

Solving a Medical Mystery with Bioinformatics: The Personal Genomics Journey of Dr. James Lupski

by

Kelly L. Brackett, Gareth R. Howell,
Charles G. Wray, and Sarah A. Wojiski
The Jackson Laboratory, Bar Harbor, ME



Part I – Seeking Answers

In this case study you will use gene databases and genome browsers to follow the path to discovery of one man’s debilitating disease. You will investigate gene function and its potential relationship to disease, assess genetic homology between species, and debate the best course of action for studying this disease. You will engage in a process that is similar to that of scientists when trying to develop models for the study of disease.

Imagine you are a family doctor ...

Today you have a follow-up appointment with Professor James Lupski, a geneticist at Baylor College of Medicine in Houston, Texas. In your initial visit, you learned that Dr. Lupski has been living with a rather troublesome disease since childhood. He presented to you with symptoms such as foot deformities, muscle weakness and scoliosis that have slowly gotten worse over the years. You also learned that three of his seven siblings have also displayed similar symptoms, but his parents never did. After some careful bioinformatics research, you believe you have a diagnosis for Dr. Lupski based upon his symptoms and family history. You will be discussing that today during his office visit.

Define the following terms.

- *Genetics:*

- *Bioinformatics:*

You and a partner will now compare two different bioinformatics resources: Wikipedia and Online Mendelian Inheritance in Man (OMIM). Each of you should select one website and follow the instructions to fill in the tables below.

Describe the website. Who writes/edits the entries? Who funds the project? What kind of content do they present?

Table 1. Website Description

<i>Wikipedia</i>	<i>OMIM</i>
Go to < http://www.wikipedia.com > and search for “Wikipedia”	Go to < http://www.omim.org > and click on the “About” link

What keywords from Dr. Lupski’s visit will you use to search for a diagnosis? Come to an agreement with your partner so that you use the same keywords on both websites.

Keywords: _____

Use these keywords to search for a diagnosis in the search bar. What are the top three hits?

Table 2. Top Three Hits

<i>Wikipedia</i>	<i>OMIM</i>

Click on one of the top three hits. What kinds of information are presented about diseases?

Table 3. Kinds of Information

<i>Wikipedia</i>	<i>OMIM</i>

Who might be the target audience of this bioinformatics resource? What about the language makes you think that?

Table 4. Target Audience

<i>Wikipedia</i>	<i>OMIM</i>

Questions

1. After reading the entries for your top three hits on both websites, what is your diagnosis? Explain how you came to this diagnosis.
2. What is the cause of this disease?
3. Provide a plausible explanation for why only three of Professor Lupski’s seven siblings have this disease.
4. What type of cure, if any, can be offered to Professor Lupski?

Part II – Finding the Genetic Cause

Professor Lupski thanks you for the diagnosis of Charcot-Marie-Tooth disease (CMT, named for the three doctors who first described the genetic disease), but because he is a scientist, he wants to know more. He seems to leave your office with more vigor than when he arrived.

A few months later, as you're sipping your coffee reading the Sunday edition of *The New York Times*, you almost fall off your chair when you see an article stating that Professor Lupski found his CMT genetic mutation by sequencing his own genome!

Define the following terms.

- *Genetic Disease*

- *Genomics*

Read the following news story and answer the questions below: "Disease Cause Is Pinpointed With Genome" by Nicholas Wade, *The New York Times*, March 10, 2010. <<http://www.nytimes.com/2010/03/11/health/research/11gene.html>>.

Questions

1. Describe what Professor Lupski found in his genome.

2. What questions do you still have about his genetic disease?

3. What information in the *Times* article would you use to find the original publication of the scientific findings about Professor Lupski?

4. Use this information to search for the original publication. List the search engine(s) you used.

5. How many people were involved in the publication of Professor Lupski's genome?

6. Read the abstract of the original publication. List the terms you are unfamiliar with. List the terms you are familiar with. Using what you understand, summarize the abstract in 100 words or less.

Part III – Finding a Cure

Professor Lupski's work was groundbreaking; never before had whole genome sequencing been used to diagnose a genetic disease. The sequencing yielded over 3.4 million single nucleotide polymorphisms (SNPs) that were unique to Professor Lupski as compared to human genomes published at that time. By focusing on genes known to be involved in CMT, Professor Lupski and his colleagues identified two recessive mutations (one missense, one nonsense) in the SH3TC2 gene.

You've never heard of this gene before, so you consult a gene database to find out more information about SH3TC2.

- Navigate to the NCBI Gene database (<<http://www.ncbi.nlm.nih.gov/gene>>) and search for SH3TC2. Click on the link to the human reference.

Questions

1. What is the official full name of this gene?
2. Is this gene protein coding? What does that mean?
3. On what chromosome is SH3TC2 located?
4. Does the summary indicate what type of CMT Professor Lupski has? (There are many types!) If so, what is the type?

Of course, your goal as a doctor is to alleviate Professor Lupski's symptoms. Since there is no readily available cure or even FDA-approved treatment for CMT, you want to give Professor Lupski peace of mind that researchers are currently looking for one.

- On the right hand sidebar, scroll down to and click on the OMIM link under the "Related Information" heading.
- Click on the "CMT Type 4C" link.
- Click on the "Animal Model" link in the left-hand menu list.

Questions

5. What type of animal is used as a model for this human disease?
6. What type of mutation in Sh3tc2 does this model have? Is that the same as Professor Lupski's mutation?
7. What do researchers think is the normal function of this gene?

Part IV – Making Connections Across Species

You begin to question how well CMT can be modeled in a mouse. Will researchers really be able to find a cure for your human patient by using a rodent? You decide to explore the genetic similarity between the mouse and human gene sequences.

- Navigate to the Ensembl genome browser, <<http://www.ensembl.org>>, to investigate gene conservation.
- Find the “Favourite Genomes” heading and select “Human.” In the new window that appears, use the search feature to find “SH3TC2” and select the first hit.
- In the left navigation panel, click “Orthologues” under the “Comparatives Genomics” heading.
- Check the box next to “Rodents and related species” to display more details about this order of the animal kingdom.
- Scroll down to find the mouse (*Mus musculus*) entry.
- Click “View Sequence Alignments” and then “View cDNA Alignment” on the pop-up menu. Each row has the human DNA sequence appearing on top of the mouse DNA sequence for 50 bases per row (there are more than 3,800 bases in the SH3TC2 cDNA, so keep scrolling down!). When the bases match, there is a (*) underneath that location.
- Record % identity and % coverage: _____
- Navigate back to the mouse orthologue entry. Click “View Sequence Alignments” and then “View Protein Alignment” on the pop-up menu. Each row has the human amino acid sequence above the mouse amino acid sequence for 50 amino acids per row (there are more than 1,200 amino acids in SH3TC2 protein, so keep scrolling down!). When the amino acids match, there is a * underneath that location. If they are similar in charge or polarity they either have a (.) or (:).
- Record % identity and % coverage: _____

Questions

1. How would you determine if this a reasonable amount of genetic conservation for DNA and protein?
2. What is the value of an animal model if it doesn't have perfect genetic conservation with humans?

- Navigate back to the mouse orthologue entry. Click “View Gene Tree.” The gene tree not only shows the phylogenetic relationships of species (as determined by SH3TC2 sequence only) but also shows a map of conservation across the gene.

Questions

3. Based on the Gene Tree, name a better animal model for studying SH3TC2 than the mouse. Explain your reasoning.
4. Name an animal model for studying SH3TC2 that is worse than the mouse. Explain your reasoning.
5. For what reasons do you think researchers chose to create a mouse animal model for CMT instead of the animal model you picked in Question 3?

Part V – The Pursuit of a Cure

Now that you are confident that the mouse is a reasonable model for human CMT, you want to know if there is active research on this disease.

- Navigate to the OMIM CMT Type 4C Animal Model Entry where the original article about the mouse model is linked.
- Click on the link to the paper, Arnaud *et al.* (2009). (Note: “*et al.*” means that there was more than one scientist in addition to Arnaud who contributed work to the project.)
- Click on the link to the PubMed ID number, 19805030.
- In the right hand column, click the “See all...” link under the “Cited by” heading. All of these articles have cited Arnaud *et al.* (2009) in the paper and are likely using the mouse model they created.

Questions

1. How many articles have cited Arnaud *et al.*?
2. Click on the first three linked articles and click the “Author Information” drop down. Where do these scientists work?
3. Can you confidently tell Professor Lupski that scientists are actively researching his disease? Why or why not? (Hint: Look for the date of publication.)
4. Do any of the article titles indicate that they may be close to a cure? Explain your answer.
5. Write an e-mail to Professor Lupski to tell him what the state of the research field is and his current treatment options.