eXplainable Artificial Intelligence (XAI) for the NOC tool in DNAxs

You have been invited to participate in a research study made by Marthe Veldhuis (TU Delft & FBDA NFI) for her MSc thesis. It will take approximately 25-40 minutes to complete. We really appreciate your participation!

The purpose of this research study is find out how informative and user-friendly different types of explanations are to help you understand how the NOC tool in DNAxs makes predictions. The explanations aim to help you see how a certain prediction is made by the tool. This could increase your trust in the prediction, or show you that a certain prediction might be wrong.

Your participation in this study is entirely voluntary and you can withdraw at any time. We believe there are no known risks associated with this research study; however, as with any online related activity the risk of a breach is always possible. We will minimize any risks by safely storing the data in a personal google drive account. To the best of our ability your answers in this study will remain confidential and anonymous.

If you have any questions at all, do not hesitate to reach out at m.veldhuis@nfi.nl (until 2-7-2021), or at msveldhuis96@gmail.com.

Do you agree with these terms? *

- (●) Yes

Demographics

We would like to get an impression of the demographics of the participants before we start the experiment.

What is your age? *

- ○ 18 - 24
- ○ 25 - 34
- ○ 35 - 44
- ◉ 45 - 54
- ○ 55 - 64
- ○ 65+
- ○ prefer not to say

How well can you understand written and spoken English? *

	1	2	3	4	5	
Not well	○	○	○	○	◉	Very well

How do you feel about incorporating a Machine Learning tool (such as the NOC tool) in your decision-making process? *

	1	2	3	4	5	
Strongly oppose	○	○	○	◉	○	Strongly support

Part A: Making decisions - Introduction (~15-20 min.)

In this section of the survey, we will ask you to determine the Number Of Contributors (NOC) based on predictions by the NOC tool. The aim is to find out if certain explanations can help you gain more insight into the predictions made by the NOC tool. The aim is not to see how well you can determine the NOC yourself. Therefore it is not necessary to look at the DNA profile.

If you feel like it is essential for you to have the profile at hand, you can view it in DNAxs (including the electropherogram). A DNAxs case has been created and named “NOC-XAI Evaluation data”. If the case is in use, you may want to load the data yourself. The data for upload in DNAxs can be found in G:\BI\BI_DigiD\0. Research\Corina\NOC-XAI Evaluation data.

We present an introductory video for each type of explanation, after which you need to answer one question to see if you understood the video. There are 3 types in total.

Hint: if the pictures are too small, right-click it and select "Open image in new tab", or use ctrl+scroll to zoom in on the survey.

Disclaimer: features

Each profile is represented by a group of 19 statistical features that were calculated for each profile. For example, counting the number of loci with 0 alleles (drop-out loci). The NOC tool makes predictions based solely on these 19 features.

We know that many of these features are not easily understood. This will need to change in the future. For now, to make it a bit simpler, please try to focus on the following features:

- TAC and MAC: you are familiar with these values; the Total Allele Count and Maximum Allele Count.
- Loci with 0 alleles and Loci with 5-6 alleles: the number of loci with 0, or 5-6 alleles. Notice that the NOC tool does not know how many loci with 1-2 / 3-4 / 7-8 / 9+ alleles there are.
- Peaks below 800 RFU and Allele count variation: interpret these values as an indication of the quality of the profile; lots of lower peaks / more variation in the allele counts per locus could indicate a lower quality profile with more drop-out etc.
- All locus-specific features: interpret all these as indications of the amount of information at each of those loci.

These should give you enough information to complete the exercises, but feel free to ask for more information if you would like to know more. The above text will be available during the exercise.

Introduction SHAP explanation

Please watch this video to learn about the SHAP explanation (< 2min.). You can open the video on YouTube to watch in fullscreen.

Summary SHAP explanation

- Shows which feature values of this profile the NOC tool used to make this prediction.
- Some values push the prediction down towards 1 or 2 (blue bars), while other values push the prediction up towards 4 or 5 (or more) (red bars). This is relation to the base value of 3.
- We only have information about the current profile and prediction.

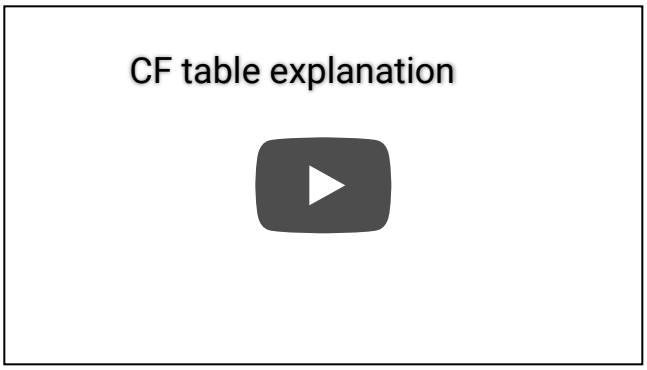
What is true about the SHAP explanation example we saw? *

- (●) The value of "Allele count variation = 0.9991" of this profile has caused the NOC tool to make a slightly higher prediction.
- ( ) The value of "Allele count variation = 0.9991" of this profile has caused the NOC tool to make a slightly lower prediction.

Introduction Counterfactual table explanation

	2.29		Run 1_Trace 1613475856276
MAC	4.000000	MAC	4.000000
TAC	68.000000	TAC	77.000000
CSF1PO min. NOC	2.000000	CSF1PO min. NOC	2.000000
D16S539 min. NOC	2.000000	D16S539 min. NOC	2.000000
D1S1656 perc. known alleles	20.000000	D1S1656 perc. known alleles	27.272727
D3S1358 allele count	4.000000	D3S1358 allele count	4.000000
D8S1179 allele count	3.000000	D8S1179 allele count	3.000000
Penta D min. NOC	1.000000	Penta D min. NOC	1.000000
Penta E min. NOC	2.000000	Penta E min. NOC	2.000000
TH01 sum of allele freq.	0.346523	TH01 sum of allele freq.	0.636930
TPOX allele count	2.000000	TPOX allele count	2.000000
TPOX min. NOC	1.000000	TPOX min. NOC	1.000000
vWa peak height variation	0.000000	vWa peak height variation	3744.780192
Allele count variation	0.999054	Allele count variation	0.698364
Loci with 0 alleles	0.000000	Loci with 0 alleles	0.000000
Loci with 5-6 alleles	0.000000	Loci with 5-6 alleles	0.000000
Peaks below 800 RFU	2.000000	Peaks below 800 RFU	11.000000
Random match proba.	0.000000	Random match proba.	0.000000
Min. NOC	1.000000	Min. NOC	1.000000
		NOC	3.000000

Please watch this video to learn about the Counterfactual table explanation (2min.). You can open the video on YouTube to watch in fullscreen.



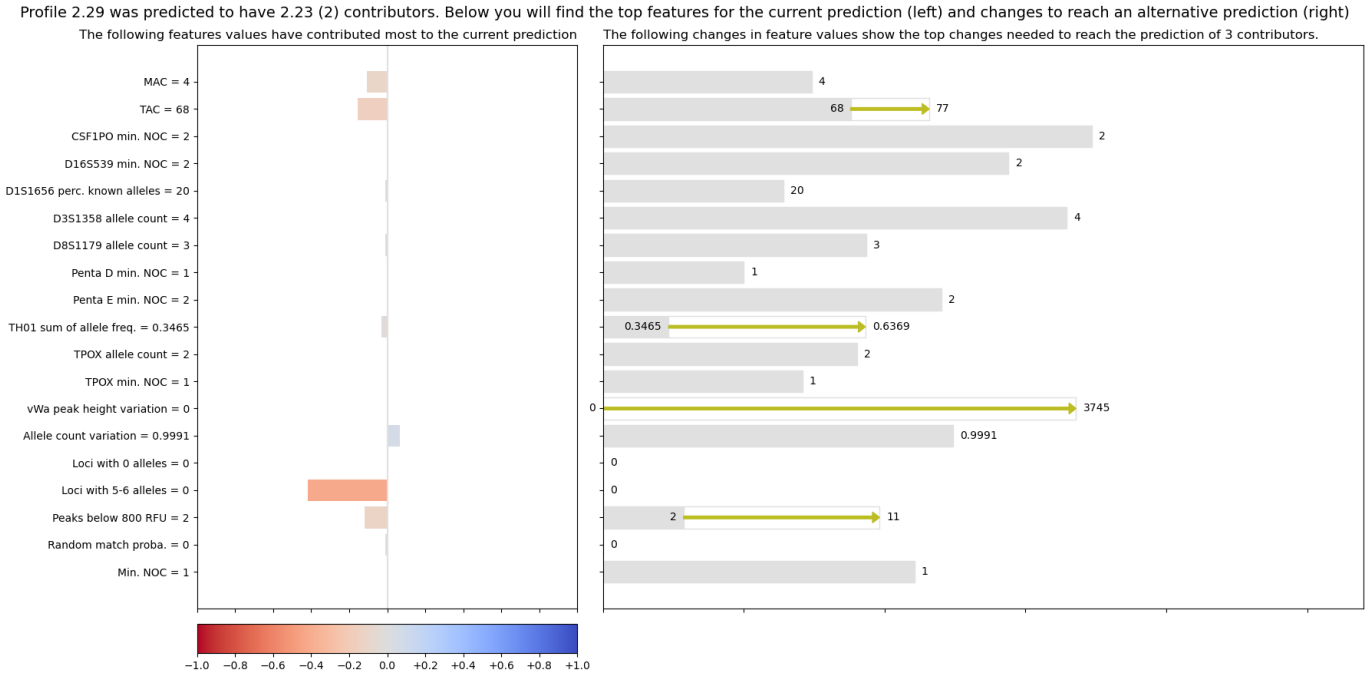
Summary CounterFactual (CF) table explanation

- Shows a comparison of the current profile with a profile that had a different rounded-off prediction.
- Differences are highlighted in red boxes.
- We do not know which differences are relevant to arrive at another prediction.

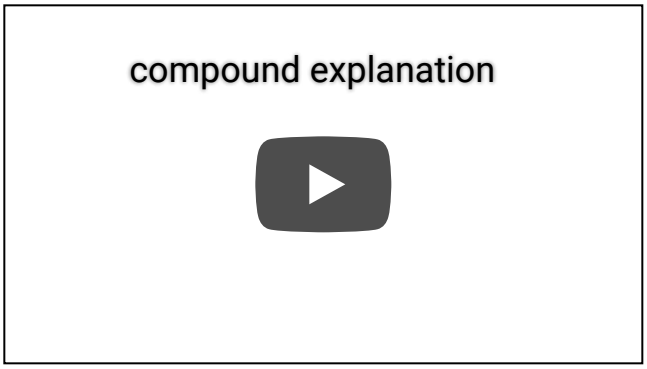
What is true about the CF table explanation example we saw? *

- ○ The feature values highlighted in red boxes have the most influence to change the prediction from 2 to 3 contributors.
- ◉ The feature values highlighted in red boxes are differences between two example profiles with a prediction of 2 and 3 contributors.

Introduction Compound explanation



Please watch this video to learn about the Compound explanation (< 3min.). You can open the video on YouTube to watch in fullscreen.



Summary Compound explanation

- Consists of two parts; one showing how the feature values have influenced the prediction, one showing the relevant feature value changes needed to reach a different prediction. The only connection the two parts share are the feature values.
- Shows how high each feature value (change) is in comparison to the total possible range of that feature.
- By looking at the right side; seeing how large and relevant the feature value changes are, you could determine how clear the NOC tool is about the prediction.

What is true about the Compound explanation example we saw? *

- ( ) The left side of the explanation shows the same information as the SHAP explanation, but the right side shows only relevant feature value changes, which is different from the Counterfactual table.
- (•) The left side of the explanation shows the same information as the SHAP explanation, and the right side shows the same information as the Counterfactual table explanation.

What is true about the Compound explanation example we saw? *

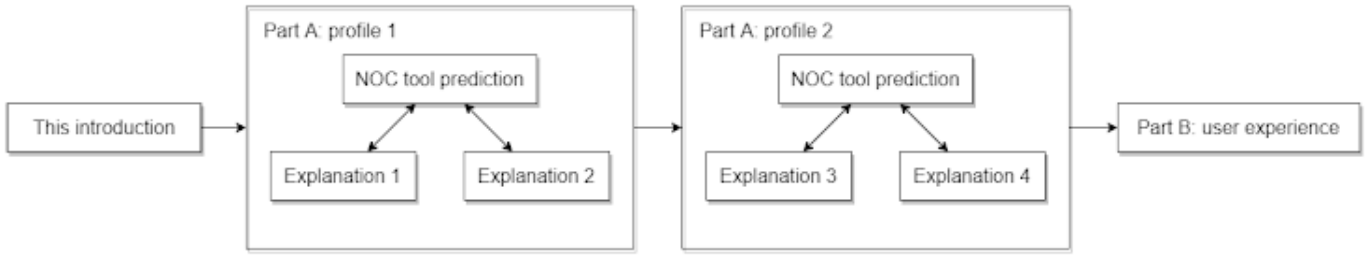
- (•) If we change the TAC value from 68 to 77, we know for certain that the TAC value's red bar on left would become blue.
- ( ) If we change the TAC value from 68 to 77, we do not know for certain how that will affect the red and blue bars on the left, because they only give information about the current profile's feature values.

Moving on to Part A: making decisions

Now that we are somewhat familiar with the types of explanations, we would like to put their usefulness to the test.

We will present 2 profiles in total. For each profile, we will show the NOC tool prediction, and 2 explanations (see figure below).
We want to see if each explanation helps you understand more about the NOC tool, in comparison to only seeing the prediction number.

We are mainly interested in 2 things:
1. When the NOC is quite clear to the NOC tool, can the explanation show this to you (take away your doubt that the prediction might be right)?
2. When the NOC is not so clear to the NOC tool, can the explanation show this to you (increase your doubt that the prediction might be wrong)?



Part A: Making decisions - Profile 1 out of 2: 1_6B.Trace#01 (5-10 min.)

Remember to open images in a new tab (right click on image + open in new tab) to view them up close.

Reminder features:
- TAC and MAC: you are familiar with these values; the Total Allele Count and Maximum Allele Count.
- Loci with 0 alleles and Loci with 5-6 alleles: the number of loci with 0, or 5-6 alleles. Notice that the NOC tool does not know how many loci with 1-2 / 3-4 / 7-8 / 9+ alleles there are.
- Peaks below 800 RFU and Allele count variation: interpret these values as an indication of the quality of the profile; lots of lower peaks / more variation in the allele counts per locus could indicate a lower quality profile with more drop-out etc.
- All locus-specific features: interpret all these as indications of the amount of information at each of those loci.

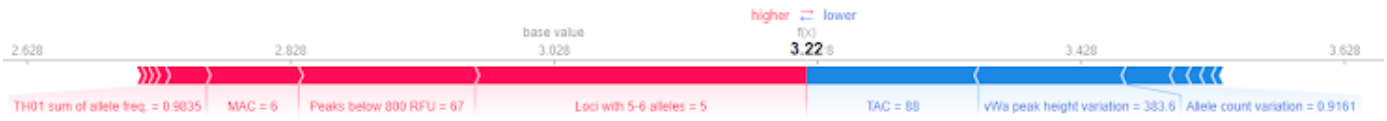
1. Please select all number(s) of contributors you would consider after seeing a prediction of 3 (3.22). *

- [ ] 1
- [ ] 2
- [x] 3
- [x] 4
- [ ] 5

2. Do you think the prediction is correct? *

- (•) Yes
- ( ) No
- ( ) I don't know

3. Please select all number(s) of contributors you would consider when you can consult this explanation (SHAP). Do you consider the same, or less options than in question 1? *



- [ ] 1
- [ ] 2
- [x] 3
- [x] 4
- [ ] 5

4. Do you think the prediction of 3 (3.22) is correct? *

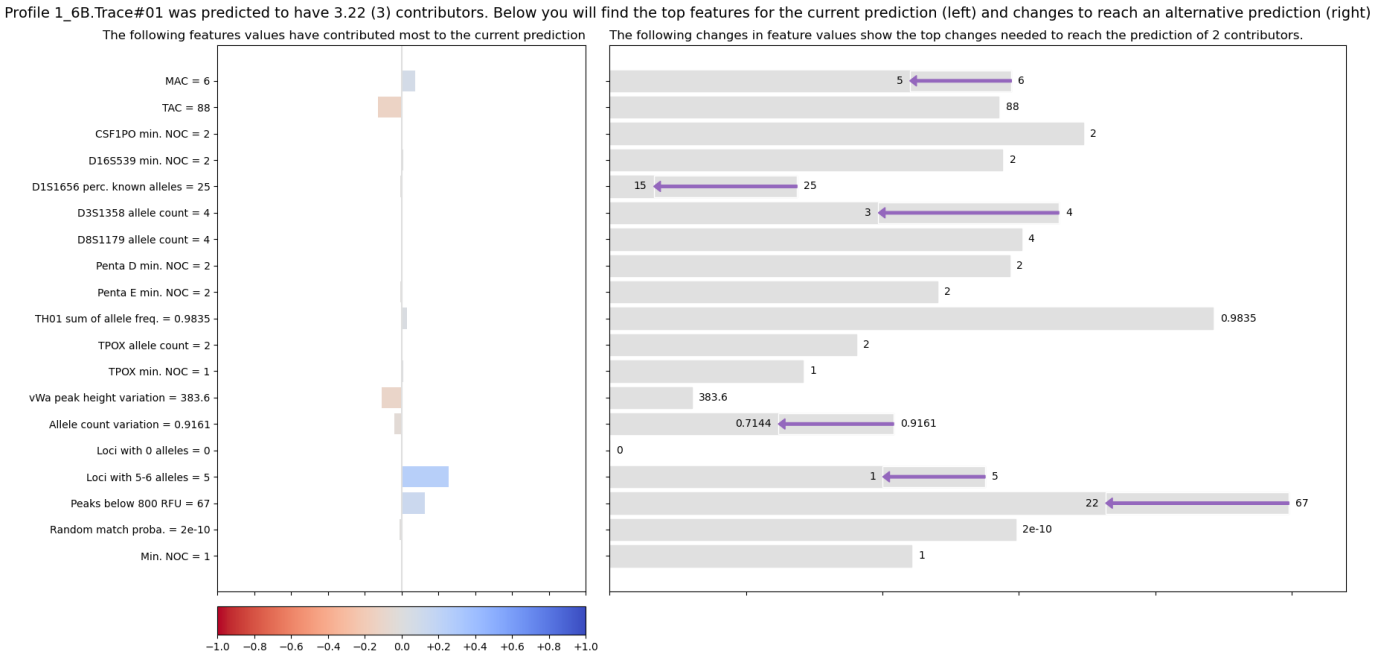
- (●) Yes
- ( ) No
- ( ) I don't know

Can you explain why you have answered questions 3 and 4 differently (or not) than questions 1 and 2 after looking at the SHAP explanation, in comparison to only seeing the prediction of 3 (3.22)? *

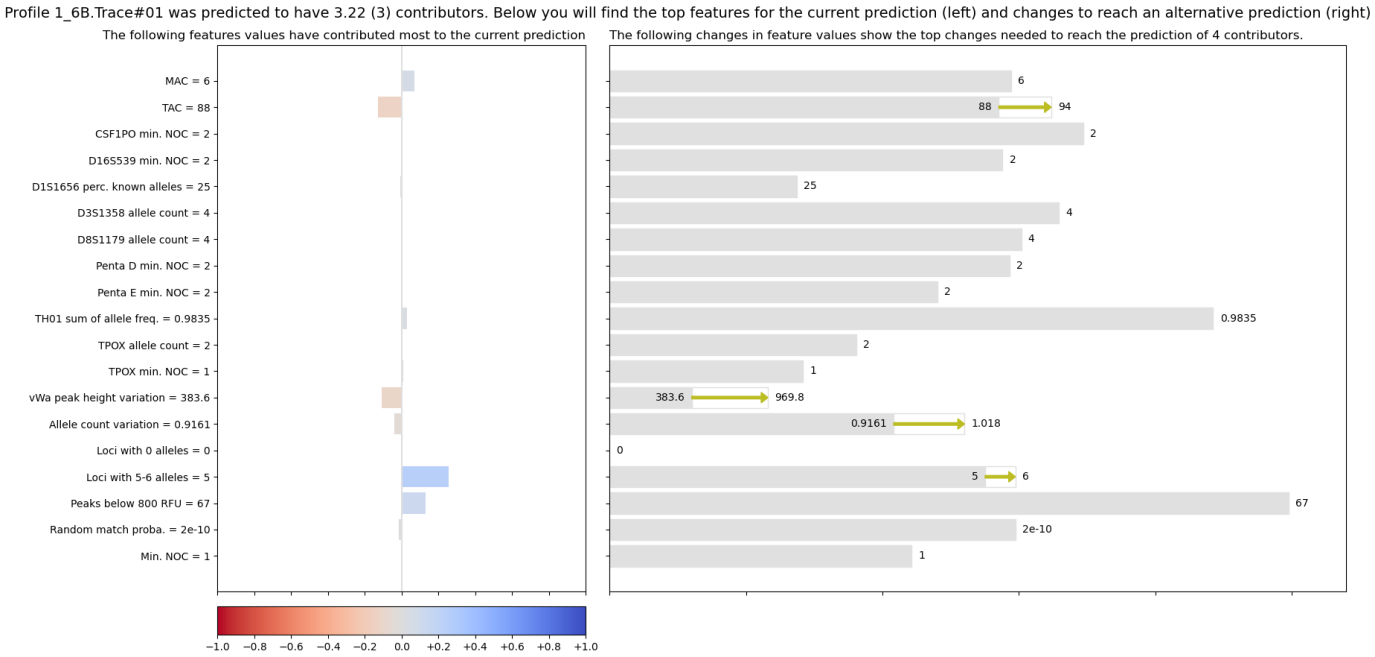
Need to see the profile to make a difference.

(Questions are below the pictures)

Right side showing how to reach a prediction of 2 contributors



Right side showing how to reach a prediction of 4 contributors



5. Please select all number(s) of contributors you would consider when you can consult the explanation above (Compound explanation). Do you consider the same, or less options than in question 1? Note: we are comparing with predictions of 2 and 4 donors. Normally, you would be able to choose which comparison you would like to make. *

- [ ] 1
- [ ] 2
- [x] 3
- [x] 4
- [ ] 5

6. Do you think the prediction of 3 (3.22) is correct? *

- (●) Yes
- ( ) No
- ( ) I don't know

Can you explain why you have answered questions 5 and 6 differently (or not) than questions 1 and 2 after looking at the Compound explanation, in comparison to only seeing the prediction of 3 (3.22)? *

Need to see the profile to make a difference.

Part A: Making decisions. Profile 2 out of 2: 2A3.3 (5-10 min.)

Remember to open images in a new tab (right click on image + open in new tab) to view them up close.

Reminder features:
- TAC and MAC: you are familiar with these values; the Total Allele Count and Maximum Allele Count.
- Loci with 0 alleles and Loci with 5-6 alleles: the number of loci with 0, or 5-6 alleles. Notice that the NOC tool does not know how many loci with 1-2 / 3-4 / 7-8 / 9+ alleles there are.
- Peaks below 800 RFU and Allele count variation: interpret these values as an indication of the quality of the profile; lots of lower peaks / more variation in the allele counts per locus could indicate a lower quality profile with more drop-out etc.
- All locus-specific features: interpret all these as indications of the amount of information at each of those loci.

1. Please select all number(s) of contributors you would consider after seeing a prediction of 4 (3.53) *

- [ ] 1
- [ ] 2
- [x] 3
- [x] 4
- [ ] 5

2. Do you think the prediction is correct? *

- (•) Yes
- ( ) No
- ( ) I don't know

3. Please select all number(s) of contributors you would consider when you can consult this explanation (Counterfactual table). Do you consider the same, or less options than in question 1? Note: we are comparing with example profiles with a prediction of 3 and 5 donors. Normally, you would be able to choose which comparison you would like to make. *

	2A3.3		2C3.3
MAC	6.000000	MAC	6.000000
TAC	98.000000	TAC	96.000000
CSF1PO min. NOC	2.000000	CSF1PO min. NOC	2.000000
D16S539 min. NOC	2.000000	D16S539 min. NOC	2.000000
D1S1656 perc. known alleles	25.000000	D1S1656 perc. known alleles	25.000000
D3S1358 allele count	5.000000	D3S1358 allele count	5.000000
D8S1179 allele count	4.000000	D8S1179 allele count	4.000000
Penta D min. NOC	3.000000	Penta D min. NOC	3.000000
Penta E min. NOC	3.000000	Penta E min. NOC	3.000000
TH01 sum of allele freq.	0.848201	TH01 sum of allele freq.	0.848201
TPOX allele count	3.000000	TPOX allele count	3.000000
TPOX min. NOC	2.000000	TPOX min. NOC	2.000000
vWa peak height variation	906.029889	vWa peak height variation	430.111334
Allele count variation	0.987636	Allele count variation	0.916144
Loci with 0 alleles	0.000000	Loci with 0 alleles	0.000000
Loci with 5-6 alleles	11.000000	Loci with 5-6 alleles	10.000000
Peaks below 800 RFU	5.000000	Peaks below 800 RFU	7.000000
Random match proba.	0.000000	Random match proba.	0.000000
Min. NOC	2.000000	Min. NOC	2.000000
		NOC	3.000000

	2A3.3		5.27
MAC	6.000000	MAC	6.000000
TAC	98.000000	TAC	101.00000
CSF1PO min. NOC	2.000000	CSF1PO min. NOC	2.000000
D16S539 min. NOC	2.000000	D16S539 min. NOC	2.000000
D1S1656 perc. known alleles	25.000000	D1S1656 perc. known alleles	25.000000
D3S1358 allele count	5.000000	D3S1358 allele count	5.000000
D8S1179 allele count	4.000000	D8S1179 allele count	4.000000
Penta D min. NOC	3.000000	Penta D min. NOC	2.000000
Penta E min. NOC	3.000000	Penta E min. NOC	3.000000
TH01 sum of allele freq.	0.848201	TH01 sum of allele freq.	0.683693
TPOX allele count	3.000000	TPOX allele count	3.000000
TPOX min. NOC	2.000000	TPOX min. NOC	2.000000
vWa peak height variation	906.029889	vWa peak height variation	302.477600
Allele count variation	0.987636	Allele count variation	0.920261
Loci with 0 alleles	0.000000	Loci with 0 alleles	0.000000
Loci with 5-6 alleles	11.000000	Loci with 5-6 alleles	10.000000
Peaks below 800 RFU	5.000000	Peaks below 800 RFU	14.000000
Random match proba.	0.000000	Random match proba.	8.60e-08
Min. NOC	2.000000	Min. NOC	2.000000
		NOC	5.000000

- [ ] 1
- [ ] 2

- [x] 3
- [x] 4
- [ ] 5

4. Do you think the prediction of 4 (3.53) is correct? *

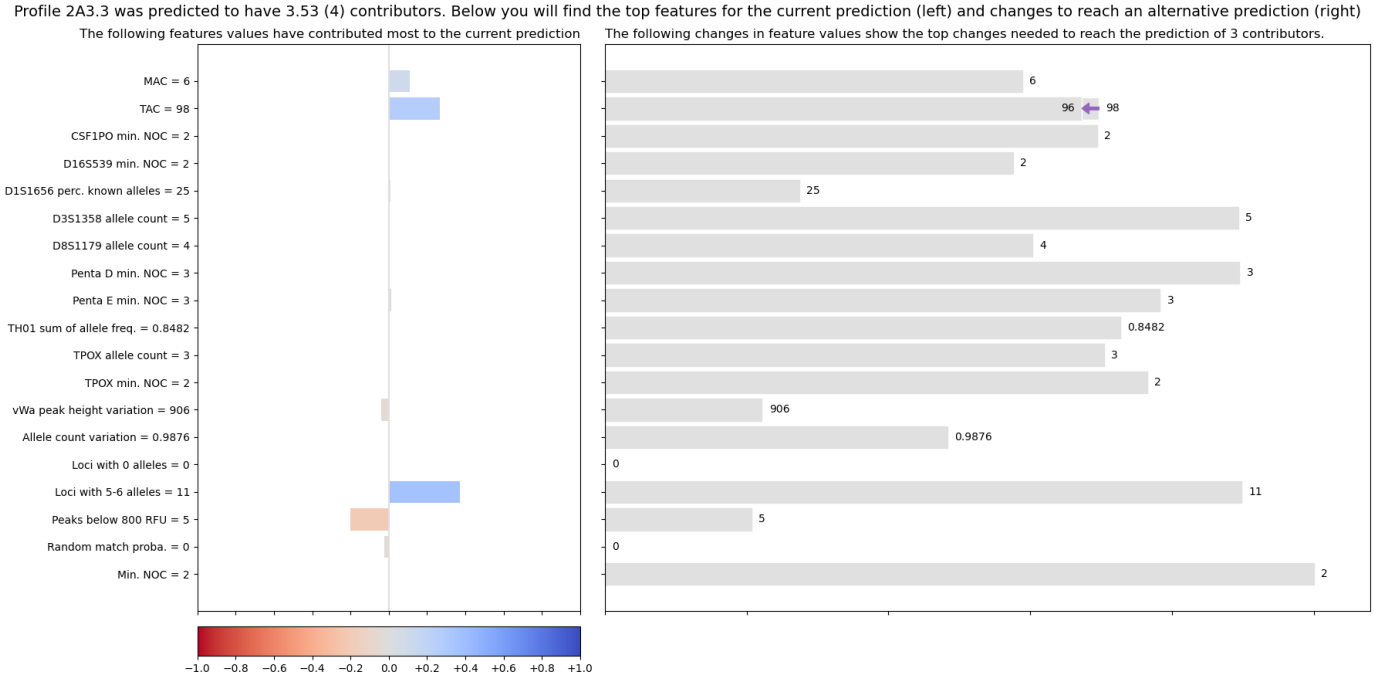
- (•) Yes
- ( ) No
- ( ) I don't know

Can you explain why you have answered questions 3 and 4 differently (or not) than questions 1 and 2 after looking at the Counterfactual table explanation, in comparison to only seeing the prediction of 4 (3.53)? *

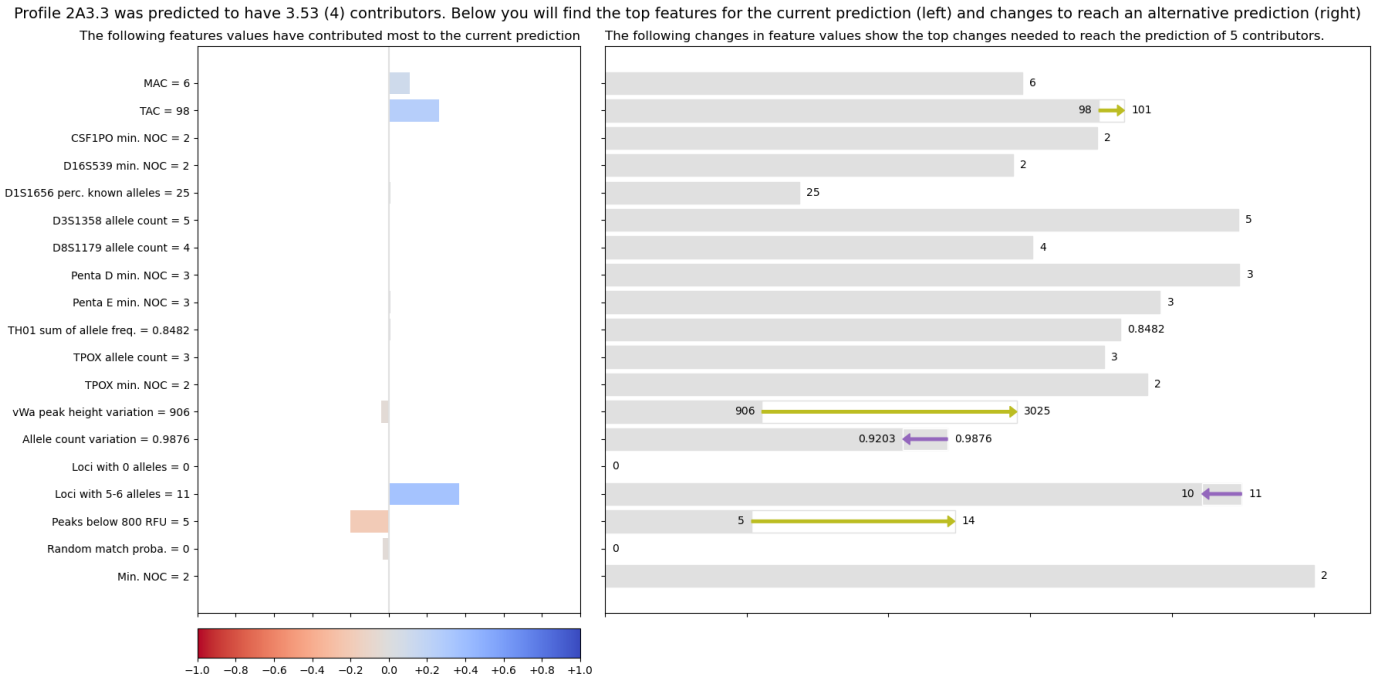
Need to see the profile to make a difference.

(Questions are below the images)

Right side showing how to reach a prediction of 3 contributors



Right side showing how to reach a prediction of 5 contributors



5. Please select all number(s) of contributors you would consider when you can consult the explanation above (Compound explanation). Do you consider the same, or less options than in question 1? Note: we are comparing with predictions of 3 and 5 donors. Normally, you would be able to choose which comparison you would like to make. *

- [ ] 1
- [ ] 2
- [x] 3
- [x] 4
- [ ] 5

6. Do you think the prediction of 4 (3.53) is correct? *

- (•) Yes
- ( ) No
- ( ) I don't know

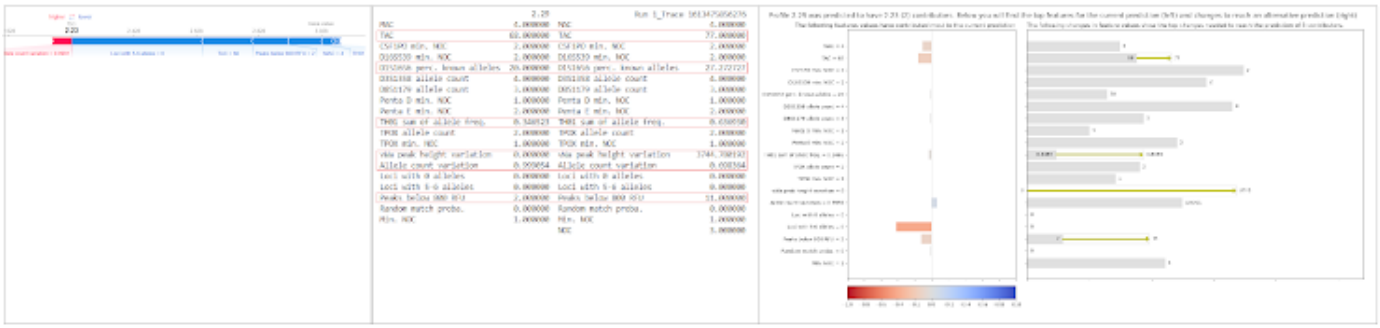
Can you explain why you have answered questions 5 and 6 differently (or not) than questions 1 and 2 after looking at the Compound explanation, in comparison to only seeing the prediction of 4 (3.53)? *

Need to see the profile to make a difference.

Part B: User experience (< 5 min.)

In this section, we will determine your preference for any of the explanations.

Please select which explanation you preferred in terms of: *



	SHAP	Counterfactual table	Compound visualization
Ease of use (how easily you could find the relevant information)	○	○	◉
Appeal (how nice you thought it was to use)	◉	○	○
Completeness (how well you can form a total picture of the prediction)	○	○	◉

What are the main reasons for picking a certain explanation over the other? *

- [x] There was more information
- [x] There was less information, but the information was more relevant
- [ ] The information was easy to find
- [x] It was presented in a visual way
- [ ] It compared two options
- [ ] It was easier to understand
- [ ] It seemed more like how I would think about this problem
- [ ] Anders: .....

Conclusion

Thank you for participating in this survey! Do not hesitate to reach out if you have more questions or remarks at m.veldhuis@nfi.nl (until 2-7) or msveldhuis96@gmail.com. Once the study has been completed, we will share the results for those who are interested.

Do you have any remarks or feedback that you would like to share?

.....