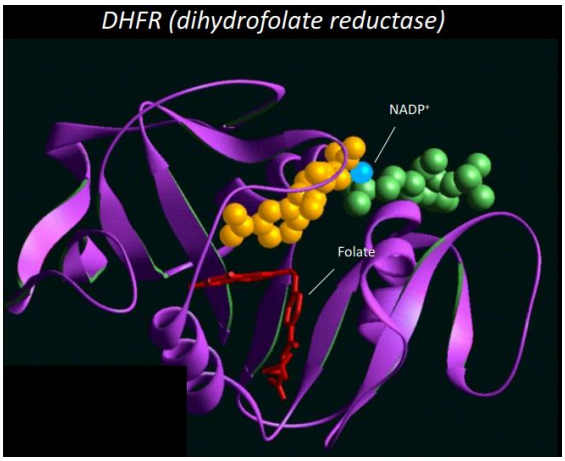


Lack of mitochondrial translation does not induce degradation of mitochondrial tRNAs

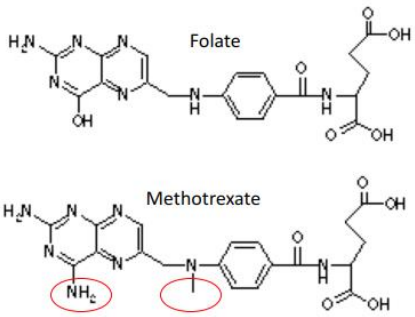


DHFR and membrane translocation



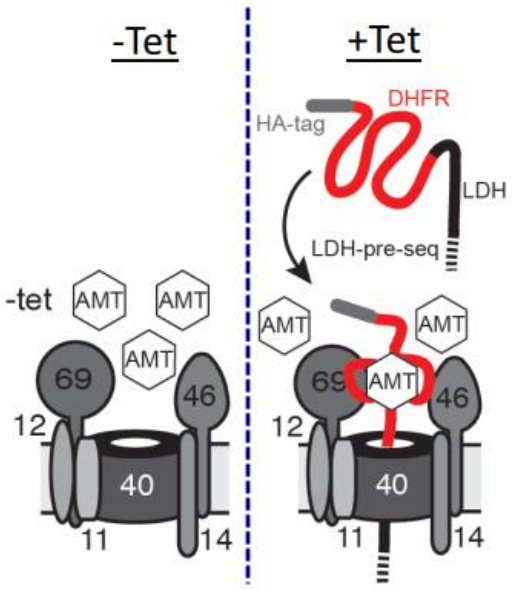


Methotrexate stabilizes tertiary structure of DHFR

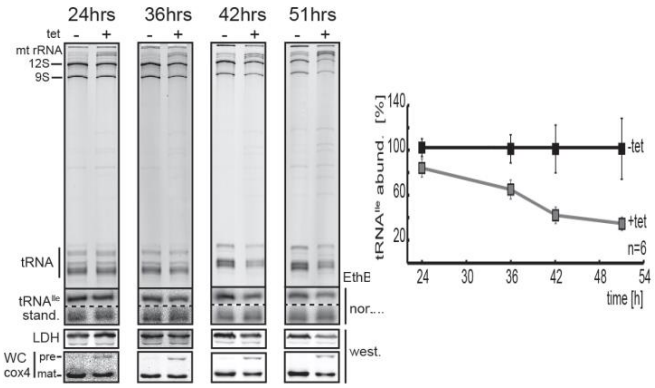


- methotrexate binds 1000fold stronger to DHFR than folate

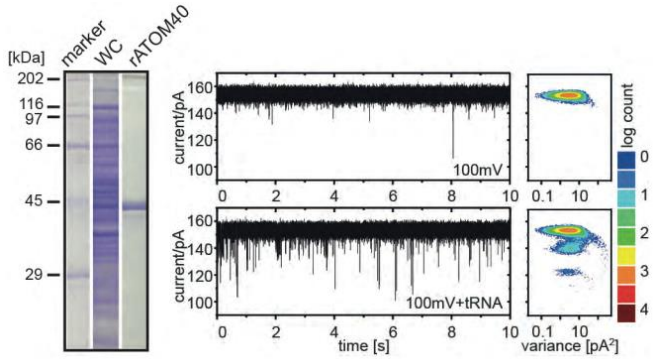
Plugging the protein import pore inhibits mitochondrial tRNA import



Plugging the protein import pore inhibits mitochondrial tRNA import



tRNAs influence gating of ATOM40 in reconstituted membranes



CONCLUSIONS

- Import of tRNAs across the mitochondrial outer membrane of T. brucei
- tRNA import required the core subunits of the ATOM complex
- tRNA import does not require the receptor subunits of the ATOM complex
- tRNAs are not co-imported with proteins
- tRNAs and protein use the same import pore

THREE MODELS FOR MITOCHONDRIAL TRNA IMPORT