**Chapter 2 Study Questions**

*Genetic Analysis: Genes, Genomes, and Networks in Eukaryotes*

Partial solutions

Preliminary Analysis of a Genome

 [Note: The placozoan genome has been sequenced, and many of these questions are based on its genome. However, knowledge about the molecular biology or the genomes of placozoa will be of no help in answering these questions since information has been adapted as needed.]

One of the “endless forms most beautiful and most wonderful” is a strange creature known as the placozoan. Placozoa are marine invertebrates that were originally found exclusively on the sides of aquariums in 19th century Europe; since then, they have been found in diverse places around the world. Next to nothing is known about them, and all specimens from the Phylum Placozoa have tentatively been classified into a single species *Trichoplax adhaerens*.

 

You are heading up the placozoa genome project. You and members of your lab have used RNA sequencing to obtain the sequences of hundreds of thousands of RNA molecules, using for your source material as many different specimens of Trichoplax as you can obtain. You are in the process of sequencing the complete genome from three specimens, although with no close relatives whose genomes can be used as an assembly scaffold, this is a daunting task.

1. While the costs of sequencing a genome are coming down, sequencing projects and annotation are still relatively costly. You need to convince some wealthy investors to support your placozoan sequence project. Based on Wikipedia or another source (or your own knowledge), give a one or two sentence justification for this project that would persuade a grant agency or a donor that this project is worth doing.

2. Transposable elements. The placozoan genome, like all eukaryotic genomes, has many different families of repeated sequences. You suspect that some of these repeat sequences are DNA transposable elements, some of them are retrotransposons, and some of them are repeats that are not associated (or no longer associated) with transposable elements at all.

a. What characteristic features would you find in the sequence of a DNA transposable element that would be absent in the sequences of a retrotransposon or a simple repeat sequence?

b. You have a candidate sequence that you believe to be an active DNA transposable element based on its sequence. What information, obtained from your sequencing project as described above, could tell you that this sequence is an active transposable element?

c. Transposable elements are thought to have shaped genome evolution and genetic diversity in many different ways. What are two ways that transposable elements and repeat sequences have shaped genome evolution?

3. Annotation of the Genome. The *Trichoplax* genome is predicted to have 11,514 protein coding genes, nearly all of which have orthologues in other animals. Remarkably, more than 80% of the introns in the orthologous genes conserved between *Trichoplax* and humans are in the same position relative to the coding region in placozoa and mammals.

a. Based on this information, what inference can be drawn about the presence of the introns in the same position in such distantly related species?

b. The 11,514 protein coding genes in *Trichoplax* can be grouped into 7800 gene families. Very briefly describe what characteristics are shared among members of a gene (or protein) family, and what characteristics might or might not be shared.

4. Analysis of genes. The most closely related species whose genome has been sequenced and annotated is a sea anemone. You can do computational and simple molecular comparisons between the genes and genomes of sea anemones and placozoa (and, of course, genomes from many other animals as well), although no mutations are available in placozoa or anemones that would allow you to investigate the functions of any genes using a mutational approach. Sea anemones and other cnidarians (like jellyfish) have stinging cells called nematocysts that are considered to be the evolutionary precursors to other excitable cells such as neurons. Neuronal cell specification in higher animals is controlled by the Lim family of transcription factors; members of the Lim family are also found in sea anemones, where they are expressed in the nematocysts.

Placozoa have only four cell types and only two layers of cells. Despite having no neuronal cells and even no cells that appear to be excitable, the placozoan genome has at least one transcription factor of the Lim family, called *lim1*.

a. Concisely but fully describe how you could determine when and where the Lim transcription factor is expressed in placozoa. Your answer should describe the starting material from your genome project that you will use, as well as the means to be used to monitor Lim expression, and an example of what results might be seen. (Assume for the sake of this question that a gene construct made *in vitro* can be re-introduced into placozoa and will function normally. Also assume that any method used to monitor gene and/or protein expression in other animals will also work in placozoa.)

b. You would also like to know which genes are targets for Lim regulation in placozoa. Concisely but fully describe a strategy that you might use to address identify Lim targets.

5. Investigators have shown that when the function of the LIM1 gene in sea anemones is knocked down, no nematocysts develop. Another gene involved in nematocyst development in sea anemones appears to be *drama queen (drq)* since when *drama queen* is deleted or disrupted, many additional cells become nematocysts and the sea anemone becomes hyper-reactive. [This part is not true, by the way.] The *drq* gene from the sea anemone has been cloned, and sequence information is available for it. There appear to be transcripts from your RNA seq experiments that are nearly identical in sequence to the *drq* gene, suggesting that placozoa may also have an orthologue of this gene, as well as *lim1*.

a. Very briefly describe the normal roles of the *drq* gene and the *lim1* gene in the development of nematocysts in sea anemones.

b. The *drq* gene from sea anemones is less than 100 base pairs in length and its predicted transcript has no apparent protein coding capacity. Based on this information, as well as your answer to part A, state a hypothesis for the mechanism by which the *drq* gene in sea anemones regulates the expression of the *lim-1* gene.

c. What are two lines of evidence that would be consistent with your hypothesis about the roles of *lim1* and *drq*? Since there are no mutants available to work with, any answer that requires the use of a mutant will be marked wrong.

Polycomb and gene silencing. These questions are based on a paper by Xiao et al. 2017 (Nature Genetics 49: 1546-1552). A more detailed discussion of the paper will be done in the Journal Club for this Chapter, so these questions will be focused primarily on the initial experiments.

1. Based on the information in the chapter and summarized in this paper, outline the biological role of the Polycomb Complex, and particularly PRC2.
2. The chapter, which is based principally on the well-studied role of PRC2 in Drosophila, suggests that the PRC2 protein itself is a sequence-specific binding protein. While PRC2 can bind to DNA, more recent results in other organisms, as well as Drosophila, indicate its binding is not sequence-specific. How then does PRC2 regulate the expression of specific genes and not other genes if it does not bind directly to the PRE sequence?
3. Using Figure 1, explain the molecular construct that was in this assay. Be sure to answer the following questions.
	1. The results in the two panels that comprise Figure 1B are fairly similar. What is the difference between what is being measured in these two panels, and why is that important (or not)?
	2. What are pMAS and pF3H-35S mini and what are their functions in this experiment?
	3. Is it significant that the PRE is inserted between pMAS and pF3H-35S mini? If so, why is it important?
	4. In Figure 1D, what is shown along the X and the Y axis, and why were these chosen?
	5. What is the importance of PC-LEC2?
	6. Nearly all of the red bars in Figure 1D are higher in the left panel than in the right panel, and the median expression as shown by the white circles is also higher in the right panel than in the left panel. Explain the importance of this result.