

Classification of Alzheimer's by DNA Analysis

1

By: Ahmed Samir, Fairuz Soufy, Omar Ehab and Sara Elbedeawi.
Supervised by: Dr. Ashraf Abdelraouf and Eng. Lobna Shaheen.
In collaboration with Faculty of Pharmacy: Prof. Lamiaa Nabil,
Dr. Omar El-Demrdash and Dr. Rawan ElKady

Agenda

2

- Introduction.
- Problem definition.
- Block diagram.
- Use-case.
- Sample of functional requirements.
- Non-functional requirements.
- Class diagram.
- Results.
- Demo.



Introduction

3

What is Alzheimer's disease?

Alzheimer's is a progressive brain *disorder* that slowly destroys memory and thinking skills and, eventually, the ability to carry out the simplest tasks. Early diagnosis of Alzheimer's may help in slowing down the progression of the disease.[1]

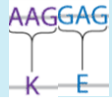

Alzheimer's can be detected by analyzing the DNA.



[1] Molinuevo, José Luis et al. "The Rationale Behind the New Alzheimer's Disease Conceptualization: Lessons Learned During the Last Decades." Journal of Alzheimer's disease : JADvol. 62,3 (2018): 1067-1077. doi:10.3233/JAD-170698

Introduction

4

Term	Description
DNA	DNA is a very long chain made up of something called nucleotides. <i>ACGTC CAGGTAGCT</i>
Nucleotides	There are 4 different nucleotides <i>A</i> - adenine , <i>T</i> - thymine , <i>C</i> - cytosine , <i>G</i> -guanine.
Gene	Is a sequence of nucleotides ,every three letters makes a codon/amino acid. 
SNP analysis	only looks at specific locations in DNA where relevant information can be gathered
Mutation	A Mutation occurs when a one or nucleotide is changed in the DNA 
Silent mutation	DNA level <i>TTC TTT</i> protein level Lys <i>Lys</i>

Fact About Alzheimer's

5

Scientists have identified four genes in which mutations cause Alzheimer's disease ,these genes are found in specific locations inside the DNA sequences^[2].

Gene	Chromosome	Start Location	End Location
PSEN-2	1	227,047,885	227,093,804
PSEN-1	14	73,593,143	73,700,399
APOE	19	45,399,039	45,422,650
APP	21	27,242,861	27,553,446

[2]“The Role of Genes in Your Alzheimer's Risk.” *Mayo Clinic*, Mayo Foundation for Medical Education and Research, 19 Apr. 2019, <https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/in-depth/alzheimers-genes/art-20046552>. Accessed 13 Jan. 2020.

Problem Definition

6

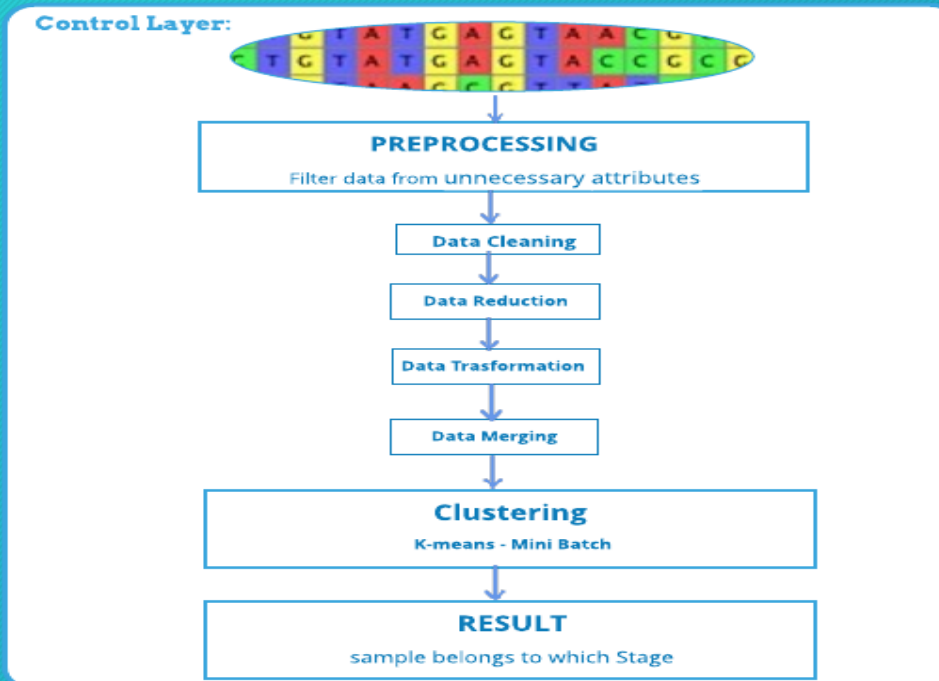
The existing classifications ^[3] of AD is either the patient is healthy, or the patient is in severe case of Alzheimer .

- We are clustering the diseased patients into three clusters mild, moderate and severe.
- Healthy patients will be classified into two stages.
 - Stage A :Healthy patients that are not going to suffer from Alzheimer's.
 - Stage B:Patients that have high risk to suffer from Alzheimer's.

[3]Finckh, Ulrich, et al. "High prevalence of pathogenic mutations in patients with early-onset dementia detected by sequence analyses of four different genes." *The American Journal of Human Genetics* 66.1 (2000): 110-117.

Block Diagram

7

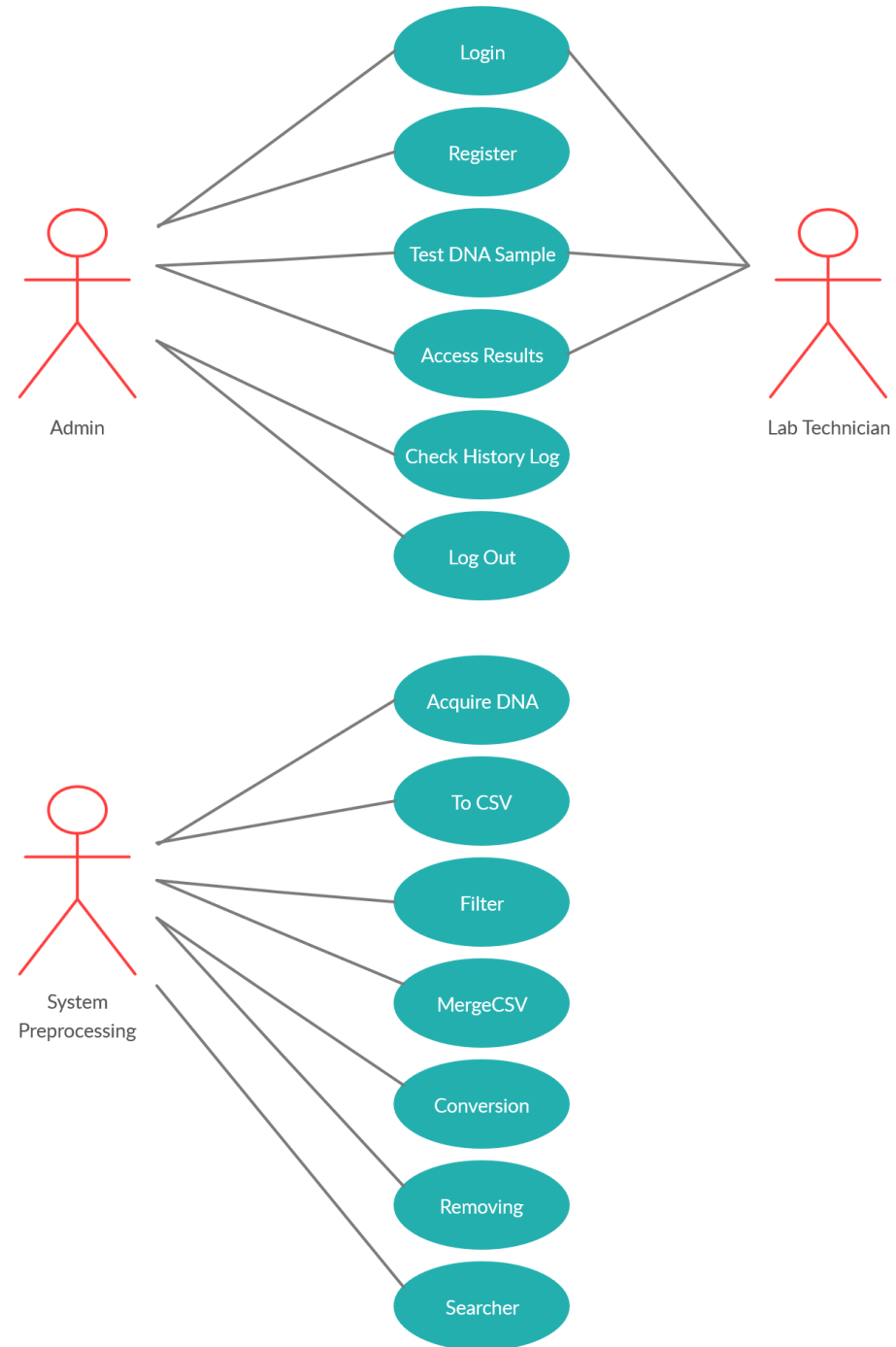


Data Access Layer :

Permissions	Admin	Lab technician
Login	✓	✓
Add Users	✓	
Test Sample		✓
Access Results	✓	✓
See History Log	✓	
Print Results	✓	✓

Use-case diagram

9



Functional Requirments

11

- The Admin should be able to register, edit and delete users.
- Users should be able to login by entering their credentials.
- Users should be able to upload a DNA sample.
- Users should be able to view result as in which stage is this patient.
- Users should be able to print result.
- Users should be able to check medical history of specific patient.

Non-functional Requirements

12

Security



Portability



Reliability

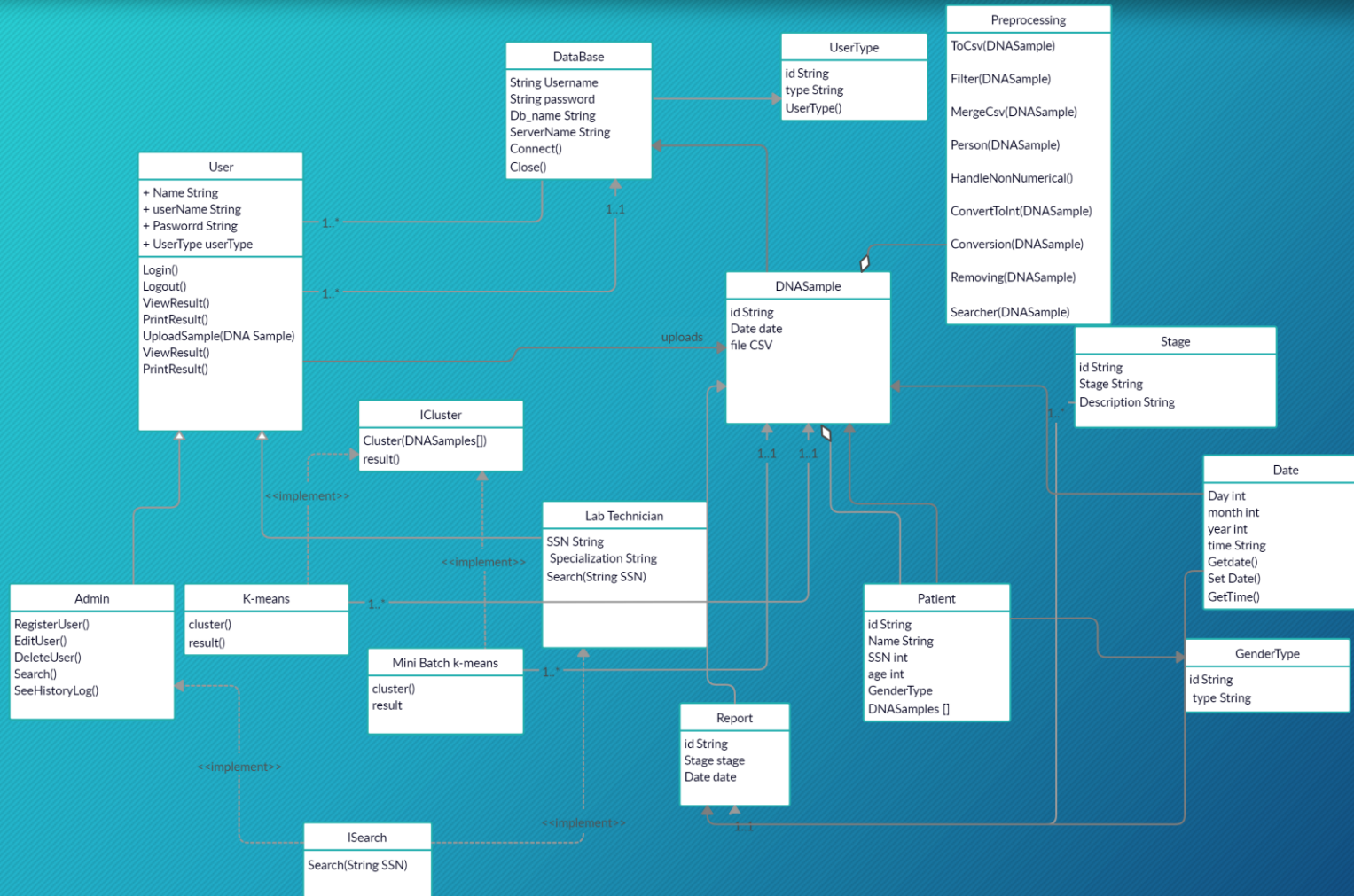


Maintainability



Class Diagram

13



Results

15

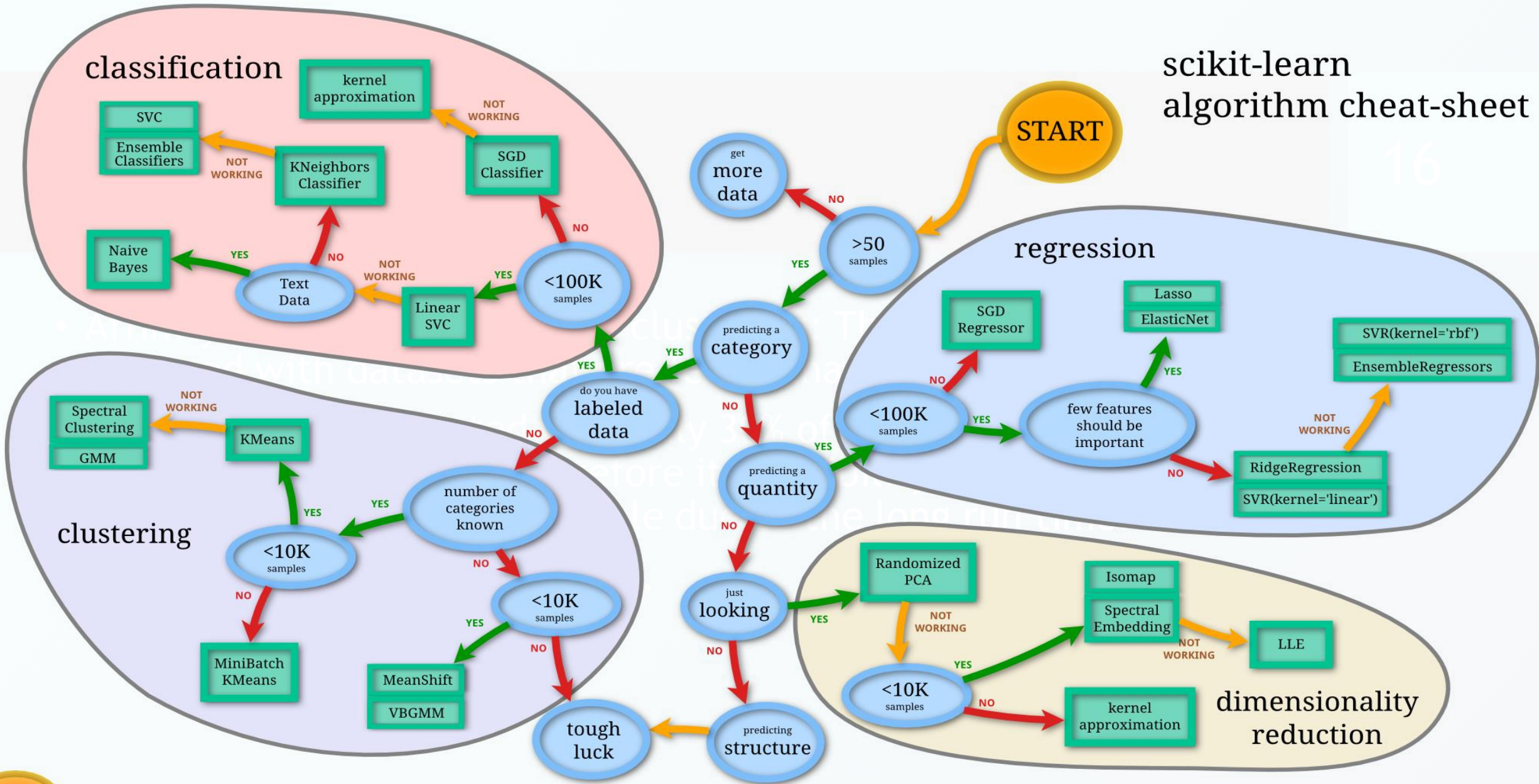
- We clustered the diseased dataset into 3 clusters here is the results

k-mean

```
k-mean
Total data = 106131 record
Cluster 1
26 %
Cluster 2
35 %
Cluster 3
39 %
-----
```

Mini batch

```
Minibatch K-means
Total data = 106131 record
Cluster 1
38 %
Cluster 2
35 %
Cluster 3
28 %
```



Back

Previous work / Present work

17

Whole Genome Sequence

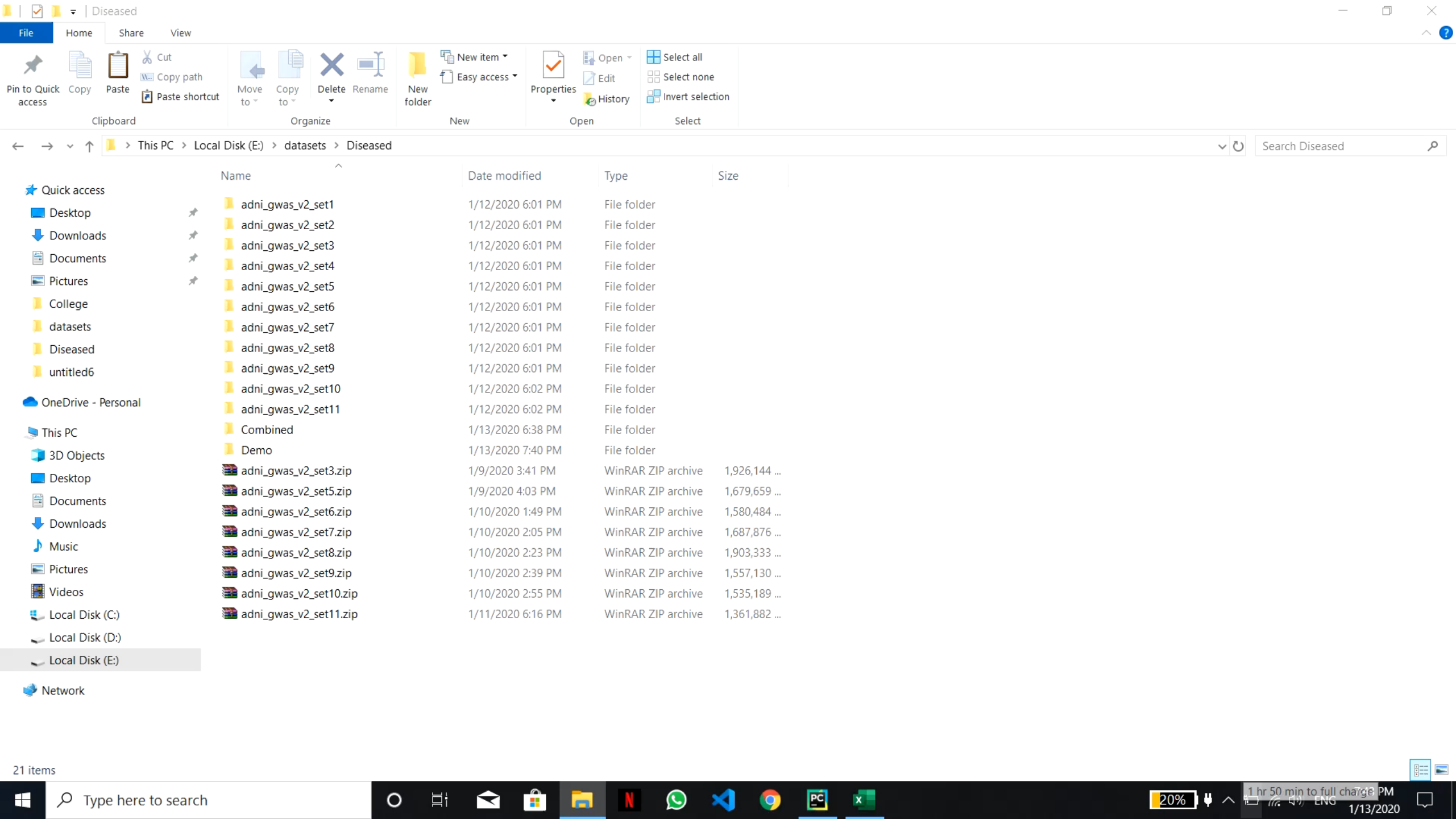
CATGACGTCGCGGACAACCCAGAATTGTCTTGAGCGATGGTAAGATCTAACCTCACTGCCGGGGGAGGCTCATAC
CTGGGGCTTTACTGATGTCATACCGTCTTGCAOOGGGATAGAATGACGGTGCCCGTGTCTGCTTGCTCGAAGCA
ATTTTCTGAAAGTTACAGACTTCGATTAAAAAGATCGGACTGCGCGTGGGCCCCGAGAGACATGCGTGGTAGTCA
TTTTTCGACGTGTCAAGGACTCAAGGGAATAGTTTGGCGGGAGCGTTACAGCTTCAATTCCCAAAGGTCCGAAGA
CGATAAAATTCAACTACTGGTTTCGGCCTAATAGGTCACGTTTTATGTGAAATAGAGGGGAACCGGCTCCCAAT
CCCTGGGTGTTCTATGATAAGTCTGCTTTATAACACGGGGCGGTTAGGTTAAATGACTCTTCTATCTTATGGTG
ATCCAAGCGCCCGCTAATTCTGTTCTGTTAATGTTTCATACCAATACTCACATCACATTAGATCAAAGGATCCCG
AGCCAGTCGCAAGGGTCTGCTGCTGTTGTCGACGCTCATGTTACTCTGGAATCTACCTGCCCTCCCTCAAC
GGTTAAGGCGTGTGATCGACGATGCAGGTATACATCGGCTCGGACCTACAGTGGTCGATCGACTGGCTACTGGCT
TCGCGGTTTCGGCGGTAGTTGAGTGCATAAACCAACCGGTGGCAAGTAGCAAGAAGACCTACCTGGGTCACTT
AGACAACCTAACTAATAGTCTCTAACGGGGAATTACCTTTACCAGTCTCATGCTCCAATATATCTGCACCGCTT
CAATGATATCGCCACAGAAAGTAGGGTCTCAGGTATCGCATACGCCGCGCCCGGGTCCAGCTACGCTCAGGAC
GACAGTAGAGAGCTATTGTGAATTCAGGCTCAGCATTCATCGACCTTTCTGTTGTGAATATTGTGCTAATGCA
TCTCGTCCGTAACGATCTGGGGGGCAAAACCGAATATCCGATTCTCGTCTACGGGTCCACAATGAGAAAGTCC
TGCGCGTGATCGTCAGTTAAGTTAAATTAATTCAGGCTACGGTAACTTGTAGTGAGCTAAGAATCACGGGAATC
ACGGGTTTCGCTACAGATGAAGTGAATTTATACACGGACAACCTATCGCCATTTGGGCGTGGGCACCGCAGATCA
AAAGTGGCAGATTAGGAGTGCTTGATCAGGTTAGCAGGTGGACTGTATCCAACAGCGCATCAAACCTTCAATAAAT
CCAAAGCGTTGTAGTGGTCTAAGCACCCCTGAACAGTGGCGCCCATCGTTAGCGTAGTACAACCTTCCCTTGTG

SNP (Single nucleotide polymorphisms)

Sample Index	Sample ID	GC Score	SNP Index	SNP Name	Allele1 - Top	Allele2 - Top	Allele1 - Forward	Allele2 - Forward	Allele1 - AB	Allele2 - AB	Chr	Position
373	002_S_0295	0.8695	30387	rs1012766	A	G	T	C	A	B	21	27329010
373	002_S_0295	0.7597	36236	rs10222075	A	G	T	C	A	B	21	27395391
373	002_S_0295	0.8036	45258	rs10482982	A	A	A	A	A	A	21	27394360
373	002_S_0295	0.9327	83580	rs11088023	A	G	T	C	A	B	21	27490373
373	002_S_0295	0.9138	94587	rs1129156	A	G	T	C	A	B	19	45410916
373	002_S_0295	0.648	97683	rs11585386	G	G	C	C	B	B	1	2.27E+08
373	002_S_0295	0.9583	105091	rs11701686	A	G	A	G	A	B	21	27384045
373	002_S_0295	0.9308	126479	rs12147044	A	A	T	T	A	A	14	73646566
373	002_S_0295	0.9337	140264	rs12481871	A	C	A	C	A	B	21	27358002
373	002_S_0295	0.8381	146427	rs12589973	A	G	T	C	A	B	14	73679172
373	002_S_0295	0.9409	163767	rs13047365	A	G	T	C	A	B	21	27326228
373	002_S_0295	0.7929	177514	rs1341715	A	A	A	A	A	A	1	2.27E+08
373	002_S_0295	0.76	179503	rs1351019	A	A	A	A	A	A	21	27453429
373	002_S_0295	0.8586	183597	rs1382005	A	G	A	G	A	B	21	27427527
373	002_S_0295	0.9515	194313	rs1468507	A	C	T	G	A	B	14	73641851
373	002_S_0295	0.9075	195609	rs1477717	G	G	G	G	B	B	21	27536742
373	002_S_0295	0.9592	197880	rs1497273	A	A	T	T	A	A	21	27278698
373	002_S_0295	0.9544	197881	rs1497290	A	G	A	G	A	B	21	27269954
373	002_S_0295	0.429	205632	rs15566	A	A	A	A	A	A	14	73597153
373	002_S_0295	0.7449	205743	rs1557365	A	G	T	C	A	B	21	27552366
373	002_S_0295	0.9105	207809	rs1571662	A	C	A	C	A	B	21	27424948
373	002_S_0295	0.9344	207810	rs1571665	A	G	A	G	A	B	21	27551612

Demo

18



Any Questions ?

Thank You 😊

20