CSI Adelaide: Who Killed The Somerton Man?

Supervisors: Professor Derek Abbott

Students: Ashley Seroka and Jiawei Chen

Outline

- Introduction
  - History
  - Evidence
  - Project aims
- Previous Study
  - Professional Attempts
  - Honours Project 2009-2015
- Specific Tasks
  - Task 1: Letters
  - Task 2: Mass Spectrometer
  - Task 3: DNA
- Future Work
- Summary
- Acknowledgement
- References
- Questions

Introduction: History
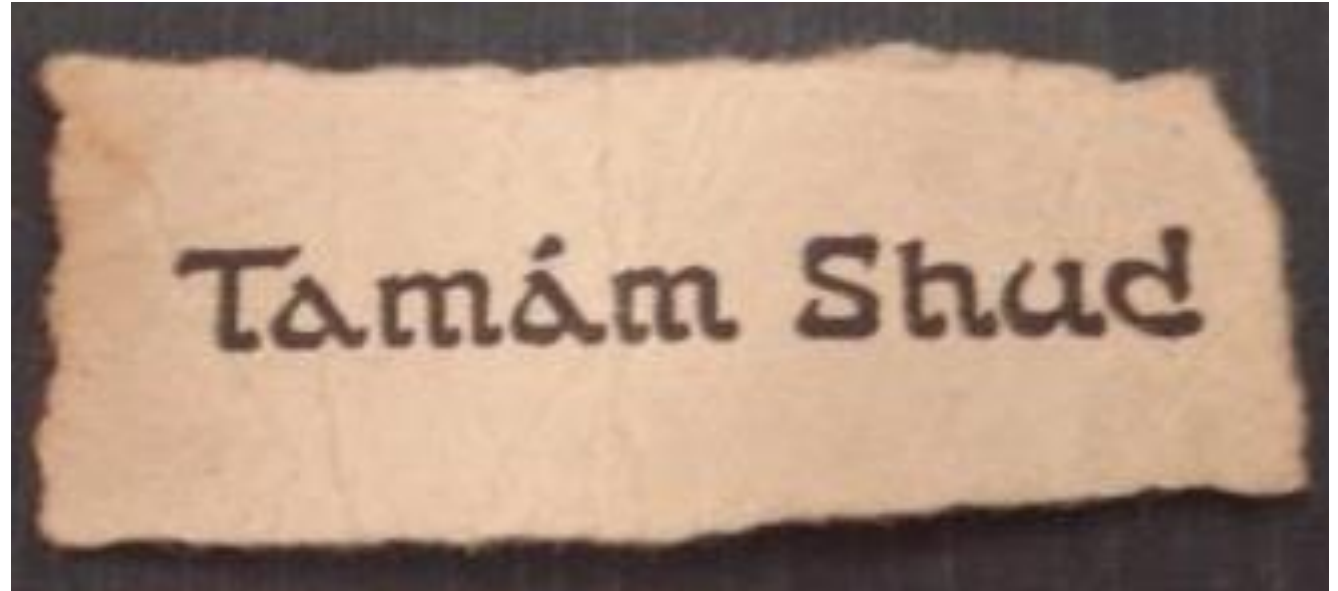
- Time of discovery:
  - Found on the 1st of December 1948
- Location:
  - Somerton Beach, South Australia
- Who:
  - Unidentified Man
- Cause of death:
  - Unknown

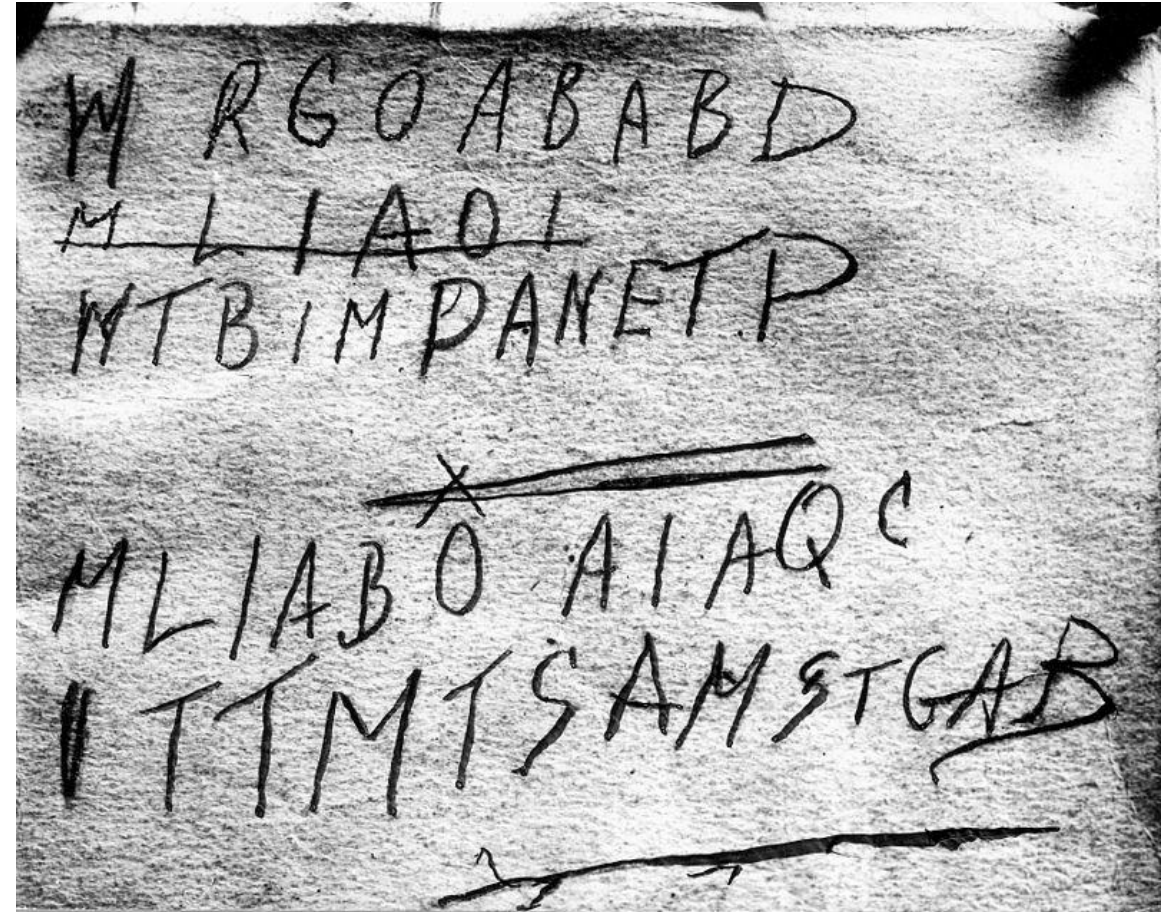<Source: www.sapolicehistory.org>

Introduction: Evidence

- A scrap of paper
  - Found in Man's trousers pocket
  - "Tamám Shud" translates to 'finished' in English
  - Proved to be part of the 'Rubaiyat of Omar Khayyam' book.

<Source: www.smithsonianmag.com/history/the-body-on-somerton-beach>

Introduction: Evidence

▶ Mysterious Code

▶ WRGOABABD
MLIAOI
WTBIMPANETP
MLIABOAIAQC
ITTMTSAMSTGAB

<Source: www.sapolicehistory.org>

Introduction: Project Aims

▶ To eliminate various possibilities, in relation to the mysterious code

▶ To use mass spectrometry on different hair samples to compare strontium-88 levels

▶ To analyse how robust DNA can be, by degrading a DNA sequence in software using MATLAB

Previous Professional Attempts to Decrypt Code

- SA police:
  - Started investigation
- Australian Navy:
  - Stated "Neither a code nor a cipher"
- Australian Defence Force:
  - Attempted to crack the code
  - Cryptographers defined code as 'unable to crack'

Previous Study: Honours Project 2009-2015

▶ Letter frequency analysis

▶ Initial letter and sentence letter probabilities

▶ Different cypher techniques

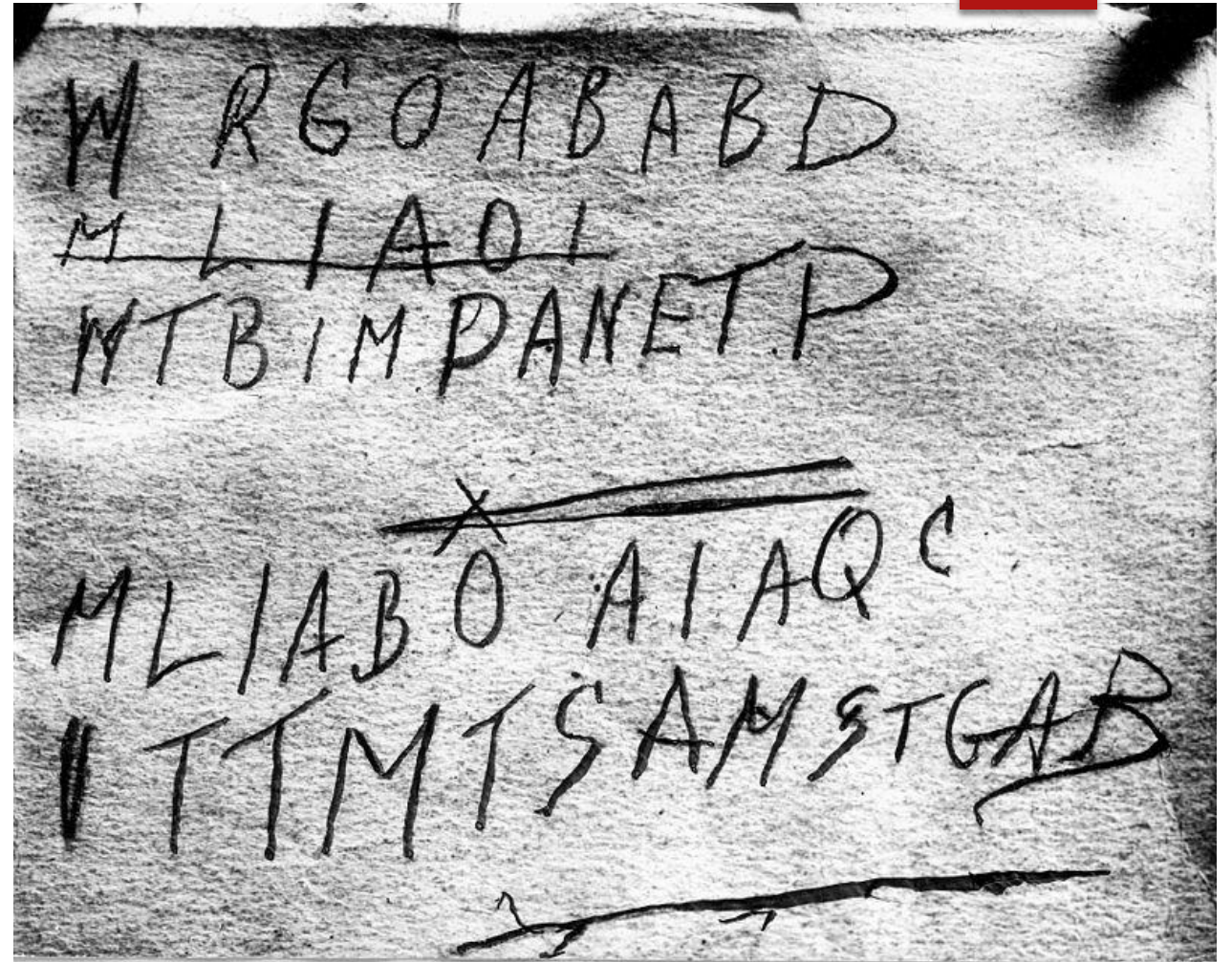▶ 3D generated reconstruction

▶ Mass spectrometer data hair analysis

<Source: www.sapolicehistory.org>

Specific Tasks:
Task 1: Letters

▶ Mysterious code

▶ Using previous studies, it is being assumed that each letter, is the first letter of a word

▶ Letters could be a collective object

▶ E.g. (Horse names, Australian City's, SA street names, Rubaiyat of Omar Khayyam' book, beach names etc.)

<Source: www.sapolicehistory.org>

Horse names

- Circumstantial evidence shows that there was connection between Somerton Man and horses
  - Somerton Beach is near Morphettville Racecourse
  - Two horse jockeys found the dead body
- Trove website for the relevant information

<Source: https://trove.nla.gov.au/newspaper/>

Statistical testing

- ► P-value
  - ► Two-tailed t-test
- ► Hypothesis test
  - ► H_0: The group of letters are horse names
  - ► H_1: The group of letters are not horse names
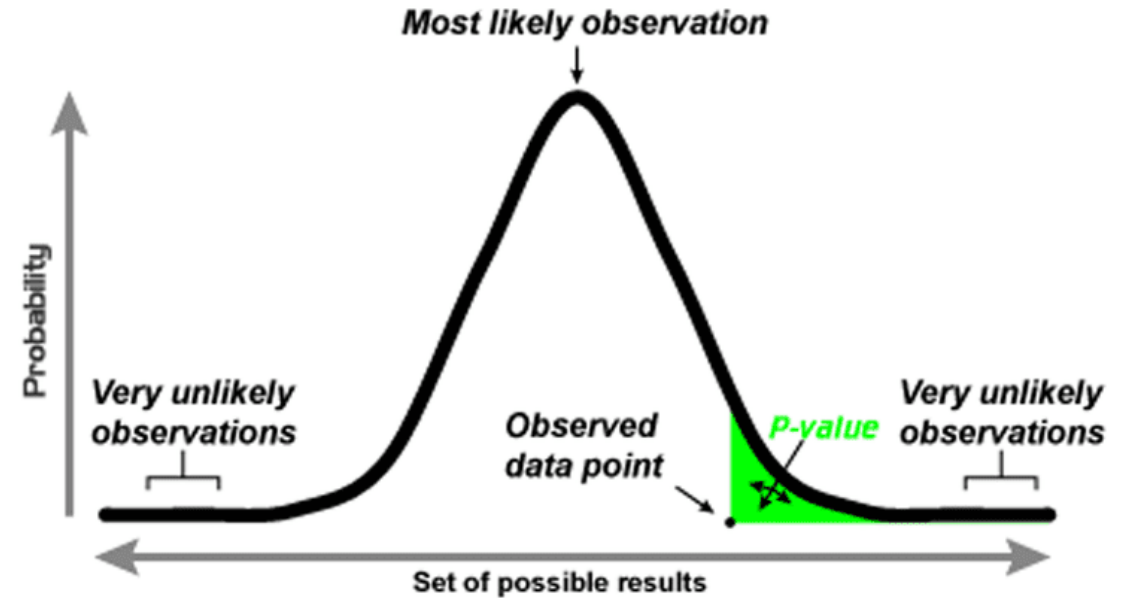
A p-value (shaded green area) is the probability of an observed (or more extreme) result arising by chance

<Source: www.thoughtco.com/definition-of-p-value-1148041>

Analysis

- ▶ The correlation between mysterious codes and horse names is not obvious

- ▶ Further proved by two-tailed test, where the p-value is lower than 0.05

- ▶ Null hypothesis is not accepted

Mysterious Code and Horse Names
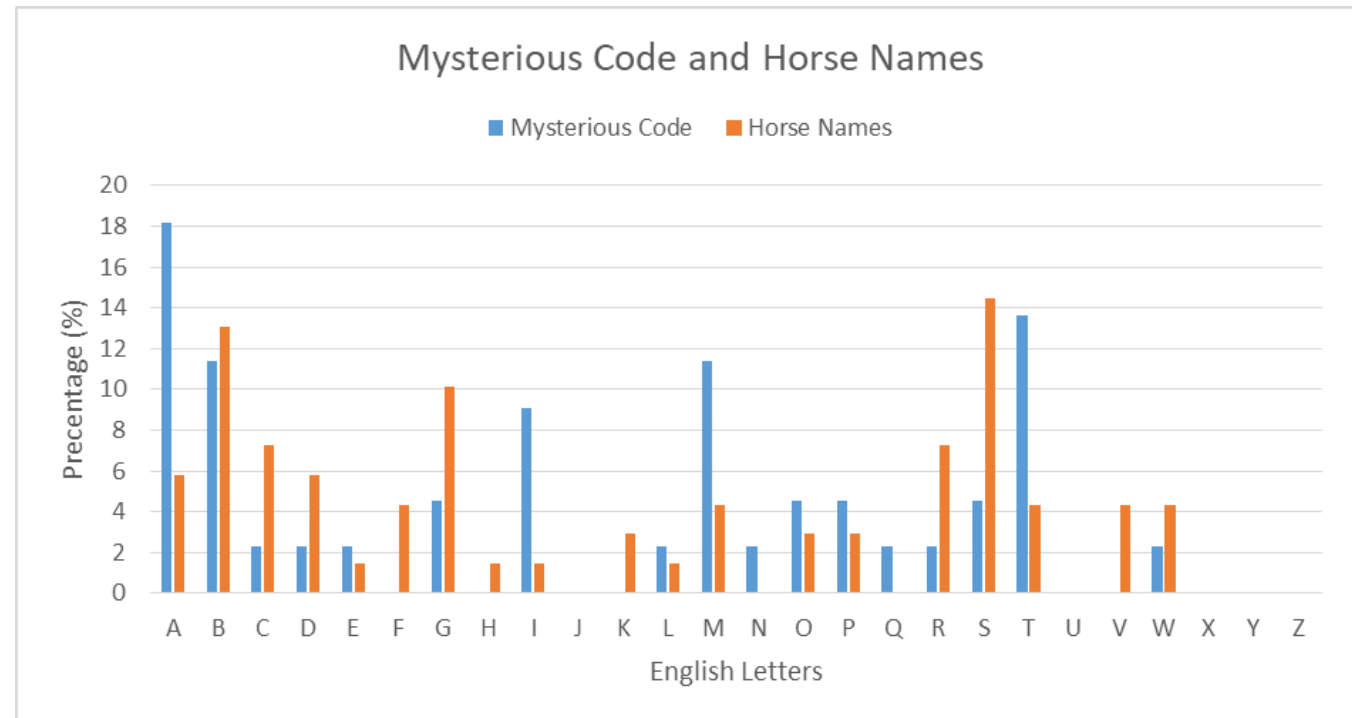
Analysis

- P-value result in 'beach name' is higher than 0.05, meaning the mysterious code can refer to beach names

- Database needs to be built up to further confirm

Mysterious Code and Australian Beach Names
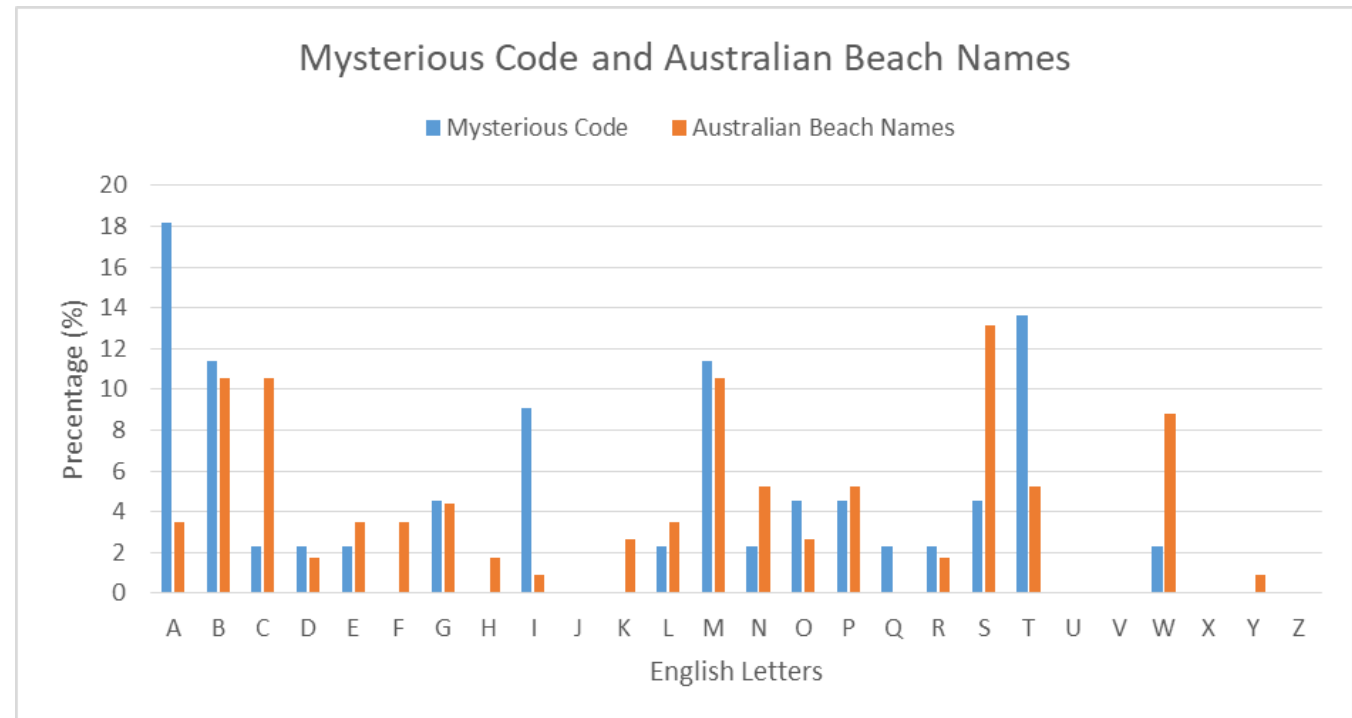
Specific Tasks:
Task 2: Mass Spectrometer

▶ Laser Ablation Mass Spectrometry (LAMS)

▶ Measures the different isotopic signatures of the sample

▶ Laser ablates the sample and records different elements

▶ The sample being used, is the shaft of the hair

<Source: https://www.adelaide.edu.au/microscopy/instrumentation/icpms>
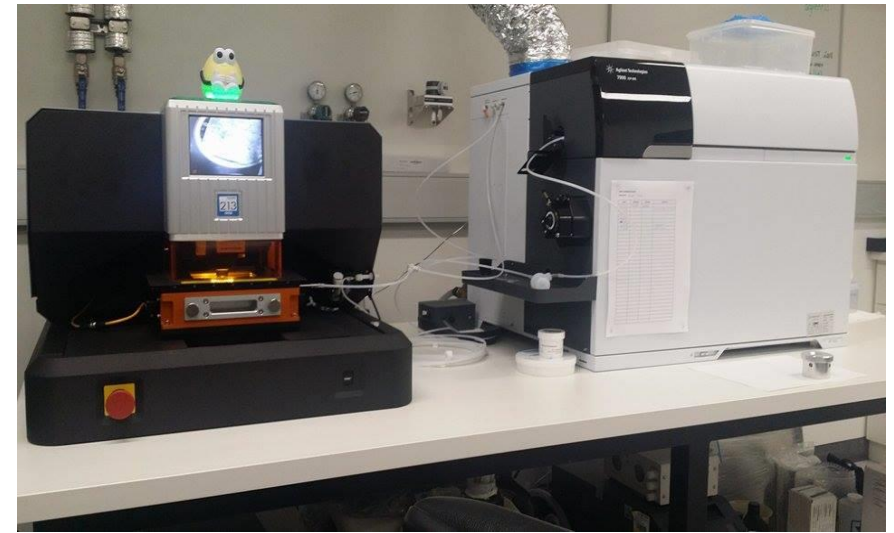
Strontium

- High strontium in Somerton Man's hair

- High strontium in Adelaide
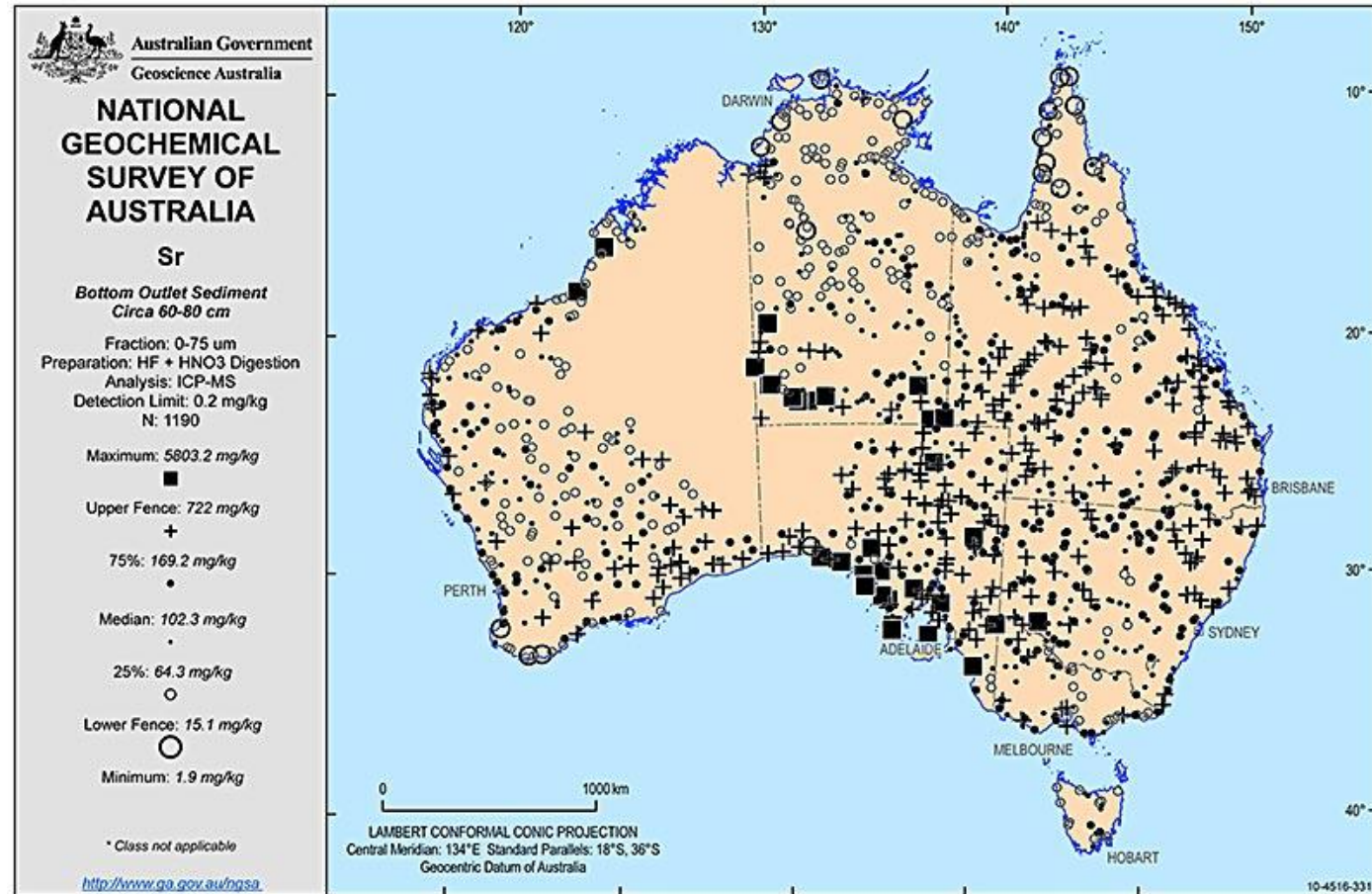
<Source: http://www.ga.gov.au>

Hair samples

- Use five hair samples from different individuals
  - Control samples
- Hair approximately grows 1 cm per month
- Stick hairs on quartz slide using silicon-based cohesive tape
- Discrete ablation on hair samples

Hair Identification	Sex	Sampling Date	Description
A	Male	17/Sep/2018	Been to Japan for nearly one week before sampling
B	Male	20/Sep/2018	Stayed in Adelaide
C	Male	20/Sep/2018	Stayed in Adelaide
D	Female	20/Sep/2018	Stayed in Adelaide
E	Female	16/Sep/2018	Been to Bali for nearly two weeks before sampling
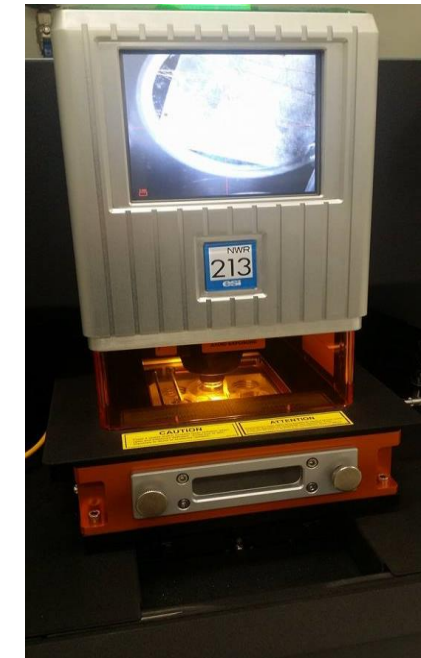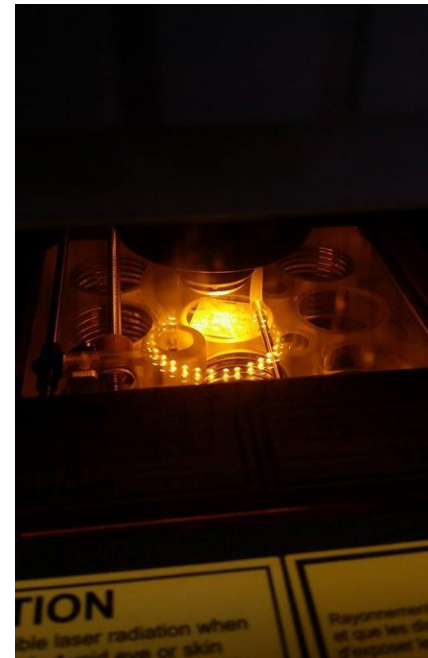
Data processing

▶ Total waveform received from Microscopy Centre

▶ Strontium-88 level is not critically high at this spot
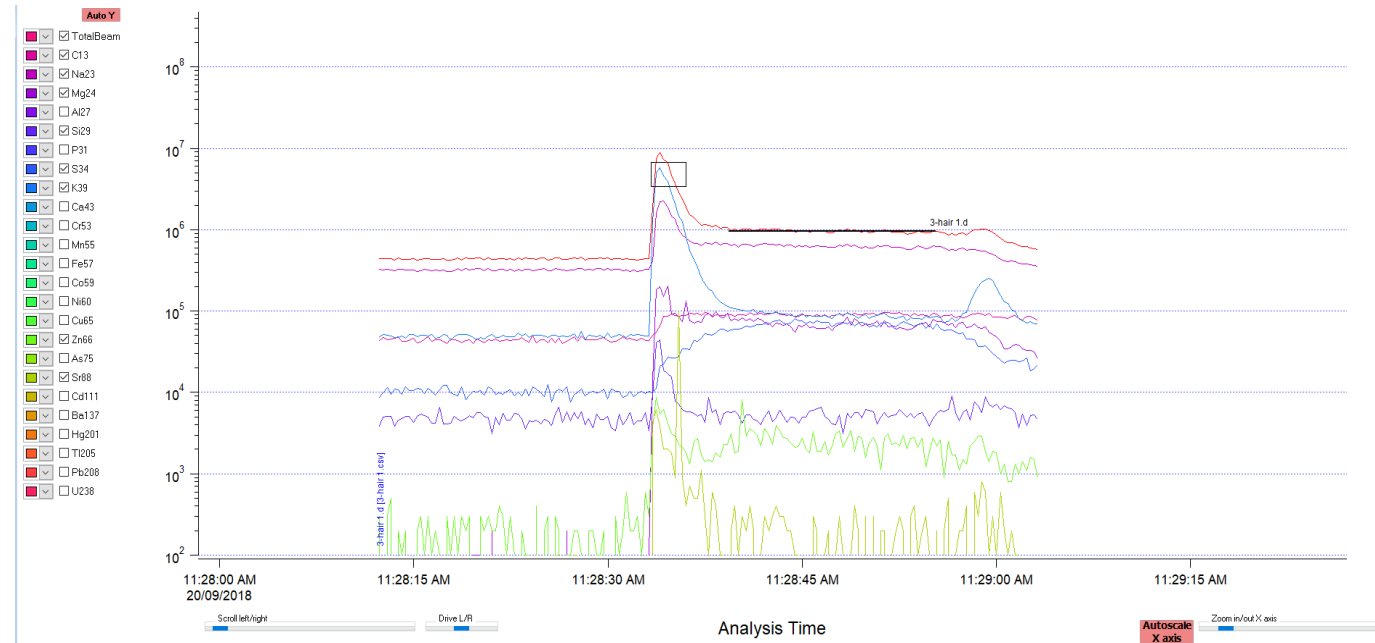
Result

- ▶ Purple box is the period when Sample A was in Japan

- ▶ Black box is the period when sample E was in Bali

- ▶ Strontium-88 in Adelaide is not significant compared to Bali

- ▶ It opposes initial assumption

Strontium-88
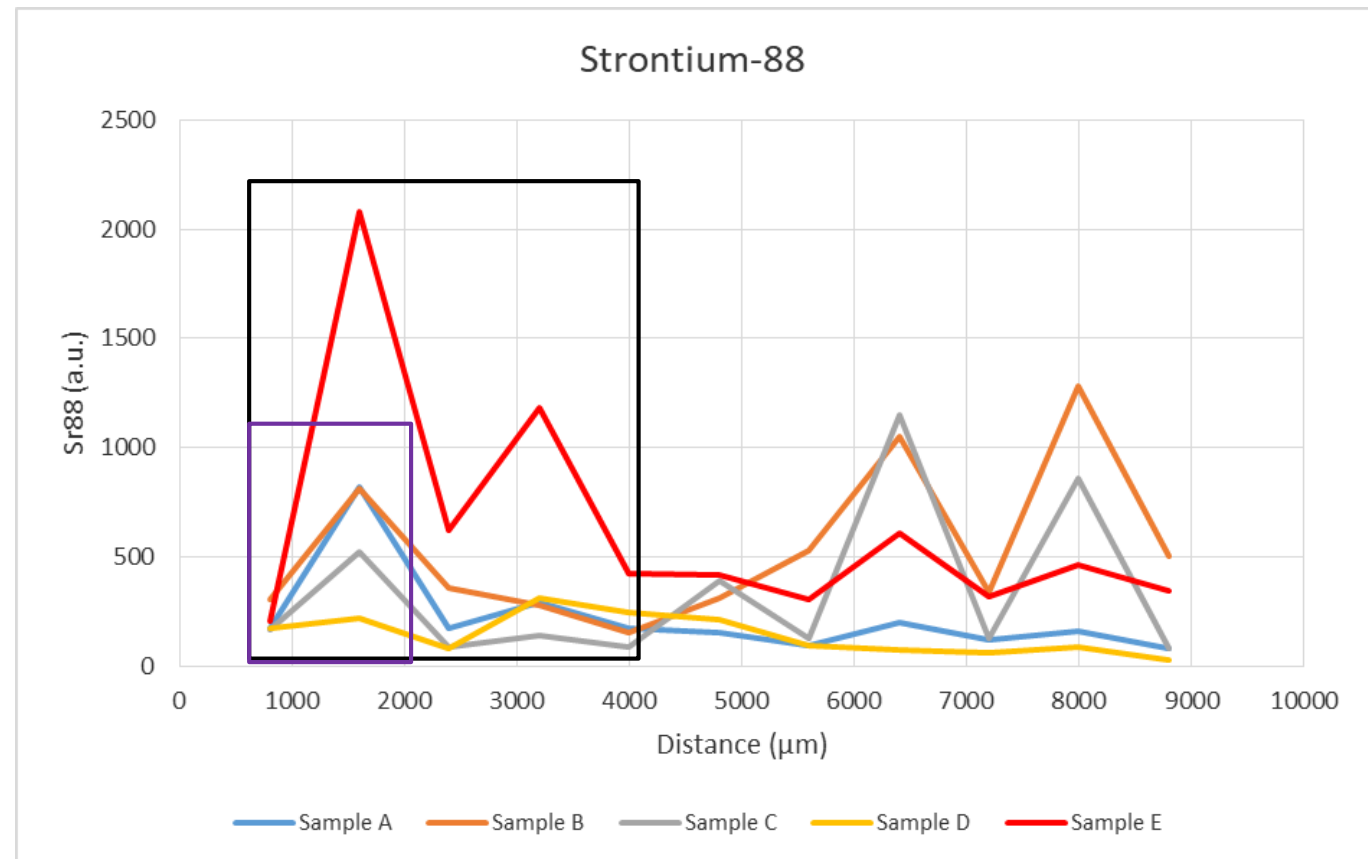
Conclusion

▶ Results not solid enough, need further investigations

Specific Tasks:
Task 3: DNA

- ▶ DNA (Deoxyribonucleic Acid)
  - ▶ Approximately 2 billion bits of information
- ▶ Single Nucleotide Polymorphism (SNP)
  - ▶ Consequences
  - ▶ Genetic similarity

<Source: www.scienceandbelief.org/tag/double-helix>

Specific Tasks:
Task 3: DNA (Method)

▶ Send DNA samples to get tested

▶ Degrade the DNA samples

   ▶ Remove SNPs randomly

▶ Check for false positives and false negatives

▶ How many SNPs can be removed, before the DNA is unidentifiable

<Source: www.socmucimm.org/single-nucleotide-polymorphism>
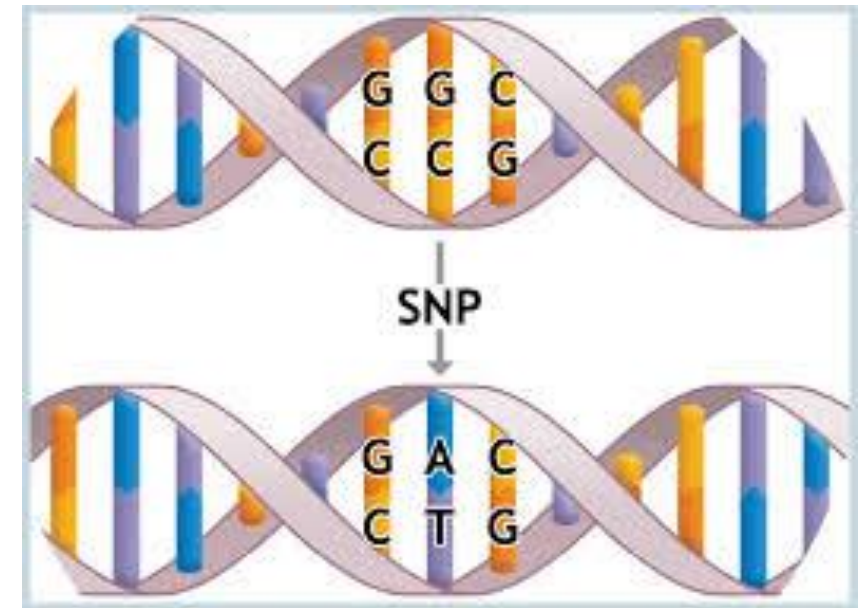
DNA sequences and DNA list

▶ Pure test file with more than 600,000 lines of SNP

▶ Matching DNA list

   ▶ 'Largest Seg' -> longest matching sequence

   ▶ 'Total cM' -> length of the same DNA sequence

   ▶ 'Gen' -> generation families

```
# rsid      chromosome      position        genotype
rs548049170     1           69869       TT
rs13328684      1           74792       --
rs9283150       1           565508      AA
i713426 1       726912      --
rs116587930     1           727841      GG
rs3131972       1           752721      GG
rs12184325      1           754105      CC
rs12567639      1           756268      AA
rs114525117     1           759036      GG
rs12124819      1           776546      AA
rs12127425      1           794332      GG
rs79373928      1           801536      TT
rs72888853      1           815421      --
rs7538305       1           824398      AA
rs28444699      1           830181      AA
i713449 1       830731      --
rs116452738     1           834830      GG
rs72631887      1           835092      TT
rs28678693      1           838665      TT
rs4970382       1           840753      CT
rs4475691       1           846808      CC
rs72631889      1           851390      GG
rs7537756       1           854250      AA
rs13302982      1           861808      AA
rs376747791     1           863130      AA
rs2880024       1           866893      TT
rs13302914      1           868404      TT
rs76723341      1           872952      TT
```
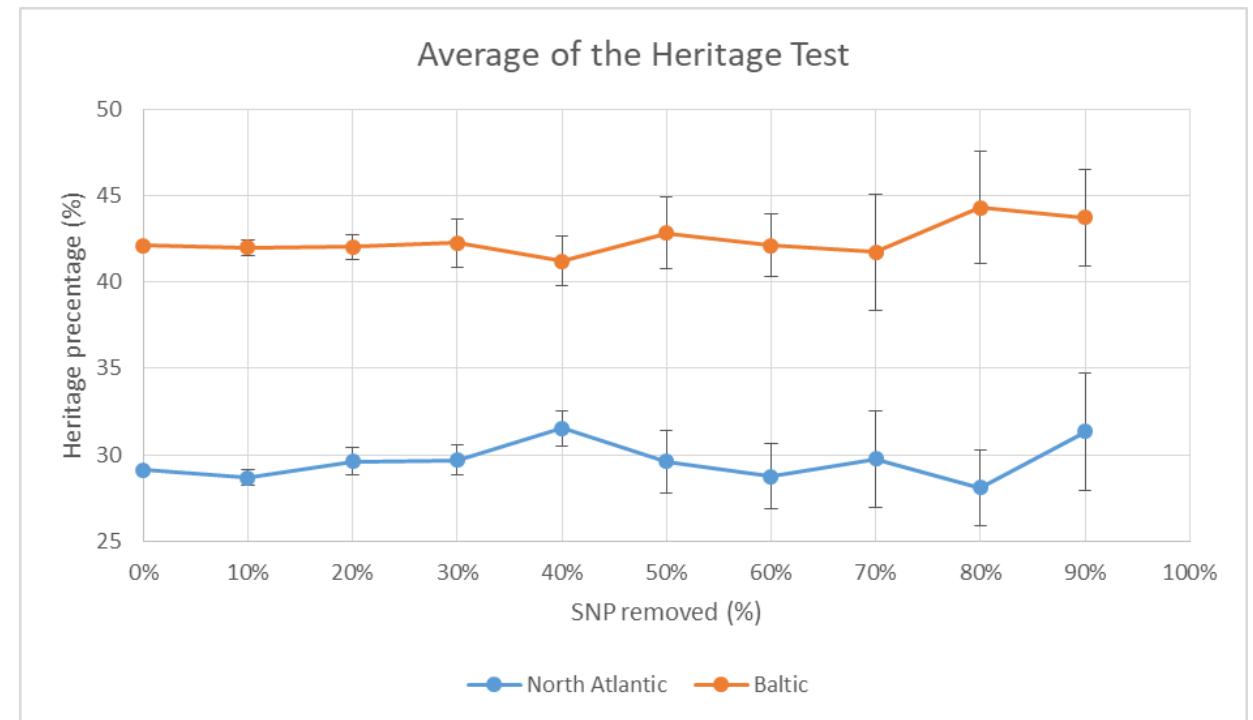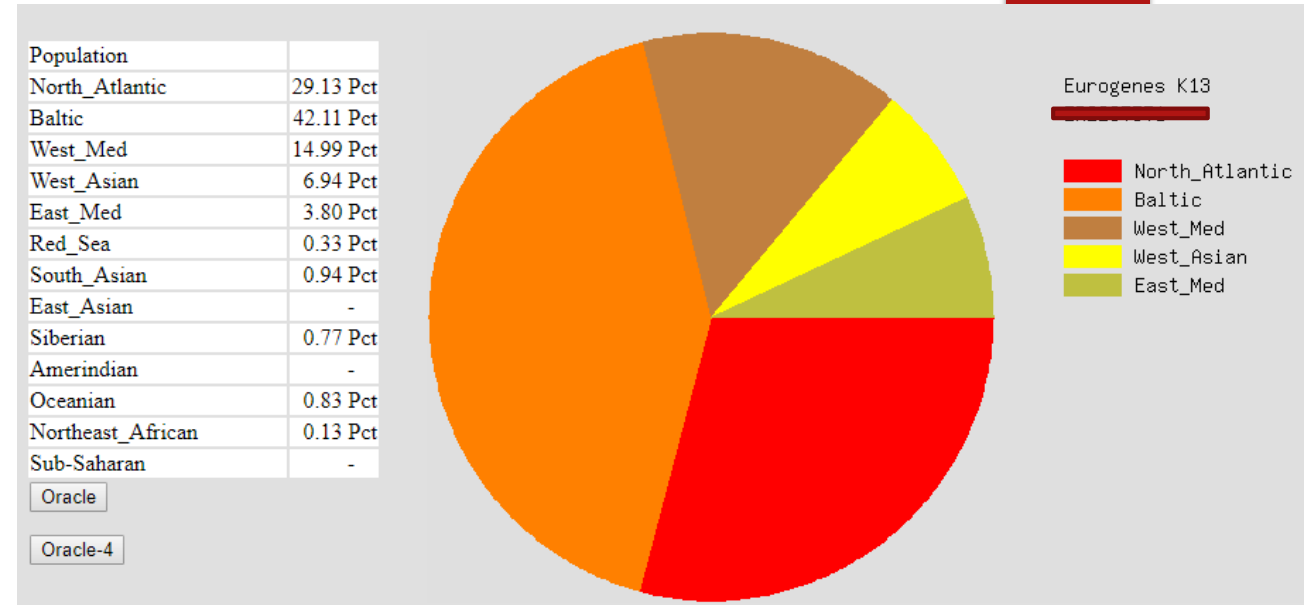
Kit	1:1	Name	Email	Largest Seg	Total cM	Gen	Overlap	Date Compared	Testing Company
	A			16.1977	89.1643	3.7	79884	2018-05-11	-
	A			14.258	68.2605	3.9	79516	2018-05-15	-
	A			14.258	68.2605	3.9	79516	2018-05-15	-
	A			15.951	65.796	3.9	74209	2018-05-25	23mofang
	A			16.7438	58.9933	4.0	73951	2018-05-11	23mofang
	A			13.9916	59.0673	4.0	74717	2018-05-11	-

Heritage test

- ▶ Pie chart result

- ▶ SNPs removed at different percentage

- ▶ Test completed 5 times

- ▶ Average value plot with error bars
  - ▶ After 50% removal, error bar is obvious with more than 1%

Population	
North_Atlantic	29.13 Pct
Baltic	42.11 Pct
West_Med	14.99 Pct
West_Asian	6.94 Pct
East_Med	3.80 Pct
Red_Sea	0.33 Pct
South_Asian	0.94 Pct
East_Asian	-
Siberian	0.77 Pct
Amerindian	-
Oceanian	0.83 Pct
Northeast_African	0.13 Pct
Sub-Saharan	-

Oracle

Oracle-4

Eurogenes K13

- North_Atlantic
- Baltic
- West_Med
- West_Asian
- East_Med

Average of the Heritage Test

False positives and false negatives

- FP and FN are medical testing terms, referring the DNA kits disappear in original and appear in reduced

- ORIGINAL: A B C D E

- REMOVED: C D F G H

    - FP: F G H

    - FN: A B E

- After 50% removal of the DNA sequence, nothing matched between the original and degraded case
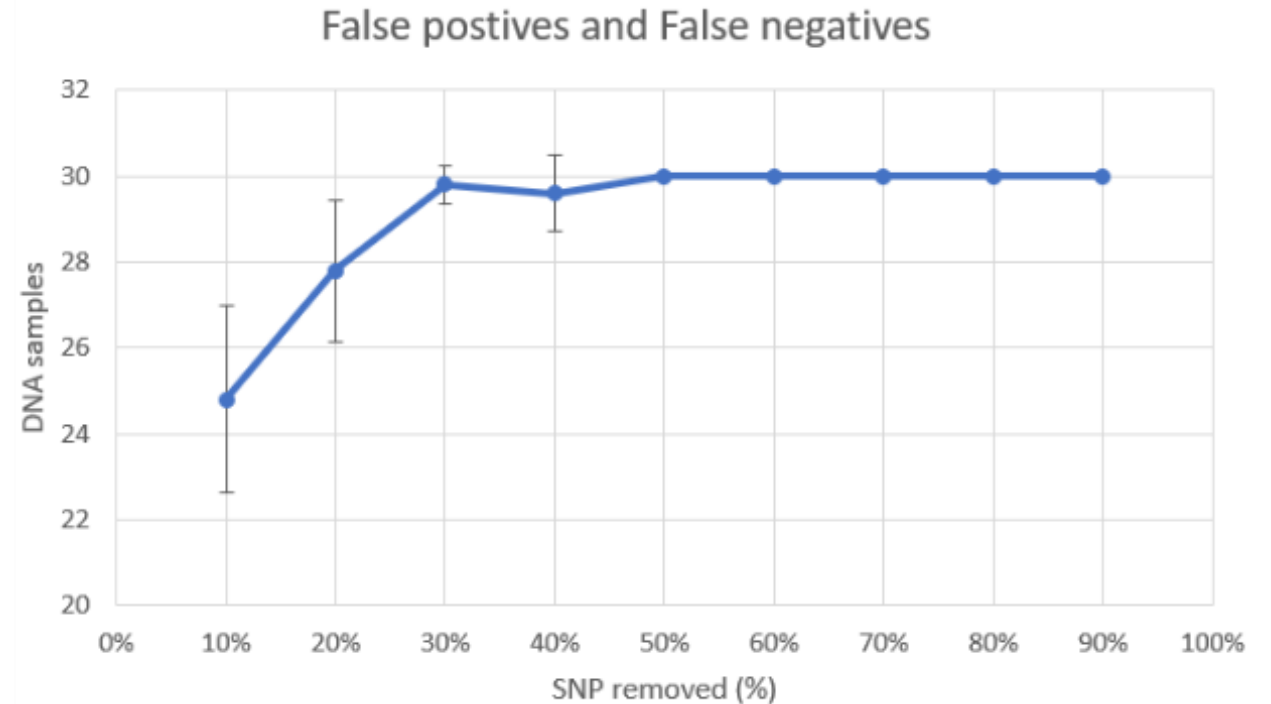
Conclusion

▶ DNA sequence will be highly unreliable if half of them are degraded

▶ Somerton Man's descendant can be found, as long as the DNA sequences are proved to be at least half matched with Somerton Man's, this could reveal where Somerton Man came from

Future work

- Cipher cracking
  - Consider other groups of names
  - Research on graphology, which is a study that can identify the writer by analysing the handwriting pattern
- Mass spectrometer experiment
  - Measure the samples continuously
  - Compare strontium level in different places of Adelaide or different cities in Australia
- DNA analysis
  - Further analyse on more accurate reduction level of DNA sequence
  - Attempt to get Somerton Man's DNA to undergo further testing

Summary

- Mysterious code is unlikely to be horse names, but Australian beach names have higher likelihood

- Further strontium testing of hair is needed to establish if it can reveal arrival date into South Australia

- DNA appears to be highly unreliable when 50% of the SNPs are removed

Acknowledgment

- Professor Derek Abbott

- University of Adelaide Microscopy Centre

References

- Author unknown (1949). "Tamam Shud", The Advertiser, 10 June 1949, p. 2

- Author unknown (1948). "Dead Man Found Lying on Somerton Beach", The News, 1 December 1948, p. 1

- Maguire, S. (2005). "Death riddle of a man with no name", The Advertiser, 9 March 2005, p. 28

- Author unknown (1948). "Dead Man Found Lying on Somerton Beach", The News, vol. 51, no. 7902, pp. 1.

- L. Griffith and P. Varsos. (2013). Semester B Final Report 2013 – Cipher Cracking .
  Available: http://www.adelaidenow.com.au/news/south-australia/somerton-man-mystery-new-details-revealed-of-jo-thomson-nurse-in-the-case/news-story/4c6bccbd2318584ad0cc6daaf3d8abd4

- Inside Story, presented by Stuart Littlemore, ABC TV, screened at 8 pm, Thursday, August 24th, 1978

- "SNP" Available: https://www.23andme.com/en-int/gen101/snps/

- https://www.adelaide.edu.au/microscopy/instrumentation/icpms

- https://trove.nla.gov.au/newspaper

- http://www.ga.gov.au

Questions

End of Presentation