Bi188 Spring 2013 Problem Set 1

Due: 3pm on April 16th, 2013

Question 1 (2.0 points)

The human genome has ∼22,000 annotated (known and confirmed) protein coding genes in the ∼3 billion base-pair genome. The average protein coding gene has its exons spread over ∼30kb of DNA. Give the major additional kinds of functional elements (including RNA-coding and non-coding) that comprise the remaining DNA sequence that is neither an exon nor an intron (give a minimum of 3 different kinds of elements that are NOT telomeres or centromeres), and explain what their functions are. This explanation should not be lengthy - three sentences and a cartoon where appropriate should do it.

Question 2 (3.0 points)

Olfactory sensitivity to specific smells (and taste, because olfaction confers a very large fraction of the taste experience) has genetic components. Olfactory sensory neurons each express only one olfactory receptor gene from among a large family of paralogous genes (several hundred) (how a gene gets selected for expression will be a vignette for a later lecture). These OR genes are distributed on many chromosomes, but they often occur within local OR clusters. The coding sequences (ORs are 7-pass G-coupled receptors) differ from one OR gene to another, so that the proteins encoded have unique binding domains for specific odorant molecules; however outside the odorant-binding extracellular domain, the genes have considerable sequence similarity to each other. This provides rich opportunity for structural variation at these loci, with individual genes being duplicated, deleted, diverged and sometimes mutated into pseudogenes. (We can also see this at work in evolution, as humans have numerous deleted OR genes and pseudogenes in places where rodent OR family members remain functional)

Now, consider variation among humans in a family being taken out for Thai food for the first time. A spirited discussion ensues about whether three dishes, tasted by all persons at the dinner, are delightful, or aggressively horrible, or incredibly bland and tasteless. The dishes are lemongrass chicken, tamarind pork, and garlic chicken. Below are subjective comments on each dish by three generations: grandparent, parents, and children. To keep you focused, all of these people are known to be OK with chicken and pork as basic foods, so that is not a critical differentiator.

More objectively, you also have independent information on how these people scored in an olfactometer test for a major olfactory component of lemongrass (L), tamarind (T), and garlic (G) (the implied chemistry here is made up for the problem). The olfactometer readings are linear in sensitivity and are arbitrarily scaled around a large population mean value that is arbitrarily normalized to 10 units. Zero units is no detection. A score for most people would therefore be ($L = 10, T = 10, G = 10$).

1. Gina is introduced to you as the matriarch, mother to Rose and grandmother to the four students. Her scores are ($G = 2, L = 26, T = 3$). Of the food, she says “Almost nothing has much taste anymore, but that Lemongrass Chicken is worth getting for take-out”.

2. Grandpa is not at dinner. No information from the olfactometer. No comments. This is typical in human genetics problem - missing data. But you must go on with what you have.

3. Gina’s daughter, Rose, mother to the students at dinner. Her scores are ($G = 10, L = 39, T = 4$). Fill in what she likely said about the three dishes.

4. Gina's Husband, George ($G = 11, L = 5, T = 4$): "The Lemongrass Chicken and the Pork are kind of blah, but the other chicken is great".

5. Student generation of 4, are introduced to you as the children of Gina and George

   (a) Sue ($L = 41, T = 1, G = 11$): "Yuk, the lemongrass chicken is inedible! But the others are.." `fill in the blank` ....

   (b) Rick ($L = 11, T = 10, G = 9$): "All three dishes are fabulous!". Interesting, subtle, and we must come back.

   (c) Jeff ($L = 4, T = 5, G = 11$): "This food is' ....

   (d) Tom ($L = 20, T = 9, G = 9$): "Its all pretty good, but the lemongrass chicken is a bit pungent"

For each person (except absent grandpa) generate a candidate map for the OR genes for L, T, and G loci assuming that the tamarind, lemongrass, and garlic receptors are located on independent chromosomes. (Use the shorthand notation of a line for the chromosome, and along it note "OR-L", "OR-T" and "OR-G" for a single functional gene encoding each of the types; for each show the maternal and paternal chromosome explicitly. Your proposed map should assume that olfactory sensitivity, as scored empirically, reflects the frequency of neurons dedicated to a given odorant, and that this in turn governs sensitivity (ie this is a simple gene dosage effect).

Obviously, you need some different explanation for Tom. (It is not, however, a bad data measurement). Give one other reasonable explanation, and propose what the resulting chromosomes look like, consistent with your explanation.

Briefly provide reasonable subjective verdicts on the three dishes for the people who have not given a recorded opinion on each dish, and say why (in a phrase).

Question 3 (2.5 points)

You observe unrelated human individuals with mutations that all alter (variously) the DNA sequence within a 500bp region arbitrarily called for this problem the USX sequence. USX is located $\sim$1 million base pairs from the SHH (sonic hedgehog) promoter. Homozygous USX mutations are associated with a phenotype of abnormal limb development. If we tell you that this phenotype is similar to the phenotype one seen after experimentally disabling the protein coding capacity of mouse shh in the developing limb (but not elsewhere), give the basic function of the region. Do you expect homozygous deletion of the protein coding domain in the mouse to have the same phenotype? If not, how and why do you expect it to differ?