Professor Hemann material:

Structure and regulation of a gene:

DNA → RNA → Protein
Transcription *Translation*

DNA has promoter to regulate transcription
RNA has Shine-Delgarno sequence to initiate translation (which is dictated by start and stop sites)
Should know how to use a codon chart

Prokaryotes: promoter/operator/operons
Eukaryotes: promoter/introns

We can use what we know about the genetic code to identify potential genes
- -look for ORFs (no stops for a set number of codons)
- -Can test for chance of finding fortuitous ORFs since there are 64 codons and 3/64 are stop codons [$p(ORF>x) = (1 - 3/64)^x$]

Investigating mutants:

Nature of the mutation?
- -Make a heterozygote to determine dominant/ recessive
- -For yeast: cross 2 haploids to get a heterozygous diploid
- -For bacteria: add in a WT copy on an F plasmid
- -For diploid organisms: cross true breeding lines and look in the F1s

How many genes?
- -Complementation test for independent mutants
- -Look at the phenotypes of F2s (diploid organisms) or do tetrad analysis (yeast only)

Autosomal or sex-linked?
- -Cross mutant females with WT males and examine F1s

Mendelian Genetics:

The expected results of F1 and F2 genotype/phenotype ratios from true breeding crosses w/ one or two genes, x-linked or autosomal

Using chi-squared to determine if observed is consistent w/ these expectations

Using pedigrees to determine disease risks when there are potential carriers in a given family; using Bayes theorem to expand on these probabilities

How to deal with autosomal dominant, autosomal recessive, x-linked diseases

Recombination and Linkage:

Recombination is a requirement of meiosis

-Should understand the physical aspect of how this creates new combinations of alleles on a chromosome

Identifying parental and recombinant classes after doing a cross

-considerations for dominant/recessive/sex-linked genes

-can also determine if dealing with multiple genes

Using recombinants to create genetic distances (in centimorgans) and order genes

-2 factor, 3 factor crosses (more accurate distances b/c can see the double crossovers)

-Genetic distance is related to the physical distance (bps) by the recombination rate in that species (genetic distance=physical distance*recombination rate)

-50 cM is the max you can do w/ not yeast. 50 cM indicates completely unlinked and the genes are assorting randomly

Mutations of interest can also be followed/mapped using DNA markers (SNPs, SSRs, etc)

Tetrad analysis (Yeast only)

-Understand how tetrads are made (mate 2 haploids → diploid → sporulate → 4 haploid spores (tetrad)

-Be able to determine the tetrad types (PD, TT, NPD) and corresponding genotypes

-Understand what each tetrad type tells you about the number of crossovers

-Use ratios of tetrad types to calculate genetic distances (centimorgans)

-Can go up to 80 cM before concluding completely unlinked/ random assortment

Linkage in humans:

Using pedigrees to link genotypes (for DNA markers) to observed traits (Mendelian diseases)

Should be able to:

-Identify informative meioses

-Determine if phased or not (and know how to apply the phase penalty if needed)

-Calculate the LOD score (log of the odds ratio for a linked vs unlinked inheritance between a marker and a disease allele) for a given θ

-Remember that if $\theta=0$, LOD score is $-\infty$ if there is a recombinant

-Add up the LOD scores of different pedigrees

We can calculate the LOD score at several loci across a given haplotype to narrow in on regions where the causative allele is located.

Types of mutations:

Nonsense/missense/frameshift/silent

Can induce mutations by using a mutagen
-Screen or select to isolate mutants

Suppressor mutations (back mutation/intragenic suppressor/extragenic suppressor)

Transposition & Transduction:
Transposons are mobile genetic elements

Transposons carry a transposase to hop in and out of genomes and often some antibiotic resistance gene (Tn5 carries Kan resistance)

For our purposes, we can use a modified lambda phage (bacterial virus) to perform transposon mutagenesis

- Infect E. coli with λ Pam int⁻::Tn5
- Select with kanamycin (because transposition rate is low). Each surviving cell will carry a Tn5 insertion in a unique place in the genome, many of which disrupt genes
- Screen or select for your phenotype of interest

Can map using transduction

- Use P1 phage to infect a starting E. coli strain
- Use that lysate to infect another E. coli strain
- Measure cotransduction frequencies (this is NOT the same as genetic distance and there is no need to try to convert it into one) of your marker of interest from Tn5, from each other

Conjugation (*E. coli)*
Terms to know:

- F plasmid (and components)
- F' plasmid
- F+ strain
- Hfr strain

Should understand the basics of cotransduction, how F plasmids can recombine (by homology) into and out of chromosomal DNA

Should be able to assign orientation of the OriT based on how markers are transferred during mating

Components carried on F or F' plasmids will transfer much more quickly that most components transferred by an Hfr strain (~100 minutes to transfer entire genome)

Prokaryotic gene regulation

Prokaryotic genes are organized into operons that contain multiple genes and are all regulated together

Should understand how the different mutants we discussed in class helped create the pictures of the Lac and Mal operons

We can do "bacterial complementation" using F' plasmids to create merodiploids

Cis/trans tests inform us if we are dealing with cis acting elements (physically nearby) or trans acting elements (diffusible protein product)

Operons can be regulated negatively (Lac) or positively (Mal)

Terms to know:

- Promoter
- Operator
- Repressor
- Inducer

Pathways and Epistasis

Know phenotypes of common regulatory mutants (e.g. operator-)

Prokaryotes – Tests to determine mutant phenotype using F' plasmids (dom/rec, cis/trans)
Yeast – Use tetrad analysis to determine mutant phenotype.

Double mutants used to order linear pathways (epistasis)
*only requires that two mutants have opposite phenotypes.

Promoter bashing is used to find functional elements in a eukaryotic promoter.
Know the various elements of a promoter, their general location in the promoter and their loss of function phenotype. (e.g. TATA box)

Genomes

What are the elements of a typical genome?
How are genomes assembled?

Pair-wise alignments – Global & Local

What assumptions do we make in order to perform such alignments and assign them a score?

Know the difference between these two types of alignments and how to interpret a dynamic programming matrix created for these two alignment types.

We can obtain the best alignment score by simply picking it out of the matrix (no need for separate calculation)

Know how to trace back within a matrix to obtain the optimal alignment(s).

Sequence Conservation

RGY Plots
-What do functional/protein coding regions look like?
-How to find an ORF or the start codon for a real gene
-Identify non-coding/intronic regions

Dotplots
-How to interpret a dotplot. (Pay attention to what the question is asking you to find in the dotplot)
-Drawing dotplots for given chromosome structures (e.g. duplications, repeats, etc.)
-If asked to interpret a dotplot in a biological way, remember to go back to the introduction of the question for any hints about the biology involved.

HWE
Understand how to calculate allele & genotype frequencies in an equilibrium situation
What assumptions does equilibrium make?

Are we in equilibrium or not? Using a chi squared test to determine this.
Understand rare allele approximations

Factors that act against equilibrium:
-mutation
-selection (against homozygotes and for heterozygotes)

Know how to calculate allele/genotype frequencies for the next generation given some mutation/selection for all four situations:
-recessive allele
-dominant allele
-X-linked recessive allele
-Balanced polymorphism

For each case identify if mutation, selection and/or heterozygote advantage is occurring and make sure to include them in your calculations.

Remember that mutation and heterozygote advantage will causes allele frequency of the minor allele to increase while selection will cause it to decrease.

Given steady state, know how to calculate mu, S and h.

Inbreeding

How to calculate F = P(homozygous by descent)*#alleles in shared ancestor(s)

Remember when calculating P(homozygous by descent) we are calculating the probability of being homozygous for the same **physically identical allele** NOT genetically identical alleles.

Fq - contribution of inbreeding to disease

LD

Not a measure of distance between genes or linkage. Linked genes tend to stay in LD for longer but being in LD does **not** imply linkage.

Know what a haplotype is and how it is different from a genotype.

Know how to calculate allele frequencies given haplotype frequencies.

Know how to calculate D, D' and r^2 and know what information these provide us with.

To determine if in LD or not you can calculate observed and expected values OR use the formula $x^2 = r^2N$. These will give you the SAME result, so unless stated otherwise in the question, use the second method as it will save you time.

Know how to calculate LD decay over time due to recombination.