



THE UNIVERSITY  
*of* ADELAIDE

Azizul Hakim Luqman Ul Hakim Ng  
Zihe Wang

**Supervised by:**

Professor Derek Abbott

Dr Andrew Allison

# **Project 141: Who Killed the Somerton Man**

[adelaide.edu.au](http://adelaide.edu.au)

# Outline

- **Introduction**
  - **Background**
  - **Project Aims and Motivation**
- **Specific Tasks**
  - **Task 1**
  - **Task 2**
  - **Task 3**
  - **Task 4**
- **Conclusion**
- **Project management**
- **Reference**

# Introduction

# Background

- Who?

**The Somerton Man**

- What?

**Found dead**

- When?

**December 1st, 1948**

- Where?

**Somerton Beach, SA**

- How?

**Unknown - mystery**

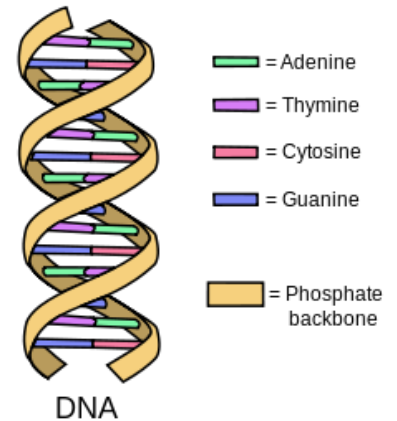


# Background (cont.)

## Somerton Man's DNA

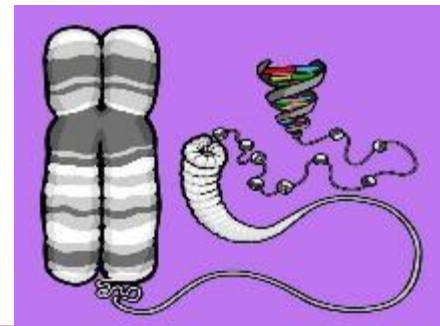
- Hair sample in the Police Museum
- Extract DNA from the hair sample
- Degraded due to being kept for long

# Background (cont.)



## Deoxyribonucleic Acid (DNA)

- Carrying genetic instructions
- A code formed from a chain of 4 chemical bases (nucleotides): adenine (A), guanine (G), cytosine (C) and thymine (T)
- DNA forms chromosomes



# Background (cont.)

## Single Nucleotide Polymorphism (SNP)

- A variation of base pairs at a specific position in a DNA sequence
- Defines characteristics of individuals (eg. eye colour)

# Background (cont.)

## Single Nucleotide Polymorphism (SNP)

Individual 1

Chr 2 *copy1* . . . CGATATTCC**T**ATCGAATGTC . . .  
          . . . GCTATAAGG**A**TAGCTTACAG . . .

Chr 2 *copy2* . . . CGATATTCC**C**ATCGAATGTC . . .  
          . . . GCTATAAGG**G**TAGCTTACAG . . .

Individual 2

Chr 2 *copy1* . . . CGATATTCC**C**ATCGAATGTC . . .  
          . . . GCTATAAGG**G**TAGCTTACAG . . .

Chr 2 *copy2* . . . CGATATTCC**C**ATCGAATGTC . . .  
          . . . GCTATAAGG**G**TAGCTTACAG . . .



# Background (cont.)

## Somerton Man's DNA file

#rsid	chromosome	position	genotype
rs548049170	1	69869	--
rs13328684	1	74792	--
rs9283150	1	565508	--
i713426	1	726912	--
rs116587930	1	727841	--
rs3131972	1	752721	--
rs12184325	1	754105	--
rs12567639	1	756268	--
rs114525117	1	759036	--
rs12124819	1	776546	--
rs12127425	1	794332	--
rs79373928	1	801536	--
rs72888853	1	815421	--
rs7538305	1	824398	--
rs28444699	1	830181	--
i713449	1	830731	--
rs116452738	1	834830	--
rs72631887	1	835092	--
rs28678693	1	838665	--
rs4970382	1	840753	--
rs4475691	1	846808	CT

# Project Aims and Motivation

- To find possibilities of who the Somerton Man was - taking a step forward to solving the unsolved mystery
- To evaluate the robustness of the Somerton Man's DNA
- To identify any possible diseases and physical characteristics of the Somerton Man
- To find the relatives, and finally find out who he actually was

# **Specific Tasks**

# Specific Task

## **Task 1: Testing with Somerton Man's DNA reference file**

### **Aims**

- Counting amounts of SNP
- Try to conduct DNA analysis on Somerton's DNA

# Specific Task

## Task 1: Testing with Somerton Man's DNA reference file

### Method

- Writing codes to count the SNPs of the DNA files with C++ language
- Upload the file to [genesis.gedmatch.com](https://www.genesis.gedmatch.com) (GEDmatch) which provide DNA analysis services



- GEDmatch is a website that has an open data personal genomics database and provide tools for DNA and genealogy research
- Tools used:
  - One-To-Many DNA Comparison
  - Eurogenes Ad-Mix Utilities for ethnicity examination
- 2000 SNPs minimum requirements for uploading DNA file

# Specific Task

## Task 1

### Results

- 613905 SNPs in the DNA data files
- 2.08% of SNPs are not empty
- Rejected by GEDmatch for DNA match test

Counting results of SNPs			
Chromosome	Total amount	Exist amount	Percentage
1	49510	1014	2.05%
2	51771	978	1.89%
3	43023	658	1.53%
4	39473	621	1.57%
5	37028	661	1.79%
6	44021	880	2.00%
7	34356	655	1.91%
8	31681	601	1.90%
9	26445	519	1.96%
10	30522	705	2.31%
11	30943	705	2.28%
12	29432	596	2.03%
13	22080	393	1.78%
14	19961	441	2.21%
15	19006	440	2.32%
16	20396	558	2.74%
17	19401	519	2.68%
18	17674	372	2.10%
19	14879	514	3.45%
20	14781	375	2.54%
21	8607	245	2.85%
22	8915	303	3.40%
Total	613905	12753	2.08%

# Specific Task

## Task 2: Artificially recover the DNA file

### Aims

- Create synthetic DNA files based on Somerton Man's file using different strategies/algorithm
- To see if there is any people in DNA public database links to the artificial DNA



# Specific Task

## Task 2: Artificially recover the DNA file

### Method

- Writing code to artificially complete the DNA file (replace the empty SNPs) for different levels (eg. 2000, 3000, 5000 SNPs for each chromosome)
- Strategies include replace empty SNPs with random base pairs or homozygous pairs (eg. AA, GG, TT, CC).

# Specific Task

## Task 2: Artificially recover the DNA file (cont.)

#rsid	chromosome	position	genotype	#rsid	chromosome	position	genotype
rs548049170	1	69869	GA	rs548049170	1	69869	AA
rs13328684	1	74792	GC	rs13328684	1	74792	AA
rs9283150	1	565508	TA	rs9283150	1	565508	AA
i713426	1	726912	TG	i713426	1	726912	AA
rs116587930	1	727841	AG	rs116587930	1	727841	AA
rs3131972	1	752721	GT	rs3131972	1	752721	AA
rs12184325	1	754105	CA	rs12184325	1	754105	AA
rs12567639	1	756268	TA	rs12567639	1	756268	AA
rs114525117	1	759036	CT	rs114525117	1	759036	AA
rs12124819	1	776546	GC	rs12124819	1	776546	AA
rs12127425	1	794332	GT	rs12127425	1	794332	AA
rs79373928	1	801536	CT	rs79373928	1	801536	AA
rs72888853	1	815421	AC	rs72888853	1	815421	AA
rs7538305	1	824398	AA	rs7538305	1	824398	AA
rs28444699	1	830181	CA	rs28444699	1	830181	AA
i713449	1	830731	CG	i713449	1	830731	AA
rs116452738	1	834830	TC	rs116452738	1	834830	AA
rs72631887	1	835092	GA	rs72631887	1	835092	AA
rs28678693	1	838665	AA	rs28678693	1	838665	AA
rs4970382	1	840753	CG	rs4970382	1	840753	AA

# Specific Task

## Task 2: Artificially recover the DNA file

### Results

- No DNA kits matched with the kits that were modified from Somerton Man's reference file.
- Even replace all empty SNPs result a 0 match.

Kit:  [23andMe]

Kit	1:1	Name	Email	Largest Seg	Total cM	Gen	Overlap	Date Compared	Testing Company
-----	-----	------	-------	-------------	----------	-----	---------	---------------	-----------------

0 is number of matches reported

# Specific Task

## Task 2: Artificially recover the DNA file

### Conclusion

- A DNA file with only 2% SNPs available cannot be used in DNA match test
- The recovery strategies introduced in 'Method' are too simple and can not help to find the related DNA

# Specific Task

## **Task 3: Investigation on ethnicity**

### **Aims**

- Conduct the ethnicity analysis on Somerton Man's DNA
- To prove that the ethnicity report of low quality DNA data is reliable

# Specific Task

## **Task 3: Investigation on ethnicity**

### **Method**

- Using Eurogenes Ad-Mix Utilities to generate ethnicity reports
- 'Calculator' model: 'Eurogenes K13'

# Specific Task

## **Task 3: Investigation on ethnicity** **Method(cont.)**

- Obtain 2 DNA samples in same format of Somerton Man's file
- Degrade these samples: remove certain percentage of SNPs
- Degrade these samples by removing 10%, 20% until 90% of SNPs.
- Also degrade the samples to the same level of Somerton Man's file (2.08% SNPs remaining)

# Specific Task

## Task 3: Investigation on ethnicity Method(cont.)

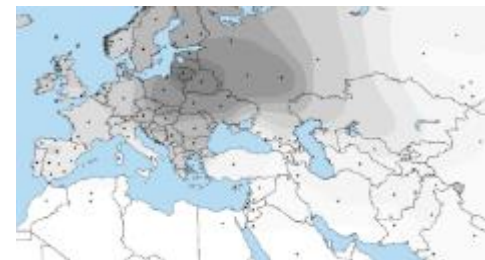
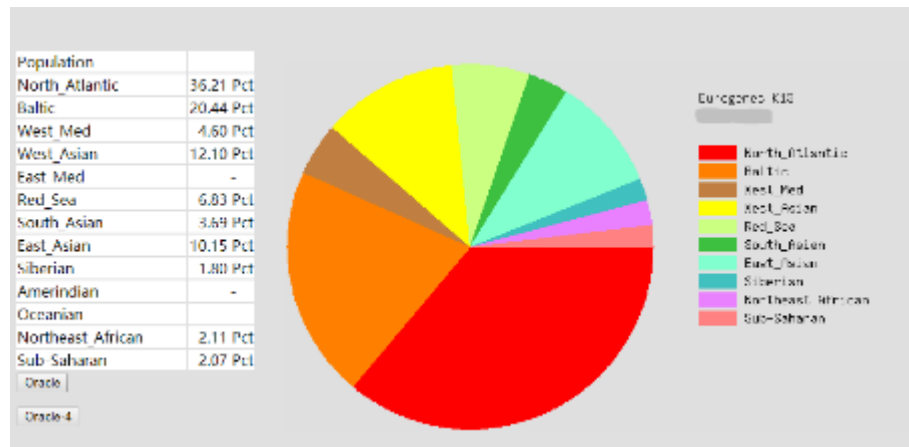
- 4 different degradation algorithms implemented:
  - For every 10 SNPs, remove first n% SNPs where n% is the percentage of SNPs we like to remove
  - For every 10 SNPs, remove last n% SNPs where n% is the percentage of SNPs we like to remove
  - Remove first n% of SNPs in each chromosome
  - Remove last n% of SNPs in each chromosome
- Conduct ethnicity analysis on each degraded file, and observe how the ethnicity proportion changes



# Specific Task

## Task 3: Investigation on ethnicity Results

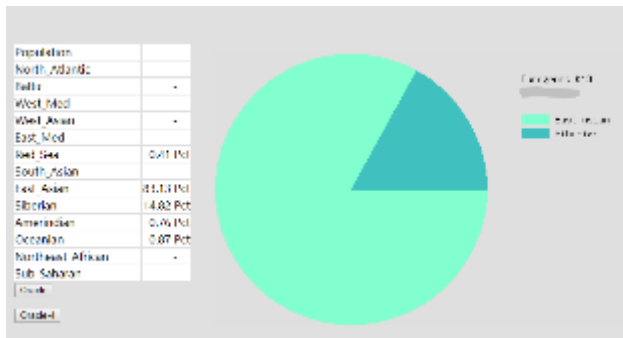
- Ethnicity report shows Somerton Man comes from North Atlantic(36.21%) and Baltic(20.44%) regions



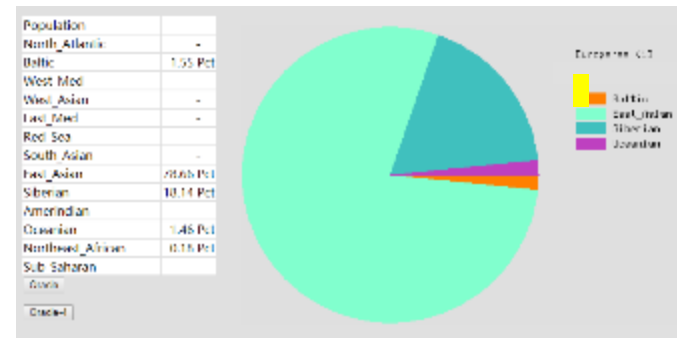
# Specific Task

## Task 3: Investigation on ethnicity Results (cont.)

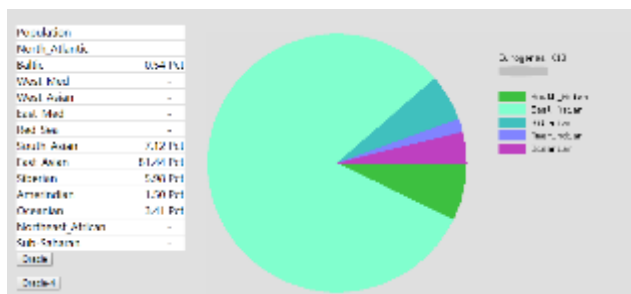
- How ethnicity changes after degradation



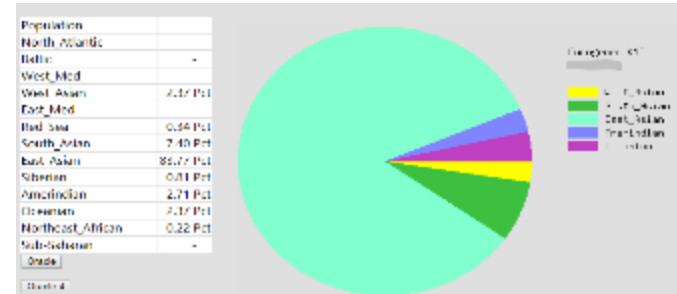
Complete DNA sample 1



Degraded DNA sample 1 (2.08% SNPs remaining)



Complete DNA sample 2

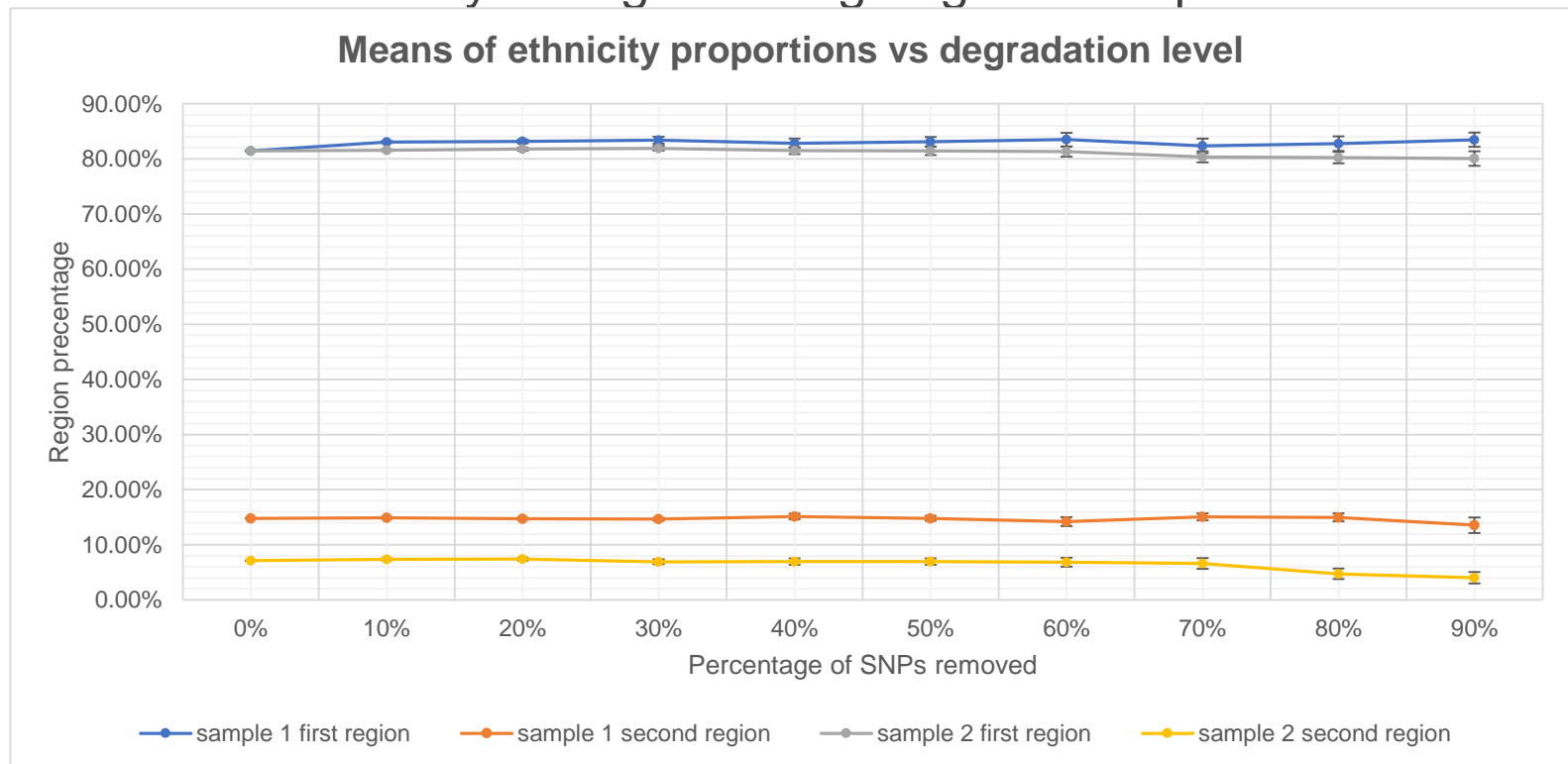


Degraded DNA sample 2 (2.08% SNPs remaining)

# Specific Task

## Task 3: Investigation on ethnicity Results (cont.)

- How ethnicity changes during degradation process



# Specific Task

## **Task 3: Investigation on ethnicity**

### **Conclusion**

- As more amount of SNPs are removed from a complete human DNA reference file, the result of ethnicity report would be less accurate.
- But for the largest and second largest ethnicity regions in the report are still reliable.
- Therefore the ethnicity of Somerton Man is most probably North Atlantic.

# Specific Task

## Task 4: Genetic diseases search

### Aims

- Search clinical effects associated with each available SNP in Somerton Man's DNA data
- Identify any possible genetic disease or physical characteristics that Somerton Man could have

# Specific Task

## Task 4: Genetic diseases search

### Method

- Use dbSNP to search genetic diseases
- dbSNP requires user to input rs number in order to observe the clinical significance of each SNP
- Python was used to extract information on the possible genetic diseases on the Somerton Man's DNA from dbSNP

# SNP database (dbSNP)

Reference SNP (rs) Report

[Download](#) [f](#) [t](#) [v](#) [?](#)

[← Switch to classic site](#)

**rs12913832**

**Current Build** 152  
**Released** October 2, 2018

**Organism** *Homo sapiens* **Clinical Significance** Reported in [ClinVar](#)  
**Position** chr15:28120472 (GRCh38.p12) **Gene : Consequence** HERC2 : Intron Variant  
**Alleles** A>G **Publications** 92 citations  
**Variation Type** SNV Single Nucleotide Variation **Genomic View** [See rs on genome](#)  
**Frequency** G=0.45329 (56919/125568, TOPMED)  
A=0.4419 (13667/30926, GnomAD)  
G=0.177 (888/5008, 1000G) (+ 3 more)

FEEDBACK

Variant Details

**Clinical Significance**

Frequency

Aliases

Submissions

History

Publications

**Allele: G (allele ID: 19784)**

ClinVar Accession	Disease Names	Clinical Significance
<a href="#">RCV000005011.4</a>	Skin/hair/eye pigmentation, variation in, 1	Association

# Specific Task

## Task 4: Genetic diseases search

### Result

- 574 potential genetic diseases were found associated to Somerton Man's DNA.
- There was no result strongly support Somerton Man's known physical characteristics such as hair colour, teeth structure or eye colour.
- There were no genetic disease related to the Somerton Man found from dbSNP such as enlarged spleen or having huge hands.



# Specific Task

## Task 4: Genetic diseases search

574 potential genetic  
diseases

#rsid	#alleles	#ClinVar	Accession	#disease	name
rs5257	[[A>A], [A>G]]		RCV000517867.2	not specified	
rs3754334	[[G>G], [G>A]]		RCV000244865.1	not specified	
rs3754334	[[G>G], [G>A]]		RCV000304354.1	Age-related cortical cataract	
rs1049675	[[G>G], [G>A]]		RCV000290348.1	Schwartz Jampel syndrome type 1	
rs1049675	[[G>G], [G>A]]		RCV000377761.1	Dyssegmental Dysplasia	
rs35669711	[[C>C], [C>T]]		RCV000306325.1	Schwartz Jampel syndrome type 1	
rs35669711	[[C>C], [C>T]]		RCV000398821.1	Dyssegmental Dysplasia	
rs2229475	[[C>C], [C>T]]		RCV000339647.1	Dyssegmental Dysplasia	
rs2229475	[[C>C], [C>T]]		RCV000375538.1	Schwartz Jampel syndrome type 1	
rs2254358	[[C>C], [C>A]]		RCV000282698.1	Schwartz Jampel syndrome type 1	
rs2254358	[[C>C], [C>A]]		RCV000374875.1	Dyssegmental Dysplasia	
rs2782643	[[C>C], [C>T]]		RCV000247399.1	not specified	
rs11577368	[[C>C], [C>A], [C>T]]		RCV000438131.1	not specified	
rs718265	[[A>A], [A>G]]		RCV000263910.1	Desmosterolosis	
rs540796	[[A>A], [A>C], [A>G]]		RCV000182574.3	not specified	
rs540796	[[A>A], [A>C], [A>G]]		RCV000256294.6	Familial hypercholesterolemia	
rs540796	[[A>A], [A>C], [A>G]]		RCV000273530.1	Familial hypobetalipoproteinemia	
rs540796	[[A>A], [A>C], [A>G]]		RCV000600317.1	Hypercholesterolemia, autosomal dominant, 3	
rs562556	[[G>G], [G>A]]		RCV000182572.3	not specified	
rs562556	[[G>G], [G>A]]		RCV000256256.6	Familial hypercholesterolemia	
rs562556	[[G>G], [G>A]]		RCV000330942.1	Familial hypobetalipoproteinemia	
rs562556	[[G>G], [G>A]]		RCV000605201.1	Hypercholesterolemia, autosomal dominant, 3	
rs505151	[[G>G], [G>A]]		RCV000030349.7	Familial hypercholesterolemia	
rs505151	[[G>G], [G>A]]		RCV000252382.2	not specified	
rs505151	[[G>G], [G>A]]		RCV000364880.1	Familial hypobetalipoproteinemia	
rs505151	[[G>G], [G>A]]		RCV000612647.1	Hypercholesterolemia, autosomal dominant, 3	
rs14008	[[G>G], [G>A], [G>T]]		RCV000352006.1	Corneal Dystrophy, Dominant/Recessive	
rs1137101	[[A>A], [A>G]]		RCV000009047.3	LEPTIN RECEPTOR POLYMORPHISM	
rs1137101	[[A>A], [A>G]]		RCV000281795.1	Monogenic Non-Syndromic Obesity	
rs1137101	[[A>A], [A>G]]		RCV000348520.1	Leptin receptor deficiency	
rs1137101	[[A>A], [A>G]]		RCV000518727.1	not specified	
rs3112831	[[T>T], [T>C], [T>G]]		RCV000085383.5	not provided	
rs3112831	[[T>T], [T>C], [T>G]]		RCV000173675.2	not specified	
rs3112831	[[T>T], [T>C], [T>G]]		RCV000297592.1	Macular degeneration	
rs3112831	[[T>T], [T>C], [T>G]]		RCV000303578.1	Retinitis Pigmentosa, Recessive	
rs3112831	[[T>T], [T>C], [T>G]]		RCV000360724.1	Stargardt Disease, Recessive	
rs3112831	[[T>T], [T>C], [T>G]]		RCV000408013.1	Cone-Rod Dystrophy, Recessive	
rs3112831	[[T>T], [T>C], [T>G]]		RCV000085382.1	not provided	
rs1801265	[[A>A], [A>G]]		RCV000086506.1	not provided	
rs1801265	[[A>A], [A>G]]		RCV000000464.3	Dihydropyrimidine dehydrogenase deficiency	
rs1801274	[[A>A], [A>C], [A>G]]		RCV000015946.3	Lupus nephritis, susceptibility to	
rs1801274	[[A>A], [A>C], [A>G]]		RCV000015947.3	Pseudomonas aeruginosa, susceptibility to chronic infection by, in cystic fibrosis	
rs1801274	[[A>A], [A>C], [A>G]]		RCV000054529.2	Malaria, severe, susceptibility to	
rs1801274	[[A>A], [A>C], [A>G]]		RCV000211160.1	trastuzumab response - Efficacy	
rs1801274	[[A>A], [A>C], [A>G]]		RCV000454909.1	not specified	
rs12129650	[[T>T], [T>A], [T>C]]		RCV000270640.1	Macular degeneration	

# Conclusion

- The quality of Somerton Man's DNA file is poor which is 2% out of 0.6 million SNPs available.
- DNA match services failed on low quality DNA file and simple recovery strategies cannot help to find his relatives.
- Somerton Man's ethnicity is North Atlantic
- 574 genetic diseases found but none was related to Somerton Man's known characteristics
- Somerton Man's identity will remain a mystery

# **Project Management**

# Risk Assessment

Risk	Likelihood	Consequences	Risk Estimation
Absence of meeting	Unlikely	Minor	Low
Miscommunication of members	Unlikely	Moderate	Medium
Loss of data	Unlikely	Severe	High
Delay of task completion	Likely	Major	High
Bugs in codes	Likely	Minor	Medium
Out of budget	Rare	Severe	Medium
Misunderstanding of tasks	Unlikely	Moderate	Medium
Unethical works	Unlikely	Major	Medium
Member drop the course	Rare	Severe	Medium
Bad quality of purchased items	Unlikely	Major	Medium

# Milestones

<b>Semester 1</b>	<b>Milestone</b>	<b>Semester 2</b>	<b>Milestone</b>
<b>week 5</b>	<b>Complete Task 1</b>	<b>week 1</b>	<b>Review works</b>
<b>week 6</b>	<b>Proposal Seminar Slides</b>	<b>week 9</b>	<b>Complete Task 3</b>
<b>break</b>	<b>Proposal Seminar</b>	<b>week 12</b>	<b>Exhibition of projects</b>
<b>Week 11</b>	<b>Complete Task 2</b>	<b>Week 12</b>	<b>Final Thesis</b>
<b>Week 12</b>	<b>Thesis Draft</b>	<b>Week 13</b>	<b>Final seminar</b>

# Reference

The University of Adelaide (2015). HSW Handbook Appendix C1 [pdf] Available at <https://www.adelaide.edu.au/hr/hsw/docs/haz-man-long-form-app-c.pdf> [Accessed 9 Apr. 2019].

Nature.com. (2019). single nucleotide polymorphism / SNP | Learn Science at Scitable. [online] Available at: <https://www.nature.com/scitable/definition/single-nucleotide-polymorphism-snp-295> [Accessed 10 Apr. 2019].

Genetics Home Reference. (2019). What is DNA?. [online] Available at: <https://ghr.nlm.nih.gov/primer/basics/dna> [Accessed 9 Apr. 2019].

Genetics Home Reference. (2019). What is a chromosome?. [online] Available at: <https://ghr.nlm.nih.gov/primer/basics/chromosome> [Accessed 9 Apr. 2019].

D. Keane, "Will the Somerton Man mystery ever be solved?," ABC News, 01-Dec-2018. [Online]. Available: <https://www.abc.net.au/news/2018-12-01/will-the-mystery-of-the-somerton-man-ever-be-solved/10420794>. [Accessed: 10-Apr-2019].

# Reference

Ncbi.nlm.nih.gov. (2019). rs12913832 RefSNP Report - dbSNP - NCBI. [online] Available at: [https://www.ncbi.nlm.nih.gov/snp/rs12913832#clinical\\_significance](https://www.ncbi.nlm.nih.gov/snp/rs12913832#clinical_significance) [Accessed 14 Apr. 2019].

“Difference Between DNA and Genes | Definition, Structure, Features, Comparison,” Pediaa.Com, 16-Feb-2017. [Online]. Available: <https://pediaa.com/difference-between-dna-and-genes/>. [Accessed: 12-Apr-2019].

**Thank you**





THE UNIVERSITY  
*of* ADELAIDE

CRICOS PROVIDER NUMBER 00123M