Reading Practice

Life code: unlocked!

{A} On an airport shuttle bus to the Kavli Institute for Theoretical Physics in Santa Barbara, Calif., Chris Wiggins took a colleague's advice and opened a Microsoft Excel spreadsheet. It had nothing to do with the talk on biopolymer physics he was invited to give. Rather the columns and rows of numbers that stared back at him referred to the genetic activity of budding yeast. Specifically, the numbers represented the amount of messenger RNA (MRNA) expressed by all 6,200 genes of the yeast over the course of its reproductive cycle. "It was the first time I ever saw anything like this," Wiggins recalls of that spring day in 2002. "How to make sense of all this data?"

{B} Instead of shirking from this question, the 36-year-old applied mathematician and physicist at Columbia University embraced it-and now six years later he thinks he has an answer. By foraying into fields outside his own, Wiggins has drudged up tools from a branch of artificial intelligence called machine learning to model the collective protein-making activity of genes from real-world biological data. Engineers originally designed these tools in the late 1950s to predict output from input. Wiggins and his colleagues have now brought machine learning to the natural sciences and tweaked it so that it can also tell a story-one not only about input and output but also about what happens inside a model of gene regulation, the black box in between.

{C} The impetus for this work began in the late 1990s, when high-throughput techniques generated more mRNA expression profiles and DNA sequences than ever before, "opening up a completely different way of thinking about biological phenomena," Wiggins says. Key among these techniques were DNA microarrays, chips that provide a panoramic view of the activity of genes and their expression levels in any cell type, simultaneously and under myriad conditions. As noisy and incomplete as the data were, biologists could now query which genes turn on or off in different cells and determine the collection of proteins that give rise to a cell's characteristic features, healthy or diseased.

{D} Yet predicting such gene activity requires uncovering the fundamental rules that govern it. "Over time, these rules have been locked in by cells," says theoretical physicist Harmen Bussemaker, now an associate professor of biology at Columbia. "Evolution has kept the good stuff." To find these rules, scientists needed statistics to infer the interaction between genes and the proteins that regulate them and to then mathematically describe this network's underlying structure-the dynamic pattern of gene and protein activity over time. But physicists who did not work with particles (or planets, for that matter) viewed statistics as nothing short of an anathema. "If your experiment requires statistics," British physicist Ernest Rutherford once said, "you ought to have done a better experiment."

{E} But in working with microarrays, "the experiment has been done without you," Wiggins explains. "And biology doesn't hand you a model to make sense of the data." Even more challenging, the building blocks that makeup DNA, RNA, and proteins are assembled in myriad ways; moreover, subtly different rules of interaction govern their activity, making it difficult, if not impossible, to reduce their patterns of interaction to fundamental laws. Some genes and proteins are not even known. "You are trying to find something compelling about the natural world in a context where you don't know very much," says William Bialek, a

biophysicist at Princeton University. “You’re forced to be agnostic.” Wiggins believes that many machine-learning algorithms perform well under precisely these conditions. When working with so many unknown variables, “machine learning lets the data decide what’s worth looking at,” he says.

{F} At the Kavli Institute, Wiggins began building a model of a gene regulatory network in a yeast-the set of rules by which genes selectively orchestrate how vigorously DNA is transcribed into mRNA. As he worked with different algorithms, he started to attend discussions on gene regulation led by Christina Leslie, who ran the computational biology group at Columbia at the time. Leslie suggested using a specific machine-learning tool called a classifier. Say the algorithm must discriminate between pictures that have bicycles in them and pictures that do not. A classifier sifts through labeled examples and measures everything it can about them, gradually learning the decision rules that govern the grouping. From these rules, the algorithm generates a model that can determine whether or not new pictures have bikes in them. In gene regulatory networks, the learning task becomes the problem of predicting whether genes increase or decrease their protein-making activity.

{G} The algorithm that Wiggins and Leslie began building in the fall of 2002 was trained on the DNA sequences and mRNA levels of regulators expressed during a range of conditions in yeast-when the yeast was cold, hot, starved, and so on. Specifically, this algorithm-MEDUSA (for motif element discrimination using sequence agglomeration) -scans every possible pairing between a set of DNA promoter sequences, called motifs, and regulators. Then, much like a child might match a list of words with their definitions by drawing a line between the two, MEDUSA finds the pairing that best improves the fit between the model and the data it tries to emulate. (Wiggins refers to these pairings as edges.) Each time MEDUSA finds a pairing, it updates the model by adding a new rule to guide its search for the next pairing. It then determines the strength of each pairing by how well the rule improves the existing model. The hierarchy of numbers enables Wiggins and his colleagues to determine which pairings are more important than others and how they can collectively influence the activity of each of the yeast’s 6,200 genes. By adding one pairing at a time, MEDUSA can predict which genes ratchet up their RNA production or clamp that production down, as well as reveal the collective mechanisms that orchestrate an organism’s transcriptional logic.

Questions 1-6

The reading passage has seven paragraphs, A-G

Choose the correct heading for paragraphs A-G from the list below.

Write the correct number, i-x, in boxes 1-6 on your answer sheet.

List of Headings

(I) The search for the better-fit matching between the model and the gained figures to foresee the
the genes

(II) The definition of MEDUSA

(III) A flashback of commencement for a far-reaching breakthrough

(IV) A drawing of the gene map

(V) An algorithm used to construct a specific model to discern the appearance of something new
effort of Wiggins and another scientist

(VI) An introduction of a background tracing back to the availability of mature techniques for detai
on genes

(VII) A way out to face the challenge confronting the scientist on the deciding of researchable dat

(VIII) A failure to find out some specific genes controlling the production of certain proteins

(IX) The use of a means from another domain for reference

(X) A tough hurdle on the way to find the law governing the activities of the genes

Example: Paragraph A **III**

1..... Paragraph B

2..... Paragraph C

3..... Paragraph D

4..... Paragraph E

5..... Paragraph F

6..... Paragraph G

Questions 7-9

Do the following statements agree with the information given in Reading Passage 1? In boxes 7-9 on your answer sheet, write

TRUE	if the statement is True
FALSE	if the statement is false

NOT GIVEN If the information is not given in the passage

7..... Wiggins is the first man to use DNA microarrays for the research on genes.

8..... There is almost no possibility for the effort to decrease the patterns of interaction between DNA, RNA, and proteins.

9..... Wiggins holds a very positive attitude on the future of genetic research.

Questions 10-13

Complete the following summary of the paragraphs of Reading Passage, using No More than three words from the Reading Passage for each answer. Write your answers in boxes 10-13 on your answer sheet.

Wiggins states that the astoundingly rapid development of techniques concerning the component aroused the researchers to look at **10**..... from a totally new way. **11**..... is the soul of these techniques and no matter what the **12**..... were, at the same time they can whole picture of the genes' activities as well as **13**..... in all types of cells. With these te scientists could locate the exact gene which was on or off to manipulate the production of the pro

Solution:

1. ix	**8.** TRUE
2. vi	**9.** NOT GIVEN
3. x	**10.** BIOLOGICAL PHENOMENA
4. vii	**11.** DNA MICROARRAYS
5. v	**12.** (MYRIAD) CONDITIONS
6. i	**13.** THEIR EXPRESSION LEVELS
7. NOT GIVEN	