Biol 213 GENETICS (Fall 2000): Problem Set 10
Mutations and Cancer[1]

0. *Special problem (not for the exam, but an example of how genetic knowledge can serve as a beacon of light for the world).* The officials in Florida still can't get things right, so they call in you, known for your prowess in genetic analysis (always a valuable commodity). You randomly sample 1% of the total ballots to see if any had been miscounted, and you find 25 erroneously tallied. From this, you estimate that there are 2500 wrong ballots in the total, but you realize that by chance your sample might have picked up a greater or lesser percentage of wrongly counted ballots. Given that, you decide to say that there are 2500 wrong ballots plus or minus one standard deviation in the number of wrong ballots (if you were able to sample a lot of times). Presuming the ballots you counted to be a completely random sample, what's your best guess as to how far off your estimate might be?

   If you have no clue how to proceed, take a look at the italicized special feature in the notes for Wednesday, November 15, p.3.

1. Match each of the following mutations to the mutagen(s) that might have produced it.

a. GGG (glycine) → TGG (tryptophan)	A. Nitrous acid
b. GCT (alanine) → GTT (valine)	B. Ultraviolet radiation
c. CCC (proline) → CAC (histidine)	C. Aflatoxin

2. The couple of hundred *lacI* mutations described in the notes of November 15 took an incredible amount of work to obtain in the 1980's. With present day tools, the job would not be nearly so onerous. After obtaining mutants on PGal plates, you could use PCR to clone the *lacI* gene and send the DNA out to an automatic DNA sequencing service to determine the site of the mutation. Suppose, as a service to science, the UR Biology Department decides to set as a requirement for graduation that each major clone and characterize 10 *lacI* mutants.

   Now fast forward fifty years and some tens of thousands *lacI* mutants later. Frustration has set in because despite the large number of mutations characterized, there are still lots of the 1080 bases comprising *lacI* at which no mutation has ever been recovered. Provide at least three explanations why this is not so unexpected a result.

3. You are a world famous geneticist testifying in a wrongful death suit brought against Philip Morris by the estate of a man who died of lung cancer. Prior to his death, the deceased worked in a chemical factory that manufactured hydroxylamine. The plaintiffs maintain that the deceased's smoking habit caused the lung cancer, while the cigarette company argues that the cancer was due to his job in the chemical factory. You have cloned and sequenced DNA encoding p53, a tumor suppressor, from the deceased. You find that the gene was indeed mutated (as is typical in lung cancers), and the mutation was an arginine residue at position 248 that mutated to a leucine. You testify that the major carcinogen in tobacco smoke is benzo[A]pyrene ("Objection! That's `suspected carcinogen', your Honor."), which acts by forming a bulky addition primarily on guanine residues. DNA replication is unable to proceed past the modified base. What will be the remainder of your testimony? What can you say regarding the claims of the disputants?

[1] See table at end of this problem set for useful information regarding mutagens

***4. You want to analyze your fly mutants to an even more detailed level, to determine what the molecular nature of the mutation each suffers from. You go back to your original mutant fly stocks and expose them to the mutagens shown below. In some cases wild-type revertants are obtained, but in some cases not. From the information below, identify the kind of mutation responsible for each mutant fly.[2]

Ability of chemical mutagens to induce reversion to wild-type

	Chemical used to induce reversion			
Mutant	**5-Bromouracil**	**Hydroxylamine**	**Proflavin**	**None**
1	-	-	-	-
2	-	-	+	+
3	+	-	-	+
4	-	-	-	+
5	+	+	-	+

Adapted from problem in Griffiths et al (1996) Introduction to Genetic Analysis, 6th Edition

5. Fanconi anemia is an inherited condition that confers on the individual a hypersensitivity to DNA-damaging agents. People with Fanconi anemia suffer from a high incidence of leukemia and other cancers. The gene responsible for the condition has recently been cloned, and Levran et al [(1997) Proc Natl Acad Sci USA 94:13051-13056] sequenced the gene from 97 ethnically diverse individuals affected by Fanconi anemia. Thirty of the patients had mutations in the same position of the gene, a deletion of three basepairs (underlined below). Suggest in less than 10 words why this position might be particularly prone to mutation, and draw a diagram illustrating your suggestion.

```
CysGluArgGluGluLeuLeuValPheLeuPhePhePheSerLeuMetGlyLeuLeu
TGCGAGAGAGAGGAGCTATTGGTTTTCCTTTTCTTCTTCTCCTTGATGGGCCTGCTG
```

6. Consider two mutations: one in Ura-DNA glycosylase, the enzyme that detects and removes uracils from DNA, and one in AP endonuclease, which detects and excises apurinic/apyrimidinic residues. If the two mutations are combined into the same strain, which phenotype is epistatic over the other (i.e., does the resulting strain have the same phenotype as the Ura-DNA glycosylase-minus mutant or the AP endonuclease mutant)?

*** 7. You wish to find the spectrum of possible *lacI* mutations, and so you borrow a bit of a culture of wild-type *E. coli* from a nearby genetics lab and grow up a large culture for the experiment. You isolate hundreds of PGal$^+$ mutants and from them clone and sequence *lacI*. You find, to your surprise, that almost all of the *lacI* sequences you obtain contain the same base substitution.

a. How do you explain these results?

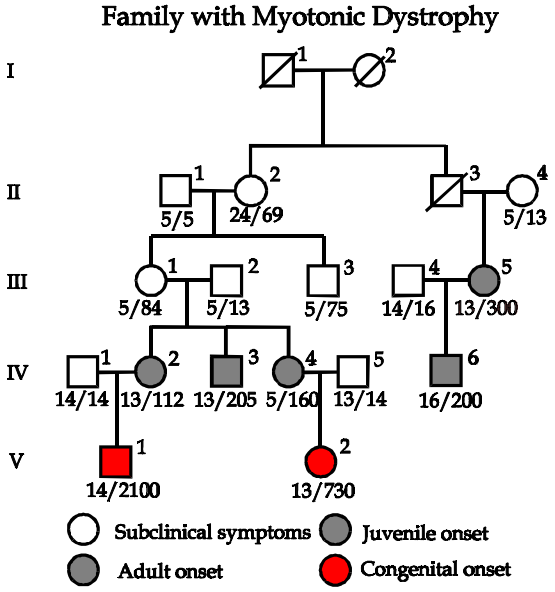
You repeat the experiment, this time being careful to start from single cells. Nonetheless, you can't get reproducible results. In most repetitions of the experiment you get about 200 *lacI* mutants per ml (per 2 x 10^9 cells), but sometimes you get several hundred mutants and occasionally many thousands!

b. Why won't this experiment sit still? (Hint: when during growth of the culture do mutations occur? Reflect on this and your answer to part a)

[2] It's true there's not enough time now to analyze the millions of flies to necessary to stand a chance of finding revertants, but there's always next summer.

8. What would be the phenotype of E. coli that lacks the ability to methylate GATC sequences? What would be the phenotype of E. coli that overmethylates GATC sequences (i.e., has such a high activity of the methylating enzyme that the sequences are methylated almost immediately after DNA replication)?

9. Several members of a family exhibit the symptoms of myotonic dystrophy, a condition typified by wasting away of muscles and a variety of other problems. The pedigree of the family is shown to the right. The gene responsible for the disease is on chromosome 19 and is preceded by a long untranslated stretch of repeated GCT nucleotides. DNA is isolated from the living members of the pedigree shown, the 5' region amplified by PCR, and the number of GCT nucleotides inferred by the size of the resulting fragment, determined by gel electrophoresis. The numbers of repeats for each individual are shown on the pedigree.

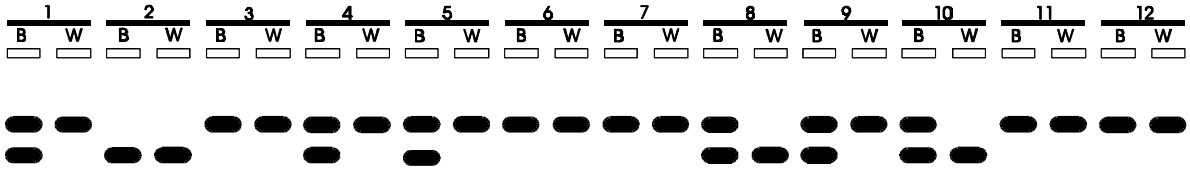


a. Why are two numbers shown for each individual?

b. What do you think is the relationship between the number of repeats and the disease state?

10. In a study aimed at understanding the basis of warts [Murray et al (1971) Nature 232:51], the authors isolated protein from blood and from warts of twelve female individuals. The twelve samples were run on a gel and stained in a way that visualized the activity of glucose-6-phosphate dehydrogenase, an enzyme encoded by a gene on the **X** chromosome. A cartoon of the stained gel is shown below. Each band represents the position of glucose-6-phosphate dehydrogenase protein, which exists in the general population in two forms. **B** indicates a sample taken from blood, **W** indicates a sample taken from wart tissue.

Glucose-6-Phosphate Dehydrogenase Phenotypes of 12 Individuals and Their Warts



a. Which of the proposed scenarios fit the data?

A. Wart viruses infect cells within a tissue. In some cases, the virus lyses the cell and spreads progeny virus particles to other cells. In most cases, however, the virus remains within the living cell and changes it into a wart cell.

B. Warts are caused by a mutation that occurs in the gene encoding glucose-6-phosphate dehydrogenase.

C. Warts are caused by mutations that occur in rare cells, causing the production of a hard case around them.

D. Wart viruses infect rare cells within a tissue. The infection causes the cell to change into a cell with a hard case and also to increase the rate of cell division of that cell and its progeny.

E. Warts are caused by a mutation that increases the rate of cell division of that cell.

b. Aren't you happier knowing where warts come from?

11. Li-Fraumeni syndrome is inherited in much the same fashion as familial retinoblastoma. It is caused by mutation in the p53 gene, a tumor suppressor gene.

a. Which of the following genotypes would you not be surprised to find in cells isolated from malignant tissue from a person with Li-Fraumeni syndrome?

b. Which of the following genotypes would you not be surprised to find in normal cells isolated from the same individual?

c. Which of the following genotypes would you not be surprised to find in germ line cells isolated from the same individual?

A. $p53^{+}$ / $p53^{+}$
B. $p53^{+}$ / $p53^{-}$
C. $p53^{-}$ / $p53^{-}$
D. $p53^{-}$ / deleted $p53^{-}$
E. $p53^{+}$ / deleted $p53^{-}$

*** 12. Light-induced skin cancer is among the most rapidly increasing cancers in the U.S. In an effort to identify the causes of skin cancer, Brash et al [(1991) Proc Natl Acad Sci USA 88:10124-10128] examined the pattern of mutation in skin tumors of fourteen unrelated individuals. DNA was isolated from each patient and amplified with oligonucleotide primers that flank the gene encoding p53. The DNA from each patient was sequenced and compared with the sequence of the wild-type gene (see next page). Note that the sequence shown is the nontemplate strand, not necessarily the strand that was altered during the original mutagenic event.

a. Do you think the mutations within the p53 gene are spontaneous or rather due to a mutagen? If the latter, then what's the identity of the mutagen? Why?

b. Argue for one of the following two propositions (Hint: What kind of mutation do you <u>not</u> find?):

A. Mutation of p53 is a harmless byproduct of the process of tumorigenesis in skin cancer.

B. Mutation of p53 is an important part of at least one route to skin cancer.

Mutations within p53 Gene from Human Carcinomas*

```
  Met Glu Glu Pro Gln Ser Asp Pro Ser Val Glu Pro Pro Leu Ser Gln Glu Thr Phe Ser Asp Leu Trp Lys
ATG GAG GAG CCG CAG TCA GAT CCT AGC GTC GAG CCC CCT CTG AGT CAG GAA ACA TTT TCA GAC CTA TGG AAA
                        C

Leu Leu Pro Glu Asn Asn Val Leu Ser Pro Leu Pro Ser Gln Ala Met Asp Asp Leu Met Leu Ser Pro Asp
CTA CTT CCT GAA AAC AAC GTT CTG TCC CCC TTG CCG TCC CAA GCA ATG GAT GAT TTG ATG CTG TCC CCG GAC

Asp Ile Glu Gln Trp Phe Thr Glu Asp Pro Gly Pro Asp Glu Ala Pro Arg Met Pro Glu Ala Ala Pro Arg
GAT ATT GAA CAA TGG TTC ACT GAA GAC CCA GGT CCA GAT GAA GCT CCC AGA ATG CCA GAG GCT GCT CCC CGC
                           T

Val Ala Pro Ala Pro Ala Thr Pro Thr Pro Ala Ala Pro Ala Pro Ala Pro Ser Trp Pro Leu Ser Ser Ser
GTG GCC CCT GCA CCA GCG ACT CCT ACA CCG GCG GCC CCT GCA CCA GCC CCC TCC TGG CCC CTG TCA TCT TCT

Val Pro Ser Gln Lys Thr Tyr Gln Gly Ser Tyr Gly Phe Arg Leu Gly Phe Leu His Ser Gly Thr Ala Lys
GTC CCT TCC CAG AAA ACC TAC CAG GGC AGC TAC GGT TTC CGT CTG GGC TTC TTG CAT TCT GGG ACA GCC AAG
                                 D

Ser Val Thr Cys Thr Tyr Ser Pro Ala Leu Asn Lys Met Phe Cys Gln Leu Ala Lys Thr Cys Pro Val Gln
TCT GTG ACT TGC ACG TAC TCC CCT GCC CTC AAC AAG ATG TTT TGC CAA CTG GCC AAG ACC TGC CCT GTG CAG

Leu Trp Val Asp Ser Thr Pro Pro Pro Gly Thr Arg Val Arg Ala Met Ala Ile Tyr Lys Gln Ser Gln His
CTG TGG GTT GAT TCC ACA CCC CCG CCC GGC ACC CGC GTC CGC GCC ATG GCC ATC TAC AAG CAG TCA CAG CAC
                        A
                           T

Met Thr Glu Val Val Arg Arg Cys Pro His His Glu Arg Cys Ser Asp Ser Asp Gly Leu Ala Pro Pro Gln
ATG ACG GAG GTT GTG AGG CGC TGC CCC CAC CAT GAG CGC TGC TCA GAT AGC GAT GGT CTG GCC CCT CCT CAG
                                    A

His Leu Ile Arg Val Glu Gly Asn Leu Arg Val Glu Tyr Leu Asp Asp Arg Asn Thr Phe Arg His Ser Val
CAT CTT ATC CGA GTG GAA GGA AAT TTG CGT GTG GAG TAT TTG GAT GAC AGA AAC ACT TTT CGA CAT AGT GTG

Val Val Pro Tyr Glu Pro Pro Glu Val Gly Ser Asp Cys Thr Thr Ile His Tyr Asn Tyr Met Cys Asn Ser
GTG GTG CCC TAT GAG CCG CCT GAG GTT GGC TCT GAC TGT ACC ACC ATC CAC TAC AAC TAC ATG TGT AAC AGT

Ser Cys Met Gly Gly Met Asn Arg Arg Pro Ile Leu Thr Ile Ile Thr Leu Glu Asp Ser Ser Gly Asn Leu
TCC TGC ATG GGC GGC ATG AAC CGG AGG CCC ATC CTC ACC ATC ATC ACA CTG GAA GAC TCC AGT GGT AAT CTA
                T         T-T                                     A
                AA

Leu Gly Arg Asn Ser Phe Glu Val Arg Val Cys Ala Cys Pro Gly Arg Asp Arg Arg Thr Glu Glu Glu Asn
CTG GGA CGG AAC AGC TTT GAG GTG CGT GTT TGT GCC TGT CCT GGG AGA GAC CGG CGC ACA GAG GAA GAG AAT
                                                    T                           A-A
                                                                                  A

Leu Arg Lys Lys Gly Glu Pro His His Glu Leu Pro Pro Gly Ser Thr Lys Arg Ala Leu Pro Asn Asn Thr
CTC CGC AAG AAA GGG GAG CCT CAC CAC GAG CTG CCC CCA GGG AGC ACT AAG CGA GCA CTG CCC AAC AAC ACC

Ser Ser Ser Pro Gln Pro Lys Lys Lys Pro Leu Asp Gly Glu Tyr Phe Thr Leu Gln Ile Arg Gly Arg Glu
AGC TCC TCT CCC CAG CCA AAG AAG AAA CCA CTG GAT GGA GAA TAT TTC ACC CTT CAG ATC CGT GGG CGT GAG
                T

Arg Phe Glu Met Phe Arg Glu Leu Asn Glu Ala Leu Glu Leu Lys Asp Ala Gln Ala Gly Lys Glu Pro Gly
CGC TTC GAG ATG TTC CGA GAG CTG AAT GAG GCC TTG GAA CTC AAG GAT GCC CAG GCT GGG AAG GAG CCA GGG

Gly Ser Arg Ala His Ser Ser His Leu Lys Ser Lys Lys Gly Gln Ser Thr Ser Arg His Lys Lys Leu Met
GGG AGC AGG GCT CAC TCC AGC CAC CTG AAG TCC AAA AAG GGT CAG TCT ACC TCC CGC CAT AAA AAA CTC ATG

Phe Lys Thr Glu Gly Pro Asp Ser Asp
TTC AAG ACA GAA GGG CCT GAC TCA GAC
```

*p53 amino acid sequence is given above the DNA sequence from the nontemplate strand. Bases in bold below the sequence represent mutations found in fourteen separate carcinomas. Adjacent bases (e.g., **AA** or **T-T** were found in the same carcinoma.

*** 13. You've hit upon an idea to reduce the incidence of lung cancer. Lung cancer is associated with the loss of p53 genes in cancerous tissue. You have cloned p53 and placed it in the DNA of adenovirus, a virus that can live within human cells, often with no ill effect but sometimes causing mild respiratory infections. You spray the recombinant virus into the lungs of several heavy smokers at risk for lung cancer but who are at present tumor-free. After allowing time for infection, you isolate RNA and DNA from the lungs of the volunteers and learn: (a) only 50% of the cells in their lung epithelial tissue has been infected with the virus, and (b) for some reason the p53 gene on the virus is not producing RNA.

a. Which of the following are appropriate responses:

A. Dismay. You were hoping that p53 would be expressed at a very high level that could not be overcome by subsequent mutation.

B. Mild amusement. The expression of p53 doesn't matter. A cell still needs to mutate all cellular copies of p53 in order to become tumorigenic.

C. Disappointment. Lack of expression of p53 means that your idea won't work.

D. Disgust. Whenever <u>you</u> inhale adenovirus, every cell in your lungs gets infected and you get a world-beating cold. Now, the cells that have escaped the virus remain at high risk to mutate the two cellular copies of p53 and become tumorigenic.

E. Frustration. You were hoping that the p53 gene you introduced would replace nonfunctional p53 genes.

To top it off, some time later a tumor is found in the lungs of one of the volunteers. DNA isolated from the cancerous tissue shows that <u>100%</u> of its cells carry the adenovirus/p53 DNA. All eyes turn to you.

b. Has your experiment somehow <u>caused</u> the tumor to occur? Can you think of some other way of explaining what has happened?

14. How likely are you to get cancer?

a. How likely is it that a cell in your body has a mutation in a gene encoding p53? Suppose that mutations arise solely from replication error, and errors occur at a rate of about 1 error per 10^9 bases per generation.

Strategy: estimate mutation rate for p53 per generation, then:
(Number of cells in body) x (mutation rate of p53/generation) x (generations)

a1. What is the mutation rate of *p53* per generation?

Strategy: estimate number of mutable bases in p53, then:
Mutation rate of p53/generation = mutations per base per generation x mutable bases in p53

a1.1. How many mutable bases in *p53*?

Strategy: estimate size of p53, then estimate fraction of bases which, when mutated, will affect function of p53.

- What is the size of *p53?*

Strategy: Presume p53 is a typical gene. Look at lengths of genes in pUR3. Alternatively, count bases in sequence given in problem 15.

- What is the fraction of bases which, when mutated will affect the function of *p53*?

Strategy: make an estimate based on shards of knowledge

· Most mutations in 3rd position of a codon are silent
· Many mutations lead to substitutions of similar amino acids

a2. How many cells are in your body?

Strategy: (Your volume)/(volume per cell) = number of cells

a2.1 What's the volume of a cell?

Strategy: approximate a cell as a cube, then volume of cell » (estimated length of cell)^3

- What's the length of a typical human cell?

Strategy: make an estimate based on shards of knowledge

· You can't see most human cells without a microscope (what size is the limit of human detection?)

· You <u>can</u> see human cells easily with even a crude microscope (say 100x magnification). So 100x the length of the cell is significantly greater than the limit of human detection.

· Even the smallest human cells are several times bigger than run-of-the-mill bacteria like *E. coli*.

· You can just barely see *E. coli* with 100x magnification. So 100x the length of *E. coli* is barely greater than the limit of human detection.

a2.2. What's the volume of you?

Strategy: volume = weight / density

- How much do you weigh?

- What is your approximate density? (Do you float or sink in water? How far away are you from equal density?)

a3.How many generations worth of cells are in your body?

Strategy: Don't know. Simplify by considering only one generation. Worry about other generations later if necessary.

Things look bleak. Fortunately, getting a mutation in a p53 gene is not by itself very serious. A specific cell needs to lose <u>both</u> copies of the gene. Furthermore, the cell has many ways of protecting itself against unbridled growth. Many genes must be lost to make a cell that divides without regulation. Suppose a single cell must suffer mutations in <u>both</u> alleles of a half-dozen genes in order to lead to a tumor.

b. How likely is it that your body will have such a cell in your lifetime?

Strategy: Calculate the probability that a one allele will suffer a mutation over the estimated number of generations of a cell, then multiply that number by itself as many times as there are alleles to be mutated.

b2. How many generations worth of cells do you have in a lifetime?

Strategy: No clue how to get a reasonable number. Go instead for a maximum number. Presume that every cell divides once a day (generous) and that the number of cells remains constant from birth to death (gross overestimate). Estimate a lifespan.

b1. What's the probability that one allele will suffer a mutation over a lifetime of cell generations?

Strategy: *presume that all alleles are about the size and mutability of p53. Multiply the mutation rate per base per generation by the number of mutable bases of p53, then multiply the product by the maximum number of cell generations over a lifetime.*

If you've gotten this far, you may conclude that getting cancer is wildly unlikely. No person in the history of the universe should have been so unlucky.

c. How do you explain the fact that cancer exists? What in the nature of cancer shoots a hole in the above calculation?

d. From the insights gained above, comment on a popular argument against evolution of life from nonliving antecedents that runs something like this:

*…how do you get the specific sequences necessary in proteins and in DNA? Consider proteins: the sequence of amino acids determines the way the molecule will "fold up", which gives it physical properties. For a particular function, an exact sequence is required. What are the odds of this occurring by accident? The odds of forming a specific molecule with 100 amino acids is (1/20) ** 100 = 10e130 (the number 10 with 130 zeros following it) to 1. Forget it!*[3]

[3]Creation Science home page: http://emporium.turnpike.net/C/cs/ol3.htm