NATIONAL CENTER FOR CASE STUDY TEACHING IN SCIENCE

Is p53 a Smoking Gun?

How Mutational Signatures Forced Big Tobacco to Change

by
Michèle I. Shuster, Joann Mudge, Meghan Hill, Katelynn James, Gabriella A. DeFrancesco, Maria P. Chadiarakou, and Anitha Sundararajan

Part I – Identify Mutational Hotspots in the p53 Protein in Lung Cancer

You will compare wild-type and mutant p53 sequences using an online tool called BLAST (blastp). Follow the instructions below.

How to BLAST Protein

1. Go to <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>.
2. Tick the "Align two or more sequences" check box, which is highlighted in the yellow box in Figure 1.

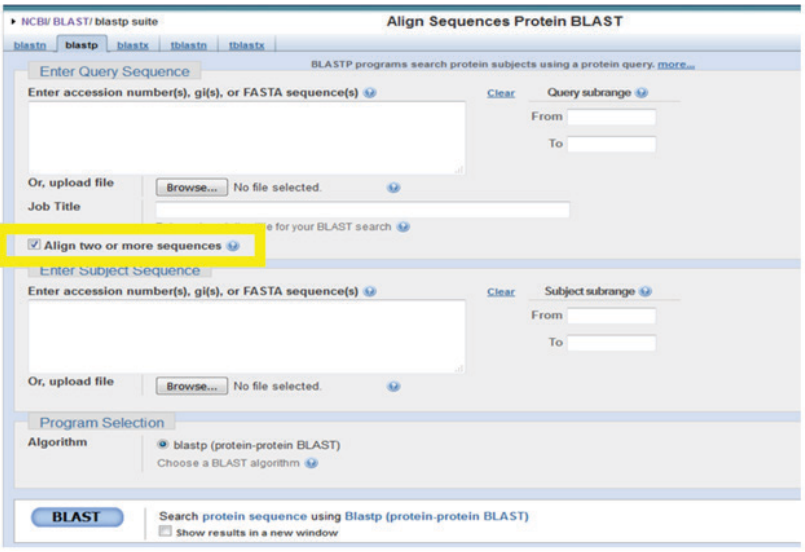

Figure 1. Screenshot of standard protein BLAST page.

3. Copy and paste one of your mutant p53 sequences into the "Enter Query Sequence" Box at the top (you can include the ">Mutation_#" as you copy and paste).
4. Copy and paste the wild-type p53 sequence into the "Enter Subject Sequence" box at the bottom (you can include the ">p53_wildtype" as you copy and paste).
5. Then hit the blue BLAST button (at the bottom).
6. Wait for a new page (with your results) to load.
7. Scroll down a bit and click on the "Alignments" tab to see your mutant p53 sequence (query) aligned with the wild type p53 sequence (subject) (see Figure 2, next page).

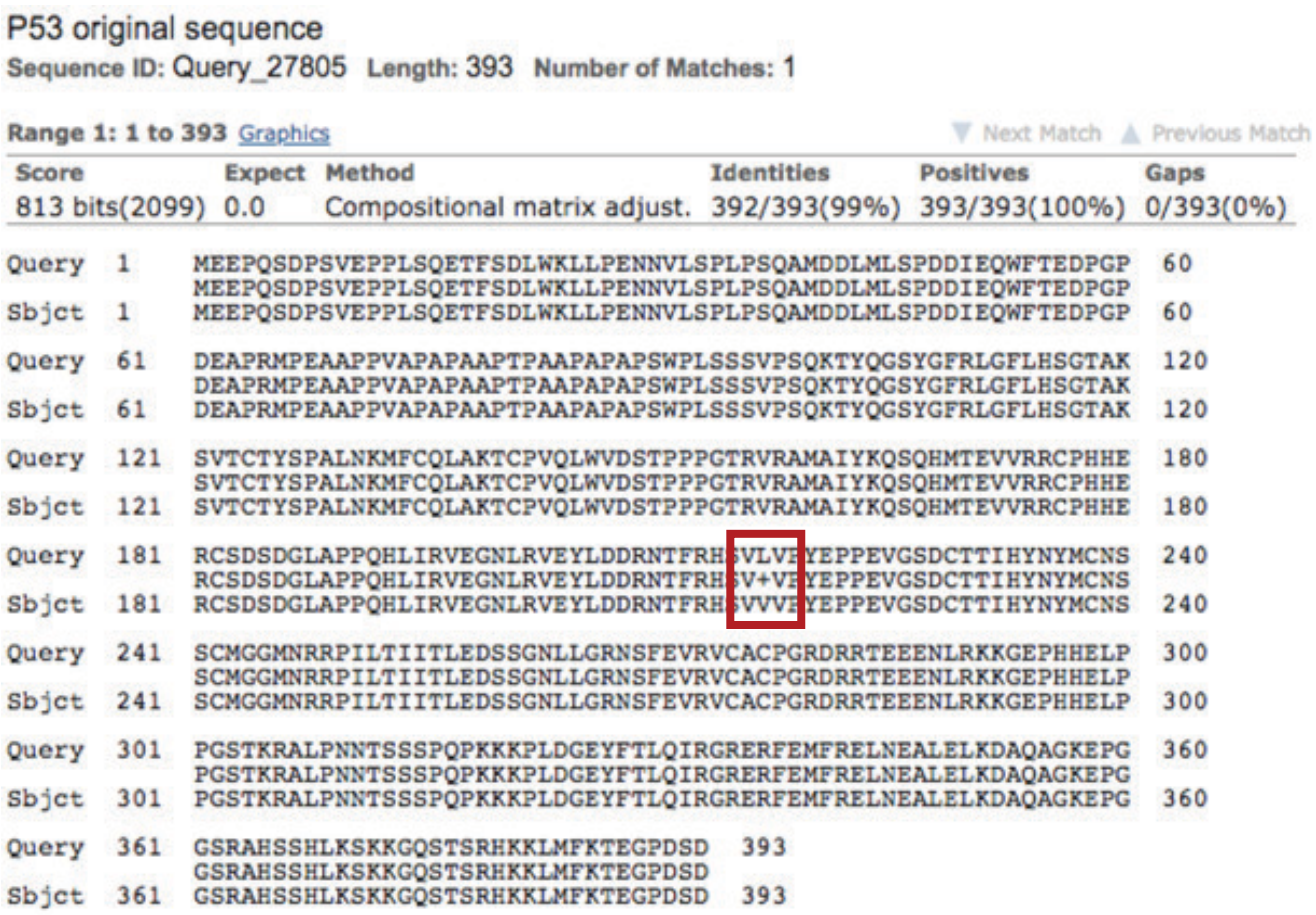

Figure 2. Example of a result page using BLAST. The red box highlights the location of a difference between mutant and wild-type.

Figure 2 shows an alignment of the two p53 sequences. The mutant is on the top (Query) and the wild-type is on the bottom (Sbjct). The middle row shows how the two align. If there is a gap or a symbol (e.g. "+") in the middle row, that is the location of a difference between the two sequences. The numbers at the end of each row are the starting and ending amino acid positions for that row. In Figure 2 above, amino acid #217 differs (see the "+" in the middle alignment row at amino acid 217.

If you only see a "partial" alignment (less than 393 amino acids lined up), this means that a stop codon was inserted at the position immediately after the last amino acid in the alignment.

For each of your mutations:

1. Record the codon position (amino acid number) that is mutated.
2. Record the corresponding domain of p53 (which domain is this amino acid in? Use the p53 diagram from your preparatory reading assignment).
3. Mark your mutations on the class histogram or on the class spreadsheet of mutations (which can be sorted by amino acid number to show the pattern of mutations across the protein).

Part II – Chemical Exposure and TP53 Gene Mutations

See PowerPoint slides for this section of the case study.

Part III – Looking for the G→T "Signature" at Hotspots in Lung Cancer

Percent of Mutations at Three Hotspot Codons that are G→T in Smokers and Non-Smokers

Using the following codon data, complete Table 1 below.

Codon 157

# (count)	*Mutation*	*Smoking*
3	Deletion	smoker
2	G→A	non-smoker
1	G→T	non-smoker
22	G→T	smoker
1	T→A	smoker
2	T→G	smoker

Codon 248

# (count)	*Mutation*	*Smoking*
1	C→G	smoker
16	C→T	smoker
4	C→T	non-smoker
1	Deletion	smoker
14	G→A	smoker
4	G→A	non-smoker
1	G→C	smoker
1	G→T	non-smoker
21	G→T	smoker
1	GG→TT*	non-smoker
1	GG→TT*	smoker

Codon 273

# (count)	*Mutation*	*Smoking*
3	C→A	smoker
3	C→G	smoker
13	C→T	smoker
4	C→T	non-smoker
1	Deletion	smoker
11	G→A	smoker
6	G→A	non-smoker
2	G→C	smoker
26	G→T	smoker
4	G→T	non-smoker
1	Insertion	smoker
1	T→C	smoker

**Note:* This is *not* a G→T signature.

Table 1. Percent of mutations at three hotspot codons.

	Total Mutations (#)	*Number G→T*	*%G→T by group*
Codon 157			
Smokers			
Non-Smokers			
Codon 248			
Smokers			
Non-Smokers			
Codon 273			
Smokers			
Non-Smokers			
Combined Codons			
Smokers			
Non-Smokers			

Part IV – Smoking, Lung Cancer, and the Tobacco Companies

Chi-Squared Test of Independence

To test whether there is a statistically significant association between smoking and mutation type (G→T vs. other mutations), we will carry out a chi-square independence test. This is a statistical test that tells us how likely it is that two variables (in our case smoking and specific mutation type) are independent and do not affect each other. We initially assume that the two variables are independent, which we call the "null hypothesis." We reject this null hypothesis and accept the "alternative hypothesis," that the two variables *are* associated, if the probability (p) of the null hypothesis is less than 0.05. In other words, if there is less than a 5% chance that the two variables are not related, we accept that there is a statistically significant relationship between the variables. So let's see whether there is a statistically significant association between smoking and mutation type (G→T vs other mutations). Remember that we are only counting mutations characterized as G→T; we are not counting GG→TT as G→T mutations (although we will count these as "other" mutations).

Step 1. Observed Data

Fill in Table 2 below using the data you collected about the mutations in the three hotspot codons. Note that in this table we are looking at G→T mutations and "Other Mutations." "Other Mutations" are all mutations that are not G→T (e.g., GG→TT is an "other" mutation). You will need to enter the number of G→T and the number of "other" mutations.

Table 2. Observed data from all three codons (combined).

	G→T Mutations (#)	*Other Mutations (#)*
Smokers		
Non-smokers		

1. How many G→T mutations were there in total (Smoker + Non-smoker)? _____
2. How many "other" mutations were there in total (Smoker + Non-smoker)? _____
3. How many mutations in smokers were there (G→T + other)? _____
4. How many mutations in non-smokers were there (G→T + other)? _____
5. How many total mutations were there? _____

Step 2. Determine the Numbers You Would Expect to Get in Each Cell if Smoking and Mutation Type Are Independent

These are the numbers we would expect under the null hypothesis, i.e., that smoking does not impact mutation type. As an example, if 3/4 of our mutations are from smokers, we would expect smokers to have 3/4 of the G→T mutations and 3/4 of the other mutations. If smokers have significantly more or less than this, it suggests that smoking and mutation type are not independent.

Follow the instructions below to find the expected number of each mutation type in smokers and non-smokers if smoking and mutation type are not related. Round to the nearest hundredth (two decimal places).

6. How many G→T mutations would we expect in smokers if smoking and mutation type are not related?

   (Total # smokers) * (Total # G→T mutations) / (Total # all mutations) _____

7. How many "other" mutations would we expect in smokers if smoking and mutation type are not related?

   (Total # smokers) * (Total # other mutations) / (Total # all mutations) _____

8. How many G→T mutations would we expect in non-smokers if smoking and mutation type are not related?

   (Total # non-smokers) * (Total # G→T mutations) / (Total # all mutations) _____

9. How many "other" mutations would we expect in non-smokers if smoking and mutation type are not related?

   (Total # non-smokers) * (Total # other mutations) / (Total # all mutations) _____

Step 3. Calculate the chi-square statistic

The chi-square statistic will allow us to figure out how likely it is that the null hypothesis (that smoking and mutation type are not related) is correct. To do this we will work through each combination of mutation type and smoking (i.e., # G→T mutations in smokers). We will calculate:

$$\frac{(\text{observed}-\text{expected})^2}{\text{expected}}$$

for each combination and then sum them all together at the end. Note that the observed values are in the table in Step 1. The expected values come from Step 2.

For the following, calculate: $\frac{(\text{observed}-\text{expected})^2}{\text{expected}}$

10. G→T mutations in smokers: _____
11. "Other" mutations in smokers: _____
12. G→T mutations in non-smokers: _____
13. "Other" mutations in non-smoker: _____
14. Now add the values you got for all four combinations. This is your chi-square value: _____

Step 4. Find the Probability that Smoking and Mutation Type Are Not Related

Use the table on the following page to look up the *p* value that matches your chi-square value (round the chi-square value to the nearest tenth or 1 decimal place). The *p* value is the probability that smoking and mutation type are not related. Express this as a percent (i.e., the percent chance that smoking and mutation type are not related). *Hint:* to convert the *p* value into a percentage, multiply it by 100.

15. What does the chi-square test tell you? What is the probability that smoking is *not* related to the frequency of G→T mutations? What does this mean about the probability that smoking *is* related to the frequency of G→T mutations?

16. In the 1990s, states sued the tobacco companies, arguing that smoking caused health problems whose costs were carried by public health systems. Are these lawsuits more likely to be successful compared to the individual lawsuits? Why or why not? Use what you have learned about G→T mutations in your answer.

Chi-square lookup table for one degree of freedom.

Chi-square	*p value*		*Chi-square*	*p value*		*Chi-square*	*p value*		*Chi-square*	*p value*
1	0.3173		3	0.0833		5	0.0253		7	0.0082
1.1	0.2943		3.1	0.0783		5.1	0.0239		7.1	0.0077
1.2	0.2733		3.2	0.0736		5.2	0.0226		7.2	0.0073
1.3	0.2542		3.3	0.0693		5.3	0.0213		7.3	0.0069
1.4	0.2367		3.4	0.0652		5.4	0.0201		7.4	0.0065
1.5	0.2207		3.5	0.0614		5.5	0.019		7.5	0.0062
1.6	0.2059		3.6	0.0578		5.6	0.018		7.6	0.0058
1.7	0.1923		3.7	0.0544		5.7	0.017		7.7	0.0055
1.8	0.1797		3.8	0.0513		5.8	0.016		7.8	0.0052
1.9	0.1681		3.9	0.0483		5.9	0.0151		7.9	0.0049
2	0.1573		4	0.0455		6	0.0143		8	0.0047
2.1	0.1473		4.1	0.0429		6.1	0.0135		8.1	0.0044
2.2	0.138		4.2	0.0404		6.2	0.0128		8.2	0.0042
2.3	0.1294		4.3	0.0381		6.3	0.0121		8.3	0.004
2.4	0.1213		4.4	0.0359		6.4	0.0114		8.4	0.0038
2.5	0.1138		4.5	0.0339		6.5	0.0108		8.5	0.0036
2.6	0.1069		4.6	0.032		6.6	0.0102		8.6	0.0034
2.7	0.1003		4.7	0.0302		6.7	0.0096		8.7	0.0032
2.8	0.0943		4.8	0.0285		6.8	0.0091		8.8	0.003
2.9	0.0886		4.9	0.0269		6.9	0.0086		8.9	0.0029

Wrap-Up Questions

Option A

Do some research to determine what "forward-looking" remedies Judge Kessler imposed on the tobacco companies, and how these are being implemented.

Option B

First, use an internet search engine to find the text of "A Frank Statement to Cigarette Smokers" that was published in many newspapers in January of 1954. Then go to the Philip Morris website, and review their statement under the heading "Smoking and Health Issues." Compare and contrast these statements, then consider what evidence contributed to this change in stance by "big tobacco."

For the text of the Frank Statement: <https://www.tobaccofreekids.org/assets/factsheets/0268.pdf>

For the Philip Morris smoking and health issues page: <https://www.philipmorrisusa.com/products/smoking-and-health-issues>