Exam 3 Material

Monday, March 30, 2020 9:52 PM

Lecture 13 - Bacterial and Viral Genetics

The Advantages of Studying Bacterial and Viral Genetics:

- Reproduction is rapid
- Many progeny are produced
- Haploid genome of all bacterial and viral genetics allows for all mutations to be expressed directly
- Asexual reproduction allows for easy isolation of genetically pure strains
- Growth in laboratory is easy and requires little space
- Genomes are smaller
- Recombinant DNA technology can be used in the medical field to produce substances of commercial value (e.g. insulin)

The Microbiome

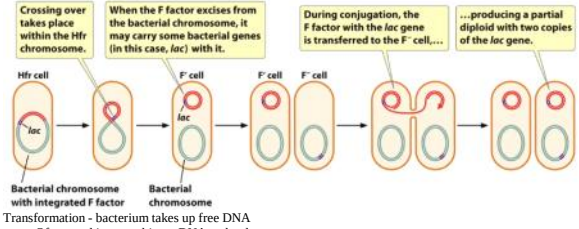
- The microbiome exists within the intestines of human bodies and is essential to health. Roles include:
  - Digestion of food
  - Regulation of immune system
  - Protect against other bacteria that cause disease
  - Produce vitamins (including B, B12, thiamine, riboflavin, and K)
- Microbiome consists mainly of bacteria, fungi, protozoa, and viruses.
  - Protozoa - single celled eukaryotes (e.g. Tetrahymena, Leishmania)
  - Fungi - heterotrophs (cannot synthesize their own food) - obtain nutrients from decomposing organic matter, reproduce with spores, and most have haploid or diploid stages
  - Viruses - infectious agent of small size and simple composition which can only multiply in the cells of animals, plants, or bacteria.
    - Viruses are composed of nucleic acid surrounding a protein coat
    - Reproduce in cells using cellular DNA replication machinery - this was exploited to study the process of DNA replication.
    - Produce viral proteins that allow for assembly then new round of protein replication, recurring cycle of replication
  - Different microbes colonize different parts of our bodies
- Laboratory growth of microbiome
  - Although some bacteria are easy to grow in a lab setting, this is only a small subset of bacteria. In fact, many bacteria cannot be cultured in a lab setting. For these bacteria, only DNA sequencing is able to reveal the bacteria.
  - Human Microbiome Project (HMP) - Studies humans as "supraorganisms," or "ecosystems" composed of human and microbial cells.
    - Goal of project is to create a "metagenome," or a genome of the whole human ecosystem including different microbes that colonize human cell types.
    - Large number of bacterial species occupying human gut first realized after DNA sequencing
    - Most common method is sequencing of the 16S ribosomal RNA (rRNA) subunit
      - First, use PCR to amplify the rDNA - despite the fact that we don't know the sequence, we can use PCR amplification because we can differentiate between conserved and hypervariable DNA sequences, thereby creating **universal primers** that match the sequences of different bacterial taxa.
        - Conserved DNA sequences are DNA sequences that have been evolutionarily selected in favor of due to natural selection.
        - Hypervariable sequences are sequences that may be very different in number of tandem repeats within an organism
      - The universal primers will amplify the variable regions in between the primers, allowing for the identification of the unique species.
      - We require this PCR step as without it, we would not be left with enough DNA to obtain a DNA sequence. This is an excellent advantage of PCR, as it can obtain sequences from very tiny amounts of DNA.
    - Alternative method includes the sequencing of all DNA within a sample without PCR:
      - Ex.) When given a fecal sample, the DNA is first extracted using a high-throughput NGS method. The computational pipelines assemble the DNA into individual genomes.

Bacterial Genomes

- Bacterial genomes are in the form of double stranded circular DNA.
- Plasmids are small circular structures of double stranded DNA
  - Maintained in cells via independent origins of replication (*ori*)
  - May be present in multiple (random number) of copies per cell.
  - Bacterial equivalent of genes that enhance growth
  - Plasmids often confer antibiotic resistance (R plasmids)
    - Used in recombinant DNA technology
- Exchange of genetic material in bacteria can be done three ways:
  - Conjugation - direct transfer of DNA from one bacterium to another
  - Transformation - bacterium takes up "absorbs" free DNA
  - Transduction - bacterial viruses take DNA from one bacterial host to another

Bacterial Gene Transfer

- Conjugation - direct transfer of DNA from one bacterium to another (bacterial mating)
  - One way transfer of DNA from the donor cell to the recipient cell
  - Nicked site in DNA at oriT (origin of transfer) is the site of transfer initiation. Since all replication begins at the oriT site, the transfer is linear and directional.
  - When the recipient cell gets the transferred DNA of the donor cell, recombination may or may not occur.
  - Requires the F plasmid, which encodes genes required for the formation of the sex pilus
    - The F+ donor plasmid donates the F factor to an F- cell as detailed by the diagram to the right. This process results in newly created F+ plasmids so that conjugation may next take place.
  - Some bacterial cells are HFR strains (High frequency recombination) strains of bacteria. In these bacteria, the newly received F plasmid is integrated into the bacterial chromosome through crossing over.
    - The difference is that these HFR strains will additionally transfer portions of their original genome to the recipient plasmid.
    - This can substantially alter the genotype of the recipient cell. A detailed diagram of the transfer of HFR genome to F- plasmids are detailed in the diagram to the right.
  - Bacterial cells may work in the opposite way to generate F' cells, or cells that excise the F genome from the HFR cell's plasmid.
    - Can also result in genetic variation, as the F factor may be excised inaccurately, bringing parts of the original HFR bacteria's plasmid with it, conferring a new genotype to the recipient plasmid. For example in the following diagram, the lac gene is conferred to an F- cell with the lac gene.



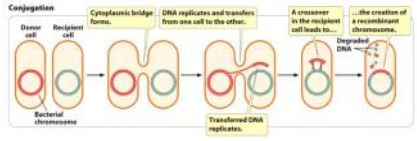
- Transformation - bacterium takes up free DNA
  - Often used in recombinant DNA technology:
    - Vector DNA is constructed using PCR and / or restriction enzymes with DNA ligase
    - Reaction mixture is transformed into bacterial cells, and bacterial cells replicate with the DNA, allowing for large amounts of the DNA to be copied.
    - Typically in recombinant DNA technology, we would want to transform a plasmid vector which replicates via its own origin of replication (ori)
  - Bacteria that can accept naked DNA are "competent" bacteria.
  - Recombination occurs between the host plasmid and the DNA that is taken up, resulting in a plasmid with the foreign DNA.
- Transduction - viral introduction of foreign DNA
  - Bacteriophages (type of virus) carry DNA from one bacteria to another
  - Recombination occurs with the chromosome, resulting in the transferring of new genes or alleles.
  - The process of viral transduction is shown to the right

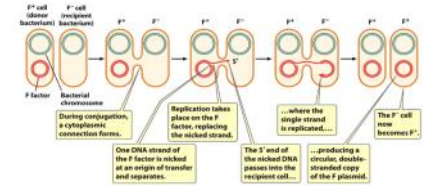
The Hershey-Chase Experiment

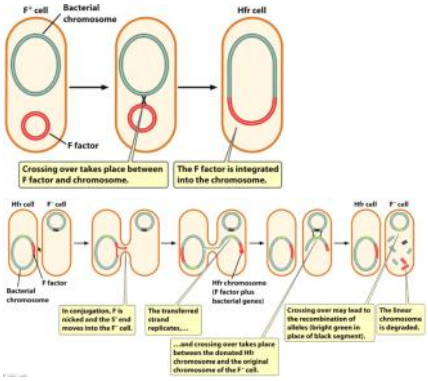
- Sought to answer the question of what carries genetic information from one cell to another. There were two hypotheses - DNA or protein.
- Grew bacteria in two media - one with S-35 labeled amino acids and bacteriophages, and a separate bacterial colony with P-32-dNTPs. Radioactivity would then be incorporated into the phage DNA.
  - If proteins were the transforming principle, the bacterial cells would end up being radioactive in the S-35 medium, as the bacteriophages would introduce the protein through their protein cap.
  - If DNA was the transforming principle, the bacterial cells would be radioactive in the P-32-dNTP media, as the bacteriophages would transform the bacteria.

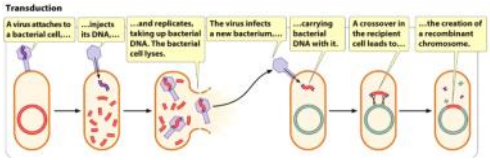
Important People to Remember

- Alfred Hershey and Martha Chase - Hershey-Chase experiment: Transduction proved that DNA is the genetic material (information carrying molecule). Shared Nobel prize with Max Delbruck and Salvador Luria.
- Baltimore and Temin - found that viruses with genomes consisting of RNA can be inserted into the DNA of host cells. (Reverse transcriptase)
- Jacob and Monod - *lac* operon model
- Francis Collins - current NIH director, among leaders in CFTR gene identification
- Craig Venter - advocate of whole genome shotgun sequencing









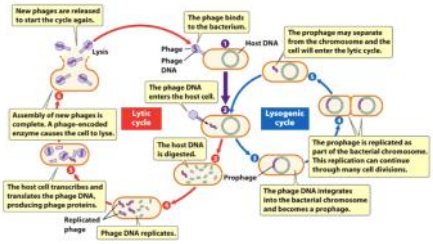
- Found that the bacterial cells in the P-32-dNTP were radioactive, which drew the conclusion that in fact the genetic material was DNA, and not proteins.

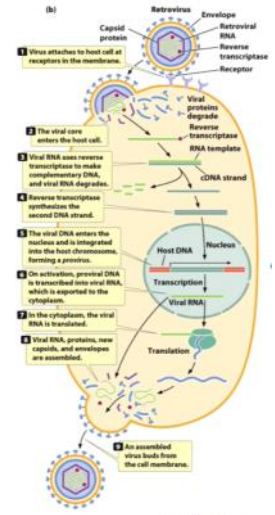
Horizontal Gene Transfer

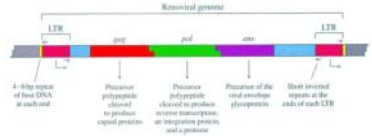
- Genes may be passed between individual members of different species by nonreproductive mechanisms. This form of gene transfer is known as horizontal gene transfer, or lateral gene transfer.
- Highly common method of gene transference between bacteria.
  - Genes that confer antibiotic resistance are of particular concern - the bacteria that have antibiotic resistant genes have a selective advantage over those who don't, resulting in rapid proliferation of antibiotic-resistant genomes throughout bacteria.
  - Antibiotic-resistance genes are typically found on R plasmids which occur naturally. Since the widespread use of antibiotics in the past 60 years, the R plasmid has rapidly evolved.
  - The transfer of R plasmids is not restricted to bacteria of the same or even related species.
- Recently, researchers have discovered the existence of horizontal gene transfer between bacterial and human genomes. This is not a mere contamination of the sample of eukaryotic genes, but an actual representation of the bacterial genome in the eukaryotic gene.

Viruses

- Viruses are defined as replicating structures of DNA and RNA with a protein coat.
- The question of whether viruses are alive or not are complex:
  - They are parasites which rely on a host to reproduce and replicate themselves, and lack enzymatic activities required for life. (Nonliving)
  - Viruses however may originate genes or at least move them around. (Living)
  - Viruses evolve, sometimes rapidly. (Living)
- A bacteriophage is a certain kind of virus, and there are two types of bacteriophages
  - **Virulent phages** reproduce through the **lytic cycle** and always kills the host cell
  - **Temperate** phages reproduce through the **lysogenic cycle**, by integrate their DNA into the bacterial chromosome, where it remains as an inactive prophage, meaning that it is not an active phage, but has the potential to become an active phage.
  - A comparison of the two cycles are shown in the diagram to the right.
  - Rapid assay of lysis in bacteria:
    - We can assess the phage concentration (viral titer) in a colony of bacteria by infecting a lawn of bacteria with bacteriophages. Each plaque (dot) on the original bacterial lawn represents the site of an original phage infection. More plaques indicate a greater concentration of bacteriophages.
- Animal and plant viruses
  - Either DNA or RNA can be the genetic material.
    - Can be single or double stranded genome
    - Specific and different host ranges result in wide variety of animal and plant viruses.
    - DNA viruses replicate DNA in the host's nucleus.
    - RNA viruses replicate RNA in the host's cytoplasm.
- Retroviruses
  - First discovered by Baltimore and Temin, discovered that viruses with genomes consisting of RNA can be inserted into the DNA of host cells. In other words, the RNA in RNA viruses are copied into DNA through an enzyme known as reverse transcriptase. This upended the central dogma of molecular biology.
  - Mechanism for retroviral integration of genetic material:
    1.) Virus attaches to host cell at receptors in the membrane.
    2.) Viral core enters the host cell
    3.) Reverse transcriptase enzyme is used in conjunction with the viral DNA to create a complementary DNA strand. In this stage, the original RNA is degraded. Note that the DNA here would have the same sequence as the original RNA strand, except of course the replacement of the U's with T's.
    4.) Reverse transcriptase continues to synthesize the second DNA strand. This one is complementary.
    5.) The viral DNA enters the nucleus and is integrated into the host DNA through integrase, becoming a **provirus**, or a virus that has integrated itself into the host chromosome.
    6.) When the proviral DNA is replicated, it is then transcribed into viral RNA, which is then exported into the cytoplasm, where the viral RNA is translated.
    7.) The viral RNA is translated in the cytoplasm, forming viral RNA, and other components of the virus.
    8.) The virus buds from the cell membrane, and the process begins again for a new cell.
  - The process of retroviral replication is detailed in the diagram to the right.
  - The genome of the retrovirus
    - Retroviral genomes encode proteins that are essential for virus proteins.
    - We will focus on three specific gene sections: *gag*, *pol*, and *env*.
      - *gag* is responsible for code proteins, or proteins that surround the nucleic acid and make up the physical structure of the virus
      - *pol* encodes the essential reverse transcriptase and integrase enzymes.
      - *env* encodes surface proteins (glycoproteins)
    - At the ends of retroviral genomes, LTRs (long terminal repeats) are present. These LTRs contain very strong transcriptional regulatory sequences, to encourage their replication in eukaryotic cells.
    - Retroviruses can remain silent in the genome, and may change over time. If the mutation makes the virus more infectious and virulent, those traits in the genome are selected for in evolution.
    - Reverse transcriptase is not as accurate as DNA polymerases, because RTs lack proofreading activity. This in turn increases mutation rate of viruses, and increases the challenge of finding a vaccine.
- Zoonotic Viruses
  - Viruses that "jump" from animals to humans - usually after a mutation in the virus.







Lecture 14 - Control of Gene Expression

- All cells in our body contain the same DNA information. However, the genes which the DNA make up are expressed, or turned on, in different cells at different times and under different conditions.

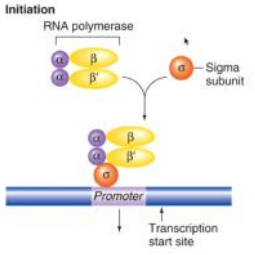
Prokaryotic Gene Regulation

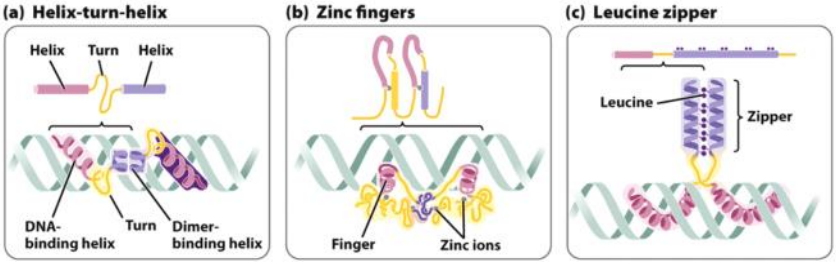
- Originally, DNA exists in a compact string, bound by nucleosomes in higher order chromatin. In order for replication to occur, The DNA must be relaxed and its structure altered for transcription, mRNA processing, and RNA stability translation, resulting in a protein, which is then modified by posttranslational modifications to activate it.
- The most control of gene regulation would, naturally, occur at the level of transcription initiation, the beginning of the formation of a working protein. Recall also that some RNAs (noncoding RNAs) may be active without translation.
- Recall the initiation of transcription:
  - RNA polymerase contains multiple subunits
  - "Core polymerase" contains alpha and beta subunits
  - "Holoenzyme" includes alp ha, beta, and sigma subunits.
  - The sigma subunit binds to the promoter, bringing the RNA polymerase enzyme to the correct position on the DNA, allowing for specific transcription to take place.
    - Recall that in transcription, upstream refers to the 5' end whereas downstream is the 3' end.
    - The sigma subunit is the initiation factor, and binds specific sequences upstream of the TSS (transcription start site).
    - Without this sigma subunit, random transcription will take place all across the DNA all over the genome, which would be a catastrophic event for the organism.
    - The sigma subunit has loose DNA binding specificity, (unlike restriction enzymes, which will cleave at exactly the same DNA sequence, every time) and may recognize one or more binding spots in the TSS.
    - Mutations may decrease affinity for sigma subunit binding. Often this won't completely stop binding, but destabilize the structure of the subunit. This then results in decreased transcription of the DNA.
      - However, the opposite may also be true - mutations may increase affinity for binding, resulting in greater transcription of the DNA.
      - Therefore, by extension, the strength of a promoter region in DNA is determined by how frequently RNA polymerase initiates transcription from it.
- DNA Binding Proteins
  - DNA binding proteins have multiple domains, known as **DNA binding domains** which are 60-90 amino acid long chains that are responsible for the binding of the protein to the DNA through hydrogen bonds.
  - The protein interaction domains may also recruit other proteins, such as the alpha and beta subunits.
  - Together, these **transcription factors** regulate gene expression
  - DNA binding proteins are grouped together in families. Each of these families have their own properties and bind to different consensus sequences, allowing them to regulate different kinds of genes. Examples of these families include:
    - Helix-turn-helix family
    - Zinc fingers family
    - Leucine zipper family

Operon Theory

- Jacob Monod and Francois Jacob were working on the analysis of the biochemical problem of enzyme induction, particularly of sugar utilization in bacteria, when they devised the *lac* operon model.
- The operon theory proposed that a single signal can simultaneously regulate the expression of several clustered genes that are involved in the same process.
  - The genes are transcribed together, and therefore by that notion are co-regulated.
- In prokaryotic organisms, the regulation of gene expression allows for the cell to modulate or adjust to changes within its environment. In this case, bacteria used gene expression to react to nutritional deprivation or alteration to optimize growth within favorable conditions.
  - Bacteria prefer glucose over lactose, but if glucose was not available and lactose was, the bacteria could also use the lactose for energy.
- Bacteria requires two enzymes to utilize lactose - permease and b-galactosidase.
  - Permease allows for the lactose to enter the cell
  - b-galactosidase breaks up the disaccharide lactose into two monosaccharides, galactose and glucose.
  - This was determined as certain mutants were isolated that did not utilize lactose. The mutations then were complemented, and mapped to different genes.

The *lac* Operon

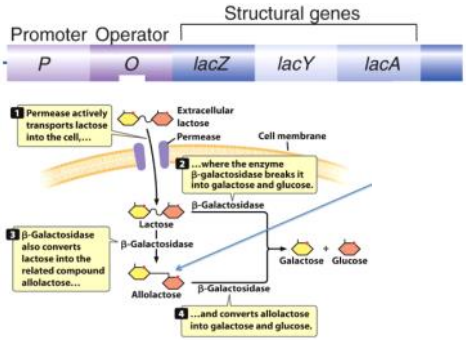


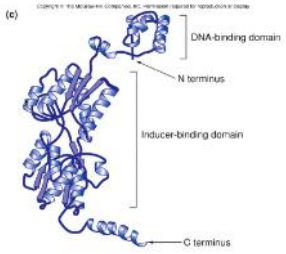


- Bacteria requires two enzymes to utilize lactose - permease and b-galactosidase.
  - Permease allows for the lactose to enter the cell
  - b-galactosidase breaks up the disaccharide lactose into two monosaccharides, galactose and glucose.
  - This was determined as certain mutants were isolated that did not utilize lactose. The mutations then were complemented, and mapped to different genes.

The *lac* Operon

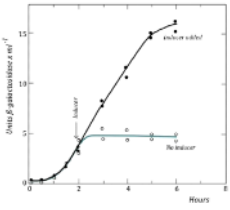
- Expression of the gene to synthesize lactose is usually off, but can be turned on by a signal.
- lac - permease and b-gal is typically at a low concentration in most cells. However, an addition of lactose to medium results in a 1000 fold increase in this concentration.
- Induction is a process by which a specific molecule stimulates synthesis of a given protein.
- The inducer is the molecule which is responsible for the induction event. In the case of the lac operon, the inducer is allolactose, synthesized by b-galactosidase. Refer to the diagram on the right.
  - The structure of the repressor protein is also given to the right. Note that the DNA-binding domain is different from the inducer-binding domain. This means that a mutation in the DNA-binding domain may not affect the binding of the inducer to the repressor and vice versa.
- There are 3 structural genes in the operon:
  - lacZ encodes for b-galactosidase
  - lacY encodes for lac permease
  - lacA encodes for thiogalactoside transacetylase (another gene)
- There are two cis-regulatory elements in the lac operon -
  - Promoter: DNA sequence that signals where RNA polymerase will begin transcription
  - Operator: DNA sequence near promoter that can bind a repressor protein.
  - Cis-regulatory means that they regulate the expression of the structural genes which come after it on the same side of the DNA molecule.
- In contrast, a trans-acting factor is a protein that may have been expressed anywhere which binds to cis-regulatory elements. In the case of the lac operon, that trans-acting factor is the repressor protein, which attaches to the corresponding site on the operator. This prevents RNA polymerase from attaching to the promoter region. The
  - The binding of the repressor to the operator keeps the expression of the gene "off". Most of the time, the repressor protein is bound to the operator, resulting in an operon being off.
  - The repressor is coded outside of the lac operon, in the regulatory gene known as lacI.
  - In the presence of lactose, the repressor is unable to attach to the operator due to a conformational change from the attachment of a product of lactose, allolactose, to the repressor, thereby releasing it from the operator. With the operator now unhindered, RNA polymerase can bind to the promoter region and start transcribing the structural genes of the lac operon, resulting in the production of permease, b-galactosidase, and transacetylase, allowing for the bacterial cell to use the lactose it finds in the environment.
- In summary, cis is on the same side and trans has effects across.
- The lac operon is a "primed" system. That is, there is no need to transcribe and translate a new gene for the induction of the lac gene. Everything is ready to go as soon as the repressor is no longer attached to the DNA.
- In summary,
  - Structural genes - encoding proteins
  - Regulatory genes - encoding products that interact with other sequences and affect the transcription and translation of those sequences. These products must be "trans-acting" - they act on DNA sequences anywhere inside of the cell.
  - Cis-regulatory elements - DNA sequences that are not transcribed but play a role in the expression of neighboring or proximal genes.





The Discovery of the *lac* operon and its model

- Jacob and Monod isolated mutant bacteria that could not synthesize glucose from lactose (could not grow in lactose)
  - Identification of genes necessary for growth on lactose is revealed by the failure of those bacteria to grow when given only lactose as sugar. If they isolated a mutant that could not synthesize glucose from lactose, that mutant gene would be responsible for galactose utilization.
- To make sure that the mutants grew, they plated the mutant bacteria on a plate of glucose to allow for growth, then replica plated the mutant colonies onto the lactose to test if the colony could grow.
  - Once they determined which of the mutants couldn't grow in lactose, they isolated the colony from the glucose plate to maintain and study the mutant.
- Jacob and Monod determined that there was a series of *lac* mutants that could not synthesize glucose from lactose. Using complementation tests, Jacob and Monod identified three genes, lacZ, lacY, and lacA.
  - They also figured out through genetic mapping that these three genes are very close to each other
- However, Jacob and Monod also found several different mutants - one series of mutants which always produced b-galactosidase and permease was known as a **constitutive mutant**, or a mutant which always expressed those genes, even in the absence of lactose. This mutant was mapped to a different gene, lacI.
  - They concluded then, that lacI must encode for a repressor. When the lacI gene was mutated, the repressor could not have been coded for,
- Assays allowed for mutant bacteria to be screened for in lacZ.
  - lacZ encodes for the enzyme b-galactosidase.
  - OPNG is a colorimetric and spectrophotometric substrate for the detection of b-galactosidase activity. This compound is normally colorless. However, in the presence of b-galactosidase, it will hydrolyze the OPNG to form a yellow color. The darker the shade of yellow, the greater the concentration of b-galactosidase.
  - Another substrate is X-gal, which similarly generates a blue color upon cleavage by b-galactosidase
- Bacterial conjugation may be used to determine the roles of different genes.
  - Bacterial conjugation transfers the genes of interest, with sets of wild type and mutant bacteria.



The PaJaMo experiment (Pardee, Jacob, and Monod)

- Involved lacI- and lacZ- bacteria, loss of function mutations involving the creation of repressors and b-galactosidase.
  - Remember that lacI encodes for the repressor and lacZ the b-galactosidase.
  - In this scenario, no b-galactosidase would be synthesized, as lacZ is epistatic to lacI.
- In the experiment itself, they started with lacI- and lacZ- bacteria, in which a plasmid of lacI+ and lacZ+ was transferred into the double mutant.
  - As soon as the lacI+ and lacZ+ were added, heavy concentrations of b-galactosidase were detected with the colorimetric assays.
  - Over time, the synthesis of b-galactosidase decreased. (Refer to the graph to the right)
    - This is due to the synthesis of the repressor, which bound to the operator preventing the further synthesis of b-galactosidase, as shown by the green line.
    - With the addition of the inducer (allolactose), the concentration of b-galactosidase gradually increased, as the binding of the inducer to the repressor caused a conformational change which prevented the repressor from binding to the operator, allowing for the lacZ gene to be expressed.
  - This experiment allowed for further conclusions of effects on mutant genes in the operon's structure.
    - Mutants of the operator, for example, could prevent the repressor from recognizing the operator's sequence and subsequently, result in the constitutive (always) synthesis of b-galactosidase and permease. The mutated operator is denoted as O^C.
      - The mutant operator O acts in cis, which means that they only affect the neighboring genes.
    - Mutants of the repressor to form superrepressor mutants, denoted as $lacI^S$.
      - These will always repress the operator, regardless of whether or not there is an inducer. Thus, b-galactosidase is never synthesized as this gene is unable to be expressed.
  - Transcription factors of genes were determined to be modular, composed of independent protein domains with independent functions.
    - Typically, when one domain binds to DNA, it is known as the DNA binding domain.
      - Recall the three examples of DNA binding domains, helix-turn-helix family, zinc fingers family, and the leucine zipper family.
      - The DNA binding domains have separate domains to interact with protein partners or regulators.

Allostery

- Allostery is defined as the process by which proteins (mainly) transmit the effect of binding at one site to another functional site on the protein, allowing for the regulation of activity. These proteins are known as allosteric proteins.
  - The conformational change is reversible, in that the repressor normally binds to the operon, but an inducer may change the conformation of R so that it no longer binds to O.

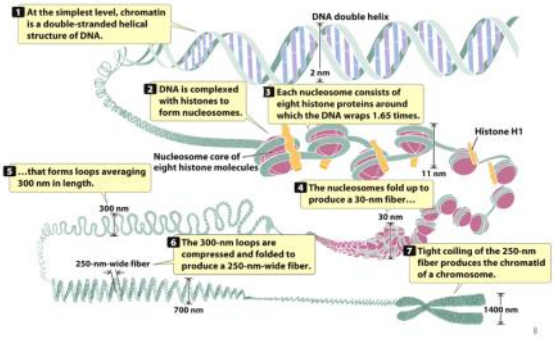
Lecture 15 - Eukaryotic Gene Regulation I

Introduction

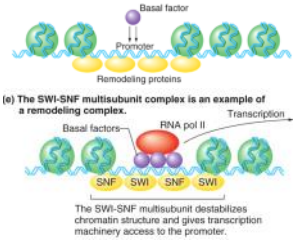
- Monocistronic - One gene is coded per one mRNA
- Polycistronic - Many genes, sometimes all, on one mRNA.
  - In eukaryotes, genes are monocistronically regulated, whereas in prokaryotes, genes are polycistronic.
  - The polycistronic nature of gene regulation in prokaryotes allows for a coordinated expression of genes involved in common pathways.
- In eukaryotes, DNA must first unwind from the histone proteins before transcription.
- Transcription and translation are separated in time and space for eukaryotes
  - Transcription takes place in the nucleus, whereas translation occurs in the cytoplasm.
- In contrast, transcription and translation are simultaneous processes in prokaryotes

The Role of Chromatin in Gene Regulation

- Chromatin has a general repressive effect on transcription.
  - Chromatin reduces binding to basal factors and RNA pol II to very low levels.
  - Eukaryotic cells keep transcription off through using chromatin
- Structure of Chromatin



- Changes in chromatin structure affect the expression of genes
  - DNase I hypersensitive sites show more open chromatin configurations.
    - DNase hypersensitive sites are regions that are freed of histones. These sites are usually found at the 5' end of the DNA, or even before the transcription start site.
    - The existence of hypersensitive sites was originally discovered with the transcription of globin genes in chick embryos.
  - Histone modification can also modulate chromatin states.
    - For example, the addition of methyl groups and acetyl groups to the histone proteins
- Chromatin remodeling allows access to transcription factors.
  - Chromatin remodeling complex can bind to DNA (a process which requires energy) and reposition the nucleosomes, which exposes a transcription factor binding site.
  - Chromatin remodeling proteins are multi-subunit protein complexes which allow for transcription factors to bind to DNA to begin transcription by exposing DNA at the promoter by removing nucleosomes.
  - A common example of a chromatin remodeling protein is the yeast chromatin remodeling protein, known as the SWI-SNF multisubunit complex (as seen to the right)
  - Chromatin remodeling proteins are targeted/recruited by specific regions by DNA binding transcription factors
  - Complexes include and/or interact with enzymes that modify histones



The evidence of DNase I hypersensitive sites and globin genes

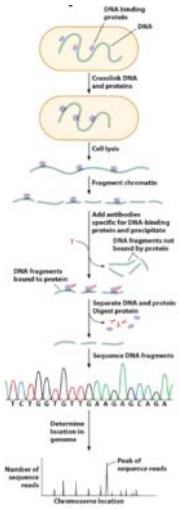
- Testable question: was chromatin structure altered in transcription?
- Researchers tested DNA's sensitivity on different tissues and at different times in development for embryonic chicken cells. Adult chicken DNA contains one section of embryonic globin genes and one section of adult globin genes. The tissues tested included:
  - Erythroblasts in the first 24 hours
    - Found that in this stage, because it is before hemoglobin synthesis, none of the globin genes are sensitive to DNase1 digestion.
  - Erythroblasts in the first 5 days
    - After globin synthesis has begun, all genes were sensitive to DNase1, but the embryonic genes were the most sensitive.
  - Erythroblasts in the first 14 days
    - At this point, only the adult hemoglobin is expressed. Unsurprisingly, the adult genes were the most sensitive and the embryonic genes became insensitive to DNase.
  - Brain cells throughout development
    - Since the brain does not produce globin, the globin genes in the brain remained insensitive and were never expressed throughout development.
- Conclusion: The sensitivity of DNA to digestion is correlated to gene expression, suggesting that chromatin structure changes in the course of transcription.

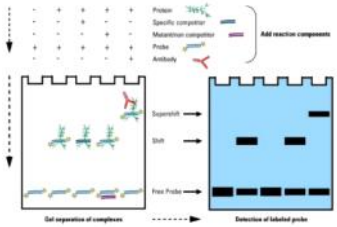
Histone Modifications

- Addition of methyl groups
  - Addition of methyl mediated by an enzyme known as histone methyltransferases.
  - Removal of methyl mediated by an enzyme known as histone demethylases.
  - Addition of a methyl group TO THE HISTONE can either activate or repress transcription depending on the site of methylation.
  - Addition of a methyl group TO DNA has a general repressive effect on transcription
    - The methylation of DNA typically takes place on cytosine bases adjacent to the guanine nucleotides (CpG)-CpG islands. HDACs then bind to the CpG islands, which results in the closing of chromatin and the repression of transcription.
- Addition of acetyl groups
  - Adding an acetyl group to the histones weaken affinity for the DNA phosphate backbone by removing the positive charge on histones.
  - Histone acetyltransferases acetylate lysine residues on histone, promoting transcription factors to bind to the DNA.
  - Histone deacetylases (HDAC) do the reverse, and inhibit transcription.
- Hypercondensed chromatin
  - Transcriptionally silent
  - Also known as heterochromatin, (as opposed to euchromatin, which is "normal" chromatin) this type of chromatin is inaccessible for transcription.
  - Common example of hypercondensed chromatin exists at the centromere.
  - Hypercondensed chromatin may result from epigenetic effects - example includes Barr Bodies (X inactivation)

What is the evidence that a particular protein binds to a certain DNA sequence?

- Genetic studies revealed protein-DNA interactions
- Biochemical methods also were used to test genetic predictions.
- Chromatin immunoprecipitation (ChIP) - method used to identify which proteins are bound to what DNA sequence (see diagram to the right)
  - Must know specific DNA sequence on the DNA to use this method
  - We must also have access to an antibody which specifically targets a protein that is on the sequence.
  - The specific antibodies of the DNA is used for recognition and binding to the protein which is crosslinked to the DNA (always attached to the DNA)
  - For a known protein, a gel shift assay is used to define binding sites.





  - Components:
    - Protein - protein of interests
    - Probe - labeled DNA fragment which we think the protein binds to (typically labeled with P32)
    - Competitor DNA - unlabeled DNA which has the same sequence
    - Mutant / non-competitor DNA - unlabeled DNA which has a different sequence
    - Antibody - attaches to the protein of interest
  - Notice in the diagram that the middle lane does not have an additional bar as the specific competitor is not marked radioactively.

Eukaryotic gene regulation

- Recall from previous lectures the general structure of a gene, including a promoter region which is bound by a transcription apparatus, proteins which control and carry out transcription.
- The promoter determines the transcription start site and positions RNA polymerase - note also that RNA polymerase does not require a primer, whereas DNA polymerase needs a primer to begin transcription
- The initiation of transcription is regulated by DNA-binding transcription factors and co-regulators - gene regulation controls cellular differentiation into hundreds of specialized cell types in multicellular organisms.
  - Note that this is similar to prokaryotic gene regulation in the case of the trp operon and the lac operon - the former responds to the concentration of trp in the cell whereas the latter responds to a presence of lactose in the environment surrounding the cell.
- Similar to prokaryotes, eukaryotes also have a promoter region to initiate gene expression.
- However, typically, eukaryotes have enhancer regions on genes which are unique to eukaryotes and are not found in

prokaryotic organisms.

Enhancers

- First found in eukaryotic viruses, enhancers can work in either direction (forwards or backwards), and are cis-acting sequences of DNA which can encourage the likelihood of transcription.
- Enhancers determine the level of gene expression and control cell type specificity
- Can be located any distance in the gene site. That is, they can be upstream, downstream, or even within the gene
- Reporter gene constructs
  - Reporter genes are genes which researchers use to attach to a regulatory sequence of another gene in bacteria, cell culture, animals, or plants.
  - Allowed for researchers to determine the way enhancers worked
  - Involved the replacement of the coding region with an innocuous gene encoding an easily detectable product using recombinant DNA technology, the insertion of the modified genome into cells of organisms, and the production of assays to organize data. A common gene which is inserted is lacZ, of which the beta-galactosidase produced is easily monitored by OPNG or X-gal.
  - The specific region at which the enhancer is found can also help to determine where the actual enhancer gene is. Refer to the diagram to the right - different segments of DNA are tested to see if they stain X-gal.
    - Researchers also subdivided DNA into regions which identified modules that direct expression in different cell types, as there may be different enhancers for different cells in the body (see example with central nervous system to the right)
  - Reporter gene constructs add DNA encoding putative enhancers which tests for enhancer function by observation of gene expression levels
  - In these reporter gene construct experiments, researchers were able to determine that enhancers functioned in a position and orientation independent fashion.
  - lacZ reporter constructs are used to identify enhancers in transgenic mice. Specific enhancers direct expression of lacZ in specific regions of the embryo.
  - Reporter genes may also be used to identify cis-acting elements and trans-acting factors
    - This may be done by constructing a recombinant DNA molecule which has an enhancer sequence fused to a reporter gene (such as GFP), generating a transgenic organism with the recombinant DNA in its genome
    - Mutations in a gene encoding for an activator would reduce the expression of the reporter
    - Likewise, a mutation in a gene encoding for a repressor would increase the expression of the reporter.
- Different core-promoters may provide specificity for different enhancers
  - Core promoters bind to the basal transcription machinery
  - The most famous example of a motif in a core promoter is the TATA box.
  - Some enhancers show preferences for different core promoters - this was determined through an experiment of placing two different reporter genes (such as RFP and GFP) using recombinant DNA technology.
  - Analysis of enhancers was also carried out with the entire genome in a recent experiment in 2015
    - Found that housekeeping genes (genes expressed in every single cell, necessary for cell viability: e.g. ribosomes, DNA polymerase) and developmental genes (genes that are expressed in specific cell types as things changes during development e.g. the HOX gene)
    - Different enhancers interacted with different promoters
- Example of cell-type specific gene expression: Insulin
  - Insulin is expressed exclusively in the pancreas, and not the brain. This is due to the presence of beta-cell specific enhancers found within the beta cells of the pancreas.
- Insulators
  - Insulators are another component which helps to explain specificity between an enhancer and a gene.
  - Insulators are DNA sequences located between an enhancer and a promoter which block access to the promoter (see diagram to the right)
    - Note that enhancers do not always interact with the promoter they are the closest to.
  - It is also believed that insulators organize genomic DNA into loops by their binding to CTCF. Enhancers therefore may only activate promoters which are located in the same loop.
    - This blocks enhancers from "cross-regulating" in other loops

Modularity of enhancers

- Modularity of enhancers provides an explanation for pleiotropy- for example, Pax6 is expressed in the embryonic neural tube, whereas Pax6 is expressed in the eye.
  - Recall that pleiotropy is when genes act at different places in different times, or when one gene encodes for or influences two or more phenotypic effects.
- Independent enhancers direct expression in different tissues.
- Eukaryotic genes also have modular cis-regulatory elements
  - Different enhancers function in different cell types to regulate expression, known as tissue specific enhancers
  - Each enhancer functions independently of the other enhancers, meaning that regulation is modular - in other words, separate cis-acting sequences control expression in different tissues.
  - Enhancers are independent of each other, and each requires the same promoter.
  - Enhancers are bound by sequence-specific DNA binding proteins that are often expressed in specific cell types. This is what allows for specific things to be expressed in different cells.
- Looping model explanation of enhancer proteins
  - Since the enhancer sequence is brought to the mediator by transcriptional activator proteins, forming a loop, it does not matter whether or not the enhancer sequence is before the transcribed DNA, after the transcribed DNA, or even within the transcribed DNA itself.

Lecture 16 - Eukaryotic Gene Regulation II

DNA Binding Transcription Factors

- Basal Factor (Core promoter region) - contains basal transcription machinery which is a complex of many different proteins, of which RNA polymerase II comprises the core of.
  - The TATA box is typically found within this basal factor
  - This basal factor alone is sufficient for basal (low) levels of transcription
  - With the binding of enhancer DNA and an activator protein to the basal factor, transcription rates increase.
- Transcriptional activator proteins - bind to enhancers or promoter-proximal elements on the DNA to stimulate transcription by stimulating or stabilizing assembly of basal transcription apparatus.
  - Stimulate recruitment of basal factors and RNA polymerase II to promoters
  - Recruit coactivators to open chromatin structure by displacing nucleosomes constricting DNA's transcription.
  - DNA binding proteins alone are not enough to activate transcription - they must have an activation domain, which interacts physically with transcription co-activators. This increases the modularity of transcription factors
    - This shows similarity with the lac operon - in the case of the lac operon, the modular part was the repressor, however, in this part, it is the activator protein.
- Some transcription factors have an additional domain which also responds to signals
  - A very well known example of this is the steroid hormone receptor.
    - Without the steroid hormone (SH), the receptor (DNA binding transcription factor) cannot bind to the enhancer site. However, with the addition of an SH, the receptor undergoes an allosteric change in the receptor, resulting in the receptor now being able to bind to the enhancer.
- Enhancer binding proteins may act as either activators or repressors
  - There is competition in binding between repressor and activator proteins for the enhancer sequence of DNA.
  - The main difference between the repressor and activator proteins is that the repressor protein may bind to the same DNA sequence as the activator, but lacks a transcription activation domain to promote transcription.
  - Enhancer binding repressor transcription factors can recruit corepressors to block transcription. These corepressors have two main functions:
    - Prevention of RNA pol II complex from binding to the promoter
    - Modify histones to close chromatin structure with HDACs (see diagram to the right)
- Similarities to lac repressor
  - Eukaryotic transcription factors are modular, similar to the lac repressor - on the DNA binding protein, there is a DNA binding domain which binds to the DNA, and an activation domain which can interact directly with the basal transcription machinery or by the recruitment/binding of coactivators which modulate chromatin.
  - Eukaryotic transcription factors can be modulated by the binding of small molecules
    - Allosteric reversible changes in conformation is a similarity shared between lac operons and ETFs.

Silencers

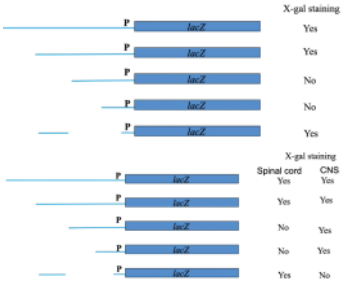
- Silencers are cis-regulatory elements that exclusively interact with repressor proteins to decrease transcription.
- Similarly to enhancers, silencers are distance and orientation independent.
- Note that silencers are not the complete opposite of enhancers, as enhancers may also be repressed by the binding of repressors to the enhancer sequence.

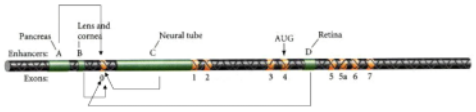
Enhancers and TFs are modular and achieve specificity by interacting combinatorically

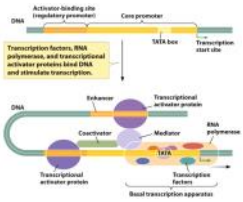
- One gene can have multiple enhancers for it and multiple transcription factors bound to it.
- Independent enhancers direct expression in different tissues of the body - that is to say, changes in one enhancer will not affect the function of other enhancers.
- Cis-regulatory changes therefore can impact expression in one tissue, but not another - specificity is key.
- Pleiotropy often reflects the function of a gene at different times and/or in different tissues
  - Genes only function when and where they are expressed in tissues of an organism
  - Different segments of genes may be re-utilized in different life stages and in different tissues
  - Example of pleiotropy: Sey gene required for neural tube development and eye development
    - Sey/Sey gene is lethal
    - Sey/+ contains small eye, but is not lethal
    - +/+ is wild type
    - Researchers determined that since the offspring of the parent heterozygote cross results in a 2/3 1/3 phenotypic mix, the Sey/Sey gene must be lethal and therefore encode for a gene in addition to eye development gene

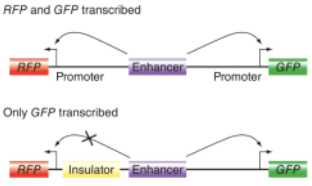
Jun-Fos, Myc-Max, and Dorsal protein of Drosophila

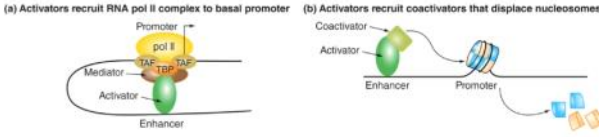
- The coordinate regulation of gene sets (without an operon) is determined by the combinatorial action of transcription factors
- A good example of ON and OFF switches are the Jun and Fos dimers, which forms a leucine zipper on a strand of DNA with one dimerization domain to recognize proteins and another DNA binding site
  - Jun - Jun is a homodimer

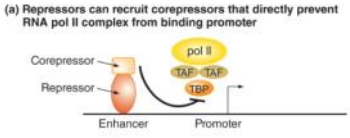


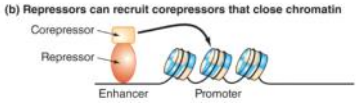


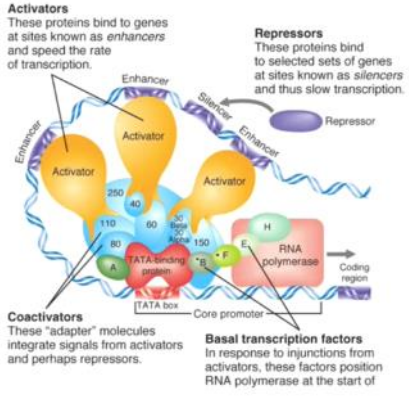


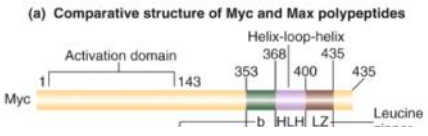


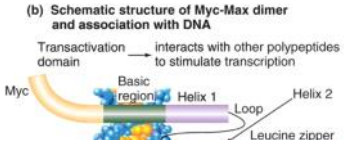












development gene

Jun-Fos, Myc-Max, and Dorsal protein of Drosophila

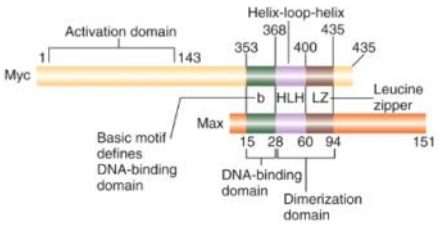
- The coordinate regulation of gene sets (without an operon) is determined by the combinatorial action of transcription factors
- A good example of ON and OFF switches are the Jun and Fos dimers, which forms a leucine zipper on a strand of DNA with one dimerization domain to recognize proteins and another DNA binding site
  - Jun - Jun is a homodimer
  - Jun - Fos is a heterodimer
  - These two different dimers have different affinities for DNA, and may bend it in different directions
- Another example of the on / off capabilities of a leucine zipper is found in the Myc and Max polypeptides, responsible for cell proliferation. Refer to the diagram to the right for a visual representation.
  - Myc and Max share a DNA binding domain. A heterodimer is formed on the LZ (leucine zipper) site, and the two polypeptides may interact with each other.
  - Myc, in addition to the DBD has an activation domain whereas Max does not have this domain.
  - The heterodimer formed by Myc-Max activates transcription
  - However, a Max-Max homodimer will not activate transcription, as recall that Max proteins don't have an activation domain. Therefore, the Max-Max homodimer will simply sit at the enhancer DNA sequence and block transcription - cell proliferation is inhibited.
    - Myc-Max heterodimers typically bind better than Max-Max homodimers - for this reason, if there is a Max-Max homodimer on an enhancer sequence, a Myc protein may be able to kick off one of the Max proteins to switch on gene expression from being off.
- Specific transcription factor example - Dorsal gene of Drosophila
  - The same transcription factor may play different roles in different cells
  - With the Dorsal activator protein alone, transcription of the gene is promoted by Dorsal's recruitment of coactivators that facilitate transcription.
  - However, in some cell types, with the addition of another transcription factor, Dorsal and the new transcription factor together now bind with a corepressor, named Groucho, which blocks gene expression.
  - This process is outlined in the schematic to the right.
- Keep in mind that with the hundreds of thousands of different proteins in the human body, well over a million transcription factors may be formed to express or repress genes.

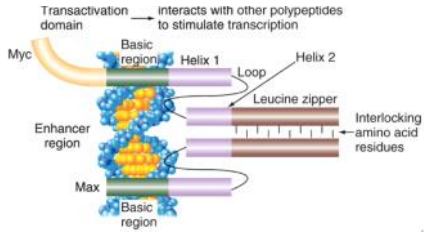
Transcription factors and coordinately regulate SETS of target genes

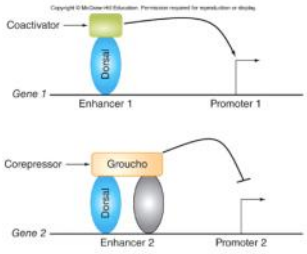
- The same transcription factor may bind to similar enhancers regulating different genes in the same pattern.
- Different sets of transcription factors may bind to different sets of DNA sequences, activating in different patterns.
- Taken together, different transcription factors can regulate many sets of genes in a coordinated fashion
- The coordinate regulation of sets of target genes by different transcription factors is akin to the prokaryotic regulation of genes as is in the example with the lac operon.

Post-transcriptional regulation of RNA

- Degradation of RNA involves the removal of the 5' cap, and shortening of poly A tail. Typically, signals for degradation are found in the 3' UTR.
- RNA interference (RNAi) in the form of Dicer and RISC complexes can also impact gene regulation
  - dsRNA may regulate by recruiting enzymes which alter chromatin structure - for example, a siRNA may attach to the DNA at a certain sequence and recruit a methylating enzyme, which methylates the DNA the siRNA is attached to, and inhibits transcription.







Drosophila Dorsal protein is an activator of gene 1.

The combination of Dorsal and another transcription factor causes Dorsal toact as a repressor.

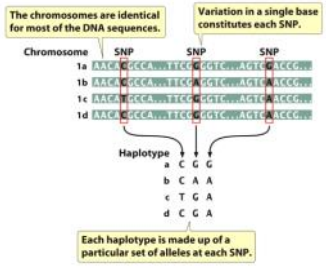
Lecture 17 - Analyzing Genomic Information - Part I

Genomics and Genetic Maps

- Structural Genomics - organization and sequence of genetic information contained within a genome
- Genetic Maps - provide approximate locations for genes relative to the location of other genes using recombination.
  - Limitations include that they are low resolution and lack detail, or may not correspond to the actual physical distances between the genes of interest
- Physical Maps - more accurate location of genes on chromosome, measured in bp, kbp, and mbp, as opposed to centiMorgans, which is the unit of measurement for Genetic Maps.
  - Generating a physical map of genes may be done by analyzing overlapping DNA sequences and linking them together (usually with the use of bioinformatic algorithms), like piecing together a puzzle
  - The longest fragment created with this method is known as a contig, while the longest possible contig would be the complete chromosome.

Genome Sequencing

- First carried out in model organisms of increasing genome size, eventually leading to the human genome project.
- Restriction Mapping
  - Restriction enzymes cut portions of DNA into fragments. The fragments are aligned, much like in the creation of a physical map of a chromosome, and assembled into the gene(s) of interest.
- Whole genome shotgun sequencing
  - Do not use restriction enzymes, but rather, cuts into numerous small overlapping fragments and is cloned in bacteria. Each fragment is then sequenced, and overlaps are used to order the clone DNA fragments, and the entire genome is assembled by computer programs.
  - Relies more heavily on computational power than map-based sequencing.
- Between individual humans, many instances of genetic variation may be seen, with four main categories:
  - SNP - single nucleotide polymorphism
    - Most common
  - DIP - also known as Indels, insertions or deletions
  - SSR / Microsatellite - Simple Sequence Repeat
  - CNV - Copy number variant
    - Rarest
- HapMap Project
  - A Haplotype is the specific set of SNPs and other genetic variants observed on a chromosome. This arose from a mutation in a single individual and spread within a population.
  - Different variants more frequent in different populations due to inbreeding
  - The HapMap project aims to catalogue these haplotypes within distinct populations of humans around the world
  - Linkage Equilibrium vs Linkage Disequilibrium
    - Linkage Equilibrium - The four haplotypes are present at equal frequencies, therefore knowledge of the SNP at site 1 of a gene would provide no additional information for another site, site 2, on the gene.
    - Linkage Disequilibrium - refers to the nonrandom association of alleles for at least two or more loci in a general population. When alleles are in linkage disequilibrium, haplotypes don't occur at expected frequencies, and instead are correlated with other SNPs.
      - Variants of two loci are correlated, and the sequence at site 1 suggests a sequence at site 2.
- Genome Wide Association Studies (GWAS)
  - Search for a linkage between markers and phenotypes in an existing population
  - Compare marker frequencies and genetic diseases or morphological traits



Lecture 17 - Analyzing Genomic Information II

Structural Genomics

- Metagenomics - the sequencing of genomes of entire communities or organisms
  - Ex.) The sequencing of the microbiome of human digestive tract (Human Microbiome Project)
- Synthetic Biology - the creation from scratch of model organisms
- Gene annotation
  - Utilize bioinformatic approaches to identify genes in genomes. This may be based on sequence information such as start / stop codons, splice sites, etc.
  - Note however that automatic annotations are not always correct
    - Ex.) Not all AUG sequences are start codons, not all splice sites have the same consensus sequence
- Predicting functions from related sequences
  - Homologous Genes
    - Genes that are evolutionarily related
  - Orthologous Genes
    - Homologous genes in different species that evolved from the same gene in a common ancestor
  - Paralogous Genes
    - Homologous genes arising via the duplication of a single gene in the same organism
  - Refer to the diagram to the right to see a clear difference
- The prediction of functions of large data sets of genes with known function
  - KEGG pathway analysis - collection of manually drawn pathway maps representing our knowledge on the molecular interaction, reaction, and relation networks for
    - Metabolism
    - Genetic information processing
    - Environmental information processing
    - Cellular processes
    - Organismal systems
    - Human diseases
    - Drug development