Case: Mapping a Promoter’s Transcription Factor Sites

Background:

### Day 1 (read before class)

One question molecular biologists want to answer is how specialized cells control patterns of RNA transcription and protein expression of tissue specific genes. **Promoter mapping** (sometimes called promoter bashing) is one approach used. In this case your team is going to try to map out how specialized cells in the cardiovascular system control transcription.

The cardiovascular system is made up of **cardiomyocytes, smooth myocytes, endothelial cells, fibroblasts, pericytes, and adipocytes** (the blood cells are a different system). Each of these cell types must transcribe RNAs from cell type-specific genes in a specific pattern AND be able to down-regulate each gene under certain conditions.

A screenshot of a cell phone

Description automatically generated

Original Image Sources:

<https://commons.wikimedia.org/wiki/File:Human_Heart.jpg>

<https://commons.wikimedia.org/wiki/File:Blood_vessels-en.svg#/media/File:Capillary.svg>

<https://en.wikipedia.org/wiki/File:Artery.png>

The BHMA9 (“Bahama-9”) gene is only transcribed by certain cells in the cardiovascular system. In class I will provide you with several pieces of data about the BHMA9 promoter. Your team’s goals are:

* Day 1: Work out which transcription factors control expression of BHMA9 by identifying what is binding to each active site in the promoter;
* Day 2: Use mutation/luciferase expression assay data to work out how these transcription factors regulate this gene

Your team's summary should outline your conclusions, and what your evidence and reasoning are supporting your conclusions.

***Hint****: there is not a “correct” answer for this case problem. You would need a LOT more data to make any firm conclusions about transcriptional regulation of this gene. What I am looking for your team to do is interpret the data, combine it with what you have learned from the class readings and videos, formulate a logical explanation, and support that explanation with a well-organized argument.*

##### DNAse I Footprint Map

This method is used to determine what regions of a DNA sequence have proteins bound to them. It is one of the first assays done to evaluate a promoter region. The mechanics of this technique are not important right now; what is important is that footprinting identifies the sequence regions in the DNA bind transcription factors and other regulatory proteins most strongly.

A close up of a device

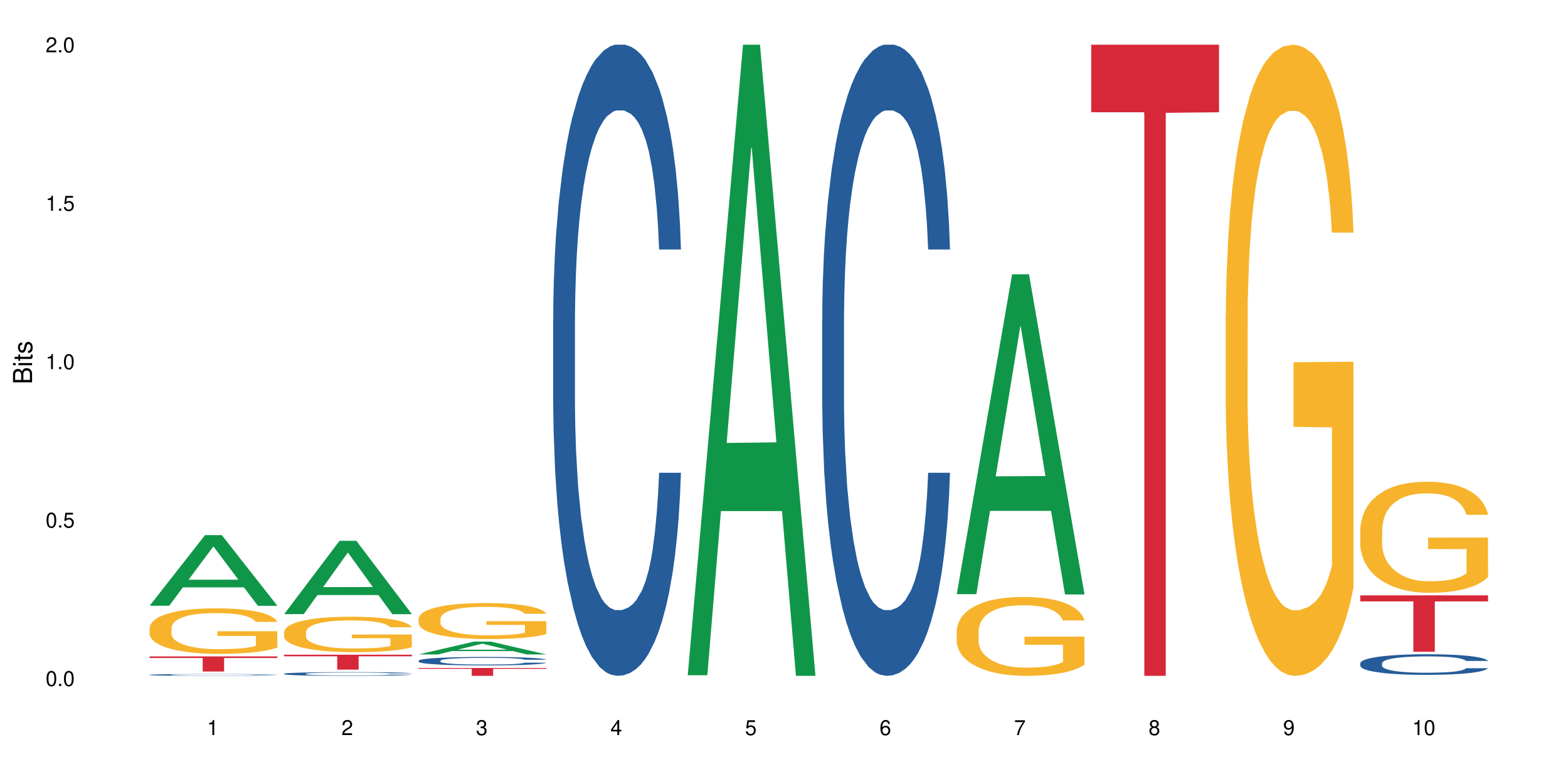
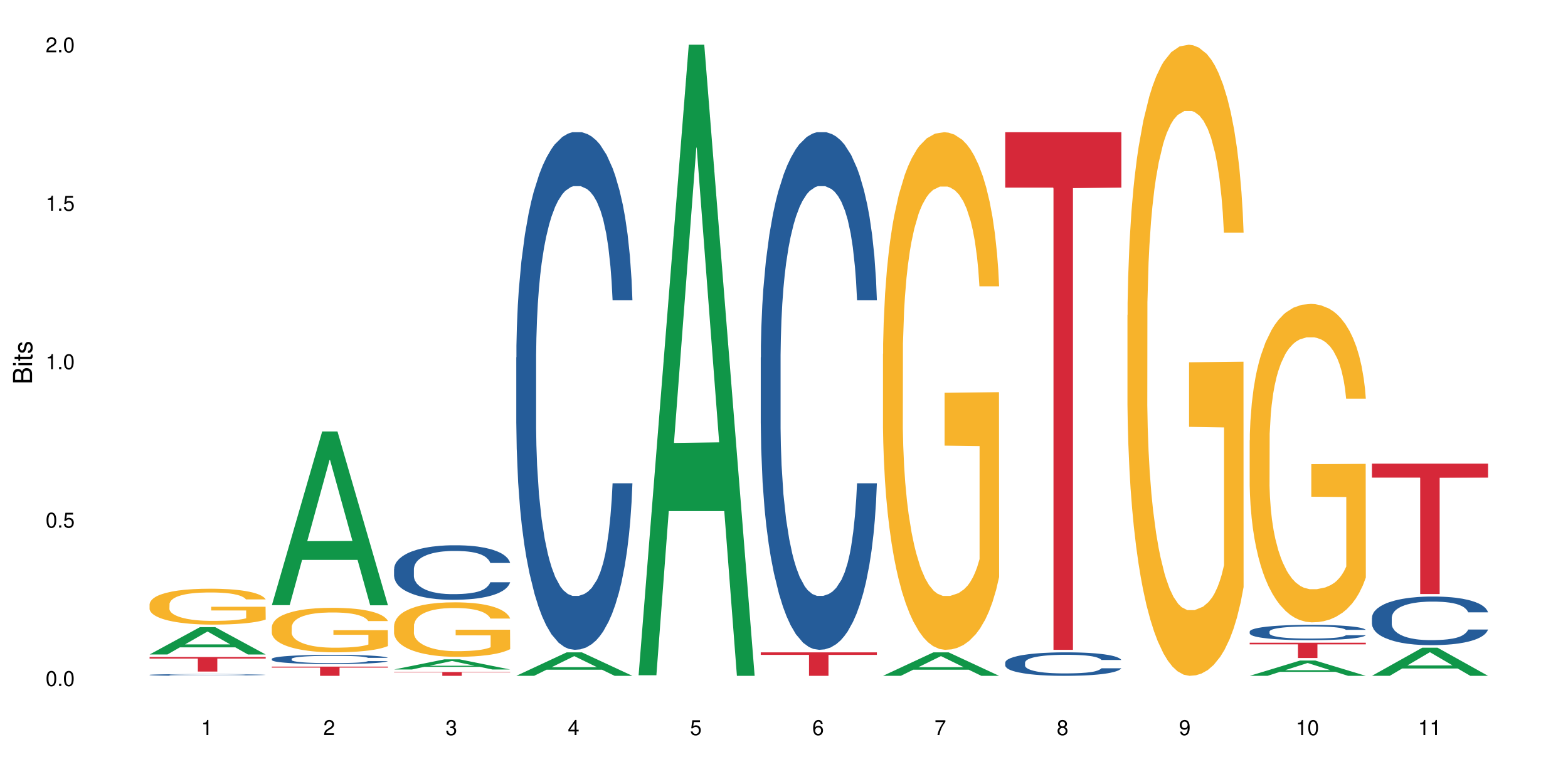
Description automatically generated

*Example of a footprint map showing which DNA bases participate in protein binding, and how far upstream of the mRNA transcriptional start site they are located. You already should be able to identify what protein(s) bind at -25. You do not have sufficient information yet to identify the proteins binding to DNA at the other three footprints.*

After footprinting, we can PREDICT what transcription factors might bind based on the DNA sequences (see next section), then use ChIP and EMSA assays (which we are not looking at) to confirm what proteins actually are binding.

##### Identifying Transcription Factor Binding Sites

Each transcription factor has a DNA sequence to which it binds best. This is called its **canonical site**. For example, the DNA sequence "5'-nnCAnnTGnn-3'" is the canonical binding site for the basic helix-loop-helix (bHLH) family of transcription factors. There are dozens of factors in this family, but they all prefer to bind to DNA at a site with that sequence. The "n" bases determine which family members bind best. For example, the bHLH called Max binds best to "nnCACaTGnn," while its cousin Myc binds best to "nnCACCTGgn."

*Sequence logos for Max (left) and Myc (right), two bHLH transcription factors.*

Many transcription factors still bind to DNA even if some bases are different from the canonical sequence. The figures are **sequence logos**. The size of the letters shows how frequently each base in the canonical binding site has been reported to differ, and still bind that transcription factor.

There are many open-access databases that catalog transcription factor binding sites and create these sequence logos. One example is the JASPAR database at <http://jaspar.genereg.net/>. Scientists use databases like JASPAR to predict which transcription factors are the most likely candidates to be binding to the DNA sites they identified via footprinting assay.

In Class, Day 1:

### Your Team’s Goals For Today

1. Identify which transcription factors are binding to the BHMA9 promoter, and where.
2. Decide which transcription factors are part of the basal/core transcription machinery, and which factors may control tissue-specific RNA transcription.
3. If there is time, discuss these questions:
   1. Are all of the binding sites pointing in the same direction?
   2. Are all of the transcription factors close enough to interact directly with each other?
   3. What are the implications of your previous two answers?

### Data: Footprint Assay

This partial footprint map of the BHMA9 gene shows the 5’-promoter region, mRNA transcription start site, and first exon and intron regions. The map ends just after the second exon starts. The top row shows the nucleotide positions relative to mRNA transcriptional start (+1); numbers -1 to -2300 are upstream, numbers +2 to +2807 are downstream.

A screenshot of a cell phone

Description automatically generated

The bottom portion shows the DNA sequences of different regions that DNAse I footprinting have identified as nuclear protein binding sites.

### Data: Potential Transcription Factor Binding Sites

The table below lists the names, canonical sequences, and sequence logos for several transcription factors that might bind and regulate the BHMA9 promoter. In the canonical sequences, R = purines (A or G), Y = pyrimidines (C or T), and [A/T] means either nucleotide will work. Immediately below the names of each transcription factor are their sequence logos. The size of the letters shows how frequently each base in the canonical binding site has been reported to differ, and still bind that transcription factor.

|  |  |
| --- | --- |
| **ETV2** (5’-GGAA[A/T])-3’ | **HAND2** (5’-CAnnTG-3’) |
|  |  |
| **Max** (5’-CAnnTG-3’) | **Myb** (5’-CAnnTG-3’) |
|  |  |
| **Myc** (5’-CAnnTG-3’) | **NF-kappaB** (5’-GGGRnYYYCC-3’) |
|  |  |
| **Sp1** (5’-GGGCGG-3’) | **SRF** (5’-CC[A/T]6GG-3’) |
|  |  |
| **TBP** (5’-TATAA-3’) | **USF1** (5’-CAnnTG-3’) |
|  |  |

Background:

### Day 2 (read before class)

Molecular biologists want to measure when and how fast mRNAs are transcribed from particular genes. This can be done by measuring the gene’s mRNAs directly, or indirectly using promoter-reporter expression assays. If we connect a DNA promoter of interest to DNA that encodes an enzyme or other marker that is easy to measure. By measuring the marker, we can estimate relative rates of RNA transcription. This is much easier and faster than measuring the mRNAs produced directly.

Today we will be looking at reporter data generated using luciferase expression assays. DNA containing the BHMA9 promoter is placed upstream of DNA that encodes the luciferase gene. If the promoter is active, mRNA that codes for the luciferase enzyme is transcribed, then translated. The amount of luciferase produced is measured by adding luciferin, which produces light when the enzyme cleaves it. The more RNA made, the more enzyme, and the brighter the flash of light.

A screenshot of a cell phone

Description automatically generated

*Example of a luciferin/luciferase assay.*

### Your Team’s Goals For Today

Use the promoter/luciferase expression assay data to work out how the transcription factors that you identified the first day of this case work together to regulate this gene. Specifically:

1. Which cells in the cardiovascular system have the highest levels of BHMA9 promoter activity?
2. How could cells in the cardiovascular system control which cells do or do not produce BHMA9 mRNAs?
3. Is the organization of the transcription factor binding sites along the DNA important? How do you think they are interacting with each other?

### Data: Directed Mutation + Luciferase Expression Assays

The figure below shows the initial BHMA9 promoter fused to the luciferase reporter. Below the “wild type” promoter/reporter are a series of mutated versions that have shorter and shorter pieces of the whole promoter.

A screenshot of a social media post

Description automatically generated

Each promoter/reporter construct shown above was inserted into each of the following cell types, and the levels of luciferase expression measured using a standard assay.

* Arterial endothelial cells
* Cardiac myocytes
* Vascular smooth myocytes
* Vascular sheath fibroblasts
* Human epithelial stem cells

### Data: Observed Results of Reporter Expression Assay

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Relative Luciferase Activity in Different Cell Types** (, n=5 trials)\* | | | | |
| **Promoter Fragment Used** | *Endothelial cells* | *Cardiac myocytes* | *Smooth myocytes* | *Fibroblasts* | *Stem cells* |
| -2300 to +200nt (wild type) | 36.3 | 44.8 | 28.4 | 3.3 | 2.3 |
| Luciferase only \*\* | n.d. | n.d. | n.d. | n.d. | n.d. |
| -40 to +30nt | n.d. | 0.12 | n.d. | 0.33 | 1.2 |
| -150 to +30nt | 0.41 | 1.2 | 0.51 | 2.4 | 8.6 |
| -200 to +30nt | 0.45 | 1.1 | 0.62 | 2.6 | 7.9 |
| -250 to +30nt | 0.62 | 50.6 | 1.0 | 2.5 | 4.5 |
| -290 to +30nt | 0.29 | 25.3 | 12.2 | 1.5 | 1.1 |
| -1000 to +30nt | 0.31 | 25.0 | 11.5 | 1.4 | 1.3 |
| -2300 to +30nt | 19.2 | 27.9 | 13.8 | 1.1 | 1.2 |

*\* Luciferase activity in different cell types was normalized to a standard expression control (not shown).*

*All values are given as averages of relative light units (RLU) for 5 replicate samples*

*\*\* n.d. = not detectable*

Final Summary

Your team's submitted summary on the LMS for today should outline your conclusions, and what your evidence and reasoning are supporting your conclusions. Your summary should address all of the questions in the Team Goals for both Day 1 and Day 2. Each team only needs to submit one summary.

***Your Team ID# \_\_\_\_\_:***

1. Which transcription factors are binding to the BHMA9 promoter, and where?
2. Which transcription factors are part of the basal/core transcription machinery, and which factors may control tissue-specific RNA transcription? What is your reasoning?
3. Is the organization of the transcription factor binding sites along the DNA important? How do you think they are interacting with each other? What is your evidence, reasoning?
4. Are all transcription factor binding sites used for controlling tissue-specific expression? Why or why not? If not, what are they for? What is your evidence, reasoning?
5. How might cells in the cardiovascular system control which cells do or do not produce BHMA9 mRNAs? What is your reasoning?

***Remember****: there is not a “right” answer for this case problem. You would need a LOT more data to make any firm conclusions about transcriptional regulation of this gene. What I am looking for your team to do is interpret the data you have, combine it with what you have learned from class readings and videos, formulate a logical explanation, and support that explanation with a well-structured argument.*

Instructor Notes

### Background

This case was developed for a 200-level cell and molecular biology course. From class students were familiar with the functions of general core transcription factors required for initiating mRNA transcription. However, students struggled to understand:

* How transcription factors find their corresponding DNA binding sites;
* How upstream transcription factors regulate the core promoter complex;
* How shared transcription factors can activate a promoter in one or a few specific cell types; and
* How the same transcription factor can act both as an enhancer and repressor depending on cell context.

Students in this course also had yet to understand a central principle of cellular and molecular biology: direct cause-effect controls are very rare. Instead, most cellular controls operate with low specificity and high redundancy. I felt the best way to help students internalize this was to give them a regulation problem that could not be explained by simple cause-effect control.

The BHMA-9 gene is fictitious but was modeled after the smooth muscle alpha actin promoter. Expression of many tissue-specific genes is controlled by a tissue-specific transcription factor. SMaA is different (and more like non-specific genes) in that there is no smooth muscle-specific transcription factor. Instead, the core promoter drives high levels of gene expression in most cell types. Adding upstream transcription factor binding sites shuts down SMaA mRNA transcription in more and more cell types, until SMaA mRNA is only produced in smooth myocytes and myofibroblasts.

The promoter fragment used in this case also has a binding site for NF-kB, which is the primary transcription factor associated with inflammatory responses. I included NF-kB as part of the case data so students could see how cells use transcriptional repression. When smooth myocytes encounter inflammatory signals, they down-regulate SMaA production as a precursor to mitosis and migration into wound sites. As NF-kB activity drops, SMaA production rises again.

### Case Data Structure

##### Figure from Day 1

A screenshot of a cell phone

Description automatically generated

ETV2 Myb (rev) HAND2 (weak) NF-kB (rev) Sp1 TBP SRF

Likely TFs Binding Each Site

##### Table from Day 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Relative Luciferase Activity in Different Cell Types** (, n=5 trials)\* | | | | |
| **Promoter Fragment Used** | *Endothelial cells* | *Cardiac myocytes* | *Smooth myocytes* | *Fibroblasts* | *Stem cells* |
| -2300 to +200nt (wild type) | 36.3 | 44.8 | 28.4 | 3.3 | 2.3 |
| Luciferase only \*\* | n.d. | n.d. | n.d. | n.d. | n.d. |
| -40 to +30nt | n.d. | 0.12 | n.d. | 0.33 | 1.2 |
| -150 to +30nt | 0.41 | 1.2 | 0.51 | 2.4 | 8.6 |
| -200 to +30nt | 0.45 | 1.1 | 0.62 | 2.6 | 7.9 |
| -250 to +30nt | 0.62 | 50.6 | 1.0 | 2.5 | 4.5 |
| -290 to +30nt | 0.29 | 25.3 | 12.2 | 1.5 | 1.1 |
| -1000 to +30nt | 0.31 | 25.0 | 11.5 | 1.4 | 1.3 |
| -2300 to +30nt | 19.2 | 27.9 | 13.8 | 1.1 | 1.2 |

The table data are organized so that the native expression pattern is in the first highlighted row. The basal promoter is from -150 to +30 nt (Sp1 and TBP binding sites.) This allows measurable luciferase expression in stem cells, fibroblasts, endothelial cells, and myocytes. Adding the HAND2 site greatly increases cardiac cell expression, which fits given HAND2 is a cardiac-specific TF. Smooth myocytes require an additional TF (Myb) that binds in the region from -250 to -290 nt. Myb also appears to block expression in fibroblasts and stem cells.

Class Management

I have attached comments to the main text of the case that describe how the case operates and the data I can extract from it.

### Let Students Struggle SOME

I find students will want to give up too soon, and simply ask you what the correct answers are. Resist the temptation and instead help them work through the logic of choosing one transcription factor or interpreting one row of the Day 2 data table. If one student seems to understand but others do not, let the students coach one another.

Reiterate that there is no correct answer to this problem; what is more important is that they provide evidence to support a logical prediction.

### Goals of the Follow-Up Questions

1. Which transcription factors are binding to the BHMA9 promoter, and where?
2. Which transcription factors are part of the basal/core transcription machinery, and which factors may control tissue-specific RNA transcription? What is your reasoning?

These two questions check whether students know the difference in basal and tissue-specific transcription, and that they understood the work from Day 1.

1. Is the organization of the transcription factor binding sites along the DNA important? How do you think they are interacting with each other? What is your evidence, reasoning?

One of the key concepts I want students to understand is that DNA is not linear; it flexes and bends to let TFs interact with one another. They should realize that the bound proteins cannot interact otherwise, because they are too far apart.

1. Are all transcription factor binding sites used for controlling tissue-specific expression? Why or why not? If not, what are they for? What is your evidence, reasoning?

No, NF-kB does not seem to affect expression in the luciferase assay. One more advanced option might be to add additional rows to the Day 2 data table, and include cells stimulated with inflammatory cytokines.

1. How might cells in the cardiovascular system control which cells do or do not produce BHMA9 mRNAs? What is your reasoning?

Can students explain how multiple TFs work together to regulate specific expression?

Supplemental Activities and Case Extensions

This case could be revised to use other tissue-specific promoters, or to explore regulation of genes under different physiological stresses. Some options for the latter might be stress responses (HSF-1), immune responses (NF-kB), hypoxia (VHL/HIF-1), or cell cycle progression (Myc/Jun/Fos).