**Epigenetics Escape Room**

Puzzle Answers

1. Bioinformatics

**Answer:** Gene 20534

**Explanation:** Of the four gene sequences, this is the only one which contains the motif we are interested in: AAT GGA TTT. In fact, it is repeated three times in Gene 20534. The first instance is highlighted below.

## Gene 20534:

GTCAGTGACAGTACGTCGATGACCGTTGACGTGCATTGATAAACTAGTCACGGTACCAGTCAGTGACAGTTCGATCTTTACGTAATGGATTTAATGGATTTAATGGATTTACGACCAGTCACTTGGTCAGT

1. Gene Targeting

**Answer:** Tube #6

**Explanation:** We can see from the image showing what the correct plasmid looks like that the enzymes would have cut it in three places, creating 4 sections of DNA. These 4 sections would all have different lengths:

The section cut by enzymes 1 and 2, containing the gene of interest, is 3000 – 1400 = **1600 base pairs (bp) long**

The section cut by enzymes 2 and 3, containing the green tag, is 3800 – 3000 = **800 bp**

The section cut by enzyme 3, containing the ampicillin resistance gene, is 6500 – 3800 = **2700 bp**

The section cut by enzymes 3 and 1, containing the red tag, is a little trickier. But using the fact that we know the total length of the circular plasmid is 7000 bp, this section must have a total length of (7000 – 6500) + (1400 – 0) = **1900 bp**.

Reading off the agarose gel, it is column 6 which shows DNA fragments at each of these lengths. In fact, as this is the only column which shows 4 fragments, it must be the right answer even without doing any calculations! However, we can confirm the lengths of these fragments by using the ladder on the left-hand side of the gel. The longest band in lane 6 (the top band) is in between the 2000bp and the 3000bp bands of the ladder, closer to 3000bp, so we can see that this matches the 2700bp fragment. The predicted 1900bp and 1600bp bands are the two middle bands in lane 6, both in between the 1000bp and 2000bp bands on the ladder on the left. Lastly, the bottom band in lane 6 is just below the 1000bp band on the ladder, which matches the predicted 800bp fragment.

1. Sequencing

**Answer:** Protein 4

**Explanation:** The completed amino acid sequence of the protein is shown in Clue 2:



The protein produced by this amino acid sequence contains the following domains: Zinc binding (Gly-Ser-Pro), DNA cutting (Gly-Arg-Leu-His), and DNA binding (Gly-Ser-Asn-Arg). This DNA sequence therefore encodes for an amino acid sequence producing a protein that contains all the domains we expect to find in protein 4.

We can confirm this by looking at each protein in turn:

|  |  |  |  |
| --- | --- | --- | --- |
| **Protein**  | **Domain protein is known to contain**  | **Sequence corresponding to this domain** | **Is this sequence in our completed protein sequence?**  |
| **1** |  Iron binding | Ala-Cys-Ser  |  No |
|  RNA binding | Leu-Leu-Gln-His |  No |
|  DNA cutting | Gly-Arg-Leu-His |  Yes |
| **2** |  Zinc binding | Gly-Ser-Pro |  Yes |
|  DNA binding | Gly-Ser-Asn-Arg |  Yes |
|  RNA cutting | Ser-Pro-Ala-Gly |  No |
| **3** | Iron binding | Ala-Cys-Ser  |  No |
| DNA binding | Gly-Ser-Asn-Arg |  Yes |
| Nucleus transport | Ala-Gly-Asp-Pro |  No |
| **4** | Zinc binding | Gly-Ser-Pro |  Yes |
| DNA binding | Gly-Ser-Asn-Arg |  Yes |
| DNA cutting | Gly-Arg-Leu-His |  Yes |
| **5** | RNA binding | Leu-Leu-Gln-His |  No |
| DNA binding | Gly-Ser-Asn-Arg |  Yes |
| DNA cutting | Gly-Arg-Leu-His |  Yes |
| **6** | RNA binding | Leu-Leu-Gln-His |  No |
| RNA cutting | Ser-Pro-Ala-Gly |  No |
| DNA cutting | Gly-Arg-Leu-His | Yes |

This sequence cannot belong to any of the other proteins as it is missing at least one domain from each of them. For example, it can’t be protein 1 as it contains no Iron binding domain or RNA binding domain.

1. Flow Cytometry & Imaging

**Answer:** Plot A – Image B, Plot B – Image A, Plot C – Image D, Plot D – Image C

**Explanation:** when a fluorescent protein is present, we can see the colour fluorescence it gives off.

Plot A – Cells on the graph measure as low Green Fluorescent Protein (GFP, bottom/X-axis) and low Red Fluorescent Protein (RFP, left/Y-axis). As these cells are neither green or red, they are therefore "non-fluorescent" as seen in Image B.

Plot B – Cells on the graph measure as high GFP and low RFP, which means they are expected to have green fluorescence, as seen in Image A.

Plot C – Cells on the graph measure high GFP and high RFP. These colours in combination will result in yellow fluorescence (as seen on the light colour wheel and in the example graph). Cells with yellow fluorescence are shown in Image D.

Plot D – Cells have low GFP and high RFP, and therefore are expected to have red fluorescence, as seen in Image C.

1. Biochemistry lab

**Answer:** 8 micromolar

**Explanation:** We want the cells to trigger the biggest amount of gene expression in the cells, more gene expression means more RNA produced, which means more protein is produced. Therefore, the amount of green fluorescence we can see in the cells is a measure of the amount of gene expression. Greener cells mean there is more expression of the GFP gene.

1 micromolar: We can see little to no green cells with 1 micromolar of the drug which suggests little to no gene expression.

2 micromolar: We begin to see more green cells with 2 micromolar of the drug, therefore there is more gene expression.

8 micromolar: We see the most, brightest green cells, and therefore the most gene expression, when using 8 micromolar of the drug.

15 micromolar: We see fewer green cells when using 15 micromolar of the drug than when using 8 micromolar, therefore there is less gene expression. This could be because the drug has reached a concentration which is unhealthy for the cells.

1. Epigenetics lab

**Answer:** Gene/enhancer pair 1

**Explanation:** We can consider each gene/enhancer pair in turn.

Pair 1: the enhancer DNA Accessibility is high (as shown by a peak in the DNA accessibility box below the enhancer) and it has markers H3K4me1 and H3K27ac. Using the enhancer rules, this means the enhancer is active. The gene DNA accessibility is high and it has markers H3K4me3 and H3K27ac. Using the gene rules, this means the gene is active. The gene and the enhancer are ***both*** active, leading to **high levels of gene expression**.

Pair 2: The gene has the same markers as for pair 1, so the gene is active. However, the enhancer has high levels of DNA methylation and marker H3K27me3. Both of these factors result in the enhancer being inactive. An active gene but inactive enhancer results in **low levels of gene expression**.

Pair 3: The enhancer is active; it has high DNA accessibility and high levels of H3K4me1 and H3K27ac. However, the gene has marker H3K27me3, making it inactive (despite all other rules). This means the gene is inactive and there will **no gene expression**, regardless of the state of the enhancer: the enhancer has nothing to “enhance”.