Outline additional chapters

1. Background interpretation number of contributors (NOC) in short tandem repeat (STR) profiles
   1. STR profiles
   2. Tools for NOC interpretation
      1. MAC
      2. nC-tool
      3. NOC Machine Learning tool
      4. Decision tree implementation
   3. Results survey NOC interpretation experts BiS from NFI
      1. Results workflow
      2. Results type of data
      3. Results type of explanations
2. Background explainable artificial intelligence (XAI)
   1. Model agnostic local explanations
   2. “Factual” methods
      1. Anchors
      2. SHAP
   3. Counterfactual methods
      1. *Wachter et al.*
      2. Multi-Objective Counterfactuals (*Dandl et al*)
      3. FACE
3. Analysis data & model
   1. Original 590
   2. Sampled 5000
      1. EuroForMix
      2. Parameters
   3. Benchmark models
4. Experiments with explanations
   1. Counterfactual visualizations
      1. Only show changed features
      2. Scale features
      3. Each feature has their own scale
      4. Interpretation by experts
   2. Anchors
      1. Research into sampling
      2. Research into *anchored counterfactuals*
      3. Interpretation by experts
   3. SHAP
      1. Multi-class classification
      2. Regression
      3. Interpretation by experts

# DNA Mixture interpretation

Experts can use DNA evidence to determine if certain people were involved in a crime by comparing the suspect DNA, victim DNA and other DNA samples to the evidence found at a crime scene. This interpretation become more difficult when the DNA profile consists of evidence from multiple people since information might overlap, or not every person contributed as much material. Even though software exists for analyzing this evidence, it is required that the expert inputs how many people contributed to the sample [1]. This chapter explains how to interpret a specific type of DNA profile, and highlights the different methods to determine the number of contributors.

## Short Tandem Repeat (STR) profiles

In forensic work, DNA evidence is often analyzed using *Short Tandem Repeat (STR)* profiles. These STR are specific tracks of repeated short DNA sequences of about two to six base pairs long that have been proven to show high variability between individuals in how many times the sequence repeats [2]. Most of these parts of the DNA or *loci* have been defined by CODIS, the United States national DNA database. We can capture the STR with a process called electrophoresis, which produces an electropherogram. In Figure 1, we see a simplified result that the electropherogram can produce for locus TH01. The y-axis shows the amount of information found in Relative Fluorescent Units (RFU), which is how the machine counts the quantity of DNA found. The x-axis shows the location of the locus on the DNA strand. Most importantly, we see two peaks, representing two alleles on this locus. These alleles are characterized by the number of repeats of the STR for locus TH01, which is [AATG]. On the right of Figure 1, we see the DNA sequence for six and eight repeats.

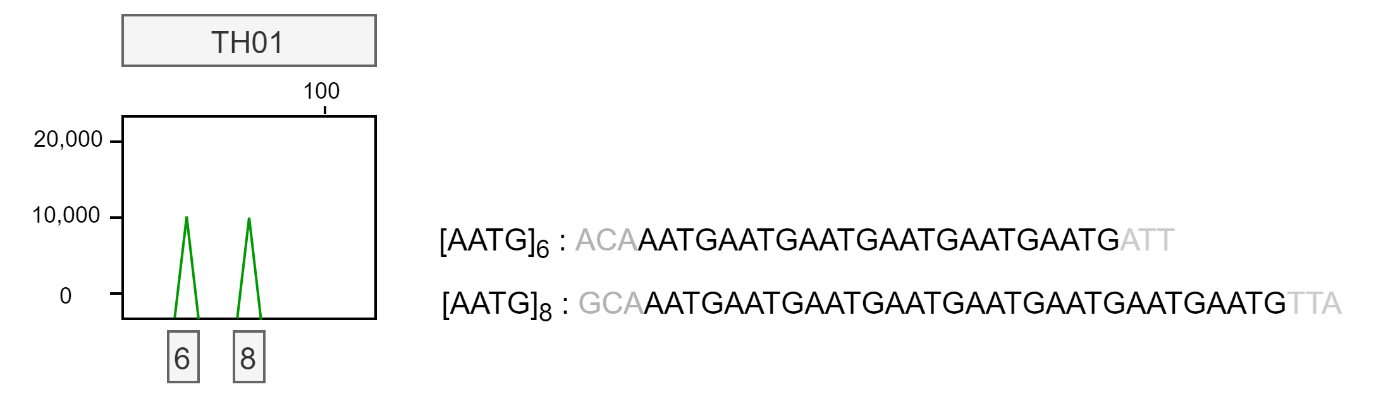


Figure 1: Simplified electropherogram result for locus TH01 showing two alleles with six and eight repeats each. The repeat sequence is shown on the right with arbitrary flanking regions.

One individual can have two different alleles for a single locus; one inherited from the mother, and one from the father. It is also possible that a person inherited the same allele from both of their parents, this means that they are homozygous at that locus. The peak will then be twice as large. We will now get into more detail of how to derive the number of contributors from an STR profile.

## Estimating the Number of Contributors (NOC)

The first step of DNA STR profile interpretation is to determine whether a sample has originated from a single source, or if the sample is a mixture [3]. This often easily discerned by checking whether or not there are loci with more than two alleles present. As we saw in Figure 1, a single person can contribute a maximum of two alleles per locus, so profiles with more alleles are considered a mixture. The next step is to determine the number of contributors. This step is necessary for DNA analysis software to calculate the weight of the evidence found [4]. When an incorrect NOC is used for further analysis involving the investigation of the DNA profiles, the results are unreliable [5]. It could make the difference between whether or not a person of interest is included in the evidence or not.

Determining the exact number of contributors is difficult. There are several obscuring factors that could make an expert underestimate the number of donors, especially when the number of donors increases [3, 6]:

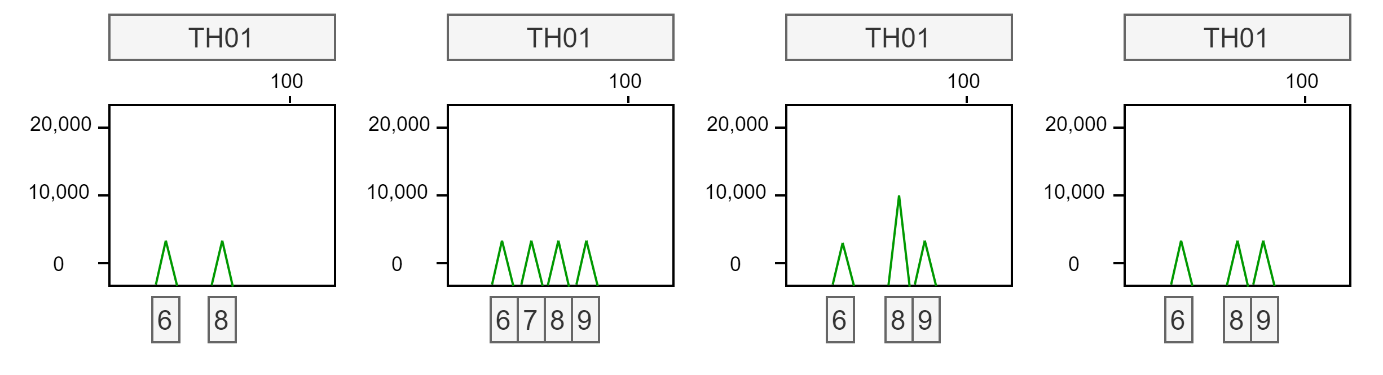


Figure 2: Four simplified electropherogram results for locus TH01. From left to right: Example of a single donor profile; Example of a 2-person mixture profile; Example of a 2-person mixture profile with allele sharing, peak 8 is twice as high compared to peak 6 and 9; Example of a 2-person mixture profile with drop-out, one peak has likely not been detected.

* Allele sharing: If two donors have the same allele at a locus, this is called allele sharing. It frequently occurs when donors are relatives, since siblings share a lot of DNA. It might be difficult to distinguish if an allele is shared between donors, or if a single donor simply is homozygous for this allele; in both cases, the peak height for that allele is higher. This can be seen in the third picture of Figure 2; allele 8 has twice as much information as alleles 6 and 9.
* Allele drop-out: If the DNA was degraded, for example due to sunlight, some parts of the DNA might not be present in the sample to measure. It is also possible that the amount of the DNA available is so small, that the alleles fall below a certain noise filter. Because of this low quality or quantity of DNA, some allele fragments might not show up in the profile at all, which is called drop-out.

These factors can decrease the number of alleles found in a certain profile, which could lead to an underestimation of the number of contributors. There are also factors that could lead to an overestimation of alleles present in a sample [3]:

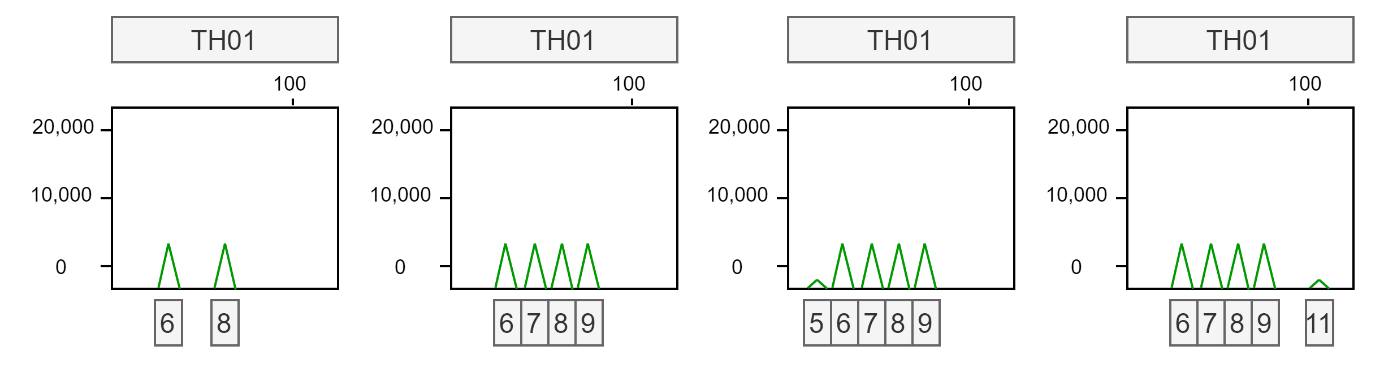


Figure 3: Four simplified electropherogram results for locus TH01. From left to right: Example of a single donor profile; Example of a 2-person mixture profile; Example of a 2-person mixture profile with a stutter peak at allele 5 caused by a STR with 6 repeats folding; Example of a 2-person mixture profile with a noise peak at location 11 caused by an error in reading.

* Stutter: During the process of measuring the STRs, a STR fragment can accidentally fold. This could cause the electropherogram to measure this strand to have one repeat fewer, since the folded part of the fragment is not correctly measured. In this way, a small stutter peak is found in the profile just before the valid peak. In Figure 3, this is
* Allele drop-in or other noise: The measuring process is not perfect, so some random noise might show up in the electropherogram, that does not contain any information about the DNA. In Figure 3, we can see that the rightmost image has a small peak at allele 11. Since it is not close to another allele, it is likely not a stutter peak.

Stutter peaks and noise are often filtered out using certain thresholds. As a result, some DNA information might also be lost due to a low-quantity donor.

### MAC method

The simplest method to get an estimate of the NOC is by using the Maximum Allele Count (MAC) [3, 7]. By taking the locus with the most alleles present, dividing that number by two, and rounding up, we can get an idea of the minimum NOC. Though this method is simple, it is unreliable due to the factors of allele sharing, drop-out, etc. Performance in general is quite poor, especially with 3 or more contributors, when there is a lot of allele sharing, or when the quality of the profile is low [8, 9]. On average, when assessing mixtures between 2-5 contributors, the MAC obtains correct predictions for about 60-70% of samples [9-11].

### nC-tool

Often experts use MAC in combination with the Total Allele Count (TAC), which measures the total number of alleles across all loci. However, this measure suffers from the same obscuring factors as the MAC. The nC-tool takes both these measures into account, as well as simulating various levels of dropout [11]. This achieves better results than using the MAC only, obtaining correct predictions for roughly 76% of 2-5 person mixtures [10, 11].

Various

# eXplainable Artificial Intelligence (XAI)

## SHAP

Even though SHAP splits the impact of correlated features, and therefore often underestimates the magnitude of the impact of these features on the model, we deem the overall value of SHAP to give a general overview of which features were used by the model for certain predictions.

For correlated features, SHAP values will turn out lower for each feature

It is an issue that feature importance methods split the impact on the model over correlated features. The result of this issue is that the importance value for correlated contributing features is underestimated, in contrast to if their importance was left undivided. However, since main goal of these values is to give an impression of the contributing factors to a prediction, the exact values are not a priority. This part of the explanation is to observe a general sense of which features contributed to the prediction in which direction. For this purpose, we deem SHAP adequate.

1. Taylor, D., J.-A. Bright, and J. Buckleton, *Interpreting forensic DNA profiling evidence without specifying the number of contributors.* Forensic Science International: Genetics, 2014. **13**: p. 269-280.

2. Boavida, A., et al., *PowerPlex® fusion 6C system: internal validation study.* Forensic sciences research, 2018. **3**(2): p. 130-137.

3. Intituut, N.F., *HBS: Interpretatie van autosomale STR DNA-profielen*. 2020. p. 16.

4. Bleka, Ø., G. Storvik, and P. Gill, *EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts.* Forensic Science International: Genetics, 2016. **21**: p. 35-44.

5. Benschop, C.C.G., et al., *The effect of varying the number of contributors on likelihood ratios for complex DNA mixtures.* Forensic Science International: Genetics, 2015. **19**: p. 92-99.

6. Coble, M.D., et al., *Uncertainty in the number of contributors in the proposed new CODIS set.* Forensic Science International: Genetics, 2015. **19**: p. 207-211.

7. Clayton, T.M., et al., *Analysis and interpretation of mixed forensic stains using DNA STR profiling.* Forensic Science International, 1998. **91**(1): p. 55-70.

8. Benschop, C., *PowerPlex Fusion 6C Profile analysis & interpretation.* 2020.

9. Haned, H., et al., *Estimating the Number of Contributors to Forensic DNA Mixtures: Does Maximum Likelihood Perform Better Than Maximum Allele Count?* Journal of Forensic Sciences, 2011. **56**(1): p. 23-28.

10. Benschop, C.C.G., et al., *Automated estimation of the number of contributors in autosomal short tandem repeat profiles using a machine learning approach.* Forensic Science International: Genetics, 2019. **43**: p. 102150.

11. Benschop, C., A. Backx, and T. Sijen, *Automated estimation of the number of contributors in autosomal STR profiles.* Forensic Science International: Genetics Supplement Series, 2019. **7**.