

**7.012 Problem Set 7 FRIDAY December 3, 2004****Not due unless you have had a prior agreement with Claudette Gardel****Solutions will be posted on the web.****Question 1**

Leukemia is type of cancer characterized by the uncontrolled proliferation of white blood cells (leukocytes). Chronic Myelogenous Leukemia (CML) is a type of leukemia primarily caused by a translocation that relocates an oncogene, called *abl*, from the long arm of chromosome 9 to the long arm of chromosome 22 in the *bcr* region (breakpoint cluster region). The resulting *bcr-abl* fusion protein encodes a constitutively active tyrosine kinase, which when expressed, leads to the CML phenotype.

a) Gleevec is an effective drug treatment of CML. What design principles were applied to the development of Gleevec? How is this approach different from that used to develop conventional cancer therapies?

b) How does Gleevec work at the molecular level?

c) How many protein targets does Gleevec have? Name each target.

d) Besides target specificity, what other characteristics of a drug should you consider when designing a therapy?

In contrast to drug treatments, gene therapy attempts to correct the defect by introducing a functional copy of the malfunctioning gene that is responsible for the disease phenotype. The functional gene copy can be introduced directly into the diseased organ or can be used to genetically modify isolated tissue that is later re-introduced into the patient.

e) What do you need to know about the target disease in order to apply gene therapy?

f) How could you deliver the functional gene into the diseased cells?

g) What do you think are some of the challenges facing gene therapy?

h) Based on your understanding of oncogenes, why would *bcr-abl* be a particularly challenging target for gene therapy?

## Question 2

Duchenne muscular dystrophy (DMD) is an X-linked recessive disorder caused by mutations in the gene encoding dystrophin, a protein involved in maintaining membrane integrity in muscle cells. The dystrophin gene spans roughly 2.5 Mb and is spliced to form a 14 kb mRNA transcript consisting of 79 exons.

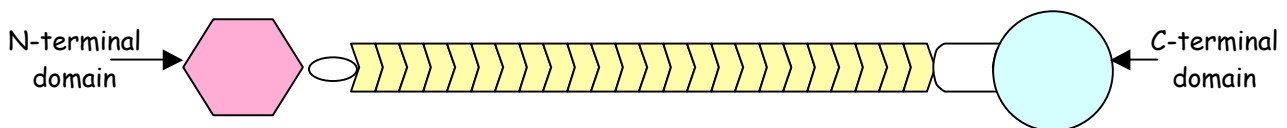
a) Is dystrophin is a typical human gene in terms of its size and exon count?

b) Why is DMD much more common in boys than in girls?

c) The DMD phenotype results from mutations that disrupt the reading frame of the dystrophin mRNA. What is the impact of such mutations on the dystrophin protein?

d) Based on what you know of dystrophin's normal function described above, what do you think are characteristics of the DMD disease phenotype?

The dystrophin protein contains several functionally distinct regions (domains). The N-terminal domain interacts with proteins inside the muscle cell, while the C-terminal domain enables dystrophin to bind to membrane-associated proteins. The long rod-like central part of the protein is composed of 24 identical rod domains (internal repeats). As a consequence of the modular structure of dystrophin, proteins missing some of the internal repeats can be fully functional or at least partly active.



One therapeutic strategy involves excluding, or skipping, an internal exon during splicing. In particular, skipping the exon that contains the frame-shift mutation in the DMD individual can result in a shorter than wild type but functional dystrophin protein. This exon-skipping therapy involves introducing a small RNA molecule that is complementary to the exon to be skipped. Binding of the RNA to the pre-mRNA prevents proper recognition of the exon by the splicing machinery and results in specific exon skipping.

e) In order for this therapy to work, what properties must this small RNA have? (Think about cellular localization.)

f) How would you deliver the RNA molecule into the target cell?

g) How is this therapeutic approach different from gene therapy?

h) What challenges do gene therapy and exon-skipping therapy have in common?

### Question 3

a) What percentage of genomic nucleotides do you expect two randomly chosen people to have in common?

Genome sequencing has revealed that the average genome nucleotide difference between two randomly selected chimpanzees is roughly four times greater than between two humans.

Images removed due to copyright reasons.

Based on population genetic theory, levels of genetic variation within species should correlate positively with population size. However, the human population numbers in the billions and the population size of chimpanzees is fewer than a hundred thousand.

b) How can you explain the comparatively little variation between human individuals?

c) What is genetic drift?

A few years ago, an international consortium was formed to uncover the locations of genetic variation in the human genome. In particular, the consortium worked to identify single nucleotide polymorphisms (SNPs) within the human population.

d) Is the genomic variation between individuals randomly distributed across the genome or does such variation occur at common sites?

e) What is a haplotype?

f) Why is an understanding of genomic variation useful for the study human health?

#### Question 4

Now that the first draft of the complete\* genomic sequence of the Chicken is available on the web (<http://www.genome.ucsc.edu/cgi-bin/hgGateway?org=Chicken>) you decide to start analyzing the sequence.

\* except for the sex chromosomes! (We're waiting on Winston for this.)

Luckily, there are a lot of resources available. After all, Chicken is an important model organism for the study of viruses, cancer, and developmental biology (and it's finger lickin' good...)

a) Genbank (available from <http://www.ncbi.nlm.nih.gov>) lists 559394 mRNA sequences derived from Chicken. Explain how you could use these sequences to find genes in the Chicken genome.

b) Does your method also provide information about the structure of genes (i.e. intron/exon boundaries, splice sites)? What are the possible limitations of your method?

c) Now that you know the locations of all of the autosomal genes in the Chicken genome, you would like to start trying to predict regulatory sequences. Unfortunately, Chicken is the only bird that has been sequenced. Why might this be a problem? How have regulatory sequences been found in other organisms, such as yeast and mammals?

d) You talk to Eric about your problem, and he offers to sequence 3 bird species to help you out. He's been thinking about sequencing some birds anyway, so he offers to let you help him decide which ones to pick. He is considering sequencing the ostrich, finch, quail, turkey, condor, pheasant, and guinea fowl. Which ones do you pick and why? (It is okay to use physical characteristics as an indicator of relatedness.)

Images removed due to copyright reasons.

### Question 5

There are several subspecies of the mouse, *Mus musculus*, living throughout the world. You decide to study their spread from an ancestral population using sequences from mitochondria and Y chromosomes.

a) Why are mitochondria and Y chromosomes often used for this purpose?

You isolate mitochondrial DNA from 10 mice from geographically diverse locations, and then sequence an essential mitochondrial gene. Dashes indicate bases that are identical to the top sequence.

|               |   |
|---------------|---|
| England       | AAAGCAGAAGAAATAGATGCAAAGGCAGAAGAAGAGTTCAA |
| Japan         | --GCAT-----C-----AATT                     |
| South Africa  | -----C-----TA-----                        |
| India         | GG-CG--C-----GGG----GG-----CGGG           |
| China         | TTT---T-----TT-----T----GGTT              |
| United States | -----C-----T-----                         |
| Philippines   | TTT---T-----TC-----T----GGTT              |
| New Zealand   | -----C-----C-----TA-----                  |
| Russia        | --GCAT-----CC-----A-TT                    |
| Yemen         | -----C-----                               |

b) Circle the region above that is most likely to be important for the function of the mitochondrial gene.

c) Construct a tree based on these mitochondrial DNA sequences. How many major subspecies of mice do you suppose that there are?

Next you decide to look at markers on the Y chromosome in mice. You study a highly polymorphic locus called Mick-Y (for Many Interesting Changes Known on the Y). You find that populations of mice in different locales have different alleles.

| Population    | Alleles of Mick-Y |
|---------------|-------------------|
| England       | A,B,C             |
| Japan         | L,M               |
| South Africa  | B                 |
| India         | G,H,I             |
| China         | J,K,L             |
| United States | C                 |
| Philippines   | K,L               |
| New Zealand   | B                 |
| Russia        | L,M,N,O           |
| Yemen         | A,B,C,D,E,F       |

d) Based on this information (and a little bit of knowledge about history) what do you think is the origin of the mouse populations in the United States, New Zealand, and South Africa?