Module 3: Genetics

Review Activity

Mouse Colouration

The Rock Pocket Mouse (Info from the video; you can skip this):

The rock pocket mouse, *Chaetodipus intermedius*, is a small, nocturnal animal found in the deserts of the south-western United States. Most rock pocket mice have a sandy, light-coloured coat that enables them to blend in with the light color of the desert rocks and sand on which they live. However, populations of primarily dark-coloured rock pocket mice have been found living in areas where the ground is covered in a dark rock called basalt caused by geologic lava flows thousands of years ago. Scientists have collected data from a population of primarily dark-coloured mice living in an area of basalt called the Pinacate lava flow in Arizona, as well as from a nearby light-coloured population. Researchers analyzed the data from these two populations in search of the genetic mutation responsible for the dark color. Their analysis led to the discovery of a mutation in the *Mc1r* gene which is involved in coat-colour determination.

The *Mc1r* Gene:

The coat colour of rock pocket mice is determined by two pigments: *eumelanin*, which is dark-coloured; and *pheomelanin,* which is light-coloured. The synthesis of these pigments is controlled by the products of several genes, including the *Mc1r* gene. The mouse *Mc1r* gene is located on mouse chromosome 16 (rock pocket mice have n=23 just like humans), and encodes a protein called the Melanocortin-1-Receptor (MC1R), which you have seen in each of the last two review activities. This receptor is found embedded in the membrane of specialized cells called melanocytes, which you have also examined. The melanocytes of wild-type (non-mutant) mice produce much more pheomelanin than eumelanin (or almost no eumlanin); the result is a sandy-coloured mouse. The mutated allele of the *Mc1r* gene, however, triggers melanocytes to increase the production of eumelanin, resulting in the dark coat-colour phenotype.

Wild-type *Mc1r* allele (light phenotype)

Below are five 15 base DNA nucleotide sequences from the wild-type (light coat colour) *Mc1r* allele template DNA strand. Use the sequences provided to determine the complementary mRNA sequence and the translated amino acid strand. Note: the actual gene contains 951 base pairs (317 amino acids). The amino acid position in the protein sequence is provided for each segment.

Extracellular Domain I
Amino Acids 015 → 019

Template Strand: 3'–TTGAGGTGGGCGTGT–5'

mRNA Strand (identify 5',3'): **5'–AAC UCC ACC CGC ACA–3'**

Amino Acid Strand (identify N,C): **N-Asn-Ser-Thr-Arg-Thr-C**

Extracellular Domain III
Amino Acids 105 → 109

Template Strand: 3'–CGGGACCGGTGGGCC–5'

mRNA Strand (identify 5',3'): **5'–GCC CUG GCC ACC CGG–3'**

Amino Acid Strand (identify N,C): **N-Ala-Leu-Ala-Thr-Arg-C**

Intracellular Domain I
Amino Acids 160 → 164

Template Strand: 3'–GCCCGAGCCACCGCC–5'

mRNA Strand (identify 5',3'): **5'–CGG GCU CGG UGG CGG–3'**

Amino Acid Strand (identify N,C): **N-Arg-Ala-Arg-Trp-Arg-C**

Transmembrane V
Amino Acids 210 → 214

Template Strand: 3'–TACGAACGTGGGGAG–5'

mRNA Strand (identify 5',3'): **5'–AUG CUU GCA CCC CUC–3'**

Amino Acid Strand (identify N,C): **N-Met-Leu-Ala-Pro-Leu-C**

Intracellular Domain III
Amino Acids 230 → 234

Template Strand: 3'–GAACAGGTGGTTCCA–5'

mRNA Strand (identify 5',3'): **5'–CUU GUC CAC CAA GGU–3'**

Amino Acid Strand (identify N,C): **N-Leu-Val-His-Gln-Gly- C**

Mutant *Mc1r* allele (dark phenotype)

The sequences below are for the mutant (dark coloured) *Mc1r* allele. There are 5 mutations in this allele (one per sequence). Compare the DNA sequences of the wild-type and mutant *Mc1r* alleles to identify the locations of these mutations. **You only need to transcribe/translate the 3-base codons with the mutations and the corresponding amino acids!**

Extracellular Domain I
Amino Acids 015 → 019

Template Strand: 3'–TTGAGGTGGACGTGT–5'

mRNA mutated codon: **5'–AAC UCC ACC UGC ACA–3'**

Altered Amino Acid: **N-Asn-Ser-Thr-Cys-Thr-C**

Extracellular Domain III
Amino Acids 105 → 109

Template Strand: 3'–CGGGACCGGTGGACC–5'

mRNA mutated codon: **5'–GCC CUG GCC ACC UGG–3'**

Altered Amino Acid: **N-Ala-Leu-Ala-Thr-Trp-C**

Intracellular Domain I
Amino Acids 160 → 164

Template Strand: 3'–ACCCGAGCCACCGCC–5'

mRNA mutated codon: **5'–UGG GCU CGG UGG CGG–3'**

Altered Amino Acid: **N-Trp-Ala-Arg-Trp-Arg-C**

Transmembrane V
Amino Acids 210 → 214

Template Strand: 3'–TACGAGCGTGGGGAG–5'

mRNA mutated codon: **5'–AUG CUC GCA CCC CUC–3'**

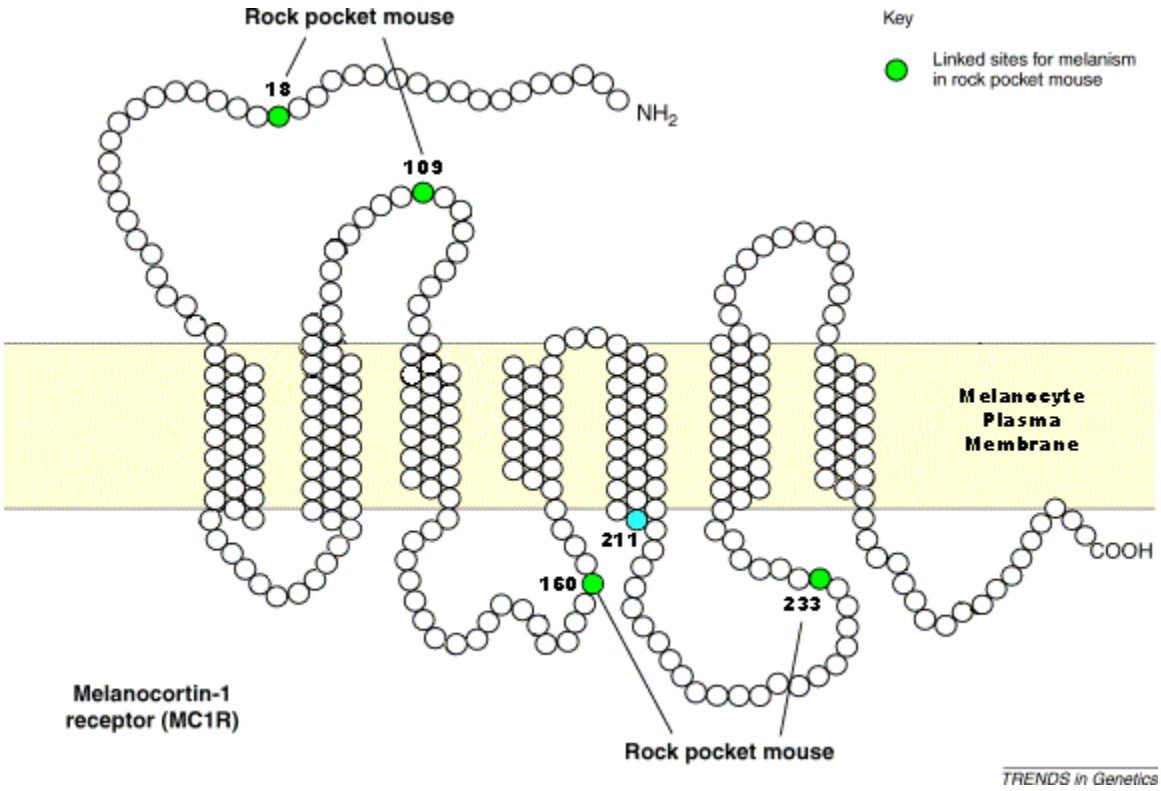
Altered Amino Acid: **N-Met-Leu-Ala-Pro-Leu-C**

Intracellular Domain III
Amino Acids 230 → 234

Template Strand: 3'–GAACAGGTGGTGCCA–5'

mRNA mutated codon: **5'–CUU GUC CAC CAC GGU–3'**

Altered Amino Acid: **N-Leu-Val-His-His -Gly- C**

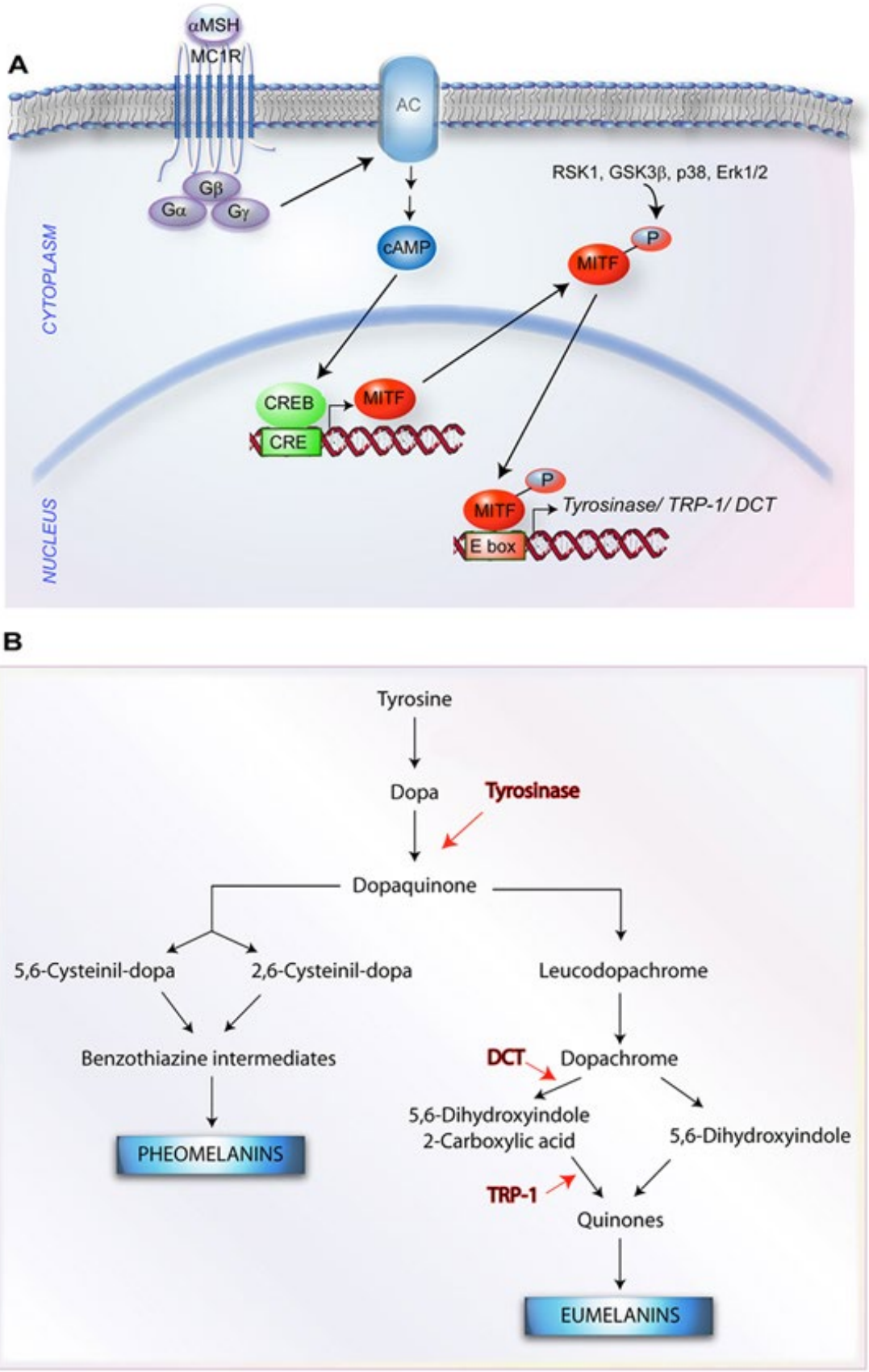


1. The five codons with mutations correspond to amino acids 18, 109, 160, 211, and 233. Explain why the mutation at codon 211 is not as significant as the other mutations.

   Mutation 211 is a silent point mutation. This means that the altered nucleic acid base doesn't actually result in a change in amino acid, and so there is ultimately no effect on the protein or on phenotype. An organism can sustain any number of silent mutations, because they ultimately have no effect one way or another.

2. Complete the table below comparing the chemistry of amino acids in the wild-type MC1R protein and the mutant MC1R protein.

Amino Acid Mutation Position Number	Wild-type MC1R Amino Acid Chemistry	Mutant MC1R Amino Acid Chemistry
211	***Non-Polar***	***Non-Polar***
18	**+ Charged**	**Non-Polar**
109	**+ Charged**	**Non-Polar**
160	**+ Charged**	**Non-Polar**
233	**Polar**	**+ Charged**


A
αMSH
MC1R
AC
Gβ
Gα
Gγ
cAMP
RSK1, GSK3β, p38, Erk1/2
P
MITF
CYTOPLASM
CREB
MITF
CRE
P
MITF
Tyrosinase/ TRP-1/ DCT
E box
NUCLEUS
B
Tyrosine
Dopa
Tyrosinase
Dopaquinone
5,6-Cysteinil-dopa
2,6-Cysteinil-dopa
Leucodopachrome
Benzothiazine intermediates
DCT
Dopachrome
5,6-Dihydroxyindole
2-Carboxylic acid
5,6-Dihydroxyindole
PHEOMELANINS
TRP-1
Quinones
EUMELANINS

As you should have determined from one of the previous review activities, the Melanocortin-1-Receptor (MC1R) protein is a trans-membrane receptor protein involved in a typical cell communication pathway (diagram A above). In other words, this protein receives signals from outside the cell, and activates molecular pathways in the cell when it is triggered. This type of receptor contains an extracellular binding site for a signal molecule, and an intracellular binding site for internal pathway molecules.

For MC1R, the signal molecule is a hormone called melanocyte stimulating hormone (α-MSH; see it bound to MC1R in diagram A). When α-MSH binds to MC1R, this protein changes its shape, and its intracellular portion is then in the proper shape to bind to and activate an internal pathway protein (the three 'G' molecule complex also shown bound to MC1R in diagram A), which, through a complex set of reactions, ultimately results in the cell producing a protein enzyme called *tyrosinase*.

This enzyme then enters into the metabolic pathway outlined in diagram B. When tyrosinase is present, it interacts with and alters a molecule called *dopaquinone*, ultimately sending this molecule along a metabolic pathway that results in the molecule becoming the dark coloured pigment eumelanin. When tyrosinase is not present, dopaquinone proceeds along a different metabolic path and eventually becomes the light pigment pheomelanin.

3. Using your knowledge of mutations, amino acids, and proteins, develop hypotheses to explain the following:
    1. How the extracellular mutations result in a dark phenotype (hint: think about the chemistry of amino acids, particularly their charge).

        The extracellular mutations result in the MC1R protein being able to bind more effectively with the α-MSH signal molecule. When these molecules are bound together, more internal signal is sent to produce eumelanin instead of pheomelanin.

    2. How the intracellular mutations result in a dark phenotype (hint: think about the chemistry of amino acids, particularly their charge).

        The intracellular mutations result in the MC1R protein being able to bind more effectively with the G-protein complex. When these molecules are bound together, more internal signal is sent to produce eumelanin instead of pheomelanin.

    3. How the wild-type MC1R proteins leads to the light phenotype (hint: it might be helpful to think of the wild-type protein *NOT* leading to the dark phenotype).

        The wild type MC1R does not bind effectively with either the α-MSH signal molecule or the G-protein complex. The result is that very little internal signal is sent to produce tyrosinase, and so the cell produces pheomelanin instead of eumelanin.

4. The creation of only a modest amount of tyrosinase in a mouse's melanocyte cells is sufficient to produce enough eumelanin for export into the hair follicles that the mouse will have dark fur. Based on this information, as well as the information presented to this point and your answers to the questions above, what do you believe is the pattern of inheritance for the mutant *Mc1r* allele at the molecular, cellular, and organismal levels?
   - Molecular:

     Melanocyte cells in heterozygous mice (mice with both the wild and mutated *Mc1r* alleles) produce both the normal and mutated versions of the MC1R proteins, and both of these protein types get embedded in the plasma membrane. Thus, if you examine the melanocytes of heterozygous mice you will find both types of proteins equally expressed, separate, and distinct from each other. Thus, on a molecular level, the *Mc1r* allele is co-dominant.
   - Cellular:

     The paragraph above states that the creation of only a modest amount of tyrosinase is enough to produce enough eumelanin to make a mouse look dark. As described for the previous answers, heterozygous mice will have both the wild and mutated MC1R proteins in their membranes. The wild protein will not bind to either the αMSH molecule or the G-protein complex, and thus will essentially not result in the production of any tyrosinase. However, the mutated version of the MC1R protein that the mice have in their membranes will bind to both molecules, and will produce tyrosinase. While the ultimate amount of tyrosinase produced in a heterozygous mouse will only be about half of what is produced in a mouse homozygous for the dark allele, this amount of tyrosinase will result in enough eumelanin for the cell to look dark. Thus, if you examine the melanocyte cells of heterozygous mice they will look equally as dark as the melanocyte cells of homozygous dark mice, meaning that, on a cellular level, the mutated *Mc1r* allele inherits as an autosomal dominant (remember that the gene is found on chromosome 16, not on an autosome).
   - Organismal:

     The answer provided above for the cellular phenotype is essentially the same answer here. If the cells of heterozygotes are the same colour as the cells of homozygous dark mice, then, on an organismal level, heterozygous mice will be dark, and the mutated *Mc1r* allele inherits as an autosomal dominant.

5. With our current technology and understanding, it's easy to derive hypotheses about patterns of inheritance based on a reasonably elaborate understanding of how certain gene products function. Now, though, imagine that you are a genetic researcher from 90 years ago who is interested in determining the organismal pattern of inheritance for the *Mc1r*. Design a study whose products should be able to indicate to you how this gene inherits.

   With numerous repeats (as many as possible, 50+ ideally), perform crosses between P generation pure breed dark mice and pure breed light mice, mixing the sexes (some crosses are dark male X light female, some are dark female X light male). These will produce an F1 generation that are hybrids. Cross these with their siblings to produce F2 generation. In this case *Mc1r* is autosomal, so sex wouldn't matter, but you would want to distinguish between the sex crosses to determine whether the gene was sex-linked.

6. The table below presents some of the F2 products of the study you described above.

Dark Mice	Light Mice
142	58

   How would you statistically analyze these data? Which statistical test should you use, what kind of output would you expect, and what would you report in a paper?
   The study above isn't an experiment with a manipulated independent variable. The dependent variable is the mice count in the different categories. This should be analyzed using a chi square test. The expected offspring ratio would be 3:1 (this is a monohybrid cross), and the observed values above would be compared to the perfect expected 3:1 ratio distribution for the number of offspring obtained (142+58=200, so 150 and 50). The Excel output would just be a p-value, and that's all you need to report in this class. In a real paper you might be required to report the test statistic (chi square value) and the degrees of freedom.

7. Suppose you performed the statistical test you indicated in the previous question and obtained a p-value of 0.76. What conclusions can you draw from this result?
   A high p-value indicates a high probability of the null hypothesis being correct. For any chi square, the null hypothesis is that there is no difference between the observed and expected distributions. Here, since the expected distribution was the perfect 3:1 ratio for a dominant/recessive monohybrid cross, which is the hypothesis, a high p-value supports the hypothesis because there is no statistical difference between this and your observed.

8. If a mouse that is heterozygous at the *Mc1r* gene locus mates with a light mouse, approximately what proportion of their offspring would you expect to be dark coloured?

   From the previous questions you should have determine that the dark allele is dominant. Thus, the heterozygote has both alleles (*Dd*), while the light mouse must be homozygous recessive (*dd*), because having two copies of the recessive allele is the only way that these mice will not be dark. So, the cross is *Dd* X *dd*, and a simply Punnet square will identify that 50% of the babies will inherit the *D* dark allele, and will thus be dark coloured.

Hemophilia

Two prospective parents are meeting with a genetic counsellor because of the presence of factor VIII deficiency hemophilia in both of their families. Factor VIII is a protein that helps the blood to clot, and when a person's factor VIII level is very low, even the smallest cuts can be troublesome, and internal bleeding is common. Complications include swelling, joint damage, and an increased likelihood of neurological complications due to intracerebral bleeding.

Neither of the two prospective parents suffer from this disorder, but both have close family members who do. Since they are now thinking about starting a family of their own, they are therefore concerned about the risks of passing on genetic diseases to their children. For example, they know that hemophilia A is an inherited disease; the prospective mother's father is also red-green colour-blind, and they know that this condition runs in families as well.

As a first step, the genetic counselor asks them to fill out a narrative history listing their relatives, relationships, and if they were affected by any genetic diseases that they know of:

NAME: Greg

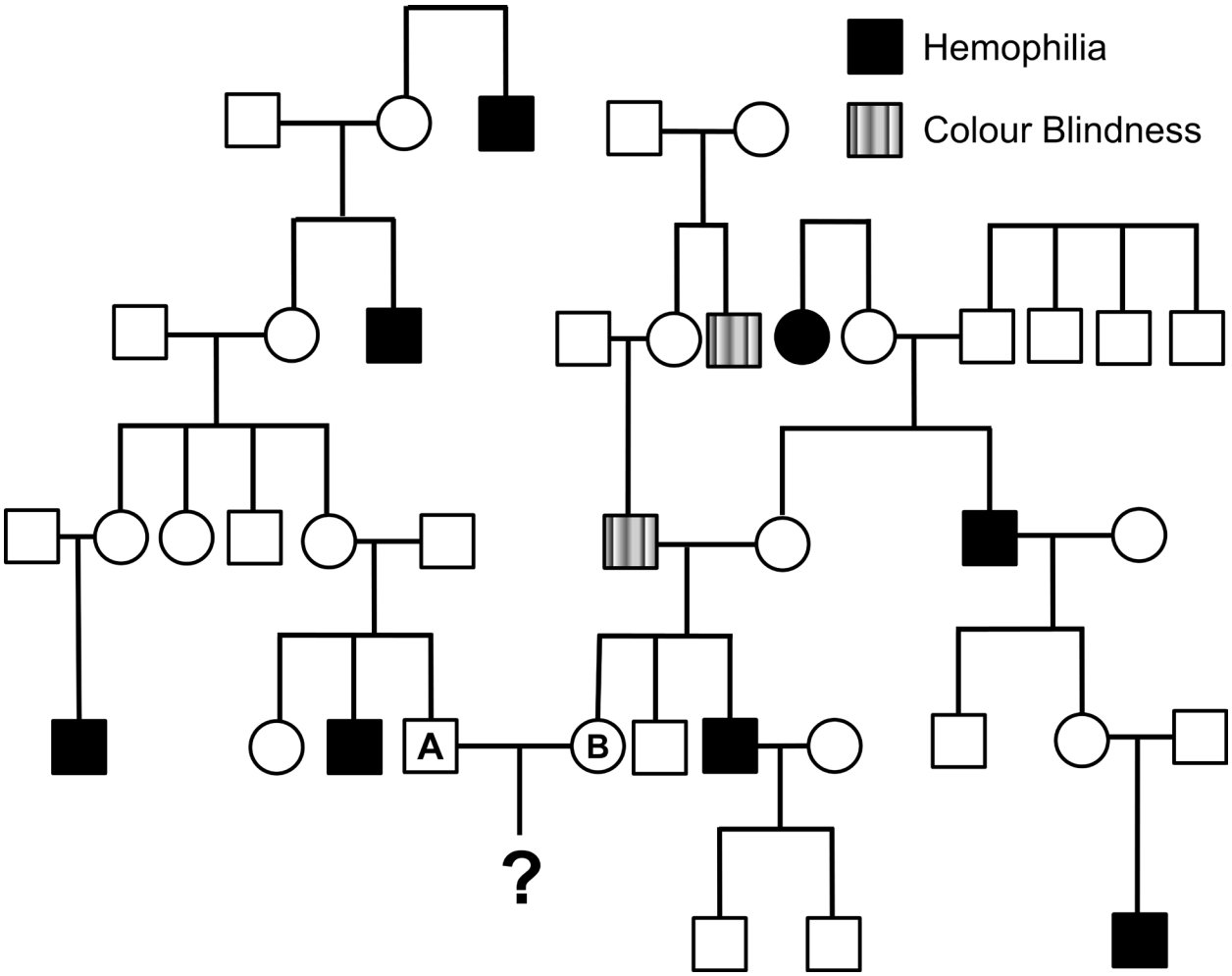
I have one brother and one sister, neither of whom are married. My brother suffers from factor VIII deficiency, but no one else in my immediate family does. My mother has two sisters and one brother, all of whom are normal, but one of my aunts has a son (my cousin Tim) who also has this disorder; all of my other cousins are normal. Both of my maternal grandparents are normal, but my grandmother had a brother who presumably had this deficiency (he wasn't able to clot properly and died very young). My great grandmother on my mom's side also had a brother who died very young because he was sick, but none of my relatives have actually been able to confirm that he suffered from this disorder. My father is completely normal. He was adopted from an orphanage and nothing is known about his family.

NAME: Olga

I have two brothers, one of whom has factor VIII deficiency. The brother with the disease is married to a woman who does not have the disease. They have two young boys, both normal. My father is an only child who does not suffer factor VIII deficiency. His father is also an only child, but his mother has a brother, none of whom suffer from any hemophilia. They are all still living. My maternal grandmother is healthy and had a sister who died from this just after birth. She married my grandfather who was one of four children, all boys, none of whom were affected by any disease that anyone is aware of. My grandparents had two children, my mother and my uncle. My uncle has hemophilia but my mom doesn't. My uncle married my (normal) aunt and they had two children, neither of whom showed any sign of any disease. Their boy is still single but their girl got married, to a normal man, and had a son, who has hemophilia A.

My dad is red-green colour blind, but neither of my brothers are. My dad's parents don't have this problem either, but his uncle does. Nobody on my mom's side has this.

The pedigrees for both of these narratives are provided below, where individual ‘A’ is Greg and individual ‘B’ is Olga. However, you can (and should) try to construct these yourself at home based on the narratives... it’s good practice.



The genetic counsellor is familiar with hemophilia and factor VIII deficiency, but decides to do her due diligence regardless and do some background research. What she finds is that the gene encoding the factor VIII protein is called *F8*, and that this gene is expressed primarily in the liver. Once exported from liver cells, the factor VIII protein circulates in the bloodstream in an inactive form, until an injury that damages blood vessels occurs. In response to injury, coagulation factor VIII is activated, and the active protein sets off a chain of additional chemical reactions that form a blood clot. The hemophilia condition arises due to a mutation in the *F8* gene which results in the production of a non-functional factor VIII protein. However, the quality control mechanisms of the cell don't identify this protein as being faulty, and so cells in the liver produce and excrete it in exactly the same manner as they do the normal factor VIII protein. The faulty protein simply circulates in the bloodstream until it is degraded, but has no discernible action during this time.

4. Based on all of the information provided to this point, including the pedigree of the two families and a description of the function of factor VIII, what conclusion do you think the genetic counsellor would come to with regard to the pattern of inheritance for hemophilia A on an organismal level? What does this imply with regard to the location of the *F8* gene within the human genome?

   The *F8* gene is located on the X chromosome, and the hemophilia allele is recessive. Thus, the pattern of inheritance is X-Linked recessive. This can be identified from the pedigree because many more males are affected than females and because parents are normal but have affected children, and also from the information provided in that the factor VIII protein doesn't do anything, and its effect will be masked by the function of a normal factor VIII protein in a heterozygous (female) individual.

5. How would you describe the pattern of inheritance for the normal and hemophilia *F8* alleles on a molecular level?

   The alleles are co-dominant on a molecular level because they both get produced and secreted into the bloodstream. Thus, in the blood of a heterozygous (female) individual, there are equal amounts of both the normal and mutated factor VIII proteins, expressed separately, distinctly, and equally.

6. The paragraph above states that the mutated factor VIII protein has no discernible action, although, of course, by not doing what it's supposed to do it results in the hemophilia phenotype. Imagine, instead, that the mutated protein still cannot form blood clots, but that its new shape results in it interacting with the proteins in myocyte muscle cells, ultimately leading to increased muscle mass in people with this *F8* mutation (this likely wouldn't be possible, but just run with this as a hypothetical). In this situation, what biological term would you use to describe the action of the mutated *F8* allele?

   This would be an example of a pleiotropic effect, as the mutated *F8* allele would have effects on multiple and seemingly unrelated phenotypes.

7. If you're reading this question without having answered number 4, note that the answer for question 4 can actually be found in the description below. Do not keep reading if you want to legitimately challenge yourself to answer question 4 correctly without assistance.

   The *F8* gene is located on the X chromosome, and hemophilia inherits in an X-linked recessive manner (on an organismal level). The gene whose mutated version results in colour blindness (*CB*) also happens to be on the X chromosome. There is a 25% recombinant frequency between *F8* and *CB*. Based on the pedigree created by the genetic counsellor (above), calculate the probability of Greg and Olga having a son who is both colour blind and has hemophilia.

   Both genes are X-linked, and the question asks about the probability of Greg and Olga having a son. In this scenario, Greg needs to pass the Y chromosome on and not the X, which will occur with a 50% probability.

   Then, because the son only gets one X and this is from mom, any alleles on this chromosome will be expressed. So, for the son to be colour blind and have hemophilia, what needs to happen is Olga needs to pass on a chromosome that has the colour blind allele and the hemophilia allele. However, it is guaranteed that she does not possess such a chromosome, as the X chromosome with the colour blind allele came from her father (colour blindness only on the father's side), while the X chromosome with hemophilia would have come from her mother (hemophilia only on the mother's side). Thus, even if she does indeed possess both alleles, they will be on homologous chromosomes.

   Of course, the possibility exists that crossing over could recombine the alleles and get them both on the same X chromosome. So, we need to figure out the probability of Olga having both alleles, AND the probability that crossing over would occur, AND the probability that Olga would pass on the chromosome with both alleles of interest.

   First, Olga's father is colour blind, which means his only X chromosome, which he gave to his daughter, includes the colour blind allele. So, she has this allele with 100% certainty. Next, we see that Olga's brother has hemophilia, which means he must have inherited the allele on the X chromosome he got from his mother. Since Olga's mother does not display the disorder, this means she is a carrier (heterozygous). Thus, she would have a 50% chance of having passed on this allele to Olga, meaning Olga has a 50% probability of having the hemophilia allele. If she does have both alleles, these would be on opposite chromosomes, but there is a 25% probability that crossing over will occur and put them on the same chromosome. Finally, if all of this happens, there is a 50% probability of passing on the recombinant chromosome with both alleles of interest.

   Final tally: 1/2 X 1 X 1/2 X 1/4 X 1/2 = 1/32

   There is a 1/32 probability of having a child with both disorders.

8. The *Mc1r* gene described in the first section of this review for mice can also be found in a similar version in humans. In our species, *Mc1r* is also a gene that affects pigmentation, and a particular mutant variant of this gene results in red hair. This particular mutant inherits as an autosomal recessive. Imagine that Greg from the story above has red hair, and that Olga has brown hair, but carries a copy of the recessive *Mc1r* red hair allele. What would be the probability of them having a red haired daughter?

   To have red hair, the daughter needs to receive a mutant *Mc1r* from both her mom and dad. Greg is homozygous recessive, with two copies of this allele, so she will get the allele from him with 100% certainty. Olga, on the other hand, is heterozygous. Thus, there is a 50% probability of her passing on the mutant allele to her daughter. That's all you need to know in terms of the *Mc1r* allele.

   However, the question is also phrased such that it is asking you the probability of the two parents having a daughter in the first place. Thus, you must incorporate the probability of this, which is 50%.

   Final tally: 1 X 1/2 X 1/2 = 1/4

Hemophilia: Extra Practice

The following section is a neat activity related to the hemophilia story above which can give you extra practice transcribing and translating genetic instructions, as well as identifying the results of mutations. Plus, it gives you a tiny peak into the real world of genetics. If you have the time, and want the extra practice, I strongly suggest running through this activity at home.

To get a better idea of exactly what mutation (i.e., what allele) is prevalent in the couple's two families, the genetic counsellor orders tests which will provide her with the exact DNA sequences that Greg and Olga have for the *F8* gene. The results of this test, for Greg (for the sake of simplicity we will just assume that Olga is the same), are provided on the following page.

Based on these results, your initial task is to take the genetic counsellor's next step, which is to compare Greg's *F8* allele sequence to that of the regular *F8* allele, to first see if you can identify any differences, and then, if you do find differences, to determine how these could potentially affect the factor VIII protein. To do this, you will need to use the National Center for Biotechnology Information site, http://www.ncbi.nlm.nih.gov/ :

- From the main page, in the dropdown box near the top-left which says 'All Databases', select 'Nucleotide'.
- Then, in the search box, enter the following code and click 'Search': NM_019863.2
- You should end up on a page titled: **Homo sapiens coagulation factor VIII, procoagulant component (F8), transcript variant 2, mRNA**
- Look over the page, but don't get too bogged down, as there is A LOT of technical information. Note that midway down the page there is a translated sequence for the factor VIII protein, and at the very bottom is the entire 2617 base pair sequence.
- From the top of the page, in the 'Analyze this sequence' menu on the right, select 'Run BLAST' (Basic Local Alignment Search Tool).
- From the BLAST page, note that the accession number you previously inputted is now in the 'Enter Query Sequence' box.
- Within the 'Enter Query Sequence' field, tick the box at the bottom that says 'Align two or more sequences'. This should open up a new 'Enter Subject Sequence' field.
- Copy/paste the entire sequence from the following page (i.e., everything on the next page, including the numbers) into the 'Enter Subject Sequence' box.
- Scroll to the bottom of the page, and click the 'BLAST' button in the bottom left corner.
- In the resulting page, scroll down to see the 'Alignments' section. Scroll through this and try to identify where the two sequences are not aligned – this will be the mutation in Greg's *F8* allele. Note that the top sequence is the normal allele, the bottom sequence is Greg's allele, and that all aligned bases have lines between them.

Greg's *F8* gene sequence:

1 gcgtcccct cggcgggctg ccgccgtgcc cgcgccggct ccccagcccg agcctgcccc
61 ttgccctgat gaggtgcaaa gagcgggatc ggaggcgggg cctggccggg ctgtgagcgg
121 cgtatgcaaa tcgagggtct cggggatgcg gatccaagac cctgggaagg tcttctttgg
181 caatgtggat tcatctggga taaaacacaa tatttttaac cctccaatta ttgctcgata
241 catccgtttg cacccaactc attatagcat tcgcagcact cttcgcatgg agttgatggg
301 ctgtgattta aatagttgca gcatgccatt gggaatggag agtaaagcaa tatcagatgc
361 acagattact gcttcatcct actttaccaa tatgtttgcc acctggtctc cttcaaaagc
421 tcgacttcac ctccaaggga ggagtaatgc ctggagacct caggtgaata atccaaaaga
481 gtggctgcaa gtggacttcc agaagacaat gaaagtcaca ggagtaacta ctcagggagt
541 aaaatctctg cttaccagca tgtatgtgaa ggagttcctc atctccagca gtcaagatgg
601 ccatcagtgg actctctttt ttcagaatgg caaagtaaag gtttttcagg gaaatcaaga
661 ctccttcaca cctgtggtga actctctaga cccaccgtta ctgactcgct accttcgaat
721 tcacccccag agttgggtgc accagattgc cctgaggatg gaggttctgg gctgcgaggc
781 acaggacctc tactgagggt ggccactgca gcacctgcca ctgccgtcac ctctccctcc
841 tcagctccag ggcagtgtcc ctccctggct tgccttctac ctttgtgcta aatcctagca
901 gacactgcct tgaagcctcc tgaattaact atcatcagtc ctgcatttct ttggtggggg
961 gccaggaggg tgcatccaat ttaacttaac tcttacctat tttctgcagc tgctcccaga
1021 ttactccttc cttccaatat aactaggcaa aaagaagtga ggagaaacct gcatgaaagc
1081 attcttccct gaaaagttag gcctctcaga gtcaccactt cctctgttgt agaaaaacta
1141 tgtgatgaaa ctttgaaaaa gatatttatg atgttaacat ttcaggttaa gcctcatacg
1201 tttaaaataa aactctcagt tgtttattat cctgatcaag catggaacaa agcatgtttc
1261 aggatcagat caatacaatc ttggagtcaa aaggcaaatc atttggacaa tctgcaaaat
1321 ggagagaata caataactac tacagtaaag tctgtttctg cttccttaca catagatata
1381 attatgttat ttagtcatta tgaggggcac attcttatct ccaaaactag cattcttaaa
1441 ctgagaatta tagatggggt tcaagaatcc ctaagtcccc tgaaattata taaggcattc
1501 tgtataaatg caaatgtgca tttttctgac gagtgtccat agatataaag ccatttggtc
1561 ttaattctga ccaataaaaa aataagtcag gaggatgcaa ttgttgaaag ctttgaaata
1621 aaataacaat gtcttcttga aatttgtgat ggccaagaaa gaaaatgatg atgacattag
1681 gcttctaaag gacatacatt taatatttct gtggaaatat gaggaaaatc catggttatc
1741 tgagatagga gatacaaact ttgtaattct aataatgcac tcagtttact ctctccctct
1801 actaatttcc tgctgaaaat aacacaacaa aaatgtaaca ggggaaatta tataccgtga
1861 ctgaaaacta gagtcctact tacatagttg aaatatcaag gaggtcagaa gaaaattgga
1921 ctggtgaaaa cagaaaaaac actccagtct gccatatcac cacacaatag gatccccctt
1981 cttgccctcc acccccataa gattgtgaag ggtttactgc tccttccatc tgcctgaccc
2041 cttcactatg actacacaga atctcctgat agtaaagggg gctggaggca aggataagtt
2101 atagagcagt tggaggaagc atccaaagat tgcaacccag ggcaaatgga aaacaggaga
2161 tcctaatatg aaagaaaaat ggatcccaat ctgagaaaag gcaaaagaat ggctactttt
2221 ttctatgctg gagtattttc taataatcct gcttgaccct tatctgacct ctttggaaac
2281 tataacatag ctgtcacagg ggggggtata gtcacaatcc acaaatgatg caggtgcaaa
2341 tggtttatag ccctgtgaag ttcttaaagt ttagaggcta acttacagaa atgaataagt
2401 tgttttgttt tatagcccgg tagaggagtt aaccccaaag gtgatatggt tttatttcct
2461 gttatgttta acttgataat cttattttgg cattcttttc ccattgacta tatacatctc
2521 tatttctcaa atgttcatgg aactagctct tttattttcc tgctggtttc ttcagtaatg
2581 taaataa aacattgaca catacaaaca aaaaaaaaaa aaaa

In case you are having difficulty identifying the location of Greg's mutation, you can try changing the results format to one that makes it easier to look for mismatches. Scroll back to the top of the BLAST results page, and click the blue "Formatting Options" link. In the "Alignment View" box, change the view from "Pairwise" to "Pairwise with dots for identities." Then click the "Reformat" button in the upper right corner. On the re-formatted results page, each base of Greg's *F8* gene (the Subject sequence) that matches normal *F8* (the Query sequence) is shown as a dot, and mismatches are shown in red.

Once you have identified the location of Greg's mutation, determine how this might affect the factor VIII protein being produced. To do this, transcribe and translate the Query and Sbjct lines where the mutation lies. Follow the instructions below, and answer the questions along the way:

9. Note that both lines start at base 2281, but that the query line ends at base 2333, while the Sbjct line ends at 2340. Based on this information, what type of mutation has led to Greg having a different allele?

   The mutation is an insertion mutation, where a small 7-base segment has been inserted into the nucleic acid strand starting after base 1298.

10. The nucleotide sequence provided is for template DNA, and not RNA. How can you tell that this sequence corresponds to DNA?

    The sequence contains T nucleotides, which are only found in DNA (RNA has U instead)

11. Copy/paste the DNA sequence from line 2281 for each allele in the space below (i.e., copy line 2281-2333, and line 2281-2340), and transcribe each into the corresponding mRNA sequence. Note that the NCBI sequence (and the one on the previous page) is provided in the 3'→ 5' direction. In other words, the base displayed at position 2281 represents the 3' end, while the base at position 2333 (or 2340) represents the 5' end.

```
3′ TATAACATAGCTGTCACAGTATAGTCACAATCCACAAATGATGCAGGTGCAAA 5′
5′ AUAUUGUAUCGACAGUGUCAUAUCAGUGUUAGGUGUUUACUAGCUCCACGUUU 3′


3′ TATAACATAGCTGTCACAGGGGGGGGTATAGTCACAATCCACAAATGATGCAGGTGCAAA 5′
5′ AUAUUGUAUCGACAGUGUCCCCCCCCCAUAUCAGUGUUAGGUGUUUACUAGCUCCACGUUU 3′
```

12. Translate the mRNA sequence you just wrote into a sequence of amino acids.

    N-Ile-Leu-Tyr-Arg-Gln-Cys-His-Ile-Ser-Val-Arg-Cys-Leu-Leu-Ala-Pro-Arg-C

    N-Ile-Leu-Tyr-Arg-Gln-Cys-Pro-Pro-Pro-Tyr-Gln-Cys-Stop

13. Based on the short segment of amino acids, how does the normal factor VIII protein differ from the mutated factor VIII protein at the primary level of protein structure? How could this difference result in an improperly folded protein?

The insertion of a 7-base segment ultimately changes the reading frame of the allele (a six base insertion would only result in two new amino acids, but 7 means every amino acid afterwards is different). As a result, only 4 codons after the insertion there is a STOP codon, meaning the protein get truncated. Clearly, missing about 1/8 of the amino acids, the protein is no longer able to fold properly.