Population Genetics

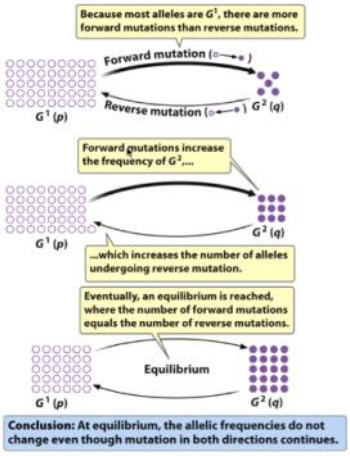
Tuesday, May 12, 2020 3:11 PM

Gene Pool
- The total set of alleles in a population, described by allele frequency
- One individual may carry 2 alleles at a single locus, but a population gene pool may consist of any number of alleles at a single locus.

Hardy Weinberg Law
- How does sexual reproduction influence the gene pool?
- The Hardy Weinberg law is a mathematical model of how reproduction impacts the genotypic and allele frequencies of a population, where p^2, q^2, 2pq represent the homozygous dominant, homozygous recessive, and heterozygous allele pair frequencies, respectively.
- Assumptions for Hardy Weinberg Law to work:
  - Mating within the population must be random
  - The population is not affected by mutation, migration, or natural selection
  - The population is sufficiently large
- If the population devates from the hardy weinberg equilibrium, denoted as the sum of the homozygous dominant and recessive alleles with the heterozygous alleles adding to 1, then this indicates that other evolutionary influences are at work.
- When the frequency of one allele in a population is high, then most of the individuals in the population are homozygous for that allele.
- The greatest frequency possible for heterozygous traits is 50%, or when p = q.

Nonrandom Mating
- Positive assortative mating - a tendency for like individuals to mate
- Negative assortative mating - a tendency for unlike individuals to mate
- Inbreeding
  - The preferential mating between related individuals
  - Causes an increase in the portion of homozugous while simultaneously decreasing the amount of heterozygotes in a population
  - Causes a departure from Hardy-Weinberg equilibrium
  - Measurement of inbreeding:
    - Uses the coefficient 'F', where 0 < F < 1 and 0 indicates random breeding while 1 indicates all alleles being identical by descent (pure inbreeding)
    - Inbreeding increases the homozygotic population at the cost of heterozygotes in the population
    - Can be factoried into the hardy weinberg equilibrium as follows:
      - $f(AA) = p^2 + Fpq$
      - $f(Aa) = 2pq - Fpq$
      - $f(aa) = q^2 + Fpq$
  - The most extreme example of inbreeding is in hermaphrodite species, where selfing (breeding within the same organism) can be common.
    - This is genetically unfavorable, as populations that exclusively breed through selfing will ultimately die off due to effects of inbreeding depression
    - An increase in the frequency of homozygotes for recessive alleles can sometimes result in detrimental alleles being propogated down generations.
  - Inbreeding may leave a species unprepared for changes in environmental conditions that may rely on a diverse set of alleles - a lack of "cryptic variation"
- Outcrossing - the preferential mating between unrelated individuals

The Effect of Mutations on Allelic Frequencies
- Two types of mutations: forward and reverse mutations
  - Forward mutation - much more common, as in a gene pool, the variant of the original allele (G1) is much greater in comparison to the variance of the mutated allele (G2).
  - Reverse mutation - uncommon, due to the rarer frequency of the mutated allele (G2)
  - When equilibrium is reached, the rate of forward mutations must equal the rate of reverse mutations.



The Effect of Migration on Allelic Frequencies
- Migration (or gene flow) has two main effects:
  - Prevents populations from becoming too different from each other (genes flow and mix with each other)
    - Therefore the amount of change in allelic frequency due to migration between populations depends on the difference in allelic frequency and the extent of migration.
    - As migration increases, gene pools become more and more similar to each other (between populations).
  - Increases genetic variation within populations
    - More migration -> greater genetic diversity

The Effect of Evolutionary Forces on Allelic Frequencies
- Genetic Drift
  - Change in gene pool due to random chance
  - The smaller the population, the greater the effect of the drift.
  - Experiment testing genetic drift in drosophila
    - Started with 107 replicate populations where the frequency of an allele was 0.5.
    - Set up crosses for 19 generations
    - Gradually, genetic frequencies became different in different populations.
    - One allele typically is 'fixed' within the population.
    - Some populations become homozygous for one allele.
    - Conclusion: as a result of genetic drift, the allelic frequencies gradually diverged from each other and often became fixed for one allele or the other.
  - Causes of genetic drift
    - Decrease in population size
      - Decrease in the size of a population results in an increase of sampling errors.
    - Founder Effect
      - The establishment of a population from a small number of individuals
      - Limits the number of alleles in a gene pool
      - There is a chance event of founder carrying certain alleles.
    - Bottleneck Effect
      - Drastic reduction in population size results in the remaining individuals being random. Therefore, a population that consisted of a small portion of a certain allele to begin with may end up representing a much greater percentage of the alleles in the pool than before the event.
      - Results in low levels of genetic variation
- Natural Selection
  - Differential reproduction of genotypes results in unequal production of offspring from individuals of one trait over another.
  - Traits providing reproductive advantages results in more offspring and therefore increases over time.
  - Those producing more offspring in a given environment are better adapted to that environment, producing more individuals which are well adapted to the environment.
  - The relative reproductive success of a genotype is represented as 'W', or the fitness of an organism, a value ranging from 0 to 1.
    - The selection coefficient is the relative intensity of selection against a certain genotype (1-W), where W = the number of offspring produced by that genotype/the most prolific genotype. See the example to the right.
    - Directional selection may occur for alleles that are favorable in a population. For example, in the case of A1A1, A1A2, and

Genotypes:	A^1A^1	A^1A^2	A^2A^2
Mean # offspring:	10	5	2

$$W = \frac{\text{\# offspring produced by that genotype}}{\text{the most prolific genotype}}$$

Fitness of A^1A^2 = 5/10 = 0.5

Selection coefficient for A^1A^2 = 1-W = 0.5

A2A2, the three possible combinations for genotypes in a population, a preference for the allele A2 would result in a greater frequency of A2 in the heterozygotes and homozygotes containing the A2 allele and a decrease in A1.

- Results of natural selection include overdominance and underdominance .
  - Overdominance - heterozygote advantage, results in both alleles being maintained
    - ex.) Sickle cell anemia
  - Underdominance - heterozygote disadvantage, may result in the fixation of one allele or another.
- If an allele is lethal, most copies of the lethal allele will be in heterozygotic populations. Therefore, in practice, there is minimal selection against alleles that cause homozygous lethality but have no effect on heterozygotes.

- Artificial Selection
  - ex.) Dogs, descendants of wolves
  - ex.) Many agricultural crops, selected over many years to generate high yielding plants
  - Desired traits are enriched through selective breeding
  - Rate of change high with artificial selection

Summary
Forces influencing population structure within and between populations

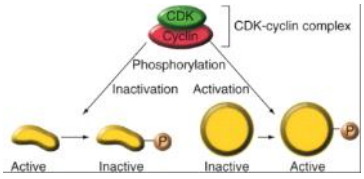
	Within Populations	Between Populations
Increase in genetic variation	• Mutation • Migration • Some types of natural selection	• Mutation • Genetic drift Some types of natural selection
Decrease in genetic variation	• Genetic Drift • Some types of natural selection	• Migration • Some types of natural selection

Cancer Genetics

Tuesday, May 12, 2020 3:26 PM

Introduction
- Cancer is a mutation and results from two different types of mutations: the germline mutation and the somatic mutation.
- Hereditary mutation - Germline mutation
- Sporadic mutation - Somatic mutations

Two Unifying Themes for Cancer Genetics
- Cancer is a collection of diseases
  - Cells proliferate out of control and acquire the ability to spread
  - The genes mutated in cancer are the genes that control cell proliferation.
  - Understanding of cancer requires an understanding of normal circumstances of cell proliferation
    - The control of cell cycles in yeast was elucidated by Nobel Prize winners Leland H. Hartwell, R. Timothy Hunt, and Paul M. Nurse for the discovery of genes regulating the cell cycle.
    - Discovered "Checkpoints" in cells - mechanisms which prevented cells from continuing to the next phase of the cell cycle until the previous phase of the cell cycle has been completed.
      - Premature passage through checkpoints allows for DNA damage to persist.
- Cancer has a different inheritance pattern than other genetic disorders
- Cancer may be a result of an accumulation of multiple mutations in a somatic cell.

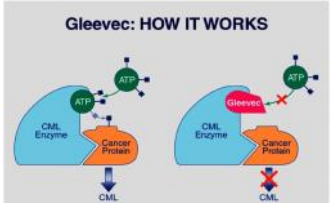


Cell Checkpoints
- Checkpoints in the cell halt progress through the cell cycle if the DNA is damaged.
- There are two main checkpoints in the cell cycle:
  - G2/M transition
  - G1/S transition
- Spindle checkpoint - mitosis is postponed if the chromosome(s) fail to attach to the spindle
  - Prevents M from
- How are cell cycles regulated?
  - Control of cell cycle is done through CDKs, or Cyclin Dependent Protein Kinases
    - Cyclin - Regulated passage of the cell through the cell cycle
      - Cyclin concentrations fluctuate during the cell cycle depending on the specific cycle of the cell.
      - Timing is key - cyclin appears at the proper time to activate the correct set of substrates, and likewise, disappears at the proper time to avoid activating incorrect substrates.
      - Regulation expression of cyclin genes -> transcriptional regulation
    - Cyclins bind to CDKs to determine the substrate specificity of the CDK.
      - Cyclins determine which proteins will be phosphorylated and when
      - Phosphorylation can activate or repress certain activities of different targets. (See diagram)
        - E2F becomes active upon the phosphorylation of the retinoblastoma protein, a tumor suppresion protein., and the cell can continue to the DNA synthesis phase.
        - Dephosphorylation activates a CDK in yeast to promote the transition to mitosis from G2.
  - Regulation of expression of cyclin genes
    - Transcriptional regulation
    - Initiated by growth factors
  - Regulation degradation of cyclin proteins
    - Proteins must be actively removed from a cell or they will continue to function: ubiquitin system actively targets proteins for degradation

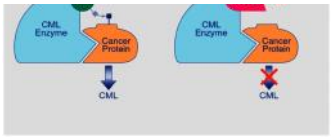
Major Chromosomal Abnormalities in Tumor Cells Results from Loss of Checkpoints
- Checkpoints allow for the cell to ensure that the M phase is not initiated if the DNA is extensively damaged. This allows for time for the cell to fix the DNA.
- Loss of telomeres: cells replicate so fast that telomeres shorten
  - Free ends of chromosomes are "read" as damage that should be repaired by the cell. Random joining of the free ends through NHEJ generates translocations. This can also generate dicentric chromosomes which break at cytokinesis and further cycles of breakage and fusion.
- Mistakes in DNA replication and repair can lead to chromosomal instability (lack of effective checkpoints)
- Homogenously staining regions (HSRs) or small, extrachromosomal pieces of DNA (double minutes).
- Aneuploidy - extra chromosomes, missing chromosomes, chromosome rearrangements
  - Results from random segregation of unattached chromatids
- HSRs and Aneuploidy are features of genomic instability which defines cancer cells.
- In summary, cell division without functional checkpoints causes cell divisions to occur without 'surveillance' of the genome. Cells with DNA damage, abnormal chromosome number, etc. can continue to divide. The resultant cells can then become highly abnormal.

p53 as the guardian of the genome
- DNA damaging agents induce p53, where p53 is a transcription factor.
  - p53 activates the repair of DNA damage and jointly halts progression to S phase of the cell cycle by blocking the phosphorylation of Rb through a CDK inhibitor protein, p21.
  - Promotes cell death if the damage is too extensive to repair.
- Cells with a large amount of unrepaired DNA damage will initiate apoptosis, or programmed cell death.
  - Apoptosis prevents an organism from replication of damaged cells.
- Loss of p53 function permits the proliferation of cells with damage.

Oncogenes
- First cancer-causing genes identified in retroviruses
  - Some retroviruses had deletions of their normal genomes replaced by a cellular gene.
  - Strong promoter sequences in the long terminal repeat region (LTR) of retroviral DNA fused to cellular DNA can result in the activation of a proto-oncogene, leading to high levels of expression of oncogenes.
- Identification of oncogenes through cell transformation as an assay
  - If viruses can carry oncogenes from cell to cell, this can also be done experimentally:
  - Researchers transfected NIH 3T3 mouse cells with fragments of DNA to identify DNA that causes cellular transformation.
- Extracellular signals trigger phosphorylation cascades leading to cell growth
- Ligand binding activates signal transduction pathways which result in the encoding of proteins which stimulate cell division.
  - RAS cascade is a key player in growth control and is an implicator of cancer.
  - When bound to GDP, RAS is inactive. However, upon the binding of a growth factor ligand to the receptor in the cell membrane, the RAS becomes activated with the phosphorylation of GTP, in which MAP kinases activate transcription factors allowing for the proliferation of cells.
  - Mutations in this RAS pathway cause cancer.
- Proto-oncogenes are normal cellular genes involved in cell growth. Oncogenes are gain-of-function mutations in proto-oncogenes that generate excessive or inappropriate activity.
  - In this case, only a single mutant allele is sufficient for an effect.
  - Ways in which proto-oncogenes transform into oncogenes:
    - Mutations leading to extra / unregulated activity can result in extra/unregulated activity.
    - Gain-of-function mutations
      - Amplification of gene copies
      - Point mutations that increase activity
      - Translocation (in the case of retroviruses, insertion near an activating enhancer can cause a proto-oncogene to become an oncogene)
        - Cause the formation of chimeric proteins, or fusion proteins, in which the proteins are created through the joining of two or more genes which originally coded for separate proteins.
        - Example of this is the Philadelphia Chromosome: a fusion of chromosome 9 and 22, resulting in chronic myeloid leukemia (CML). The translocation produces an abnormal tyrosine kinase which is constitutively active, resulting in uncontrollable cell growth. Gleevac therapy is a tyrosine inhibitor, preventing ATP from phosphorylating the target protein attached to the brc/c-abl complex, and stopping CML at the source of the problem. This treatment is found to be 98% effective for patients with CML.



genes which originally coded for separate proteins.
- Example of this is the Philadelphia Chromosome: a fusion of chromosome 9 and 22, resulting in chronic myeloid leukemia (CML). The translocation produces an abnormal tyrosine kinase which is constitutively active, resulting in uncontrollable cell growth. Gleevac therapy is a tyrosine inhibitor, preventing ATP from phosphorylating the target protein attached to the brc/c-abl complex, and stopping CML at the source of the problem. This treatment is found to be 98% effective for patients with CML.
- Another example of genetic treatment of cancer is in the case of Her2, which is a gene responsible for breast cancer. Herceptin is a monoclonal antibody therapy for this form of breast cancer. Her2 receptors in the cell membrane of a tumorous cell dimerize, resulting in a signal transduction pathway which results in uncontrollable cell proliferation. However, Herceptin stops this by preventing the dimerization from occuring through ordering the immune system to attack the cancer cell. With the signal transduction pathway deactivated, the tumor cell is no longer able to replicate uncontrollably.



- Activation by point mutation
  - Typical for RAS proteins
  - Point mutations typically map to specific positions in RAS that decrease GTPase activity, which consequently result in over-active promotion of growth through the uncontrollable production of growth proteins.

Knudson's Two-Hit hypothesis

- First genetic explanation of cancer in 1971 by Dr. Alfred Knudson specifically on retinoblastoma, a heritable cancer of the eye.
- There were two types:
  - Unilateral (one eye effected) - not usually heritable
  - Bilateral (both eyes effected) - almost always heritable, as dominant with incomplete penetrance
- The two cases led to Knudson's two hit hypothesis, which states that disease onset requires "two hits" (biallelic) - both copies of the same locus had to be hit with a mutation. In the familial cases, one hit of the disease was already inherited in the RB mutation.
- "Loss of Heterozygosity" is commonly found for mitotic recombination or nondisjunction.
  - The loss of function of a second allele invovled in mitotic recombination is the trigger for cancer.
  - Individuals heterozygous for a tumor suppressor gene are predisposed to cancer.

Impact of microRNAs on cancer

- BRCA1 is a tumor suppressor gene, impacted by upstream microRNAs.
- One mutant copy of a tumor suppressor gene will not result in overgrowth. Therefore, inherited mutations in tumor supressors are recessive at the cell level, but result in a dominant susceptibility to cancer

Steps to Malignancy

- Through clonal evolution, tumor cells acquire multiple mutations which allow them to become increasingly aggressive and proliferative.
- Hanahan and Weinberg (2000) have defined the following steps to malignancy, in increasing severity:
  - Independence of external growth signals
  - Insensitivity to external anti-growth signals
  - Ability to avoid apoptosis
  - Ability to replicate indefinitely
  - Ability of mass of cells to trigger angiogenesis (formation of new blood vessels) and vascularize
  - Ability to invade tissues and establish secondary tumors (metastasize, which requires angiogenesis)

How can causal genes be identified in cancer?

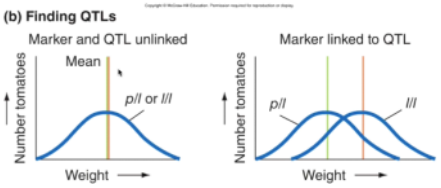
- Microarray experiments: changes in gene expression between cancer and non-cancer cells.
- Whole genome or transcriptome comparison.
- Hold promise for identifying gene clusters of genes that are repeatedly affected in particular types of cancers.
- Major problem in these analyses:
  - How do we distinguish between the "driver" and the "passenger" mutations?
  - Driver mutations: contribute to oncogenesis and have been positively selected in the evolution of the cancer.
  - Passenger mutations: incidental consequences of genomic instability and multiple cell divisions

Quantitative Genetics

Tuesday, May 12, 2020 3:26 PM

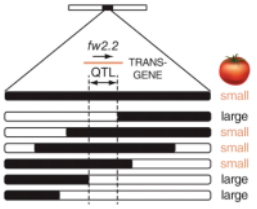
Quantitative Genetics

- Quantitative genetics is defined as the analysis of complex traits: multiple genes and the environment an individual is in.
- Quantitative vs. Qualitative traits:
  - Qualitative traits are discontinuous traits
    - Small number of distinct phenotypes
    - Ex.) black or white with maybe one gray intermediate (easily seen)
  - Quantitative traits are continuous traits
    - Many overlapping phenotypes, with a range of phenotypes varying along a scale.
    - Ex.) plant height varies on a range.
- R.A. Fisher created the landmark paper "The Correlation between Relatives on the Supposition of Mendelian Inheritance" which showed that quantitative traits result from multiple genes acting to produce a continuous range of phenotypes.
  - Kernel color in wheat - Nilsson-Ehle experiment
- General idea is that the greater the number of loci for a specific trait, the greater the number of phenotypic classes.
  - To determine the number of individuals in the F2 progeny that resemble each of the homozygous parents, the formula (1/4)^n is used.
  - The general pattern as number of loci present -> infinity for a specific trait is a normal distribution curve.
- Complex traits may also be due to the environment - for example, adult human height may have increased in the recent years due to better nutrition.



QTLs

- Quantitative Trait Locus (QTL) is a locus which correlates with a variation of a quantitative trait in the phenotype of a population of organisms.
- Most traits are affected by multiple genes including most diseases
  - Genetic and environmental factors effect penetrance and expressivity of genes.
- Identifying QTLs:
  - Direct QTL mapping - requires crosses between individuals that differ in the phenotype of interest.
    - Example: crosses between large tomatoes and small tomatoes
      - Experimental scheme to identify the genes involved in tomato size starts with the isogenic (pure -breeding) lines of small and large tomatoes. The F1 generation is medium in size. We then back cross the the F1 generation to the parental line, noting the changes to the size.
      - The SNPs in the tomatoes were used to determine if the mean weight of the tomatoes were different for each marker. Difference s in mean weight of the I/I allele and p/I allele for a given marker would indicate that the marker is linked to a QTL.
      - Fine mapping is a technique used to identify a causal gene:
        - The boundaries for nearly isogenic lines (NIL) of tomatoes is determined by comparing the SNPs from each parent of the progen y. The trait value is then determined for each. (See diagram to the right)
    - Generally this method of identifying QTLs results in the narrowing down of genes to <100 or maybe even <10 genes. At this level, the DNA can then be directly sequenced. Often the coding regions may be identical. However, this does not mean that it is the wrong gene, as often QTL reveals variation in expression of candidate genes. Variation in cis regulatory elements (CREs) may also result in different levels of expression and candidate genes.Changes in cis regulatory elements resulting in changes in expression of genes results in greater biodiversity.
    - An experiment to test the candidate gene is by inserting the particular QTL of a small tomato that is thought to encode for size as a transgene into a large tomato gene during development. If the resultant tomato is smaller than its large counterparts, one can confirm that the QTL in questiondoes in fact play a role in determining size.
  - Association mapping - takes advantage of the history of a random breeding population.
  - DNA markers (SNPs, DIP or Indels, SSRs & microsatellites) are used for mapping QTLs.
    - The more markers there are in a genome, the greater the likelihood of finding the gene.
  - Each gene in the QTL region may contribute to a trait at a different extent
  - A trait may be determined by many small effect QTLs, or a small number of large effects on QTLs or a combination of the two.
- The variant effects of QTLs on traits:
  - Some QTLs have more influence on a trait than other QTLs. The effects of the QTLs are denoted as major effect QTLs, and minor effect QTLs. For one single gene trait, the trait is entirely determined by one major effect QTL.



Relative Contribution of Genes and the Environment

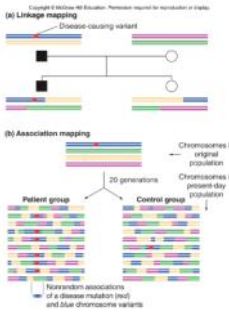
- A goal of quantitative genetics is to separate the genetic and environmental effects.
- Heritability - the proportion of the total phenotypic *variation* that is due to genetic difference (and not due to environmental factors)
  - If heritability is 0, no difference would be observed between monozygotic (MZ), dizygotic (DZ), or unrelated by adoption (UR) in humans
  - Conversely, if heritability of a certain trait is high, then differences would be observed in comparing the MZ, DZ, and UR children.
  - Phenotypic variation - defined as Vp, and genetic variation - Vg.
  - The heritability of a trait is the proportion of the total phenotypic variation that is ascribable to the genetic variation.
  - This can be emperically sorted out for some traits.
- The components of phenotypic variance are as follows:
  Vp = Vg + Ve + Vge
  where Ve is the environmental variance and Vge is the interaction between the genetics of an organism and its environment.
- The components of genetic variance includes:
  Vg = Va + Vd + Vi
  where Va is the additive genetic variance, Vd is the dominance genetic variance, and Vi is the genic interaction variance.
  Therefore Vp = (Va + Vd + Vi) + Ve + Vge.
- Studies of the environment and the genetics of dandelions can help to sort out the effects of genes and the environment.
  - To compare the contributions of genes and the environment, we much test each one separately - i.e. do one experiment in which the genes are the same while the environment is different, and another in which the environment is the same but the genes are different.
  - Any variance in these two experiments of growth must be due to the environment or the genes, whichever is not held constant.
- The total phenotypic variance
  - Growing of genetically diverse organisms in different environments results in a much broader variance, as both the environment and the genes are contributing toward the total variance.

GWAS

Tuesday, May 12, 2020 3:26 PM

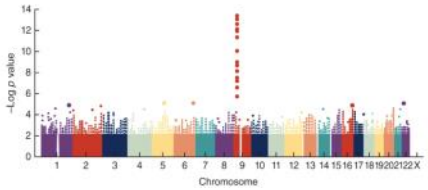
Genome Wide Association Studies (GWAS) Intro

- Involve the search for linkage between markers and phenotypes in an existing population, correlating the haplotype with a tra it.
  - Of these markers, SNPs are the most common.
- Human diseases, morphological traits, etc. can be isolated from these GWAS.
- As opposed to setting up crosses to map genes based upon recombination, association mapping depends on recombination that occ urred over many generations.
  - In the diagram to the right, the different colors of the chromosome represent the same chromosome pair on two different organisms which have different SNPs which distinguish them from each other.
  - We see that the blue variant of the chromosome is strongly associated with the disease. Therefore, the marker most likely will not be the "red" mutation itself but rather a linked polymorphism. Fine mapping or positional cloning is the next step to identify exactly the region of the causal gene.
    - The red marker is not necessarily the cause of the disease.
    - Recall that multiple causal genes may contribute to a single phenotypic output.
    - Therefore, results may apply to one specific population but not necessarily other populations.
- Genes may be causal but very rare in real populations. For example, mutations in the leptin receptor can cause morbid obesity with early onset during childhood.
  - Loss of function of the leptin receptor results in failure to transmit satiety signal. However, very few humans carry this mutation.



Manhattan Plot

- Significant associations in SNPs may be observed by chance.
- Example of Manhattan Plot to the right: the -log p value is determined from the chi-square test which shows a significant difference between control and disease groups. Peaks in the Manhattan plot determines a significant association. In the particular plot given, we see a region on ch romosome 9 significantly associated with a trait. Note, however, that peaks between two chromosomes are unlikely to indicate association with each other, as they are on completely different chromosomes.



The Limits of GWAS and QTL Mapping Studies

- The number of QTL depends on the amount of variation in the population.
- Many genes can influence a trait but if they don't vary in the population studied, they cannot be identified.
- Rare alleles are difficult to identify.
- Alleles with small contributions are also difficult to identify.
- Some may argue that few important genes at all have been identified in humans to date.

Haplotypes

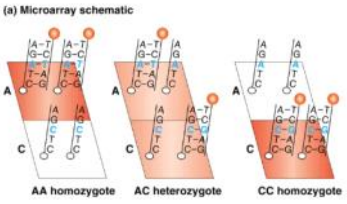
- Haplotypes are polymorphisms in a region of the genome that continue to be inherited together as a unit over several or many generations.
  - Certain polymorphisms may be co-inherited because they are near each other on the chromosome and recombine away from each other less commonly.
- Mapping is done through the presence of known DNA markers (polymorphisms) in humans based on segregation patterns of allelic differences.
- Linkage
  - Linked genes or linked markers are on the same chromosome, where the strength of the linked genes are dependent on the distance between the genes. The recombination frequency increases as the distance between the genes or markers increase.
    - If two genes are not linked, the markers will segregate independently.
  - Linkage equilibrium - the four haplotypes are present at equal frequencies, therefore knowledge at site one of the sequence does not provide any information about the sequence at site 2.
    - Markers will be present in different individuals in proportion to their frequency in a population.
  - Linkage disequilibrium is when the SNP variants of two loci are correlated, therefore knowledge at site one of the sequence will provide additional information about the sequence at site 2.
  - In the example to the right, a1's frequency is 0.7, whereas b1's frequency is 0.8. this would mean that a2's frequency is 0.3and b2's frequency is 0.2. Since the results of the expected values at linkage equilibrium are not the same as the observed values, we can conclude that the two genes are not in complete gene equilibrium. Instead, there seems to be an element of linkage disequilibrium.
    - We can also infer the positioning, or arrangement, of the alleles based on the population frequencies. The order is most likely a1,b1/a2,b2.
  - We can conduct a genome-wide microarray experiment to identify known SNPs across the whole genome.

	Expected	Observed
a^1, b^1	0.7 x 0.8 = 0.56	0.75
a^1, b^2	0.7 x 0.2 = 0.14	0.07
a^2, b^1	0.3 x 0.8 = 0.24	0.03
a^2, b^2	0.3 x 0.2 = 0.06	0.15
Total =	1.0	1.0

- a^1, b^1 are found together more often than expected
- a^1, b^2 are found together less often than expected
- a^2, b^1 are found together less often than expected
- a^2, b^2 are found together more often than expected

Microarray

- A genome wide microarray may be conducted to identify known SNPs across the entire genome



Developmental Genetics

Tuesday, May 12, 2020 3:27 PM

Cell Potency

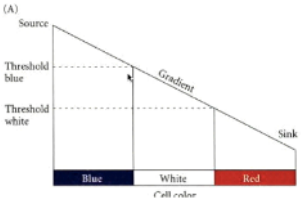
- Cell Potency is restricted over time during animal development by three main factors:
  - Differential expression of genes
  - Cascades of gene expression that increasingly specify cell types
  - Epigenetic changes accompanying / controlling these processes.
- On the other hand, plant cells retain their totipotency throughout their development.
  - For example, a carrot's cells may be isolated and placed in a nutritive medium, and will give rise to an entire carrot plant.
- Experiments by embryologists suggested that all embryonic cells are totipotent, or have the potential to generate a whole (to tal) embryo.
- Totipotent cell - the cell which has the potential to develop into any type.
- Determination - a cell becomes committed to a particular cell fate.
  - Differentiation occurs when a cell takes on that fate.
- Hematopoetic cell lineages illustrate the sequential specification / restriction of potential.
  - The cell's fate becomes more and more restricted until differentiation occurs.
- Stem cells retain potency through cycles of self-renewal and differentiation. Terminally differentiated cells will never go back to being a stem cell (unipotent)
- The successful cloning of animals demonstrated that the genetic material is not lost from the nuclei of differentiated cells.

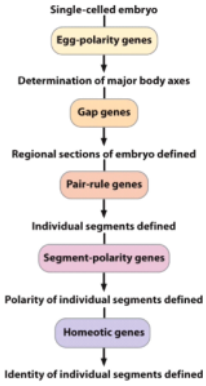
Stem Cells and Therapy

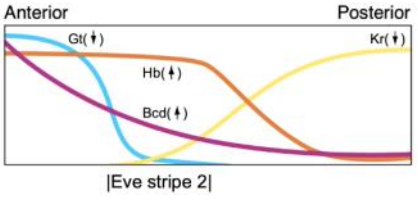
- The goal is to isolate the stem cell and culture the stem cell in vivo to generate specific cell types and deliver the differ entiated cells to a normal organism.
- The ideal stem cell is the totipotent cell, and this has many applications in regenerative medicine, where damaged tissue can be recovered through the introduction of these totipotent cells that can recover the damaged tissue.
  - Currently stem cell therapy is only being used for bone marrow transplantation.
- Embryonic stem cells (ES) cells are an ideal choice for regenerative medicine, as the zygote is totipotent. In the blastocyst , there are inner mass cells which are pluripotent (not completely totipotent), and able to develop into many different cell types such as heart cells, nerve cells, or immune c ells.

Drosophila and Developmental Genetics

- Researchers found that in drosophila cells, the fate of the cells were determined within the first hours of embryonic develop ment by regulatory genes.
- In the early drosophila embryo, the major body axes are established, the number and orientation of the body segments are dete rmined, and the identity of each individual segment is established. These are as a result of the expression of several regulatory genes:
  - The establishment of the main body axes is determined by egg polarity genes, whereas the determination of the number and pola rity of body segments is determined by segmentation genes. The establishment of the identity of each segment is regulated by the homeotic genes.
    - Egg polarity genes
      - The egg polarity genes are believed to be of maternal origin
      - Segmentation genes - determines number and polarity of body segments
      - Homeotic genes (HOX genes) - identity to each segment
- Morphogens
  - Morphogenic chemical agents and their concentrations are vital for determining the developmental fate of a region - this is the case with egg polarity genes.
  - Morphogen gradients can specify both "color and pattern" in the French flag model -> different levels of a morphogen would produce different colors at different thresholds. In the example to the right, a high concentration of a morphogen could mean a blue color, whereas a lower concent ration of morphogen would indicate a white color, and the lowest concentration of morphogen would result in a red color. The morphogens must pass thresholds in or der for the right colors to be expressed.
  - We see that the bcd protein forms a concentration gradient from anterior to posterior, and binds to DNA in a sequence -specific fashion and activates transcription, setting up the initial differences in an embryo along the A-P axis - a high concentration of bcd would result in one axis, whereas a lower concentration would result in a different axis, etc.
  - bcd also works in a concentration dependent fashion, in which it acts from anterior to posterior to determine the "headness" of a segment.
  - Multiple morphogenic gradients are set up in the axes of the embryo, and may explain the intial patterning of embryo. The hu nchback morphogen is another example of a gradient vital for embryonic development.
- 1995 Nobel Prize awarded to Christiane Nusslein Volhard and Eric Wieschaus for the identification of the regulation genes for development in drosophila.
  - Their experiment involved generating random mutations in the entirety of the genome to identify ALL genes required for normal development, then examining the embryo for visible defects, and classified the mutant embryos into complementation groups, identifying around 50 genes with u nique phenotypes.
  - The researchers then were able to classify genes into groups based upon the type of defect.
  - Drosophila embryo with a bcd (bicoid) mutation (mutated egg-polarity genes) were unable to specify head development, and therefore the embryo mutated with bcd was unable to successfully develop the head.
    - Interestingly, the transplantation of wild type cytoplasm to the anterior end of a bcd mutant rescued the mutant embryo - indicating the presence of some substance in the cytoplasm facilitating this process. That substance later turned out to be bcdmRNA.
- Drosophila embryo are sequentially subdivided by sequential action of regulatory genes, in an order as displayed in the diagr am to the right. Segmentation genes actually determine segment formation and additionally, regulate each other, hierarchically. (Gap genes regulate pair rule genes which regulate segment polarity genes) The regulation of "downstream" target genes encoding product involved in a physical process is required to build body segments.
  - Segmentation genes require a combination of maternal (egg-polarity genes) and gap genes to function, not just one dedicated regulator.
  - Evidence of segmentation genes regulating each other using mutant animals and an assay to detect expression:
    - The expression of genes in wild type and in mutant embryos can be compared - if gene a regulates gene b, then the expression of gene b will be impacted in a mutant version of gene a. We can then analyze cis regulatory elements in vivo using reporter genes to distinguish direct from indirect effects.
  - A mutation in the gap genes causes the elimination of anterior segments (Kruppel) resulting in a shorter larva.
  - A mutation of pair rule genes can result in the deletion of even-numbered segments, resulting in a shorter larva.
  - A mutation of the segment polarity can result in many hairs facing all directions due to the absence of the segment polarity morphogen.
  - The hunchback mutation (hb) is a gap gene mutation.
- Segmentation genes regulate downstream target genes, encoding product involved in physical process required to build body seg ments.
  - Downstream target genes encode products involved in differentiation.
  - Strong evidence of a correlation between when and where segmentation genes are expressed and when and where the genes functi on.
- Dissection of eve-cis regulatory elements using reporter constructs
  - Researchers generated transgenic flies with a candidate enhancer DNA promoter lac-Z reporter, and found one region of DNA which directed expression in one stripe. This stripe was named eve stripe 2
  - In gap mutants, the expression of eve stripe 2 is expanded. However, the expression of eve stripe 2 is lost in bcd or hunchback (hb) mutants, with Eve stripe 2 being the only stripe which was expressed in the middle of the organism (see diagram to the right)







Hox Genes and Evo-Devo

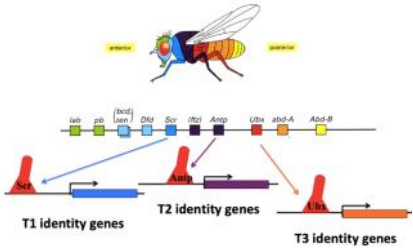
Tuesday, May 12, 2020 3:27 PM

Hox Genes or Homeotic Genes

- Hox genes each give a unique identity to the different segments of the organism.
- Homeotic mutations - transformation of one body part into fully formed alternate body parts.
- Edward Lewis was awarded the Nobel Prize in 1995 for his studies on the Hox genes of drosophila.

Hox Genes in Drosophila

- The homeotic genes are expressed in different regions along the body axis and determine segment identities.
- Antp is a regulatory gene encodes a transcription factor which determines leg formation. Antp would regulate genes encoding products that make T2 Hox gene legs. Antp may be mis-expressed in the head, caused by a chromosomal translocation which brings Antp under the control of a head-specific enhancer. This causes Antp to be expressed in the head, where it does its job of making a T2 leg, but now being expressed in the wrong place.
- Homeotic genes are expressed in the primordia of the segments they specify, well before any of the parts of the segment are developed.
  - Scr is expressed in the primordia of T1, where it specifies the formation of the first pair of legs.
- A homeotic mutation in the gene Ubx results in a bithorax, four winged fly.



Mechanism of Action of Hox Genes

- Hox genes encode hox proteins which act as transcription factors that determine developmental pathways. Like other transcription factors, Hox proteins are modular and consist of both a DNA binding domain and an activation or repression domain. As a result, each hox protein regulates a unique set of 'target' or 'downstream' genes unique to its segment (body parts).
- All hox genes contain a homeobox domain with a very similar DNA sequence encoding a 60 amino acid homeodomain

Evo-Devo

- Hox genes are highly conserved in all animals throughout the animal kingdom. The genes controlling the development of very different animals are the same.
- Thus, the hox genes are known as the genetic toolkit, a shared set of developmental regulatory genes which can be utilized indifferent ways and in different species.
- Despite hundreds of years of divergence, Hox orthologs retain their sequence, colinearity, and function.
- Do Hox proteins retain their function across broadly divergent taxa?
  - Comparison of mouse Hox genes and Drosophila Hox genes yielded a surprising result - orthologs between the scr hox gene of the fly and the a5 hox gene in mice were detected,
  - A functional test was done to see if the T1 encoding scr gene of the drosophila indeed was functionally conserved in mice. In the fly with the scr gene expressed in the head, T1 legs were present in the head region, which definitively showed that Mouse Hox-a5 mimics Dm-Scr antennal to leg transformation.

The Eyeless Gene

- In 1994, cloning of Drosophila eyeless phenotype happened, which drew similarity between the eyeless gene and the mammalian Pax6 gene.
- Transcription factors expressed in the anterior region of the embryo activated a cascade of regulatory genes. and once eyeless is expressed, the fate of the cells are locked in - the cells are determined to become eye cells.
- The eyeless gene is known to be a master regulator gene, which determines entire developmental pathways. Once the eyeless gene is expressed, the cell will become an eye cell. The ectopic (abnormal place) expression of the eyeless gene results in formation of eyes in that abnormal place.
  - The same master regulatory gene is present in humans and in mice as orthologous genes.

The study of development and link to evolution

- The hox gene, as the genetic toolkit, is an ancient developmental feature which has been shared by countless animals.
  - Genetic changes in development may lead to a change in morphology, but the same underlying Hox gene toolkit remains the same. Major changes in evolution actually occur through gene expression.
- Example.) The evolution of a snake
  - The snake is thought to have lost its forelimbs from a lizard-like ancestor. A shift in Hox expression must have accounted for the loss of a forelimb in snakes, particularly in Hox C6 and Hox C8 - an anterior shift in Hox expression resulted in the loss of a limb in a region that previously developed the limb.
    - Modern day lizards retained the expression of Hox C-6 and Hox C-8, which repressed the expression of limbs only in the flank. However, limb formation in snakes was repressed by the extension of Hox C-6 and Hox C-8 into the forelimb region of the snake's genes, suggesting that a change in the regulation patterns of Hox C6 and Hox C8 must have resulted from a change in their respective cis regulatory elements.
    - We can hypothesize that if we misexpress the Hox C6 and Hox C8 genes in the forelimbs of a lizard, the anterior limbs would not grow.
- Alterations of key regulatory sequences often bring about major developmental changes.
  - Example.) Marine sticklebacks vs. Freshwater sticklebacks
    - Marine sticklebacks are seen to possess large pelvic spines, whereas in freshwater sticklebacks these pelvic spines are largely reduced in size.
    - Researchers used QTL mapping to identify the genes responsible for the loss of spines, and resulted in the identification of a cis regulatory change - the candidate gene, Pitx1, which causes reduced limbs in mouse mutant orthologs, was isolated. The researchers, however, found no difference in coding region between the marine and freshwater Pitx1 genes. This could have meant that the region isolated was wrong- however, researchers noted that marine fish expressed Pitx1 in the pelvic region whereas freshwater fish did not.
      - Identified a tissue-specific enhancer region responsible for pelvic expression.
      - Freshwater fish are homozygous for a deletion of that pelvic specific enhancer.
    - Enhancer modularity allows for evolutionary changes in gene expression, as the loss of one enhancer does not impace expression in other tissues, and a change arising in one enhancer will only affect one tissue.
      - If this change is beneficial, it may persist due to drift or selection.
  - Another example.) Blind cavefish and their loss of eyes
    - Lens cells undergo apoptosis early in development, and other portions of the eye fail to develop as a result.
    - A microarray experiment was conducted to see which genes are responsible for eye loss in cavefish. Later, QTL mapping and genome sequences were used. The results found that the pax6 hox gene was responsible.
      - Multiple regulatory gene changes actually explain the loss of cavefish eyes. The orthologs of the drosophila hedgehog gene (hh) segment polarity gene play a role in cavefish as well. shh and twhh paralogs are expressed more broadly in blind cavefish than in surface fish. The expanded expression of these genes result in activation of genes involved in apoptosis. This was proved by an injection of shh and twhh in surface fish, which resulted in eye degeneration.
    - Most likely the evolutionary change in expression of hh-genes of cavefish initiated a cascade leading to the loss of eyes.