Questions and Answers for Genetics and Genomics in Medicine Chapter 4

Question 1
Most of the constitutional variation in our DNA comes from endogenous sources. What are they?

Answer
- Errors in recombination.
- Errors in chromosome segregation.
- Errors in DNA replication.
- Damage to DNA from chemicals within cells including water molecules (hydrolysis) and reactive oxygen species.
- Errors in DNA repair.

Question 2
A single-strand DNA break is a problem for the cell, but a double-strand DNA break is often an emergency. Explain why.

Answer
In the case of a single-strand break, there is a transient problem but no immediate emergency because base-pairing with the unbroken DNA strand keeps the broken DNA strand in the correct place until it gets repaired, and the repair is helped by having the complementary unbroken DNA strand available to act as a template.

For a post-replicative double-strand DNA break (occurring after DNA replication when the two sister chromatids are held in close proximity for most of the time until mitosis), the chromatid with the double-strand break can be held in place, and the repair is helped by using a neighboring non-sister chromatid to act as a template.

For a pre-replicative double-strand DNA break, two double-stranded DNA fragments can readily drift apart; if the DNA repair is not done very quickly, it might be impossible to fix. That can be lethal if the break occurs in an important gene; if the fragments do drift apart, there is pressure to join the broken ends to other chromosomal DNA molecules nearby as an emergency solution, but that can lead to major problems and so if the repair is not fixed quickly the cell is often induced to commit suicide.

Question 3

In what circumstances do cells use the nucleotide excision repair pathway?

Answer

Generally when DNA damage takes the form of a bulky lesion on one DNA strand that distorts the double helix (and can block transcription and/or DNA replication). The DNA repair is conservative and involves breaks some distance on either side of the lesion to generate a moderately long oligonucleotide (30 nucleotides or so) that is discarded. The resulting gap is filled by DNA synthesis and ligation (see Figure 4.3B).

Question 4

Why should the repair of double-strand breaks in DNA be more accurate in cells after the DNA has replicated than before it has replicated?

Answer

After DNA replication, the sister chromatids are held closely together (aligned along their lengths) all the way up to metaphase, preventing drifting of the broken DNA fragments on the chromatid that sustained the double-strand break. The close presence of normal double-stranded DNA on its sister chromatid provides a perfect template for copying its sequence to make an accurate repair (homologous recombination-mediated DNA repair; see Figure 4.4).

If a double-strand break occurs before DNA replication (or in non-dividing cells at the G_0 phase), there is no sister chromatid to help maintain the integrity of the broken chromosomal DNA molecule and provide a template to direct the correct base insertion during DNA repair; if a repair is to be made, it needs to be done quickly before the broken ends drift apart. The repair is made by fusing the two broken ends together and is usually inaccurate.

Question 5

With regard to human genomic DNA, what is meant by a DNA polymorphism and an allele, and how is the term allele used differently when it refers to genetic variation at the protein level?

Answer

In practice, a DNA polymorphism means a locus where there are two or more alternative genetic variants that have a population frequency of 0.01 or greater. These variants would be referred to as alleles (in contrast, rare variants with a population frequency of less than 0.01 would be simply referred to as variants). The cutoff at 0.01 is somewhat arbitrary and is intended to exclude variants that have become more frequent simply by recurring mutation.

In the case of genetic variation at the protein level, the term allele is used to describe a variant irrespective of its frequency in the population.

Question 6
The pattern of single nucleotide substitution in the human genome is not random: as in the genomes of other vertebrates, there is a marked excess of C⟶T substitutions. Why?

Answer
DNA methylation in vertebrate genomes is very largely directed at the 5' carbon atom of cytosines that have a neighboring guanine on the same strand toward the 3' side (that is, occurring in the dinucleotide CG). The resulting 5-methylcytosine (5-meC) base pairs with guanine. In both cytosine and 5-meC an amino group is attached to the 4' carbon atom by a covalent bond that is vulnerable to hydrolysis. Loss of the 4' amino group from cytosine produces uracil, a base that is not normally present in DNA and is easily recognized as an unnatural base by the DNA repair machinery. However, loss of the 4' amino group from 5-meC produces thymine, a natural base in DNA that can go unrecognized by the DNA repair machinery (see Figure 4.5).

Question 7
Single nucleotide polymorphisms (SNPs) are distributed across the genome but occur at only a small minority of nucleotide positions. Why should only certain nucleotides be polymorphic and be surrounded by stretches of nucleotide sequences that rarely show variants?

Answer
In general, the nucleotides found at SNP sites are not particularly susceptible to mutation, and SNPs are stable over evolutionary time (some SNPs in our genome are known to have arisen as polymorphic sites before the divergence of humans from the great apes). Instead, the alternative nucleotides at SNP sites mark alternative ancestral chromosome segments that are common in the present-day population.

Question 8
Explain the general concept of natural selection. What type of selection is involved when we talk about selection pressure to conserve sequences and their functions, and how does it operate?

Answer
Natural selection means that a phenotype that confers increased ‘fitness’ (increased survival rates and increased reproductive success, resulting in a higher proportion of healthy progeny) leads to an increase in population frequency of alleles carried by the organism with the advantageous phenotype. When the advantageous phenotype is conferred by some allele or combination of alleles, allele(s) with this effect will increase in frequency within the population (simply because they will be transmitted more effectively to subsequent generations).

Natural selection often works to conserve an important sequence. That happens because the functions of genes with essential or important roles needs to be conserved to ensure that the organism is healthy and reproductively successful. To preserve their function, the sequence of the gene product must not change too much. Natural selection will ensure that harmful alleles (which cause loss of function or disrupted function of the important gene) will be reduced in frequency because they will

be associated with a lower fitness: people who carry them will have a reduced chance of reproductive success leading to healthy progeny. As a result there will be 'selection pressure' to conserve the sequence of the gene product. The form of natural selection that results in sequence conservation is known as *purifying selection* (or sometimes *negative selection*).

Question 9

What is positive selection? Give examples of individual genes in which positive selection appears to have led to an increase in heterozygosity, an increase in gene expression, or an increase in gene copy number.

Answer

Positive selection is a form of natural selection. Unlike in purifying (negative) selection (in which selection works to reduce the frequency of harmful genetic variants in the population), in positive selection natural selection works to produce an increase in frequency of some genetic variant that confers some phenotypic advantage. People with this variant will generally have higher survival rates and increased reproductive success compared with people who do not possess the variant, leading to increased transmission of the genetic variant.

The extraordinarily high level of polymorphism of classical HLA proteins (which function in presenting foreign antigens to lymphyocytes) is believed to be due to positive natural selection of individuals with individual alleles at different HLA loci. Positive selection is also believed to have led to an increased expression of the lactase gene *LCT* in adults to protect against lactose intolerance in populations that developed a cultural tradition of lifelong drinking of animal milk. Positive selection also appears to have led to an increased copy number of the salivary α-amylase gene, *AMY1A*, in populations that have developed starch-rich diets (see Figure 4.11).

Question 10

The relationship between genetic variation at the level of genes and proteins is complicated and is not simply due to whether genetic variants in a single gene cause amino acid substitutions or not. What other factors are involved?

Answer

Alternative forms of a protein (isoforms) are not always alleles: different isoforms of the same protein can be produced by non-allelic genes at duplicated gene loci. For example, isoforms of the HLA-DRβ protein are made by both the polymorphic *HLA-DRB1* and *HLA-DRB5* loci.

Alternative transcription or processing of an individual allele of a protein-coding gene can produce different protein isoforms. That can happen by the alternative use of promoters (to produce multiple dystrophin isoforms, for example), alternative splicing, RNA editing, and so on.

Protein variation reaches its greatest complexity in the case of immunoglobulins/antibodies and T-cell receptors because of specialized mechanisms that are required to enhance the diversity of these proteins. Post-zygotic genetic variation at each of the three immunoglobulin gene loci, *IGH*, *IGK*, and

IGL, and at each of the four T-cell receptor gene loci, *TRA*, *TRB*, *TRD*, and *TRG*, results in a huge number of different isoforms produced by an individual allele.

Question 11

What is genetic mosaicism, and how widespread is it?

Answer

Post-zygotic (somatic) genetic variation means changes in the DNA sequence that arise after the very earliest stages of development. When a mutation arises in a cell in early development, all cells that descend from that mutant cell will carry that mutation, and so a person will be a genetic mosaic with two different cell populations, one carrying the mutation in question and one carrying the equivalent normal sequence.

The mosaicism described above refers to a single mutation, but if we generalize and simply ask what percentage of people are genetic mosaics, the answer is 100%. That is, all of us carry some cells descended from a cell that received a post-zygotic mutation and other cells that lack that mutation. Just on the basis of mutation frequency alone and the numbers of cell division required in our development, we are mosaics for many different mutations.

For a disease-associated mutation, the chance that clinical symptoms will develop in a genetic mosaic is generally low, unless mutation occurs quite early in development so that there are very many cells with the mutation. If, when constitutional, the mutant allele were to be highly penetrant and expressed in the heterozygote, a key factor would be the proportion of mutation-containing cells in tissues in which the disease phenotype was normally expressed. Or if the mutation were to confer some cancer-associated growth advantage, it would be more likely to cause cancer when expressed in relevant progenitor cells.

Question 12

In human zygotes there are only six immunoglobulin genes, two alleles each at the *IGH*, *IGK*, and *IGL* loci, and yet each of us is able to make huge numbers of different antibodies. How is that possible?

Answer

Various post-zygotic diversity-generating mechanisms take place in maturing B lymphocytes to ensure that different B cells in a single person make different antibodies (soluble immunoglobulins). A major contribution is made by certain types of somatic recombination that occur at individual immunoglobulin loci in maturing B cells. In germline cells, immunoglobulin genes have multiple repeated gene segments of different types (V, J, and C for all immunoglobulin genes, plus additional D-gene segments in the case of the *IGH* gene locus). Each of many functional gene segments of a particular type, say V-gene segments, can be used to specify a defined component of the antibody, and recombinations occur at these loci in maturing B cells but in a cell-specific way to bring different combinations of the gene segments together. Because there are multiple repeats of the different gene segments, a large number of different combinations are possible.