Topic 22 and 23 (Transcription)

Why bother with RNA step (transcription)?

1. the RNA step provides amplification which can contribute to differential gene expression (increase the number of a gene in different cells)
2. RNA can be rapidly degraded, expression of a gene can be stopped quickly. You don't want to degrade the gene permanently but sometimes you'ld like to turn it off.
3. RNA provides additional opportunities for regulation of gene expression (RNA processing, RNA export from the nucleus, etc.)

The first in a biological process is often highly regulated as it **saves energy!** Don't have to waste ATP.
-Transcription is the 1st step in the expression of any gene and is often highly regulated.

Procaryotic transcription
Promoter: DNA sequences required to initiate transcription of a gene or operon
Terminator: the DNA sequence required to STOP transcription
Operon: a set of genes transcribed from a single promoter and thus expressed from a common RNA.

What are the key features of a bacterial promoter?
+1 site where trans begins (where the RNA starts) *the base on the right (downstream) is +2, and on the left (upstream) is -1, there's no zero.
-10 sequence positions 10 base pairs upstream from the +1 site
-35 sequence which is 35 base pairs upstream from the +1 site

How were -10 and -35 sequences identified?
-finding common sequences in bacterial promoters by aligning E.coli
-If a sequence is found a lot it's probably important
-They aligned the e.coli at their +1 sites and looked for common sequences and found they had shared common sequence at -10 (TATAAT) and at -35 region (TTGACA); the exact sequence isn't always exactly the same but they are very similar most of the time, e.g., the first base at -10 is T 77% of the time.

?? Having found these how would you test to see if these sequences were important?
-mutate one of the sequences, e.g., mutate -35 sequence and see how transcription occurs.

These sequences are called **consensus sequences;** The most frequent base at each position in a group of functionally related DNA elements.
-the actual concensus sequence doesn't actually have to show up, it's the collection of the most frequent bases you see in a number of examples.

Bacterial RNA polymerase

-The enzyme that makes RNA trans using DNA as a template and nucleoside ts as a substrate.
-Others can do the job but RNA polymerase is the only one that uses DNA as a template
-multi subunit enzyme (5) but don't memorize
The core enzyme can make RNA, but doens't recognize promoters. The promoter specificity of RNA polymerase is determined by the **sigma subnunit**
The core subunit + sigma subunit = RNA polymerase Holoenzyme

-Sigma recognizes the -10 and -35 sequences and allows polymerase to come to bind to the promoter.

STEPS IN THE INITIATION OF TRANSCRIPTION:

1. RNAP holoenzyme binds the promoter. **Closed complex** is the RNA polymerase bound to the promoter
2. RNA polymerase unwinds the DNA strands (all by itself) around the start site - **open complex**
3. The first NTP is brought to the template - no primer is required! (This is at sequence + 1)
4. Using the NTPs as substrates, chain elongation proceeds in a 5'-3' direction, following base pairing rules. Replication and transcription both follow these rules. Phosphodiester formed- pyrophosphates released but the complex is still sitting there
5. After 5-10 nucleotides the sigma falls off, the core enzyme is freed up and races down making RNA transcript.
6. RNA sunthesis continues until ta terminator is reached and RNAP falls off and you're left with the transcript.
7. Sigma can rebind to RNAP and the cycle is repeated. Sigma can rebind to another RNA polymerase too not just the same one.

GENE REGULATION
How are different bacterial promoters transcribed at different levels?

1. Some genes have better -10 and -35 sequences (better at binding RNAP whole enzyme), if you have a weak binder, that promoter won't bind well and the gene won't be well expressed. But this is not dynamic. The cell can't change it's -10 and -35 sequences so this is a static regulation, can't be changed over time.
2. There is more than one sigma factor. Each recognizes different promoter sequences. The normal sigma factor used most of the time is called sigma 70, sigma 54 is used in Nitrogen metabolism, sigma HS is used under heat shock. Works well but very specific for bacteria.

3. Gene specific regulatory proteins. Very dynamic!! Specific proteins that interact with specific genes turning them on/off. If they inhibit (negative reg) if they induce (positive reg). Found in both prok and euks and is how most genes are regulated.

Example of negative reg: trp operon

- trp operon is a set of 5 genes required for tryptophan biosynthesis are transcribed from a common promoter.

??? when is the trp operon expressed?

-when theres little tryptophan in the cellular environment (bc this is when you need to make tryptophan).

How is the trp operon regulated?

-between -10 and -35 is a trp operator sequence which binds a protein called the trp repressor.

-At high tryptophan, tryptophan binds to trp represser, and this complex binds to the operator. Then RNAP whole enzyme cannot recognise the -10 and -35 thus blocking transcription. The trp repressor doesn't bind well to the operator on its own.

-At low tryptophan, theres no tryptophan bound to the repressor and so doesn't bind to the operator and transcription can proceed.

Trp repressor is a protein, expressed as a monomer, and it's mostly helical (6 helices in total). Helix 4 and 5 make up the helix-turn-helix motif. It funtions as a dimer, has 2-fold symmetry; one side is mirror image to other. When they bind to tryptophan the helix 5 fits into the major groove (so there are two 5 helices and both bind to adjacent major groove sections on DNA).

Tryptophan binding causes a conformational change in Trp repressor which allows DNA binding. This is the active repressor. THe change is that the helix 5 are flipped up so that they can bind to the DNA.

General themes to take from the trp operon:

1. trp repressor is a site specific DNA binding protein.
2. There is a binding site for trp repressor within the Trp promoter
3. trp repressor inhibits transcription by blocking RNA polymerase from the promoter
4. trp repressor is responsive to an environmental signal (tryptophan)

Six differences in Transcription between proks and eucs

1. three RNA polymerases in eucs
    1.1. Pol 1 most rRNA genes
    1.2. Pol 2 mRNAs (protein encoding genes)
    1.3. Pol 3 tRNAs, 5s rRNA
2. No operons in eucs. Each gene is transcribed as a single unit. 1 gene per promoter. Procs can have multiple genes transcribed from 1 promoter.

3. Eucs do not have -10 and -35 sequences - there is NO sigma factor. Promoter recognition is determined by a set of proteins, on of which recognizes a TATA box. Don't know details just know the recognition is not through sigma, or -10 and -35 but a distinct set of proteins.
4. In eucs regulatory proteins often bind DNA several thousand base pairs form the start site, far away. (compared to the -10 and -35 in proks). How can they influence one another when they're spaced apart. DNA looping.
5. **Combinational Control:** groups of proteins work together to determine the expression of a gene. You can have a mixture of events, positive or negative regulators, some on, some off, at different times, etc. For proks are regulated pretty much just on and off but eucs are more like a dimmer switch, off/on and anything in between, a huge range of expression. (as much as 1000 fold)
6. Nucleosomes and higher order chromatin structure regulate transcription. These tend to repress transcription. *sometimes the promoter is wrapped up around the nucleosome core inhibiting binding of RNA polymerase to the promoter and thus transcription, and the cell would have to get rid of that nucleosome to expose the promoter or the cell can shift the nucleosome.

Regulation of the genes required for galactose metabolism in yeast as a model for pos control and neg control. (this is in eucs)

-The GAL10 gene of yeast is regulated by galactose and glucose.
-Cells always prefer to use glucose
-in the presence of galactose and absence of glucose we express the genes needed for yeast to use galactose but if there's glucose around, they won't be expressed.
-The regulatory protein in this case (GAL4) binds specifically to sites in the GAL10 promoter. So there are GAL4 binding sites within the GAL10 promoter. In galactose media Gal4 binding "recruits' RNA polymerase 2 (enhancing transcription).
-The GAL10 promoter is not expressed in the absence of galactose. Gal80 (another protein) which sits on GAL4 and inhibits gal4's activity in the absence of galactose. The TATA is still there so you get a little bit of transcription still.
-Glucose turns off the GAL10 promoter even with galactose present. This isn't through GAL4 which is positive reg. This is neg reg.
-The Mig1 protein represses the GAL genes in glucose.
-So this promoter has both positive regulation (via gal4) and negative (via mig 1)

KEY POINTS:

1. Activation of galactose regulated genes requires a site specific DNA binding protein (gal4)
2. Gal4 helps recruit RNA polymerase 2 to the promoter
3. Activity of Gal4 can be regulated by the Gal80 protein. (transcriptional regulators themselves are also regulated, regulating the regulators)
4. The GAL genes are subject to pos and neg regulation (simple example of combinatorial control)
5. The system is sensitive to environmental signals (glucose and galactose).

TOPIC 24 - Formation of mRNAs: RNA processing (eucs cells)

5' capping

7-methylguanosine at the end.

-Importance of the cap: marks the 5'end of the mRNA as being intact.

required for:

- -mRNA export from the nucleus
- -translation of the RNA

3'Polyadenalation

-protein factors recognize a sequence near the end at the 3' end and adds AAAAAAAAA…..

-Importance:

- -marks the 3' end of the mRNA as being in tact.
- -protects the mRNA from degradation (if nucleases are chewing mRNA from the end, the A tail acts as a buffer.

Required for:

-mRNA export

-translation

RNA splicing

-Protein encoding sequences (exons) are interrupted by non-coding sequences (introns)

-Introns are spliced out of the primary transcript to give the mature mRNA.

*Exons = expressed, Introns-interruptions

-We have a lot of introns. Most of the gene is introns, not expressed at all, not just total junk tho.

Why splice?? Splicing increases the coding capacity of the genome

Differential splicing!!!

An RNA can be spliced in different ways to create related but distinct proteins.

-Splicing pattern is often tissue specific (e.g., all smooth muscle will have the same exons.

Mechanism of splicing

-Splicing requires specific sequences in the RNA

a) At the 5' splice junction; includes some sequences in both exon and intron
b) At the 3' splice junction
c) Branch point; internal; includes an A which is required for splicing

After splicing you'll have AG from the 3' end and G from the 5' end joined together AG-G.

1st step:

Attack by the 2'-OH of the adenine at the branch point with the 5' splice junction. Frees the 5' exon and forms a lariat structure.

2nd step:

Attack by the 3'-OH of the free 5'splice junction with the 3' splice junction and this joins the exons and releases the lariat which is then quickly removed.

The spliceosome: the enzymatic machinery for slicing
-composed of snRNPs which contain both RNA and protein.
-Through RNA-RNA and protein-protein interactions, the spliceosome positions the RNA for splicing. THe spliceosome brings everything together so the freed 5' end doesn't just float away, etc.

*Self-Splicing RNAs:
-RNA can have catalytic activity
-Some RNAs can self splice
This is why RNAs are thought to be the first (before DNA and proteins).

Splicing and human disease:
e.g., Abnormal splicing of the b-globin RNA can result in hemoglobin deficiency. Problems with transporting O2 in the blood.
-mRNA with extended exon 3; In some cases, people have a mutation where theres another splice site (another junction) so an additional piece of protein is incorporated into the final sequence, this doesn't work well and can cause disease.
-another problem, mRNA with exon 2 missing; important sequence is then missing in the protein

*RNA processing occurs when transcription is ongoing. The splicing happens while the polymerase is still on the DNA.

mRNA EXPORT
-Requires proteins that interact with the 5'cap and the polyA tail, and with specific protein carriers. Some mRNA get out faster than others.
-transport occurs through complex structures called nuclear pores. In doing so it ineracts with the pore components. So the transport isn't diffusion, it involves interaction with nuclear pores which are selective gates. It's a double bi-layer, with a lot of proteins involved. Proteins in the pores "basket" recognize the 5' cap, etc. Transports large molecules, smaller ones can diffuse through the nucleus.

It is important to recognize that most tools recombinant DNA technology use chemical quantities of the DNA. While I will often present DNA as a single line, usually there are many identical molecules in the reaction. Check question on the forum!

TOPIC 25 TRANSLATION

GENETIC CODE

-"spells" out the amino acid sequence in 3 "letter" "words" called codons.
??? how many different codons does the cell need? 20 for the 20 amino acids. But there are 64 codons in total but some code for the same amino acid.
???Why is each codon 3 bases?
-A one letter code with A,G,U,C would provide 4 codons
-A two letter code would provide 4 x 4 = 16 codons
-A three letter code provides 4 x 4 x 4 = 64 codons
So a three letter code is sufficient for 20 amino acids, more than enough but 2 is not enough.

Key features:
-The code is **Universal** - with only rare exceptions, meaning the code has evolved ONCE!
-The code in non-overlapping; overlap would restrict which amino acid residues could follow each other and that would be very restrictive.
-There's no gaps in the code
-Redundancy; some codons specifiy the same amino acid and the variation usually happens at the 3rd position but the 1st 2 bases will be the same for all of them coding for the same amino acid.
-Generally amino acids found less frequenctly in proteins have fewer codons, e.g., Met - AUG
-3 stop codons (UAA, UAG, UGA) and 1 start codon (AUG- which codes for methyionine)
-Functionally related amino acids have similar codons, eg, Asp and Glu.

??? why do functionally related amino acids have similar codons?
-increases the change of a functional protein in the case of a single base mutation

TYPES OF MUTATIONS

1. Missense mutation

-results in a single amino acid change in the protein.

2. Frameshift mutation

-insertion or deletion of bases; can dramatically change the protein sequence.
-the only time if a frameshift doesn't cause a huge problem is if it's in a multiple of 3

3. Nonsense mutation

-prematurely terminates protein. When a stop codon is created. The cell has special machinery that looks for these and if found the RNA is removed.

*#2 and #3 are definiltey no good but sometimes #1 can be tolerated.

tRNA (TRANSFER)

-all have a similar clover leaf type structure with stem loops, ~80 bases

-stem loop structure due to internal base pairing, where there's no base pairing you get a loop, base pairing makes a stem
-In 3D the tRNAs look like an L or a gun.
-they all have nonconventional bases (not just U,T,C,A), these are bases that have been modified and put in post-transcriptional.
-2 crucial single strand regions

a) 3' acceptor site, CCA is added post-transcriptionally, accepts the amino acid
b) Anticodon (base pairs with the codon of the mRNA)

Wobble

Most organisms have fewer than 45 different tRNAs. How can 61 codons be used?
-Some tRNAs must pairs with more than one codon.
-for some tRNAs base pairing between the anticodon and the codon only requires matching at the first two positions of the codon (last two of the anticodon).
This is why we can get by with only 45 different tRNAs when there are 61 codons.

Aminoacyl tRNA synthetases

-Couple the 3' end of a specific tRNA to its correct amino acid. At least one for each amino acid. Often called "charging"; the bond formed is a high energy bond. :Store energy from ATP in a high energy ester linkage. They recognizes the amino acid, the tRNA, and have a site for ATP. Coupling is highly specific! Aminoacyl-tRNA synthetases have proofreading activity in order to check that the right amino acid is being incorporated into the protein.
*specificity
*high energy conserved in the ester linkage

Ribosomes

Macromolecular machinery (enzyme) that catalyzes protein synthesis. Composed of rRNAs and proteins. Huge molecule! The RNAs play key roles (the enzymatic catalytic activity).
-has 2 subunits; large and small
-Three sites for the binding of tRNAs on the ribosome (A, P, E)
    -A site: bnds aminoacyl-tRNA
    -P site: binds peptidyl-tRNA
    -E site: the site from which tRNAs exit
-Only 2 sites are occupied by tRNA at a time
-mRNA is bound close to the A and P sites

STEPS IN TRANSLATION

Fig (7-34) assumes that translation has already started - starting with a peptidyl-tRNA is in the P site of the ribosome.
Step 1. An aminoacyl-tRNA binds to the A site of the ribosome. Requires base-pairing between the tRNA and the codon, only the complementary anticodon (on the tRNA) will pair.
Step 2: The energy from the ester bond of the peptidyl-tRNA in the P site is used to form a new peptide bond between the amino acids in the A and P sites.

Step 3: Peptide bond formation is coupled to a conformational change in the ribosome that shifts the large subunit "forward". This conformational change repositions the tRNAs. The tRNA that was in the P site moves into the E site and the tRNA in the A site moves into the P site.
Step 4: The small subunit moves forward (5' - 3") exactly 3 bases. At the same time the tRNA that was in the E site exits.
Cycle repeats

??? Specificity of translation comes from?
-correct charging of the aa-tRNA synthetase
-base pairing of tRNA and codon in the A-site of the ribosome
*the ribosome doesn't recognize the amino acid

Key points for translation:

1. Translation occurs in the cytoplasm
2. Energy for peptide bond synthesis comes from the high energy aa-tRNA ester bond-only indirectly form ATP
3. Specificity comes from the :
    3.1. aminoacyl tRNA synthetases
    3.2. requirement for base pairing in the A site of the ribosome
4. Complex reaction involving:
    4.1. both RNA and protein molecules
    4.2. conformational (shape) changes in the ribosome
5. occurs in a 5'-3; direction along the RNA making a protein from N to C terminus
6. The mRNA is decoded on codon at a time

Initiating translation
Determining the reading frame is important. Start site.
in Euks
Starts with the initiator tRNA interacting with the small subunit. Small ribosomal subunit with translation initiation factors bound. This interacts with the 5' cap of the mRNA. Initiator tRNA scans/moves along RNA searching for the first AUG (start codon). When it hits that, the large sun-unit comes in, now the MET (coded by AUG) is in the P site and this is when were ready to start; can bring in the next aminoacyl-tRNA into the A site.
??? what might happen if there was a hairpin loop in the mRNA between the 5'end and the AUG?

a) it would have no effect on translation
b) it might decrease translation
c) it might increase translation

answer: b, in some cases it might stall at the hairpin and not be able to reach the AUG. Things like this cause different mRNA to be translated differently.

??? WHy must the translational initiation in bacteria use a mechanism other than 5' scanning?

a) there are no small ribosomal subunits
b) there are multiple proteins encoded from common mRNAs in bacteria
c) the mRNAs are tranlsated by RNA polymerase in bacteria
d) bacterial genes are monocistronic

Answer: b, if it used a scanning mechanisms it would only get the first start codons for the first genes.

In prokaryotes

-polycistronic mRNAs
-Ribosomes recognize internal ribosome binding sites found just upstream from each functional AUG
-Ribosome recognition involves an RNA-RNA interaction

Termination of translation

Requires: one of the three stop codons and specific termination factors.
Mechanism: when you get a stop codon in the A site, instead of tRNA coming in, a special release factor comes in (termination factor) causes the ribosome to catalyze the addition of water to the peptide chain rather than amino acid and that releases the peptide and though subsequent conformational changes of the ribosome everything dissociates.

***Antibiotics and translation**

-Many antibiotics block bacterial translation.
-Its intricate molecular interactions and importance make it a prime therapeutic target.

Topic 26: Recombinant DNA technology (genetic engineering)

-Def: The techniques which recombine DNA fragments from different sources into new molecules with unique features.
Synthetic Biology: Engineering of biological pathways or organisms, with the goal of carrying out novel functions or performing existing functions more efficiently. Ex, Craig venter is trying to produce oil.
-Manipulating genomes of organisms in not new. Humans have been doing this for 1000s of years through classical genetic selection (roses, dogs). Plant/crop and livestock breeding are examples of humans manipulating the DNA of other organisms. Chickens are HUGE now!

Comparison of classical genetic techniques and RDT:

Classical genetics:

a) slow; limited by the breeding time of the organism and chance genetic events.
b) exchange of genetic materials is limited of breeding species.

Recombinant DNA techology:

a) rapid; as quick as a few days in some organisms.
b) No limitations

Significance of RDT

1. Research - key tool to understand cell and molecular structure/function
2. Biotechnology - major impact on society

In Medicine:
-drug production and design (human insulin, etc.)
-Diagnosis of disease - detect pathogens and disease causing genes through their DNA signatures
-Genetic counseling - does an individual carry a disease related to an allele?
-Potential for gene therapy; Many diseases result from a defective gene - we should be able to add back functional genes to cure disease.

It is important to recognize that most tools recombinant DNA technology use chemical quantities of the DNA. While I will often present DNA as a single line, usually there are many identical molecules in the reaction. Check question on the forum!
??? In a DNA sequencing reaciton yu genearlly add 100ng of DNA. Assuming the DNA is 3000bp and one base pair has a mass of 600daltons, approximatley how many molecules of template are added?
-1 mole of DNA of 3000 bp has a mass of 3000bp x 600g/bp = 18 x 10^5 = 2 x 10^6g
x = 1 x 10^-7 / 2 x 10^6
= 5 x 10^-14 moles

this x 6.02 x 10^23 molecules
=30 x 10^9 molecules
= 3 x 10^10 molecules

Agriculture:
Producing crops with unique features
-Vit A enhanced rise (golden rice)
-Cold and drought resistant crops
-Pest resistant corn (BT corn restitant to european corn borer)
-Nonspoiling tomatoes
-Nonallergenic peanuts
-Production of novel molecuels in plants, e.g., large scale production of therapeutic drugs

Manufacturing:
Novel products e.g., spider silk, proteases for detergents, biofuels

Environment:
Bacteria that "eat" oils, PCBs, styrofoams. Just add the bacteria to the spill and they clean it up for you.

Forensics:
DNA fingerprinting - CSI

Biotechnology remains one the fastest growing segments of the North American economy.

recombinant DNA tech case study
You have identified a human protein that will be of tremendous value (e.g., human factor VIIIm HGH, INsulin,)
Your goal is to mass produce this protein your favourite protein (YFP).
How do we do this? how do we mass produce a protein from a gene?

Why is expression of YFP using RDT important?
-many proteins are difficult to obtain from their native source.
-RDT can dramatcially increase expression and facilitate purification.

The process of expressing YFP will require:
1. cloning YFG
2. introducing YFG into e.coli
3. purifying the protein from e.coli (covered in dr. deroo's section)

E.coli are often used to express YFP
-grow quickly and inexpensively
-protein extracts are easily made

-genetic engirneering is simple (there are many developed techqniques)
-multipcopy plasmids and strong promoters can drive expression

The first issue we need to think about:
-We need a source of the gene. Should it be genomic DNA or mRNA?
??? if our goal is to express YFhumanG, can genomic DNA be used as the source?
-Only for the few genes that lack introns; e.coli bacteria can't deal with introns, they can't splice so you can't use genomic DNA (if there's introns which theres probably is).

We will use mRNA (cDNA) to close YFG
-mRNA/cDNA does not contain introns
-each mRNA/cDNA encodes only 1 gene!

Do all tissues express YFG as mRNA? no. How do we find a tissue or human cell line that expresses a lot of YFG? use a hybridization technique called northern blot.

Nucleic acid hybridization
-Denaturation/melting
-Renaturation/annealing/hybridization

DNA denaturation: if you heat a DNA double helix, the strands will melt/dissociate and you get two independent strands because the Hydrogen bonds ebtween the strands is disturbed.

Melting tmeperature: TM Given a DNA at a defined concentration, the temperature at which 50% of the molecules are single stranded is the TM. THe temp needed to melt half the molecules. Not 50% of one molecule, 50% of the population of molecules.

Factors that determine the TM :
Intrinsic features of the moecule itself.

1. G:C content (the more GCs the higher the melting temperature; because G:C has 3 hydrogen bonds between them where AT only has 2, also theres more base stacking for GCs, stronger base stacking)
2. The longer the DNA molecule, the higher its TM.
3. Degree of complementaririty between the two nucleic acid strands the higher the TM .Some mismatch can be tolerated in a hybridization but it does significatly lower the TM .

Extrinsic factors that effect the TM

1. salt concentration; the higher the salt concentration, the higher the ™. Positively charged ions shield the repulsive interactions of the phsophaste backbone.
2. Solvents that disrupt hydrogen bonding, interfere with base pairing and lower the TM.

Melting is reversible (process is called renaturation, hybridization or annealing).

Hybridization requires complementary strands. DNA will hybridiza to RNA or DNA, just has to be complementary and it doens't have ot be exact just close.

Hybridization identifies complementary nucleic acid sequences.
Hybridizations are often done after transferring the nucleic acid to a membrane.
-Southern Blot: DNA to DNA
-Northern Blot: DNA to RNA (RNA is on the membrane)

<u>Northern blotting ex</u>
Goal: identify which tissues express YFG

1. Isolate RNA from a group of tissues; grind up the tissue and extract/isolate the RNA
2. Separate the RNA on the basis of size by agarose gel electrophoresis
3. Transfer the RNA to a nitrocellulose membrane (botting); very simple, takes about 24 hours. End up with an exact replica of the gel
4. Label the probe. Need a probe to detect YFG. Radioactively label DNA that encodes a portion of YFG.
5. Incubate the HOT PROBE with a filter. complementary strands will anneal (hybridize)
6. Wash away the nonspecifically bound probe.
7. Expose the nitrocellulose filter to an x-ray film to determine where the hot probe has hybridized.

???Why might there be two bands in liver but only 1 band in heart?

OTHEr hybridization techiniques:
Southern plot: Hybdrization of DNA to DNA on a membraine. commonly used to address whether YFG is expressed in different **organisms**, instead of different cells/tissues. fig 10-5

Microarrays: DNA to DNA on glass slides. Used for profiling RNA/cDNA in different cells fig 10-27.

Where do we go from here? **CLoning!**
mRNA can't be cloned, so we have to convert it cDNA using reverse transcriptase. cDNA is complementary DNA copy of mRNA.

1. isolate RNA from the tissue (heart of liver, etc).
2. we know it ends with a polyA tail so we can add a primer that is a polyT primer (at the 3' end) and hybridize.
3. Add the enzyme reverse transcriptase and dmtps and it will make a complementary cDNA.
4. Remove the RNA using alkali so you're left with a cDNA single strand

How do we get enough of YFG to be able to clone it? POLYMERASE CHAIN REACTION!
-amplifies DNA from a small amount

*Synthetic Oligonucleotides
-Fragments of single stranded DNA with defined sequence that are made synthetically.
-When they are short (20-30 bases), they are often referred to as primers or oligos.

PCR

Exponential amplification of any DNA from a source in which it is found as little as one time.
Reagents required for PCR:

a. Template DNA
b. Two oligonucleotide primers which flank YFG (in both directions) - these you design/order
c. dNTPs
d. DNA polymerase

PCR cycle (repeat 25-30 times)

1. denature the strands (with heat ~ **95** degrees), the strands separate
2. lower the temp to **~50** degrees so the primers anneal to the strands (the primers will only hybridize to those specific strands).
3. extension reaction by DNA polymerase done at **72** degrees. polymerase will extend the primers to the end of the DNA.

After 1 cycle, where you start with 1 DNA molecule, you end with 2 DNA molecules
After 2 cycles, each of those 2 molecules are cloned = 4 DNA molecules
After 3 cycles, each of those 4 are cloned = 8 DNA molecules
Number of copies = 2^ (#of cycles); after about 30 cycles we have 2^30 which = 1 billion

TWO KEY TECHNICAL ADVANCES:

1. Discovery of thermostable DNA polymerase (Taq polymerase)
2. Thermocyclers (machine) that oscillate between the 3 required temperatures (50, 94, 70)

Gel electrophoresis

Next, we have to purify the DNA (isolate the specific gene from the DNA) using gel electrophoresis:
-In an electric field, DNA (bc it's negative phosphate backbone) will migrate towards the positive field.
-DNA is sieved through a matrix of agarose being pulled by an electric field.

Agarose : after melting it hardens into a form like jello. It has small pores through which the DNA can pass through as it migrates towards the positive end.
The agarose gel is in a chamber submerged in buffer. An electric field is applied.
Small molecules move faster than the big ones bc they can sneak through the pores easier.
You also separate RNA by electrophoresis but you have to denature it first because RNA often folds back on itself.

Detection of the DNA in the gel: best 2 ways

1. When stained with ethidium bromide DNA fluoresces red under UV light.
2. Autoradiography (used if theres very small amounts of DNA)
    a. radioactively label the 5' ends of the DNA frags using polynucleotide kinase and [32P] = ATP; adds a phosphate group on the end
    b. Expose your gel to an x-ray film

How do we get YFG into E.coli? We insert in into a plasmid!

E.coli expression plasmid

Plasmid is fragment of DNA that replicates independently from the host chromosome.
-copy number varies depending on the plasmid from 1-100 copies
-3000 bp (much smaller than the host chromosome)
-circular

Key features required to make plasmid useful for cloning:

a. Origin of replication
b. A selectable marker: must be a way to detect that plasmid when its in the e.coli; usually a gene encoding resistance to an antibiotic, e.g., ampicillin
c. Insertion site: A site(s) into which YFG can be inserted; somewhere to clone YFG

How to get YFG into the plasmid: **Restriction Enzymes**

-DNA is a long relatively homogeneous polymer. Looks like the same thing, it's not in workable units. Genes can't be easily identified. Restriction enzymes allow you to cut DNA into defined workable units.
-Restriction enzymes are like site specific DNA scissors. Site specific DNA binding proteins that recognize and cleave palindromic DNA sequences usually of 4,6, or 8 base pairs.
-Palindromic sequences have a 2-fold symmetry. The restriction enzymes also have two fold symmetry, they are dimers.
-Approximately 300 restriction enzymes are commercially available
-Isolated from bacteria where their native role is to act as anti-viral agents.

3 types of restriction enzymes:

1. 5' over hang, e.g., EcoR1-ex RE.,
2. 3' overhang, e.g., Kpn1
3. blunt end, e.g., Ssp1

-restriction enzymes do not have to be specific at every single position that they cut.
-Some restriction enxymes cut identical sequence but leave different TYPES of overhangs (blunt vs 3' or 5')
-some RE cut different recognition sequences but leave identical overhanging ends
-RE are enzymes; like any other enzyme EACh RE has a preferred conditions in which it functions (temp, pH, salt concentration, etc.), so you have to incubate them with the DNA under their specific conditions for them to cut

How do we get the molecules back together after cutting them?

DNA LIGASE!

-will reseal compatible sticky ends (ends that will anneal together) and much less efficiently blunt ends
-requires an energy source (ATP)
-reseals the nicks (the phosphodiester backbone of the DNA)
-requires phosphates on the 5' DNA ends

Inserting YFG into a plasmid - making a recombinant molecule

Steps in cloning:

1. Digest ~100 ng of YFG and a plasmid vector with the same restriction enzyme or one that gives compatible overhanging sequences. Results in YFG and the plasmid left with compatible sticky ends.
2. Clean/purify your DNA via electrophoresis
3. Incubate plasmid + insert (YFG) in the presence of DNA ligase and ATP.
    a. the sticky ends anneal
    b. ligase seals the ends with covalent phosphodiester bonds

How do you get the restriction enzyme sites onto the ends of YFG???? order them with them like that!

- Gibson assembly:
    - Allows construction of recombinant molecules without using restriction enzymes
    - based on annealing of complementary ends
    - very important in synthetic biology

So now, introducing YFG into E.coli (Transformation):

-kind of like magic
-The process by which cells take up DNA from their environment; natural property of some bacteria. E. coli aren't so good at it and must be treated with chemicals to do it. Tho don't really know exactly how this works.
-mix e.coli with ligated DNA
-incubate them together and give them a heat shock
-some of the e.coli will take up the plasmids (1 in 10 000)
-Plate cells onto agar-plates containing antibiotic.
-Cells with the plasmid will divide, forming visible colonies.
-leave over night and e.coli colonies containing YFG will grow
-those that haven't taken up with plasmid won't grow on the plate because of the antibiotic

??? not all of the bacteria on the plate contain a plasmid with YFG inserted. Why?

1. contaminating amilcillin resitant bactiera may appear

2. The plasmid vector can re-circularize without YFG (reform the circle without incorporating the DNA insert) you can avoid this by de-phosphoralting the ends so they dont re-criculize (this is the most likely problem)
3. Contaminating DNA may be ligated into the vector

Checking that clones contain YFG

Amplify bacteria in individual colonies. Isolate the plasmid DNA. Verify which "clones" contain YFG.

-restriction map
-hybridization
-PCR
-sequence

Restriction mapping:

A DNA molecule can be defined by the positions of its restriction enzyme cut sites. YFG will have specific RE cut sites that are diagnostic.

DNA sequencing:

To really be sure it's right, you have to sequence it!

-Two methods were developed in the early 70s to sequence DNA
    -Chemical sequencing (maxam and gilbert)
    -Dideoxy or chain termination sequencing (sanger); method of choice.

Chain termination sequencing:

-Involves the enzymatic synthesis of a DNA strand in the presence of base specific chain terminators.

-The bases in DNA synthesis has an OH at the 3' positions. The dideoxynycleotides used in chain termination doesn't not, so the chain stops. Regular DNA has deoxy at 2'. Dideoxy has deoxy at 3' and 2'.

Materials required to sequence DNA:

1. Dna to be sequenced template
2. Oligonucleotide primer that anneals to the template
3. DNA polymerase
4. dNTPs (G,A,T,C)
5. ddNTPs (small amount (of one), 0.2% of dNTP)

Mix all that together and let DNA polymerase extend 5' to 3' and at some point the ddNTP will be incorporated and terminate the chain. There is a 1/500 chance a ddNTP will be inserted. And you get a collection of DNA molecules with an identical 5' end but differ in their length depending on when the ddNTP was added. And this will give you the pattern of the As or Gs or Cs or Ts, etc. that you are analyzing for. So you do 4 separate reaction to sequence each base.

Then, denature the DNA molecules. Analyze the products by gel electrophoresis. [Make the strands radioactive (5' label)]. Read the DNA sequence as a ladder from 5'-3'.
-Too little ddNTP would mean the chain wouldn't be terminated and you wouldn't read any sequence. If you add too much, the chains would terminate to quickly/often and few bases would be read, you would just have short strands.

-In the mid 90s automated DNA sequencing became common place. Each ddNTP is labeled with a different dye that fluoresces at different wavelength. This way you don't need radioactivity.

*Next generation DNA sequencing
-Numerous novel methods have been devised to obtain large amounts of DNA sequence at minimal cost
-$1000 genome and this allows personalized medicine
-RNA profiling (RNA-seq)
-replacing microarrays
-p.343

*E.coli are the most commonly used organism for expressing YFG. Easy to use, cheap, etc. Which of the following need to be considered when expressing YF(Human)G (protein) in E.coli?

a. Human and E. coli promoter structures are different
b. Introns are found in human genes (have to use cDNA)
c. They differ in their mechanism of translational initiation
d. Different post-translational modifications occur in proc and eucs.

Answer is all of them!
We haven't talked about d. If theres post translational modification necessary for function, don't use bacteria.
These two issues (a and c) are handled through the design of the plasmid used for cloning.
-These plasmids are called expression vectors; they have a promoter and a ribosome binding site just in front of the cloning site. Insert YFG just downstream from the promoter.

Producing YFG; purify protein from E.coli, check protein for purity and activity and
YFG is NOW ready for use!

Let's say you want to alter the sequence of the gene……
Site-directed Mutagensis; creating a gene with altered sequence and in turn altered function
-You believe that an aspartate within YFP may lead to aggregation. Want to overcome this by changing a codon to alanine. Converting a codon for Asp to a codon for Ala.
-Asp codon is ACG and Ala codon is CCG. We have to change an A to a C.

- make mismatch oligonucleotides; anneal them to the plasmid, introduce the plasmid into the bacteria and grow them.

Transgenic organisms

GMOs

-An organism that has had its genome permanently altered through genetic engineering.

-You create a GMO by integrating DNA into the organism's genome through the porcess of homologous recombination.

Three different types of genetic changes are possible:

1. gene replacement
2. gene knockout
3. gene addition

Gene knockout is taking the gene right out. Gene replacement is altering the gene, slightly modify it. Or add a brand new gene to the genome is "addition".

Gene knockout is used a lot in research; if you want to know what a gene does in an organism, knock it out and see what happens to the phenotype. This is how we know what all 6000 genes in yeast do. Also used in biotechnology, e.g., non-spoiling tomato; knock out the genes that are responsible to ethylyne production so they don't spoil.

Gene addition is the basis for biotech industry, e.g., golden rice; adding genes required to produce a vitamin, or genes that make the plant resistant to cold.

Gene replacement important in research.

Uses of PCR

1. Rapid isolation of YFG
2. Analysis of bacteria or viruses in clinical or environmental samples
3. Diagnosis of genetic disease - or in genetic counseling
4. Forensics - DNA fingerprinting

DNA fingerprinting

-In many regions of our genome there are repetitive sequences, e.g.,

-CACACACACACACA…… or CAGCAGCAGCAG

-The repeat varies from 4-40x.

-These are called **STRs** for short tandem repeats.

-If a PCR reaction is performed with genomic dNA as a template;, and a pair of primers flanking one STR locus, you will see 2 bands bc our genome is diploid. Paternal and maternal.

??? Are the primers for PCT amplification X and Y the same sequence or different?

answer: different;

??? There is a chance that any two individuals will have the same two bands when any one STR locus is examined by PCR.

answer: true

-Two individuals might have the same two bands simply by chance. For this reason multiple STR loci are analyzed in a DNA fingerprint. Each STR locus requires a distinct pair of primers.

CSI London

1. You're a member of CSI and investigating a robbery.
2. From the scene you carefully remove a hair that may be the culprit's.
3. The crook was wearing a mask, but witnesses noticed the get away car.
4. You track down 3 possible suspects and obtain biological samples
5. in the lab you isolate DNA from the forensic and suspects' samples
6. you have primers for 3 STR loci; you PCR with the 3 primer pairs, using each of the suspects' DNA and the forensic DNA sample as templates.
7. you analyze the PCR products by agarose gel electrophoresis

Points to consider with DNA fingerprinting
The sensitivity makes the approach powerful but means extreme care must be taken not to contaminate the samples.
It is based on probability. That is two individuals could have the same pattern by chance. You have to PCR multiple loci.

Topic 27: Genomics

Sequencing the code!
Sequencing genomes involves:

1. creating a genomic DNA library
2. many independent sequencing reactions
3. aligning the independent sequences into a continuous sequence that represents the genome.

<u>Genomic library</u>
A collection of cloned DNA frags that rep all of the DNA in an organisms genome. Each cloned DNA frag is like a different book. There's also cDNA libraries; repping all the RNA in an organism.
How to construct a DNA library:
-You've got human DNA (from billions of cells)
-Cleave that with restriction nuclease and you get a bunch of frags
-clone these as a pool into a plasmid and introduce into bacteria and each of the colonies on the plate will have a different frag in them. Need millions of clones to cover the whole genome.

H. influenza genome's was the first sequenced:
Gram negative bacteria. Genome ~2 million bp (small genome).
??? The H. influenza genome is 2 million bp's. How many independent clones must be in the genomic library to cover the genome at least once, assuming an average insert size is 2000bp?
=1000 clones minimum (2 mil/2000)

What they did:
Start with millions of cells from the h.influenza, extract the DNA.
-To fragment it they sonicated instead of using restriction enzymes bc the sonication is random, but same principle; get a bunch of frags of different sizes.
-They sized the frags using gel electrophoresis purifying for DNA frags of size 2000bp's.
-prepared the genomic library from that DNA using 20 000 clones (20 fold excess) each representing a fragment of the genome.
-Isolated the plasmids from each of the clones and sequenced.
-Used conventional dideoxy sequences and read about 500 bases at a time.
-Did 25 000 sequencing runs from the 20 000 clones.
-Resulted in 12 million bp of sequences; 6 fold excess.

Now putting all 25 000 sequences together in the correct order to make the genome.
-sequence contig: a contiguous DNA sequence representing a portion of the genome. Contigs are not physical entities. They are sequences.
-A computer searches for overlaps between the 25 000 sequences runs.
-Overlapping sequences are arranged into contigs.
-From 20000 clones, 25000 sequences; computer aligned them into 140 contigs
-WIth 140 contigs, you will have 140 gaps in a circular genome. We need to fill in the gaps to order the contigs and to complete the sequence.

How to deal with the gaps?:
2 types of gaps

1. Sequence gaps: are represented in the cloned sequences. They can be closed by completing the sequence of clones in the library.
2. Physical gaps: are not represented by clones in the original library. Filled in using PCR, with genomic DNA as the template.

Annotating the genome (finding out the functions of the genes)
Where the functions of the 2 million bp of sequence?
-What types of info are in the genome?
Expressed regions:
-Protein encoding genes
-tRNAs
-rRNAs
-other functional RNAs (eg those in snRNPs)
-small regulatory RNAs

Key sequences that are not expressed:
-Regulatory sequences: promoters, terminators
-positioning elements that orient the spatial positions of the chromosomes
-nucleosomes positionings elements (for eucs)
-origins(s) of replication

-Telomeres (in some organisms)

How are protein encoding sequences identified? What are the features of a protein encoding gene?
Open reading frame: A series of codons starting with an initiation codon and ending with a termination codon. Not always protein encoding but potentially protein encoding.
-start site
-long string codons
-stop codon
Promoter

The first step in annotating a genome is for a computer program to scan looking for open reading frames.
In any given double stranded DNA sequence how many frames will the computer have to scan? 6. 3 on the top strand and 3 on the bottom.

What criteria will we use to decide if an ORF encodes a protein?
-ORFs that encode proteins usually:

1. contain 100 codons or greater. Although there are some small proteins but mostly >100 codons.
2. Show a codon usage typical for the organism (codon bias); in any organism there are preferred codons used for each amino acid. So if you are checking an organism and the open reading frame had a lot of the codons that aren't used much you know that this properly isn't coding.
3. Related sequences, encoding similar proteins are found in other species. You can do a "BLAST" search. If you see the same pattern in many organisms than it's probably a protein.
4. Are expressed as mRNA. You can do experiments to see if your ORF has a corresponding mRNA.
    a. Northern blot
    b. Microarray
    c. RNAseq
5. Contain appropriate regulatory sequences.
    a. -10 and -35 bacteria
    b. TATA for eucs
6. Contain the chromatin signatures of expressed genes
    a. nucleosome acetylation and methylation patterns
7. Are found as proteins. You have to find the protein using:
    a. Western blot
    b. Mass spectrometry

So this is for protein encoding genes. Non-coding genes are often much harder to identify but they do have their own characteristics.

TOPIC 28 CANCER

Cancer is :
-The leading cause of death in North America (used ot be heart disease)
-Strikes victims of all ages
-Is becoming more prevalent as the population ages

Tumour types (2)
-benign (cells are dividing out of control but are contained) can be removed
-malignant (cancerous cells not contained and many cases can move around) hard to remove bc they move. Malignant tumours can break through the basal lamina and invade the capillary moving around the blood stream and escaping the blood stream and goes into other tissues. Less than 1 in 1000 cells survive to form metastasis.

How do we know that cancer is a genetic disease?

1. Susceptibility to cancer can be inherited.
   -Retinoblastoma (cancer of the eye) - Rb gene
   -xeroderma pigmentosa (a cancer of the skin), nucleotide excision repair. See this in children very early.

-can be inherited, some forms of : breast, prostrate, and intestinal

2. DNA damaging agents cause cancer. e.g. radiation (UV, x-rays), chemical mutagens (benzene when it exposed to it can damage your DNA). Thus cancer must be related to DNA damage.

What are some of the avoidable causes of cancer?
-agents that cause DNA damage, radiation (xrays, UV) mutagens, etc.
-chemicals that cause cells to divide: tumour promoters - asbestos. These agents cause cells to divide may not directly damage the DNA but division can lead to tumours.
-Certain viruses: Hep (liver cancer), HPV (cervical), HIV (kaposi sarcoma, non-hodgkins lymphona).

??? What is the single most preventable cause of cancer?
-smoking! if people stopped smoking we'd cure 30% of all cancer (90% of lung cancer are due to smoking).

Cigarette smoke contains 69 chemicals that are known to cause cancer:
-Arsenic - rat poison
-Benzene - chemical mutagen
-Formaldehyde - used in mortuaries, breaks down RNA and DNA
-Radioactive Polonium 210
+ lots more

Cancer risk increases with age!
A single mutation is not sufficient to cause cancer.
-Tumour progression involves successive rounds of mutation and selection
-At each round the descendent cell acquires another mutation allowing it to grow faster or in abnormal faster.
Most cancers derive from a single abnormal (mutated) cell.
First mutation allows a cell to grow more quickly
Second mutation may allow the cells to grow in the absense of the basal lamina
a third mutation may allow the cells to penetrate the basal lamina..
etc….
This process takes some time so thats why the risk of cancer increases with age.

Properties of cancer cells:

1. Divide in the absence of growth factors. (normal cells need a growth factor to signal division)
2. Are immortal, do not respond to signals that normally trigger cell death. A normal cell has a set number of times it can divide, has a shelf life.
3. Have lost cell cycle control.
4. Are genetically unstable;
    a. more point mutations
    b. copy number variations (things that are duplicated)
    c. major chromosome abnormalities (pieces of 1 chrom onto another, etc)
5. Can multiply in abnormal places

<u>The cell cycle:</u>
G1 before synthesis, Gap 1, S (synthesis), G2 (gap 2), Mitosis phase
Checkpoints:

1. At G1 - Checks for cell size, nutrients, growth factors, DNA damage, this all occurs before S
2. After synthesis (G2 checkpoint) - Checks cell size, DNA replication completeness (no errors)
3. In mitosis (spindle assembly checkpoint) - checks chromosome attachment to spindle

A cancer cell ignores these checkpoints; doesn't care if theres nutrients, if it has errors, etc.

Types of cancer causing genes:

1. Oncogenes -
2. Tumour suppressors

Ongogene: A mutant form of a normal gene whose presence causes cancer (dominant - gain function), they do something and they do it badly and this results in cancer.

What does abnormal mean?
-Protein is expressed at too high a level

-Protein has altered activity - substrate recognition
-Protein is expressed at the wrong place or time
-What causes this:

a. mutation in coding sequence (point mutation)
b. protein amplification
c. Chromosome rearrangement (can get placed downstream from a strong promoter, get a lot of copies of the RNA and protein) or fusion proteins that do bad things.

Tumour suppressor
A gene whose absence causes cancer. (recessive, loss of function). Lack of activity causes cancer.
In cell division, one allele can get knocked out and everything still cool but after another division if you inherit 2 knock out alleles that you probably get cancer. BOTH copies must be mutated to have a problem. These are recessive. Ongogenes are dominant.

<u>Functions of cancer causing genes:</u>
-Most oncogenes and tumour suppressors code for proteins that act in:
cell division or cell differentiation. Oncogenes are the *accelerator* pedal of a vehicle.
-Tumour suppressors are the *brakes.*

1. Growth Factors and cell receptors for growth factors
   -growth factors are overexpressed or growth factor receptors are stuck in the on position. So you can have too much or it thinking its always on.

2. Molecules involved in cell-cell interactions.
   -the cell-cell interactions that control growth and differentiation are lost. When cells see another cell, it inhibits division. Cell surface proteins that recognize other cells are defective.

3. Regulators of normal/programmed cell death (apoptosis)
   -Cells do not respond to the normal signals that trigger cell death
   -Cancer cells express telomerase; their telomeres don't shorten

4. Transcription factors
   -Cancer can be caused by too much or too little expression of the genes that regulate cell growth, differentiation or cell death.

5. DNA repair proteins
   -Cells accumulate more and more DNA damage and thus mutations to key genes.

classic cancer treatments:
-Surgery (remove the tumor)

-Radiation - stops cell from replicating by damaging DNA (hitting with a high dose of radiation, can't divide). Non-specific, just damages DNA
-Chemotherapy: Stop cells from replicating by damaging DNA just like radiation. Or interfering with the mitotic machinery (taxol) or reducing replications substrates (methotrexate), inhibits the production of DNTPs.

Difficutles with treatment:

1. Not specific!
   -Radiation and classic chemo treatments also affect normal cells, which accounts for the hair loss and digestive problems, etc.

2. Different cancers have different causes
   -differente genes involved
   -different originating tissues
   -different locations
   so they are all different diseases, we need different treatments

3. Heterogeneity of the tumour
   -Not all cells within the tumour are identical
   -The tumour cells are always "evolving" (they change with time)

4. Cancer cells develop resistance to the drug
   -drug target acquires mutations or is lost
   -cells overexpress transporters that pump out the drugs

5. Inaccessibility of the tumour to drug
   -blood-brain barrier; very hard to get drugs into the brain, hard to cure brain cancer.

Goals of modern cancer treatment:
-Earlier detection
-Greater drug specificity (antibodies) antibodies specific to cancer cells, target specific cancer cells, the genes that are involved
-Combination therapies, using multiple drugs so they can't evolve to resist them all
-identifying drugs for the "undruggable" targets
-Prevention!

CML is caused by the fusion of Bcr (on chromosome 22) and Abl (on chromosome 9)
Fusing Bcr to Abl kinase alters the substrate specificity of Abl. This activates cell division and inhibits apoptosis of hemopoietic cells and that drives the cancer. So how do we block the activity of Bcr-Abl?
-Geevac is a drug that binds and blocks where the ATP would bind to Bcr-Abl and therefore inhibits activity of the molecule thus stopping the disease.
-Knowing the structure of Abl allowed an inhibitor to be designed.