Detecting miRNA Biomarkers for Alzheimer's Disease Using Next Generation Sequencing Data

Group:
E/15/081 Dinuwanthi S.L.I
E/15/362 Thilakarathna K.A.H.P
E/15/345 Sripadi D.V.

Supervisors:
Dr. Damayanthi Herath
Prof. Roshan Ragel

Background:

- What is Next Generation Sequencing ?
    - Sample preparation
    - Sequencing by machines
    - Data output
- What are miRNAs ?

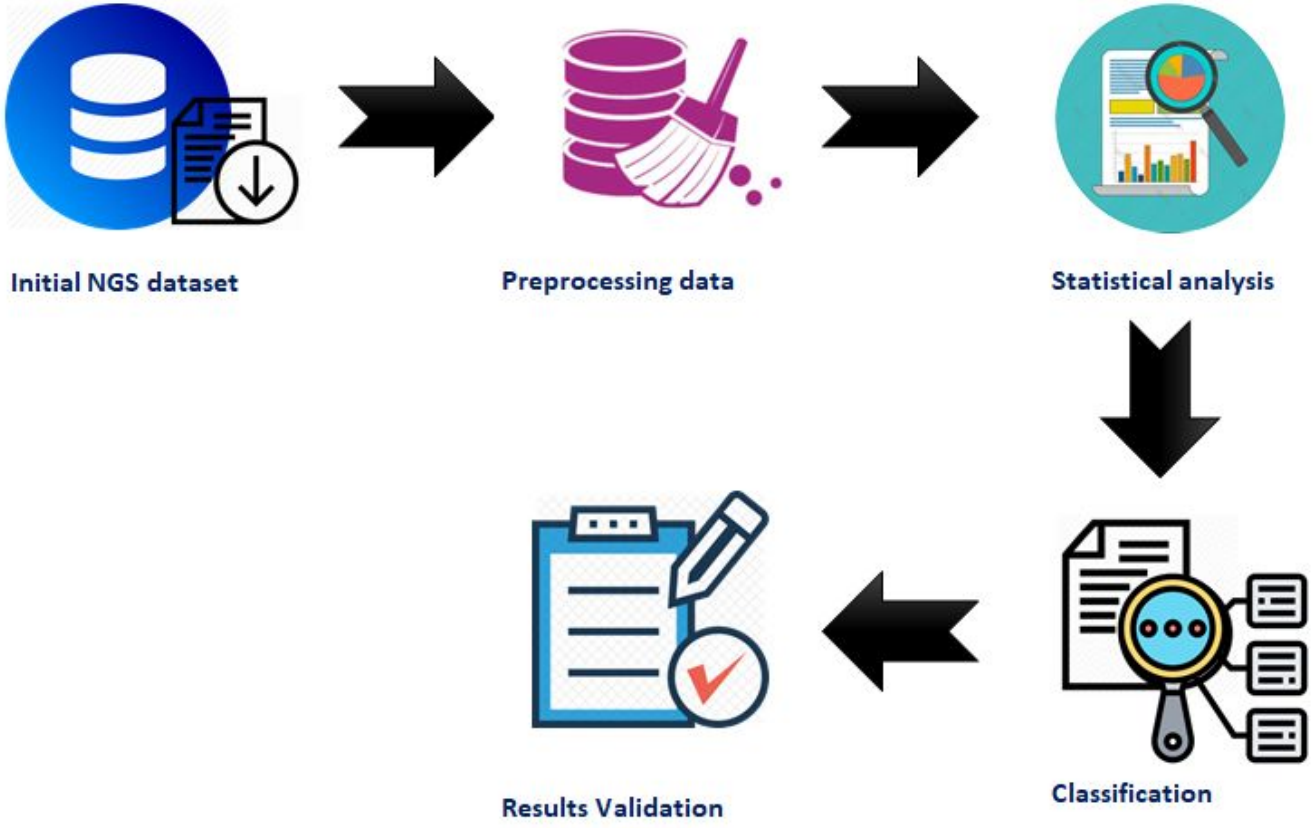
Problem definition:

- Medical history of Alzheimer's Disease
- Drawbacks of previously introduced methods
  - sample selection
  - preprocessing methods
  - statistical analysis
  - machine learning approaches

Design justification

- Why miRNA biomarkers?
- Why Next Generation sequencing technology?
- Quantile normalization as normalization technique?
- Selection of methods used for statistical analysis?
- Use of machine learning algorithms?

Methodology:



Implementation choices: Methodologies

Preprocessing

- Trimming sequence data - Adapters, indexes, low quality reads
- Filtering - short read sequences

Statistical analysis

- Quantile normalization - remove unwanted variations
- P value & Fold change - identify most significance miRNAs
- AUC - identify dysregulated miRNAs

Implementation choices: Tools

NGS data analysis

- Galaxy platform - web based, open source platform
- Galaxy tools
  - FastQC - Quality check
  - Trim Galore - remove adapters and low quality reads
  - Filter FASTQ - filter short read sequences
  - Bowtie2 - map sequence against reference genome
  - Htseq-count - identify read counts

Implementation choices: Tools

Statistical Analysis

- Python - clean syntax, straightforward semantics & Third-party toolkits
- Python libraries
  - Numpy
  - Pandas
  - Scipy
  - Scikit learn
  - Matplotlib

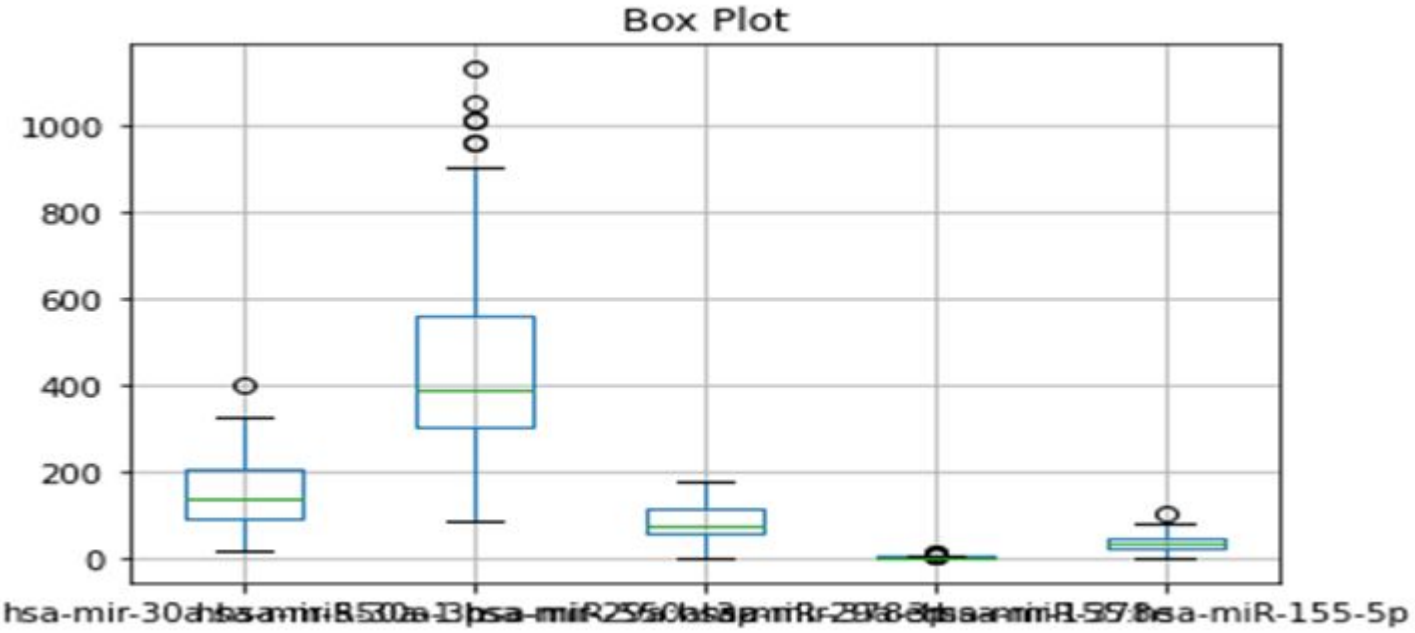
Implementation choices: Models

Classification

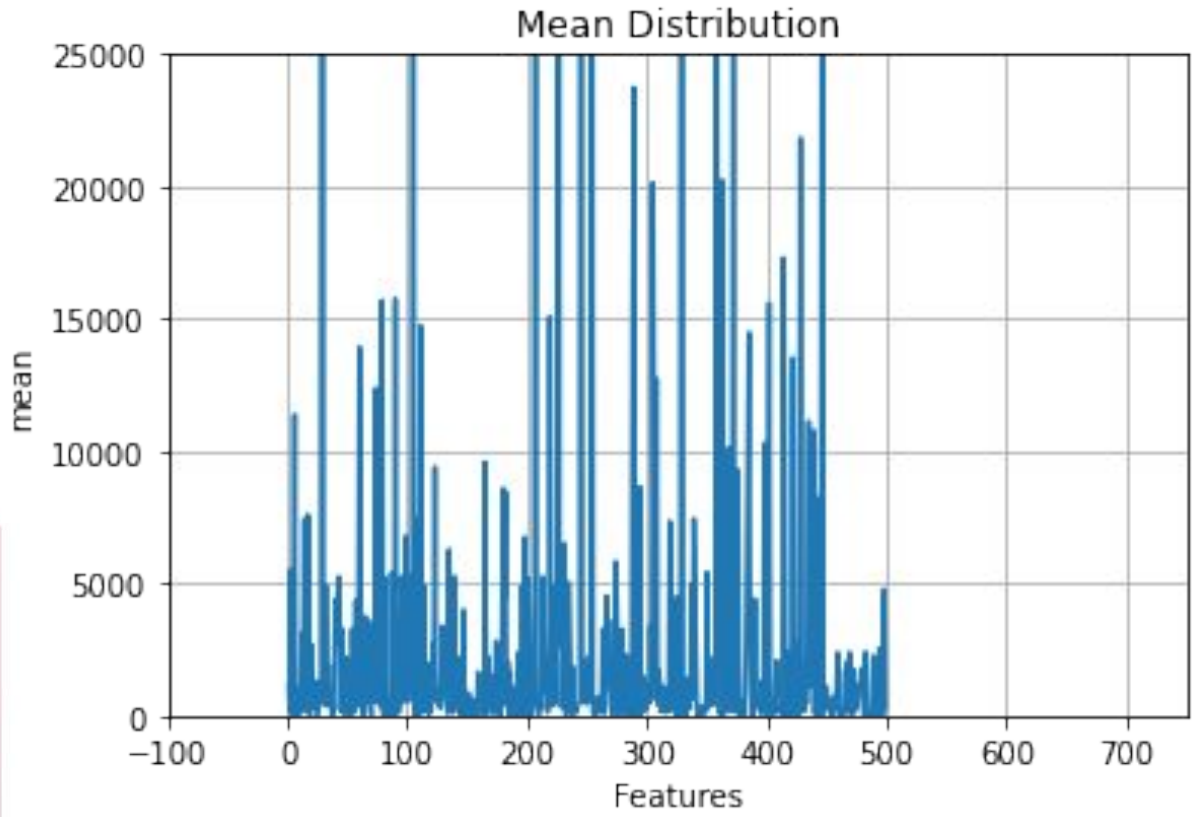
- Machine Learning models
  - Logistic regression
  - Linear SVM
  - Gaussian SVM
  - Naive Bayes
  - K Nearest Neighbour
  - Random Forests

Results Obtained From Each Stage of Analysis

Box Plot For Five Random Features From Normalized Dataset



Mean Distribution of Normalized Dataset

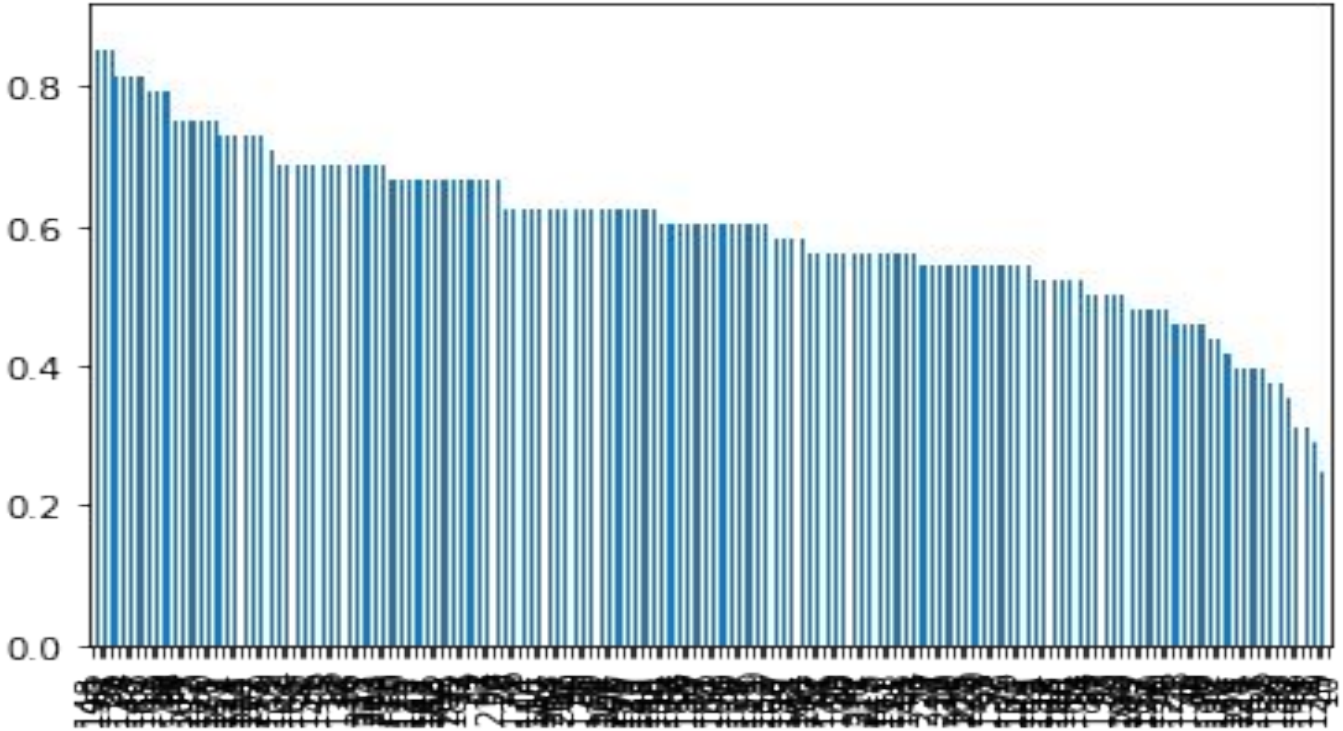


Filtered miRNAs Using Significance Values and Fold Change

- 228 miRNAs

```
Filtered miRNAs using significance value and fold change are: ['hsa-mir-30a:hsa-miR-30a-3p', 'hsa-mir-550a-1:hsa-miR-550a-3p',
'hsa-mir-29a:hsa-miR-29a-3p', 'hsa-mir-628:hsa-miR-628-3p', 'hsa-mir-26a-2:hsa-miR-26a-5p', 'hsa-mir-106b:hsa-miR-106b-5p', 'hs
a-mir-4781:hsa-miR-4781-3p', 'hsa-mir-10b:hsa-miR-10b-5p', 'hsa-mir-215:hsa-miR-215', 'hsa-mir-548aj-2:hsa-miR-548g-5p', 'hsa-m
ir-181a-1:hsa-miR-181a-3p', 'hsa-mir-548x:hsa-miR-548ar-5p', 'hsa-mir-548k:hsa-miR-548av-5p', 'hsa-mir-199a-1:hsa-miR-199a-3p',
'hsa-mir-30e:hsa-miR-30e-3p', 'hsa-mir-4508:hsa-miR-4508', 'hsa-mir-548aj-2:hsa-miR-548x-5p', 'hsa-mir-371b:hsa-miR-371b-5p',
'hsa-mir-5001:hsa-miR-5001-3p', 'hsa-mir-16-2:hsa-miR-16-2-3p', 'hsa-mir-128-2:hsa-miR-128', 'hsa-mir-486:hsa-miR-486-3p', 'hsa
-mir-4482-1:hsa-miR-4482-3p', 'hsa-mir-941-4:hsa-miR-941', 'hsa-mir-550a-1:hsa-miR-550a-5p', 'hsa-mir-199a-2:hsa-miR-199b-3p',
'hsa-mir-144:hsa-miR-144-5p', 'hsa-let-7f-2:hsa-let-7f-5p', 'hsa-mir-126:hsa-miR-126-5p', 'hsa-mir-191:hsa-miR-191-3p', 'hsa-mi
r-10a:hsa-miR-10a-5p', 'hsa-mir-98:hsa-miR-98', 'hsa-mir-548x:hsa-miR-548x-5p', 'hsa-mir-363:hsa-miR-363-3p', 'hsa-mir-548h-1:h
sa-miR-548h-5p', 'hsa-mir-223:hsa-miR-223-3p', 'hsa-mir-5690:hsa-miR-5690', 'hsa-mir-199b:hsa-miR-199b-3p', 'hsa-mir-3200:hsa-m
iR-3200-3p', 'hsa-mir-424:hsa-miR-424-3p', 'hsa-mir-644b:hsa-miR-644b-3p', 'hsa-mir-548h-5:hsa-miR-548h-5p', 'hsa-mir-18a:hsa-m
iR-18a-5p', 'hsa-mir-548g:hsa-miR-548x-5p', 'hsa-mir-548g:hsa-miR-548g-5p', 'hsa-mir-21:hsa-miR-21-5p', 'hsa-mir-99b:hsa-miR-99
b-5p', 'hsa-mir-25:hsa-miR-25-3p', 'hsa-mir-937:hsa-miR-937', 'hsa-mir-1180:hsa-miR-1180', 'hsa-mir-30c-1:hsa-miR-30c-5p', 'hsa
-let-7a-1:hsa-let-7a-5p', 'hsa-mir-941-1:hsa-miR-941', 'hsa-mir-660:hsa-miR-660-5p', 'hsa-mir-421:hsa-miR-421', 'hsa-mir-374a:h
sa-miR-374a-5p', 'hsa-mir-328:hsa-miR-328', 'hsa-mir-151a:hsa-miR-151a-5p', 'hsa-mir-548x:hsa-miR-548aj-5p', 'hsa-mir-101-2:hsa
-miR-101-3p', 'hsa-mir-28:hsa-miR-28-3p', 'hsa-mir-139:hsa-miR-139-5p', 'hsa-mir-2110:hsa-miR-2110', 'hsa-let-7g:hsa-let-7g-5
p', 'hsa-mir-550a-3:hsa-miR-550a-3-5p', 'hsa-mir-548aj-2:hsa-miR-548ar-5p', 'hsa-mir-144:hsa-miR-144-3p', 'hsa-mir-548e:hsa-miR
```

Plot for ROC AUC values for selected miRNA



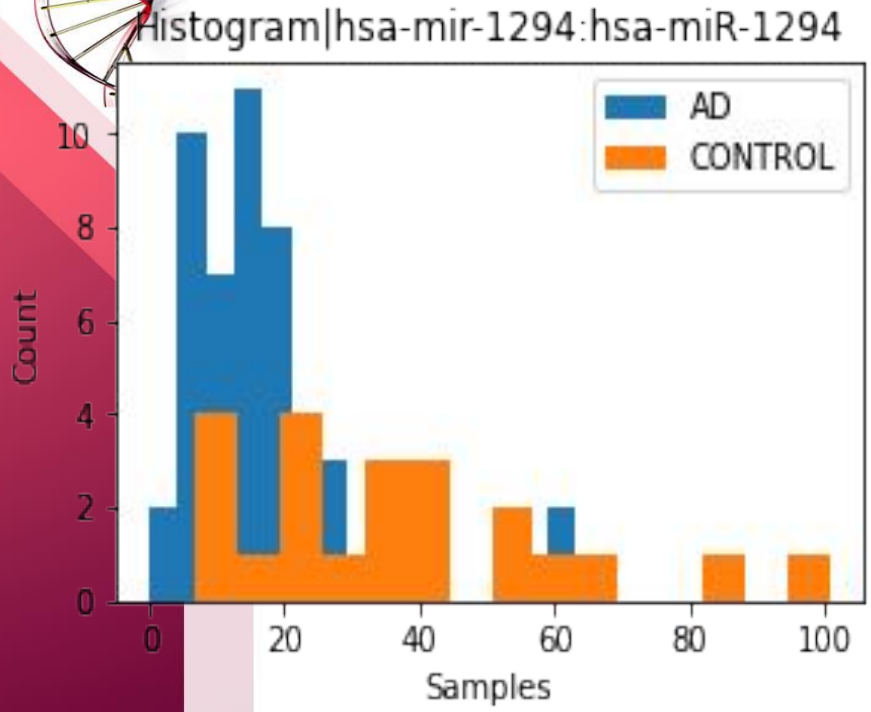
miRNAs Selected Using ROC AUC values

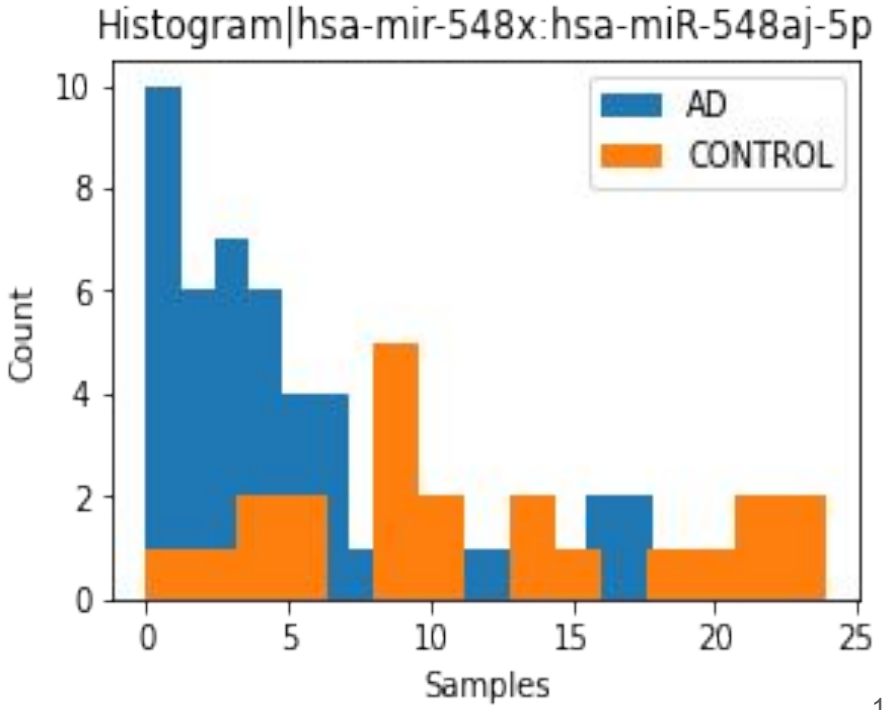
- 154 down regulated miRNAs
- 32 up regulated miRNAs

```
Down regulated miRNAs: ['hsa-mir-30a:hsa-miR-30a-3p', 'hsa-mir-29a:hsa-miR-29a-3p', 'hsa-mir-155:hsa-miR-155-5p', 'hsa-mir-26a-
2:hsa-miR-26a-5p', 'hsa-mir-106b:hsa-miR-106b-5p', 'hsa-mir-4781:hsa-miR-4781-3p', 'hsa-mir-204:hsa-miR-204-5p', 'hsa-mir-10b:h
sa-miR-10b-5p', 'hsa-mir-1260a:hsa-miR-1260a', 'hsa-mir-215:hsa-miR-215', 'hsa-mir-548aj-2:hsa-miR-548g-5p', 'hsa-mir-3613:hsa-
miR-3613-3p', 'hsa-mir-1226:hsa-miR-1226-3p', 'hsa-mir-7-3:hsa-miR-7-5p', 'hsa-mir-1303:hsa-miR-1303', 'hsa-mir-196a-1:hsa-miR-
196a-5p', 'hsa-mir-181a-1:hsa-miR-181a-3p', 'hsa-mir-548x:hsa-miR-548ar-5p', 'hsa-mir-548k:hsa-miR-548av-5p', 'hsa-mir-199a-1:h
sa-miR-199a-3p', 'hsa-mir-4448:hsa-miR-4448', 'hsa-mir-30e:hsa-miR-30e-3p', 'hsa-mir-3177:hsa-miR-3177-3p', 'hsa-mir-4508:hsa-m
iR-4508', 'hsa-mir-548h-4:hsa-miR-548z', 'hsa-mir-548aj-2:hsa-miR-548x-5p', 'hsa-mir-378a:hsa-miR-378a-3p', 'hsa-mir-548o-2:hsa
-miR-548au-5p', 'hsa-let-7a-1:hsa-let-7a-3p', 'hsa-mir-486:hsa-miR-486-3p', 'hsa-mir-4482-1:hsa-miR-4482-3p', 'hsa-mir-4511:hsa
-miR-4511', 'hsa-mir-1270-1:hsa-miR-1270', 'hsa-mir-132:hsa-miR-132-3p', 'hsa-mir-941-4:hsa-miR-941', 'hsa-mir-877:hsa-miR-877-
5p', 'hsa-mir-5189:hsa-miR-5189', 'hsa-mir-144:hsa-miR-144-5p', 'hsa-let-7f-2:hsa-let-7f-5p', 'hsa-mir-378b:hsa-miR-378b', 'hsa
-mir-126:hsa-miR-126-5p', 'hsa-mir-1538:hsa-miR-1538', 'hsa-mir-191:hsa-miR-191-3p', 'hsa-mir-181b-2:hsa-miR-181b-5p', 'hsa-mir
-196a-2:hsa-miR-196a-5p', 'hsa-mir-98:hsa-miR-98', 'hsa-mir-330:hsa-miR-330-3p', 'hsa-mir-548x:hsa-miR-548x-5p', 'hsa-mir-363:h
sa-miR-363-3p', 'hsa-mir-424:hsa-miR-424-5p', 'hsa-mir-223:hsa-miR-223-3p', 'hsa-mir-5690:hsa-miR-5690', 'hsa-mir-548am:hsa-miR
-548au-5p', 'hsa-mir-1976:hsa-miR-1976', 'hsa-mir-199b:hsa-miR-199b-3p', 'hsa-mir-548ah:hsa-miR-548ah-3p', 'hsa-mir-3200:hsa-mi
R-3200-3p', 'hsa-mir-192:hsa-miR-192-5p', 'hsa-mir-424:hsa-miR-424-3p', 'hsa-mir-644b:hsa-miR-644b-3p', 'hsa-mir-548h-5:hsa-miR
-548h-5p', 'hsa-mir-548aa-2:hsa-miR-548aa', 'hsa-mir-196b:hsa-miR-196b-5p', 'hsa-mir-93:hsa-miR-93-5p', 'hsa-mir-548g:hsa-miR-5
48x-5p', 'hsa-mir-548g:hsa-miR-548g-5p', 'hsa-mir-21:hsa-miR-21-5p', 'hsa-mir-652:hsa-miR-652-3p', 'hsa-mir-25:hsa-miR-25-3p',
'hsa-mir-937:hsa-miR-937', 'hsa-mir-625:hsa-miR-625-3p', 'hsa-mir-1180:hsa-miR-1180', 'hsa-mir-30c-1:hsa-miR-30c-5p', 'hsa-mir-
548o:hsa-miR-548o-3p', 'hsa-let-7a-1:hsa-let-7a-5p', 'hsa-mir-941-1:hsa-miR-941', 'hsa-mir-548o-2:hsa-miR-548c-5p', 'hsa-mir-36
```

Up regulated miRNAs: ['hsa-mir-550a-1:hsa-miR-550a-3p', 'hsa-mir-378e:hsa-miR-378e', 'hsa-mir-628:hsa-miR-628-3p', 'hsa-mir-194-1:hsa-miR-194-5p', 'hsa-mir-4732:hsa-miR-4732-3p', 'hsa-mir-183:hsa-miR-183-3p', 'hsa-mir-486:hsa-miR-486-5p', 'hsa-mir-5001:hsa-miR-5001-3p', 'hsa-mir-16-2:hsa-miR-16-2-3p', 'hsa-mir-128-2:hsa-miR-128', 'hsa-mir-4753:hsa-miR-4753-5p', 'hsa-mir-326:hsa-miR-326', 'hsa-mir-10a:hsa-miR-10a-5p', 'hsa-mir-4732:hsa-miR-4732-5p', 'hsa-mir-4286:hsa-miR-4286', 'hsa-mir-99a:hsa-miR-99a-5p', 'hsa-mir-151a:hsa-miR-151a-5p', 'hsa-mir-548x:hsa-miR-548aj-5p', 'hsa-mir-4685:hsa-miR-4685-3p', 'hsa-mir-139:hsa-miR-139-5p', 'hsa-mir-3661:hsa-miR-3661', 'hsa-mir-342:hsa-miR-342-5p', 'hsa-mir-30d:hsa-miR-30d-3p', 'hsa-mir-431:hsa-miR-431-5p', 'hsa-mir-140:hsa-miR-140-3p', 'hsa-mir-1299:hsa-miR-1299', 'hsa-mir-1306:hsa-miR-1306-5p', 'hsa-mir-500a:hsa-miR-500a-3p', 'hsa-mir-3615:hsa-miR-3615', 'hsa-mir-4746:hsa-miR-4746-5p', 'hsa-mir-1301:hsa-miR-1301', 'hsa-mir-92a-1:hsa-miR-92a-3p']

Distribution of The Most Up Regulated And The Most Down Regulated





Deliverables Addressed in Phase 1

Milestone 01 : Background study

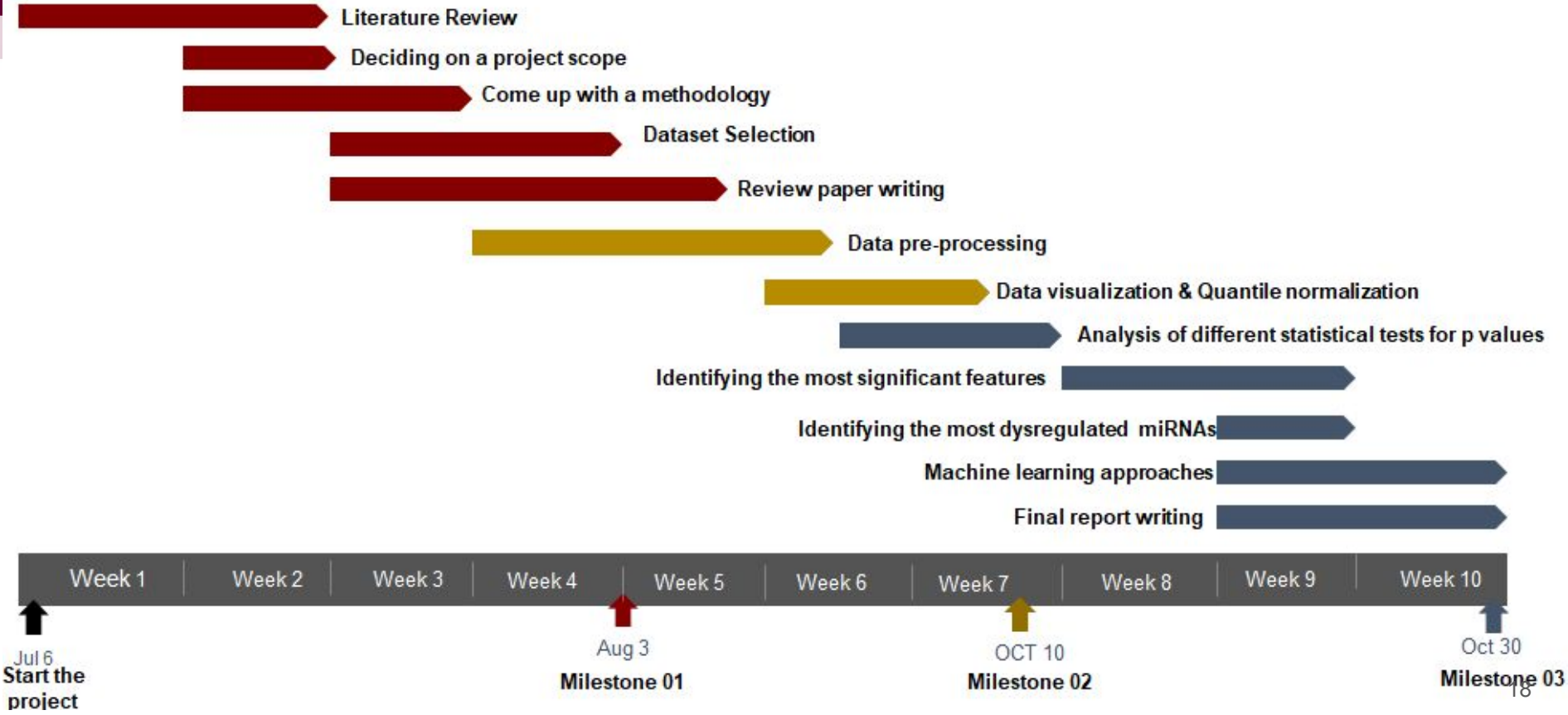
Dataset selection

Milestone 02: Preprocessing dataset (Galaxy tool)

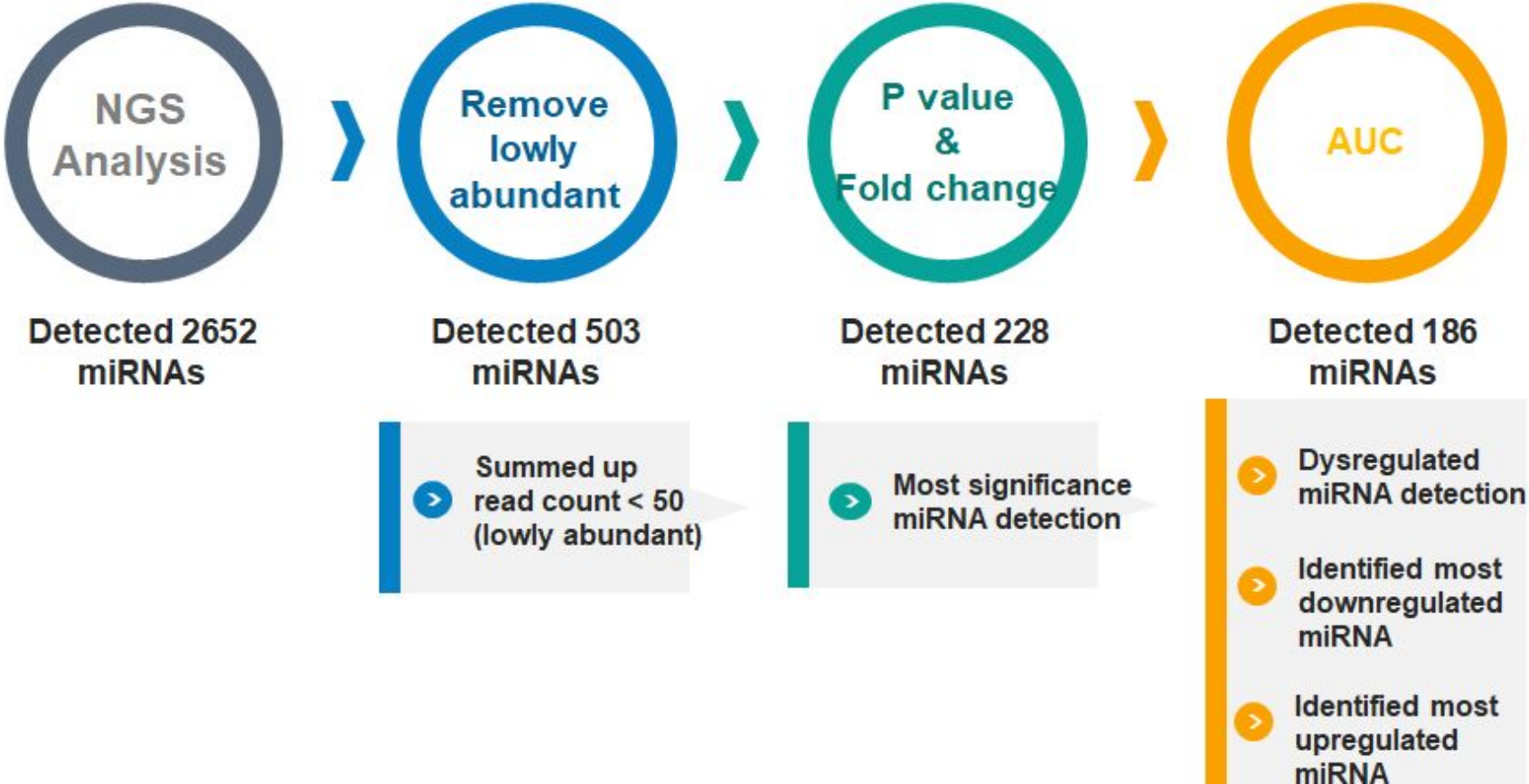
Data visualization and normalization

Milestone 03: Statistical analysis

Phase 1 Timeline



Conclusion



Plan for The Next Phase

Milestone 01: Background search on validation methods

Selection of a validation dataset

Milestone 02 : Validation of the biomarkers

Milestone 03 : Developing the implemented solution to be used in clinical use

Demonstration

Thank you

Q & A