

Decoding the Epigenome of Neuronal Networks in Health and Disease

Dissertation

for the award of the degree
"Doctor rerum naturalium" (Dr.rer.nat.)
of the Georg-August-Universität Göttingen

within the doctoral program
Theoretical and Computational Neuroscience (PTCN), GGNB
of the Georg-August University School of Science (GAUSS)

submitted by
Gaurav Jain
Born in Satna (M.P.), India

Göttingen 2018

Thesis Committee

- Prof. Dr. André Fischer
(Director, Dept. Systems Medicine and Epigenetics in Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), Dept. for Psychiatry and Psychotherapy, University Medical Center, Göttingen, Germany)
- Prof. Dr. Martin Göpfert
(Dept. of Cellular Neurobiology, Schwann-Schleiden Research Centre, Georg-August-Universität, Göttingen, Germany)
- Dr. Dr. Oliver Schlüter
(Psychiatry and Psychotherapy, University Medical Center, Molecular Neurobiology, European Neuroscience Institute, Göttingen, Germany)

Members of the Examination Board

1st Reviewer: Prof. Dr. André Fischer

(Director, Dept. Systems Medicine and Epigenetics in Neurodegenerative Diseases, German Center for Neurodegenerative Diseases (DZNE), Dept. for Psychiatry and Psychotherapy, University Medical Center, Göttingen, Germany)

2nd Reviewer: Prof. Dr. Martin Göpfert

(Dept. of Cellular Neurobiology, Schwann-Schleiden Research Centre, Georg-August-Universität, Göttingen, Germany)

Further members of the Examination Board:

- Dr. Dr. Oliver Schlüter
(Psychiatry and Psychotherapy, University Medical Center, Molecular Neurobiology, European Neuroscience Institute, Göttingen, Germany)
- Prof. Dr. Tiago Fleming Outeiro
(Head of the Department, Dept. of NeuroDegeneration and Restaurative Research, University Medical Center, Göttingen, Germany)
- Prof. Dr. Melanie Wilke
(Head of the Department, Dept. of Cognitive Neurology, University Medical Center, Göttingen, Germany)
- Prof. Dr. Lutz Walter
(Head of the Department, Dept. of Primate Genetics, German Primate Center, Göttingen, Germany)

Date of oral examination: 15th October 2018

Declaration

I hereby declare that I have written the dissertation “**Decoding the Epigenome of Neuronal Networks in Health and Disease**” independently with no other sources or aids than stated.



Gaurav Jain, Göttingen, 16th October 2018

1. Table of Contents

1. Table of Contents.....	4
2. Thesis abstract.....	10
3. Acknowledgements.....	11
4. List of Figures and Tables.....	13
5. List of Abbreviations.....	19
6. Introduction	20
6.1 Introduction to neurodegenerative disorders.....	20
6.1.1 Potential classification of neurodegenerative diseases	20
6.1.2 The partnership of aging and neurodegeneration.....	22
6.1.3 Effects of neurodegeneration: The big picture.....	23
6.2 Overview of dementia.....	23
6.2.1 Mild Cognitive Impairment (MCI)	23
6.2.1.1 MCI: normal aging or dementia?.....	23
6.2.1.2 Symptoms of MCI	23
6.2.1.3 Potential causes of MCI	24
6.2.1.4 Clinical criteria for the diagnosis of MCI and treatment ...	24
6.2.2 Dementia	25
6.2.2.1 Dementia – Information and Statistics	25
6.2.2.2 Symptoms of Dementia.....	25
6.2.2.3 Dementia - Types and Causes	26
6.2.3 Alzheimer’s Disease (AD).....	27
6.2.3.1 An overview of Alzheimer’s disease	27
6.2.3.2 Disease pathology of Alzheimer’s disease	28
6.2.3.3 Stages and Symptoms of Alzheimer’s disease.....	29
6.2.3.4 Diagnosis and treatment of Alzheimer’s disease	31
6.3 Exosomes.....	31
6.3.1 Overview of Exosomes.....	31
6.3.2 Biogenesis and functions of exosomes	32
6.3.3 Role of exosomes in Alzheimer’s disease	34
6.3.4 Exosomes as potential biomarkers for Alzheimer’s disease.....	35
6.4 Biomarkers	35
6.4.1 Definition of a biomarker.....	35
6.4.2 Types of biomarkers	35
6.4.3 Pros and cons of biomarkers.....	36

6.4.4	Potential biomarkers for MCI and Alzheimer's disease	36
6.5	Epigenetics and role of Non-coding RNAs	37
6.5.1	Introduction to Epigenetics	37
6.5.2	A general overview of small non-coding RNAs (sncRNAs)	38
6.5.3	Biogenesis of microRNAs (miRNAs) and piRNAs	40
6.5.3.1	miRNAs biogenesis.....	40
6.5.3.2	piRNA biogenesis	41
6.5.4	Role of miRNAs and piRNAs in Alzheimer's disease	41
6.6	Machine learning in biomarker development.....	42
6.6.1	Overview of machine learning	42
6.6.2	Types of learning	43
6.6.2.1	Supervised learning	43
6.6.2.2	Unsupervised learning	43
6.6.2.3	Semi-supervised learning	43
6.6.3	Alternative categorization of learning algorithms	44
6.6.3.1	Generative learning.....	44
6.6.3.2	Discriminative learning.....	44
6.6.4	Prior knowledge in learning	44
6.6.5	Handling heterogeneous data.....	45
6.6.6	Data munging and normalization	45
6.6.6.1	Data munging and wrangling	45
6.6.6.2	Data normalization	45
6.6.7	Random forest and other machine learning algorithms used in the analysis.....	46
6.6.8	Feature selection	47
6.6.9	Evaluation of learning models.....	49
6.6.9.1	Classification metrics	49
6.6.9.2	Regression metrics	50
6.6.9.3	Unsupervised learning metrics.....	51
6.7	Three dimensional organization of the genome	51
6.7.1	General features of the 3D genome organization.....	51
6.7.1.1	Chromosome territories	52
6.7.1.2	Compartments	53
6.7.1.3	TADs.....	53
6.7.1.4	subTADs	54
6.7.1.5	Chromatin Loops.....	54
6.7.2	Tools to explore 3D genome organization	54

6.7.2.1	Imaging based tools.....	54
6.7.2.2	3C-based technologies	55
6.7.3	Role of 3D genome organization in diseases	55
7.	Objectives	56
8.	Materials and methods.....	58
8.1	Cerebrospinal fluid (CSF) from humans.....	58
8.1.1	CSF from Alzheimer’s disease patients and control humans	58
8.1.2	Extraction of CSF exosomes	58
8.1.3	Total RNA isolation.....	58
8.1.4	SmallRNA library preparation	59
8.2	Processing of next generation sequencing data.....	59
8.2.1	Generation of raw fastq files	59
8.2.2	Quality control (QC) for raw sequencing data.....	59
8.3	SmallRNA mapping pipeline (GJSrMap)	60
8.3.1	Prepare custom reference genome	61
8.3.1.1	Small noncoding RNA and full genome sequences source	61
8.3.1.2	Pre-processing of the sequences	62
8.3.1.3	Custom reference genome indexes	62
8.3.2	Pre-processing and quality control before alignment.....	62
8.3.2.1	Quality control for FastQ files.....	62
8.3.2.2	Low-quality read filtering	63
8.3.2.3	Adapter trimming and read size selection.....	63
8.3.3	Iterative mapping of the filtered fastQ reads.....	63
8.3.3.1	Iteration 1: mapping to pimiRNA reference genome.....	63
8.3.3.2	Iteration 2: mapping to sncRNA reference genome.....	64
8.3.3.3	Iteration 3: mapping to full genome.....	64
8.3.4	Read counting and summary statistics	64
8.3.4.1	Read counting and summary quality control report	64
8.3.4.2	Summary mapping and smallRNA statistics	65
8.4	Data filtering and normalization	66
8.4.1	Normalization of smallRNAs	67
8.4.2	Filtering of non-expressed smallRNAs	68
8.5	Statistical and machine learning analysis.....	68
8.5.1	Variable ranking and removal of low ranked variables	70
8.5.2	Multivariate analysis of covariance	72
8.5.3	Model selection and performance.....	72

8.6	Handling of mice and nuclei preparation	74
8.6.1	Ordering and upkeep of mice	74
8.6.2	Tissue collection	74
8.6.3	Nuclei isolation and sorting.....	74
8.7	Chromosome Conformation Capture Sequencing (3C-Seq)	75
8.7.1	Preparation of 3C library	75
8.7.2	Quantification of 3C library	76
8.8	Sequencing of 3C-seq library	79
8.8.1	Preparation of 3C library.....	79
8.8.2	Quantification of 3C library and library preparation	80
8.9	Processing of 3C-seq sequencing data.....	80
8.9.1	Generation of raw fastQ files and quality control	80
8.9.2	Iterative mapping to the reference genome	80
8.9.3	Reads assignment to restriction enzyme fragments and filtering	82
8.9.4	Assign reads to genomic bins and perform iterative correction ...	83
8.9.5	Defining topological regions and domains	84
9.Results	86
9.1	CSF exosomes and small non-coding-RNAome	86
9.1.1	Analysis of isolated RNA from CSF exosomes and exosomal free CSF fraction.....	86
9.1.2	miRNAs and piRNAs in CSF exosome small non-coding- RNAome	87
9.1.3	miRNAs arm expression in CSF exosomes and brain tissues	89
9.1.4	Cellular small non-coding-RNAome and exosomal small non- coding-RNAome	90
9.2	A CSF miRNAs/piRNAs signature to diagnose Alzheimer's disease patients along with the available clinical markers.....	91
9.2.1	Demographic information for the samples included in the biomarker signature analysis	91
9.2.2	Performance of clinical signatures on replication cohort	92
9.2.3	Identification of small non-coding RNAs signature	93
9.2.4	The role of clinical and small non-coding RNA signatures	95
9.2.5	The performance of small non-coding RNAs signatures on replication cohort	96

9.3	CSF smallRNAs signature along with the available clinical markers as a promising biomarker to predict the progression from mild cognitive impairment (mci) to Alzheimer’s disease.....	99
9.3.1	Demographic information for the samples included in the biomarker signature analysis.....	99
9.3.2	The role of small non-coding RNA signatures in predicting the onset of disease progression.....	99
9.3.3	The performance of small non-coding RNAs signatures on predictive replication cohort 3 for the predicting the onset of disease progression.....	100
9.4	CSF smallRNAs signature in blood and brain	101
9.4.1	Demographic information for the samples included in the biomarker signature analysis.....	101
9.4.2	Role of CSF exosomes small non-coding RNA signatures in blood and brain.....	102
9.4.3	Performance of CSF exosomes small non-coding RNAs signatures in plasma exosomes blood samples (cohort 4) and from the post mortem brain tissue samples from the Prefrontal Cortex region (cohort 5).....	103
9.5	Functionally enriched piRNAs and miRNAs gene target pathways.....	106
9.5.1	Pathway analysis of miRNAs signature functional targets.....	106
9.5.2	Functional annotation of piRNAs signature targets	107
9.6	3C-seq Quality control and mapping statistics	109
9.6.1	3C-seq library quality	109
9.6.2	3C-seq iterative mapping and iterative correction	112
9.7	Three dimensional landscape of neuronal and non-neuronal populations	113
9.7.1	Chromatin interaction patterns in the two types of CA1 cell population	113
9.7.2	TADs in neuronal and non-neuronal population	114
9.7.3	Local interaction patterns in chromosome 9 (mm9) for neuronal and non-neuronal populations	115
9.7.4	Interactions pattern near the BACE1 gene	117
10.	Discussion.....	119
10.1	miRNAs and piRNAs as biomarker for Alzheimer’s disease	119
10.1.1	Characterization of CSF exosomes.....	121
10.1.2	Characterization of NGS analysis and small noncoding RNAs ..	123

10.1.3	Role of clinically established CSF biomarkers for Alzheimer's disease diagnosis	125
10.1.4	The CSF miRNA/piRNA signature as biomarker for Alzheimer's disease diagnosis.....	126
10.1.5	The predictive power of CSF miRNAs/piRNAs signature from MCI to Alzheimer's disease	128
10.1.6	Diagnostic properties of CSF miRNA/piRNA signature for plasma and brain tissue samples	129
10.1.7	Functional characterization of CSF miRNAs/piRNAs signature	130
10.2	Three dimensional architecture of neuronal and non-neuronal cells.....	131
10.2.1	3C-seq library quality assessment, reads alignment and balancing	132
10.2.2	The conformational differences and identification of TADs in the neuronal and non-neuronal CA1 populations	132
10.2.3	Interaction pattern differences on a single chromosome (chr9) and near a single gene (BACE1)	133
11.	Conclusions	134
12.	References.....	137

2. Thesis abstract

Alzheimer's disease (AD) is the most prevalent form of dementia that has vast emotional and economic implications in our society. There is no cure for this neurodegenerative disorder as the pathological changes occur years before the manifestation of the clinical symptoms. Thus, there is a great need for the development of effective and non-invasive biomarkers allowing the identification of patients at risk. During my Ph.D., I used next generation sequencing to study the small noncoding RNAome in the exosomes derived from cerebrospinal fluid (CSF), the majority of which are microRNAs (miRNAs) and piwi-interacting RNAs (piRNAs). Statistical and machine learning methods were able to identify putative miRNAs and piRNAs signature that can classify AD and controls with an AUC of 0.83. The piRNAs signature was suitable to predict conversion of patients suffering from mild cognitive impairment (MCI) to AD with an AUC of 0.86. The putative signature performed even better in the brain region with an AUC of 0.89 suggesting that we can use the smallRNAs signatures to perform a good diagnosis and prognosis between AD and controls.

To better understand the mechanism that disrupts the human homeostasis leading to several neurodegenerative disorders, in a pilot study, I looked into the dynamic changes in higher order chromatin structure that control gene expression programs in synaptic plasticity, memory function, and neurodegenerative disorders by the use of Chromosome Conformation Capture (3C) based technique (3C-seq). One finding was related to the hallmark of AD (A β plaques). There was a preference of looping interactions involving BACE1 gene (initiates the A β generation that leads to the formation of A β plaques) in the neuronal population compared to the non-neuronal population. The results, however, for this pilot study should be interpreted cautiously due to small sample size and availability of the low resolution data.

My study thus aims to provide further evidence that circulating small noncoding RNAs could be a suitable biomarker to detect the Alzheimer's disease. As these small noncoding RNAs are extremely stable both longitudinally and during the experimental procedures, they make excellent candidates for biomarkers for the prediction of the disease onset. The study also focuses on standardization and replication of the results by providing an open source access to the statistical and machine learning pipelines that were developed during the course of this study. This work also provides new insights to the genome stability, functions and the underlying mechanisms that are responsible for the correct gene expression in the genome and disruption of which causes these neurodegenerative disorders.

3. Acknowledgements

Alzheimer's disease (AD) is a topic that has always remained so close to my heart. The irreversible, progressive nature of the disease made its first big impression when my grandfather was diagnosed with this brain disorder that slowly destroyed his memory and thinking skills. I was young at that time and was made to understand that there is no cure for this disease. The emotional and physical suffering associated with this disorder gave me the push I needed when I decided on shaping my bioinformatics career. It would not have been possible for me to convert this personal aspiration to professional research that can potentially help identify and manage this dreadful disorder, had it not been for Andre Fischer, my thesis advisor. He believed in my ability to channel my personal curiosity to scientific research that can help thousands of people in the future. I thank him for giving me the freedom and support to shape my research and actively guiding my projects to achieve quantifiable results.

I also want to thank Prof. Dr. Göpfert and Dr. Dr. Schlüter for exciting discussions and critical yet positive and honest feedback during my thesis committee meetings. My gratitude goes to all members of my extended examination board Prof. Dr. Tiago Fleming Outeiro, Prof. Dr. Melanie Wilke and Prof. Dr. Lutz Walter for being present for my defense.

I would also like to thank Dr. Farahnaz Sananbenesi for her belief in me and guiding me through all the stressful times and always smiling and being positive. It was very helpful.

A big thanks goes to Ulrike and Daniel, who has been since the beginning of our association with Germany were the most supportive, kind and sincere people I have come to know. All their tips, suggestions, prompt paperwork, and most importantly, just their presence alone was really helpful. They really helped me with every step here in Göttingen.

This thesis would not have been possible without the generous support of old and current Fischer lab team members Susi, Nicole, Christian, Lalit, Henning, Robert, Sasha, Rezaul, Carmelo, Sakib, Cemil, Tonatiuh, Vakhtang and, especially, Eva, Rashi, Tea, Magda, Hendrik, Maryam, Jonas and Vincenzo for being there and supporting me immensely in my research. Magda and Hendrick you guys really helped me a lot in finalizing my thesis. I am immensely indebted for your timely help and support.

An extended thanks to my discussion partners Dr. Anna-Elisa Roser, Lucas A. Caldi Gomes and Melanie Nuesch Germano, who from time to time, gave me ideas that helped me develop key modules for my pipelines.

I would also like to extend my thanks to my previous lab team members, Dr. Amartya Sanyal and Dr. Noam Kaplan, for channelling my interest with the right biological and machine learning techniques. Above all, Dr. Job Dekker for introducing me to Chromatin Conformation Capture and techniques like 3C, 5C and HiC as well as new and exciting field of Epigenetics. Bryan for giving positive encouragement and having discussions on how to analyze data effectively, faster and better. Johan for sharing peculiarities of living in Europe even before we thought about moving to Germany.

I would like to thank all my friends in Goettingen – Rashmi, Abhay, Chaitanya, Aishwarya, Ashish, Amrita, Sneha, Eric, Orr, Ellen, Raza, and Abhi for helping us transition smoothly in this country and supporting us in times of need. I would also like to thank Lynel and Dennis for being wonderful friends.

In the end, I would like to thank my parents, sister, in laws, wife (Gitanjali) and kids (Riya and Naitik) for being patient with me in times of stress. It is their dream, their hard work, their trust and support in me that I am standing at this place. The words are not enough for what they have sacrificed for me. I am eternally grateful to them.

4. List of Figures and Tables

Figure 6.2-1 Brain shrinkage. The MRIs shows reduction in hippocampal region responsible for memory. Brain with normal cognitive function (Left) and suffering from Mild Cognitive Impairment (MCI) on the right. Inset shows the an overview of the right hippocampal region.....	24
Figure 6.2-2 Axial proton density MRI showing white matter lesions and rating. A. Normal aged matched individual used as a control. B. Individual suffering from Vascular Dementia. C. An individual suffering from Lewy Body Dementia (Adapted from Barber et.al. [26])	27
Figure 6.2-3 Types of Dementia	27
Figure 6.2-4 Estimated yearly cases of Alzheimer's disease. Source World Alzheimer Report (2016)	28
Figure 6.2-5 Amyloid plaques and tau neurofibrillary tangles in a human Alzheimer's disease brain.....	29
Figure 6.2-6 A model of progression of Alzheimer's disease. A hypothetical model showing the progression of a normal individual from healthy to Alzheimer's disease with the decline in cognition. The solid line representing aging shows a normal decline in cognitive functions compared to a demented individual (represented by dotted lines)	30
Figure 6.3-1 Exosome Electron micrograph. Exosomes secreted from Epstein–Barr virus-transformed B cell and multivesicular bodies (MVB). The MVBs content is either degraded in lysosomes or more exosomes are released after their fusion with the cellular membrane. Adapted from Edgar et. al. [61]	32
Figure 6.3-2 Schematic of exosomes biogenesis. The surface receptors buds to smallRNAs and proteins in cytoplasm ILVs after budding with MBVs. From there they can either release exosomes or get degraded in lysosomes. Adapted from Schorey et. al. [64].....	33
Figure 6.3-3 A β proteins clearance through exosomes. Amyloid- β precursor protein (APP) gets cleaved to form Amyloid- β (A β) proteins. They are then	

released into the extracellular space through exosomes. Adapted from Yuyama, K et.al. [71].....	34
Figure 6.5-1 Epigenetics landscape in health and disease.....	37
Figure 6.5-2 Key epigenetics mechanisms (Adapted from Fischer et.al. [105]).....	38
Figure 6.5-3 miRNA biogenesis (Adapted from Devaux et.al. [121])	40
Figure 6.5-4 The biogenesis of piRNAs (Adapted from Watanabe & Lin et.al. (2014) [125])	41
Figure 6.6-1 A typical machine learning application used in biological data analysis (Adapted from Camacho et.al. (2018) [133]).....	42
Figure 6.6-2 Schematic diagram of internal working of RF algorithm	47
Figure 6.6-3 Relationship between a machine learning algorithm's performance and the number of features	48
Figure 6.6-4 Modified relationship between a machine learning algorithm's performance and the number of features	49
Figure 6.7-1 Common organizational structures in the genome adapted from Szalaj et.al (2018) [199].....	52
Figure 6.7-2 Chromosome territories (Adapted from Bolzer et.al. (2005) [203]).....	53
Figure 8.3-1 The smallRNA mapping pipeline (GJSrMap) [253].....	61
Figure 8.3-2 Sequencing quality control, mapping and smallrnaome distribution.....	65
Figure 8.4-1 Sample filtering, feature normalization and filtering.....	66
Figure 8.4-2 Various normalization strategies	67
Figure 8.5-1 Feature selection	69
Figure 8.5-2 miRNAs and piRNAs	70
Figure 8.5-3 Optimal model selection and performance evaluation.....	73
Figure 8.7-1 Genomic location of primers and restriction enzymes	77

Figure 8.7-2 PCR amplification curve	78
Figure 8.7-3 Melting curve and 3C interaction on a gel	79
Figure 8.9-1 3C-seq mapping artifacts (Adapted from Fides Lay's HiC workshop [291])	83
Figure 8.9-2 Schematic to identify topologically associating domains (TADs) by two competing methods.....	85
Figure 9.1-1 Profile of exosomal RNA from human CSF A. Exosomes isolated from human CSF was analyzed via EM. B. Fragment size using a nanosight instrument and via immunoblot for exosomal marker proteins. C. Electropherogram showing the profile of RNA isolated from exosomes. D. Electropherogram showing the profile of RNA isolated from exosome free CSF. E. Electropherogram showing the profile of RNA isolated from exosomes treated with DNAase (F) and RNAase (G).....	87
Figure 9.1-2 miRNAome and piRNAome in CSF exosomes	88
Figure 9.1-3 Heat map showing expression values of the 3-p and 5-p arms of all miRNAs detected in human CSF and in the human cortex (Brodmann Area 9)	89
Figure 9.1-4 Pearson correlation between miRNA and piRNA expression values of hippocampal and cortical neurons against the miRNA and piRNA expression in the exosomes released from these cells. Here x-axis represents the normalized expression of exosomal small RNA and the y-axis represents the normalized expression of cellular small RNA. The Pearson correlation coefficient and significance level are mentioned in the plot	91
Figure 9.2-1 showing summary statistics about the clinical markers used in the study. Abbreviations: Alzheimer's disease, Alzheimer's disease; A β 40 and A β 42, amyloid beta peptide 40 and 42 amino acids; pTAU is phosphorylated TAU. B. Receiver operating characteristic (ROC) plot was obtained during the performance testing of the clinical markers on replication cohort 2 with mean test AUC of 0.87. The training was done on cohort 1 with a tenfold cross validation. An inset plot showing the variable importance which explains how good a variable in classifying the data. Here x-axis represents the mean decrease in Gini and the y-axis shows each variable	

which are ordered top-to-bottom as most- to least-important. In this case, pTAU is most informative. 92

Figure 9.2-2 Identification and quantification of small non-coding RNAs signature 94

Figure 9.2-3 Heatmaps showing Pearson correlation coefficient between the normalized expression of putative biomarker signature and levels of clinical markers in cohort 1 (top) and in the replication cohort 2 (bottom). The Pearson correlation coefficients and a two-sample paired t-test were calculated and the significance is marked as (*) with p-value < 0.05. 96

Figure 9.2-4 Performance of signature on replication cohort. A. ROCs showing the performance of miRNA signature on separate replication cohort (cohort 2). The training was done on cohort 1 with tenfold cross validation. Mean AUC of 0.56. Barplot (right) shows the variable importance explaining how good a variable is in classifying the data. Here x-axis represents the mean decrease in Gini and the y-axis shows each variable which are ordered top-to-bottom as most to least important. In this case, hsa-miR-30a-5p is most informative. B. ROCs showing the performance of piRNA signature on separate replication cohort (cohort 2). Mean AUC of 0.82 was obtained. Barplot (right) show the variable importance explaining how good a variable is in classifying the data. Here x-axis represents mean decrease Gini value and the y-axis shows each variable which are ordered top-to-bottom as most- to least-important. In this case, hsa_piR_019949 is most informative. C. ROCs showing the performance of combined miRNA and piRNA signature on separate replication cohort (cohort 2). Interestingly the mean AUC of 0.83 was obtained suggest that the combined signature performed way better than the miRNA signature and marginally better than the piRNA signature. An inset plot showing the variable importance which explains how good a variable in classifying the data. Here x-axis represents the mean decrease in Gini and the y-axis shows each variable which are ordered top-to-bottom as most- to least-important. In this case, hsa_piR_019949 is most informative. D. ROCs showing the performance of combined clinical markers and smallRNA signature on separate replication cohort (cohort 2). The mean AUC of 0.98 was obtained showing that the combined signature performed massively better than the individual clinical markers or the smallRNAs signature 97

Figure 9.3-1 A. Barplot (left) showing the expression of the putative signature in training cohort 1 and predictive replication cohort 3. The expression of mciStable samples was scaled to one and the expression of mciAD samples was scaled appropriately. Dark color shows control/mciStable samples and light shade of the same color represents the Alzheimer's disease/mciAD samples. Three signatures are downregulated in mciAD compared to mciStable and remaining three signatures are unregulated.....	100
Figure 9.3-2 Performance of signature on predictive replication cohort.....	101
Figure 9.4-1 Barplots showing the expression of the putative signature in training cohort1, plasma exosomes cohort 4 and PFC brain cohort 5. The expression of control samples was scaled to one and the expression of Alzheimer's disease samples was scaled appropriately. Dark color shows control samples and light shade of the same color represents the Alzheimer's disease samples.....	103
Figure 9.4-2 Performance of signature in blood and brain samples	104
Figure 9.5-1 Signature miRNAs-gene targets regulatory network	106
Figure 9.6-1 Quality score distribution over all sequences and duplicate sequences.....	111
Figure 9.6-2 Overrepresented Sequences.....	112
Figure 9.7-1 Chromatin interaction contact maps for non-neuronal and neuronal population.....	114
Figure 9.7-2 Number of Topologically Associating Domains (TADs) for neuronal and non-neuronal populations.....	115
Figure 9.7-3 Figure 9.7-4 Local interaction patterns in chromosome 9 (mm9) for neuronal and non-neuronal populations.....	116
Figure 9.7-5 CA1 interactions looping profiles in neuronal and non-neuronal population....	117
Figure 9.7-6 The interaction pattern in the vicinity of the BACE1 gene	118
Table 6-1 A small description of selected list of neurodegenerative disorders	20

Table 6-2 Common classification metrics [181-183]	49
Table 6-3 Common regression metrics [185-189].....	50
Table 8-1 Description of small noncoding RNAs using in the mapping	61
Table 8-2 The ligation master mix.....	75
Table 8-3 3C primers details.....	76
Table 8-4 Reaction usage.....	77
Table 8-5 PCR tubes	77
Table 8-6 Standard curve range	78
Table 9-1 showing demographic information for the samples included in the biomarker signature analysis. Abbreviations: AD, Alzheimer’s disease; F, female; M, male.	91
Table 9-2 Showing demographic information for the samples included in the biomarker signature analysis. Abbreviations: mciAD, mild cognitive impairment patients that went on to develop Alzheimer’s disease; mciStable, mild cognitive impairment patients that did not develop Alzheimer’s disease; F, female; M, male.....	99
Table 9-3 showing demographic information for the samples included in the biomarker signature analysis. Abbreviations: PFC, Prefrontal Cortex; AD, Alzheimer’s disease; F, female; M, male	102
Table 9-4 Detailed Annotation of CSF piRNAs signature with their Alzheimer’s disease associated alterations	108
Table 9-5 General sequencing statistics.....	109
Table 9-6 Relationship of base call accuracy and Phred quality score.....	110
Table 9-7 Iterative mapping results.....	112

5. List of Abbreviations

3C	Chromosome Conformation Capture
3D	Three-dimensional
5C	Chromosome Conformation Capture Carbon Copy
AGO	Argonaute
APP	Amyloid-Beta Precursor Protein
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operating Characteristic Curves
A β	Amyloid-Beta
BACE1	Beta-Secretase 1
BAM	Binary Alignment Map
CA1	Cornu Ammonis1
CSF	Cerebrospinal Fluid
CSV	Comma Separated Values
CT	Computed Tomography
DNA	Deoxyribonucleic Acid
FACS	Fluorescence Activated Cell Sorting
FTD	Frontotemporal Dementia
ILV	Intraluminal Vesicle
JSON	JavaScript Object Notation
MANCOVA	Multivariate Analysis of Covariance
MCI	Mild Cognitive Impairment
miRNA	MicroRNA
MM9	Mus Musculus 9 (Build 37)
MRI	Magnetic Resonance Imaging
MVB	Multivesicular Bodies
ncRNA	Non-coding RNA
NGS	Next Generation Sequencing
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PET	Positron Emission Tomography
pimiRNA	Piwi-interacting/Micro-RNA
piRNA	Piwi-Interacting-RNA
pTAU	Phosphorylated TAU
QC	Quality Check
qPCR	Quantitative Polymerase Chain Reaction
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic Curves
SAM	Sequence Alignment Map
sncRNA	Small non-coding RNA
TADs	Topologically Associating Domain
TSS	Transcription Start Site
TSV	Tab Separated Values
XML	eXtensible Markup Language

6. Introduction

6.1 Introduction to neurodegenerative disorders

6.1.1 Potential classification of neurodegenerative diseases

Neurodegenerative disorders belong to the category of disorders and diseases that are mainly associated with the neurons. They are identified by the escalating loss of neuronal structure and/or functions mostly in the brain and spinal cord [1, 2]. At present there are various treatment strategies that had seen different levels of success, but there are no cures for these disorders. These neurodegenerative disorders can broadly be classified into two categories on the basis of their effects. They can either affect the movements of the muscles or can cause cognitive decline [3, 4]. A selected list of few neurodegenerative disorders are mentioned in Table 6-1.

Table 6-1 A small description of selected list of neurodegenerative disorders

Disorders	Description
Alzheimer's disease (AD)	<ul style="list-style-type: none">• The most common causes of dementia and characterized by the accumulation of amyloid plaques and neurofibrillary tangles• Symptoms include memory and cognitive decline• Disease onset starts around age 40
Schizophrenia (SCZ) [5]	<ul style="list-style-type: none">• Psychotic disorder resulting in neurodegeneration• Symptoms include hallucinations, delusions and disorganized behavior• Disease onset starts in the early 20s
Bipolar disorders (BD)	<ul style="list-style-type: none">• Majorly identified by extreme mood swings• Symptoms range from low depressive episodes to high maniac episodes• Disease onset starts around age 25
Prion disease (CJD)	<ul style="list-style-type: none">• Also known as Creutzfeldt-Jacob disease and mainly characterized by amyloid plaques and

	<p>spongiform</p> <ul style="list-style-type: none"> • Symptoms include fast progressive dementia and changes in behavior • Usual disease onset is around age 60
Amyotrophic Lateral Sclerosis (ALS)	<ul style="list-style-type: none"> • Also known as Lou Gehrig's disease and mainly characterized by progressive degeneration of the muscles which results in paralysis • Symptoms include muscle weakness and paralysis • Disease onset starts in early teenage
Parkinson's Disease (PD)	<ul style="list-style-type: none"> • Chronic progressive disorder characterized by bradykinesia, tremors, impairment of postural balance and rigidity • Apart from above mentioned characteristics, symptoms also include stress, anxiety, depression, memory loss and dementia • Usual disease onset is around age 60
Spinocerebellar ataxia (SCA)	<ul style="list-style-type: none"> • An autosomal dominant neurodegenerative disorder and mainly caused by ATNX1 gene mutation on chromosome 6 • Symptoms include spasticity, dysarthria, gait difficulty and loss of coordination
Dystonia	<ul style="list-style-type: none"> • Characterized by agonizing and extended muscle contractions causing irregular postures and movements • Major symptoms include muscle cramps and spasms • Disease onset starts as early as the age of 12
Multiple System Atrophy (MSA)	<ul style="list-style-type: none"> • Also known as "Shy Drager", "Striatonigral degeneration" or "Olivopontocerebellar Atrophy" • Symptoms mainly include Parkinsonism, autonomic failure, and cerebellar findings • Disease onset starts around age 40
Huntington's Disease (HD)	<ul style="list-style-type: none"> • An autosomal dominant progressive neurodegenerative disorder caused by a defect

	<p>in HTT (Huntington) gene on chromosome 4</p> <ul style="list-style-type: none"> • Symptoms include impaired gait, involuntary movements or rigidity in the muscles and cognitive decline • Disease onset starts around age 40
Rett Syndrome	<ul style="list-style-type: none"> • A neurodevelopmental disorder primarily affecting girl child • Symptoms include diminished motor skills, chorea, microcephaly • Disease onset starts around age of 6 months
Progressive Supranuclear Palsy (PSP)	<ul style="list-style-type: none"> • Also known as Steele-Richardson-Olszewski syndrome affecting stance and cognitive abilities • Symptoms include progressive hindrance of speech and swallowing, limited mobility and imbalance • Disease onset starts around age 55
Tourette Syndrome	<ul style="list-style-type: none"> • Mainly characterized by involuntary tics, twitching and uncontrollable vocal outbursts • Symptoms include motor and vocal tics and do not stop during sleep • The disease onset typically starts around age 7 years

6.1.2 The partnership of aging and neurodegeneration

The advancement in modern medicine has increased the average lifespan with an unwanted side effect in the form of neurodegenerative diseases. These diseases are a major concern all over the world as they are one of the leading causes of disability in the elderly and ultimately lead to death. With the aging, the human brain starts to aggregate and accumulate abnormal and misfolded proteins causing lesions to the affected brain regions. With the progressive loss of neurons over the course of many years, it is very hard to estimate the real reason for the cell deterioration and ultimately the death of the neurons. Some important neuronal cell death mechanisms include Autophagy, apoptosis, excitotoxicity, and necrosis [6]. It is still a big challenge to understand how aging only affects the growth of the neurons and can be distinguished solely from the neurodegeneration.

6.1.3 Effects of neurodegeneration: The big picture

The worldwide threat to human health from these disorders is increasing every day disturbing the lives of the patients, families and the community in general. Almost all neurodegenerative disorders are incurable to date and a huge amount of resources are being allocated to find the cure for these diseases. Developing a deep understating of the pathology and onset of disease progression is very important. For example, use of animal model systems to understand the working of certain disorders can play a major role in identifying key mechanisms of these neurodegenerative disorders. The final aim is to develop biomarkers that can identify, classify and ultimately predict the onset of the disease progression.

6.2 Overview of dementia

6.2.1 Mild Cognitive Impairment (MCI)

6.2.1.1 MCI: normal aging or dementia?

The decline of cognitive abilities due to aging remains an increasing problem in today's world. Further, considering everyone with a diminished cognition to be diagnosed as an early form of Alzheimer's disease (Alzheimer's disease) is misleading. With the improvement in the disease diagnosis, it is easier to classify most of these cases as mild cognitive impairment (MCI) [7]. In MCI, a person is suffering from diminished cognition typically related to judgement, thinking, vocal skills and memory is worse than the average age-matched individual. The individuals with MCI may lead a relatively normal life as these symptoms are not enough to be classified as dementia [8]. However, there is a decent chance that an individual suffering from MCI may go on to have dementia. Proper care at this stage in term of lifestyle may help individuals to keep the symptoms same or even get better [9].

6.2.1.2 Symptoms of MCI

An individual suffering from MCI display some or all of the below-mentioned symptoms [10, 11]:

- Memory complains which relates the diminishing of memory. This is mostly related to forgetting events like appointments
- Diminishing of communication skills
- Showing poor judgement and becoming more spontaneous

- Trouble keeping up a line of thoughts and taking extra time during the decision making

All of the above mentioned symptoms are worse compared to normal aging but not enough to be classified as dementia [12].

6.2.1.3 Potential causes of MCI

There are multiple causes to MCI some of which are treatable. It becomes more challenging to identify the real cause of MCI as the symptoms may improve over time or stay as it is or may deteriorate to dementia. As most of the causes of MCI are not understood, some of them have shown an improved understanding. Most of these causes are similar to dementia [13]. These include accumulation of amyloid plaques, Lewy bodies and misfolded protein tangles, neuronal injuries (motor or sensory signs), decreased levels of the CSF and shrinkage of the brain region mainly hippocampus, which is associated with episodic memory [14]. Figure 6.2-1 [15] left shows the normal hippocampal region while the shrunken hippocampus related to MCI is shown in the right.

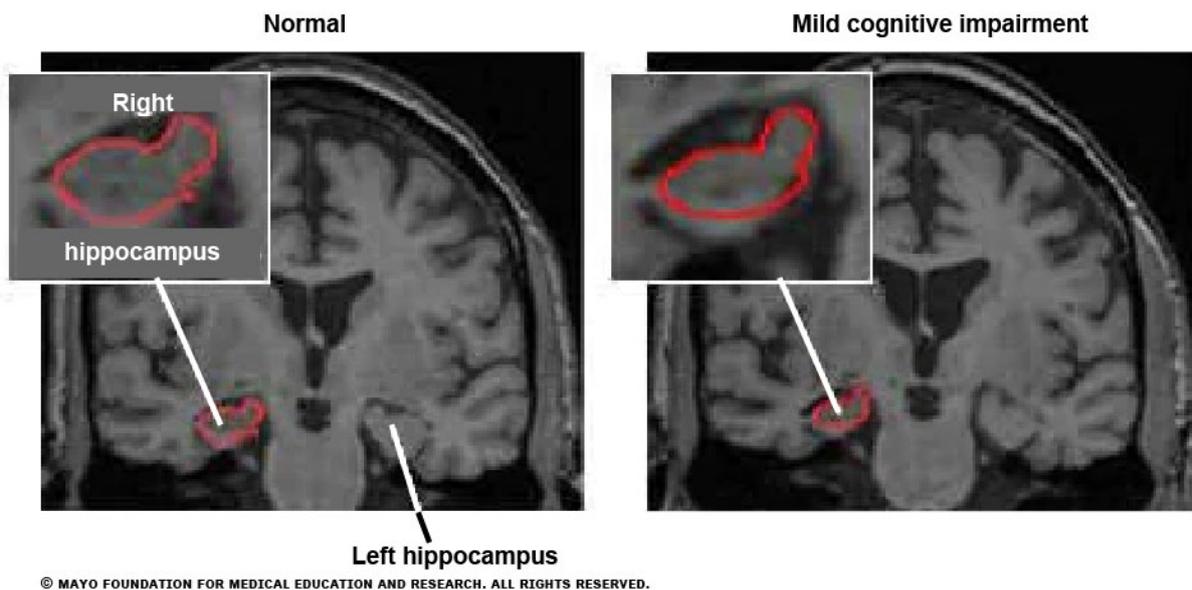


Figure 6.2-1 Brain shrinkage. The MRIs shows reduction in hippocampal region responsible for memory. Brain with normal cognitive function (Left) and suffering from Mild Cognitive Impairment (MCI) on the right. Inset shows the an overview of the right hippocampal region.

6.2.1.4 Clinical criteria for the diagnosis of MCI and treatment

Individuals suffering from MCI are difficult to diagnose as there are no specific protocols defined. However, there are certain clinical guidelines that are defined and followed by clinicians all over the world to perform the MCI diagnosis [16]. It is also important that these

guidelines also distinguish individuals suffering from MCI to dementia and Alzheimer's disease [17]. These guidelines include:

- A decline in cognitive performance over time
- Memory and other reasoning abilities are impaired
- A lookout for brain shrinkage, tumors and strokes using brain imaging techniques like MRI or CT scan
- All the diagnosis of dementia and Alzheimer's disease should be negative
- Activities of daily living (ADL) should be intact

There are currently no treatments available for MCI. However, few medications that are used for dementia and Alzheimer's disease are used for MCI patients, but they do not always work. There is major research going on the development of biomarkers for MCI and a better understanding of MCI will definitely lead to a better understanding of dementia and Alzheimer's disease [12].

6.2.2 Dementia

6.2.2.1 Dementia – Information and Statistics

Dementia can be broadly stated in terms of mental decline of an individual that negatively affects their life on a day to day basis [18]. In order to classify an individual to have dementia, there should be significant impairment in their mental abilities, communication skills, and long & short term memory. Dementia affects about 50 million people worldwide with an annual increase of 7.7 million cases every year and the number of people with dementia are expected to triple by the year 2050. The economic burden caused by dementia is expected to be around 818 billion USD per annum and is the 7th leading cause of death worldwide [19, 20].

6.2.2.2 Symptoms of Dementia

Dementia cannot be classified as a disease in clinical terms, but can be referred to as an umbrella term for a set of symptoms that collectively describes the decline in the cognitive abilities of a person that impairs their daily activities [21, 22]. Some common symptoms include:

- Memory loss, disturbing daily life
- Consistently displaying poor judgement
- Irregular moods and behavior
- Communication and language disturbances

- Overall increase in disorientation and confusion
- Inability to show proper reasoning and thinking

The above mentioned symptoms are generally known as primary symptoms and often followed up with more in-depth and through check-ups in the clinics [23].

6.2.2.3 Dementia - Types and Causes

Dementia is majorly caused by loss of neurons in the brain. This affects the thinking, judgement and emotional balance of the individual [24]. Dementia is mainly classified into five subtypes [25]. They are:

- Alzheimer's Disease (Alzheimer's disease) which is the most common form of dementia showing memory, language and spatial decline
- Vascular Dementia (VaD) also known as multi-infarct dementia experiencing an abrupt onset and associated with vascular risk factors like stroke
- Lewy Body Dementia (DLB) with individuals experiencing hallucinations
- Fronto-Temporal Dementia (FTD) exhibiting behavior and personality changes
- Other Dementias associated with other neurodegenerative diseases like Parkinson's disease (PD), Huntington's disease (HD), Creutzfeldt-Jakob disease, Progressive supranuclear palsy, Prion diseases etc.

Figure 6.2-2 shows the MRI of the brain from different individuals who observed the changes in the white matter of three different individuals, one suffering from Vascular Dementia, one suffering from Lewy Body Dementia and an age-matched control. The changes progressed to worse from the control to Vascular Dementia [26].

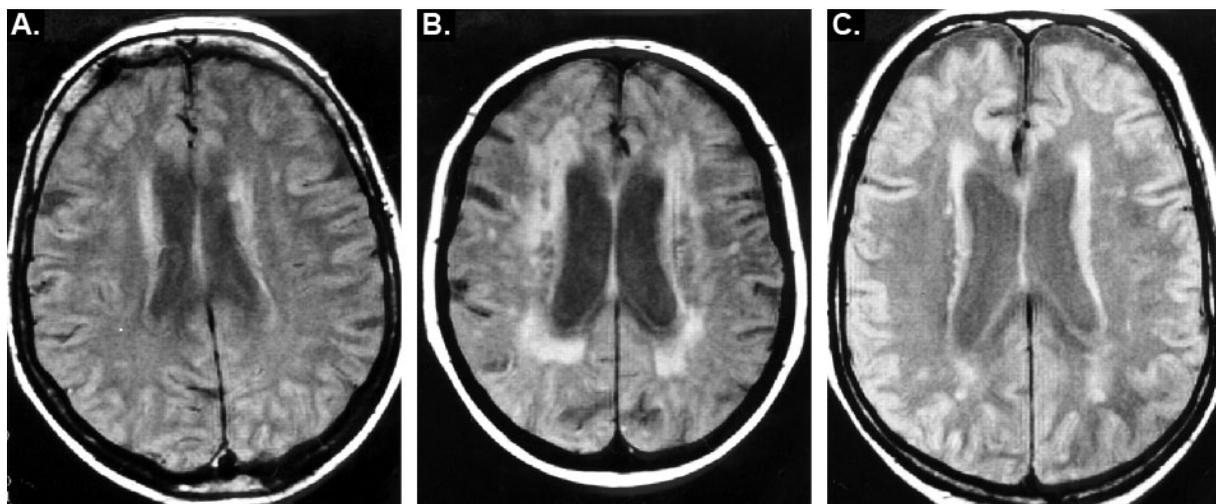


Figure 6.2-2 Axial proton density MRI showing white matter lesions and rating. A. Normal aged matched individual used as a control. B. Individual suffering from Vascular Dementia. C. An individual suffering from Lewy Body Dementia (Adapted from Barber et.al. [26])

Figure 6.2-3 displays the most common forms of dementia. Alzheimer's disease accounts for almost two-thirds of all forms of dementia, contributing approximately 11.6% of all recorded deaths worldwide [25, 27].

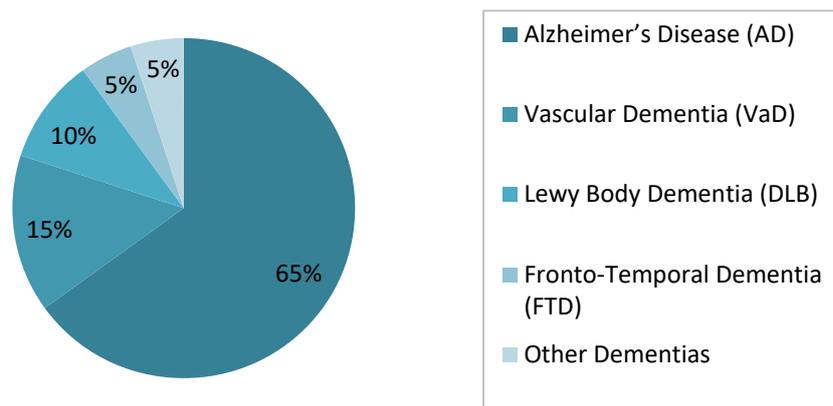


Figure 6.2-3 Types of Dementia

6.2.3 Alzheimer's Disease (AD)

6.2.3.1 An overview of Alzheimer's disease

As mentioned in the previous section 6.2.2.3 above, Alzheimer's disease is the most common cause of dementia, accounting for almost two-thirds of dementia cases. It is a progressive neurodegenerative disorder where the symptoms get worse and worse over time. Figure 6.2-4 shows the progression of Alzheimer's disease patient over time and distinguishes between normal aging and Alzheimer's disease. Approximately 45 million individuals are suffering from Alzheimer's disease worldwide with one new case of dementia is occurring every 3.2 seconds and expected to be around 131 million by mid of the 21st century [28] (Figure 6.2-4).

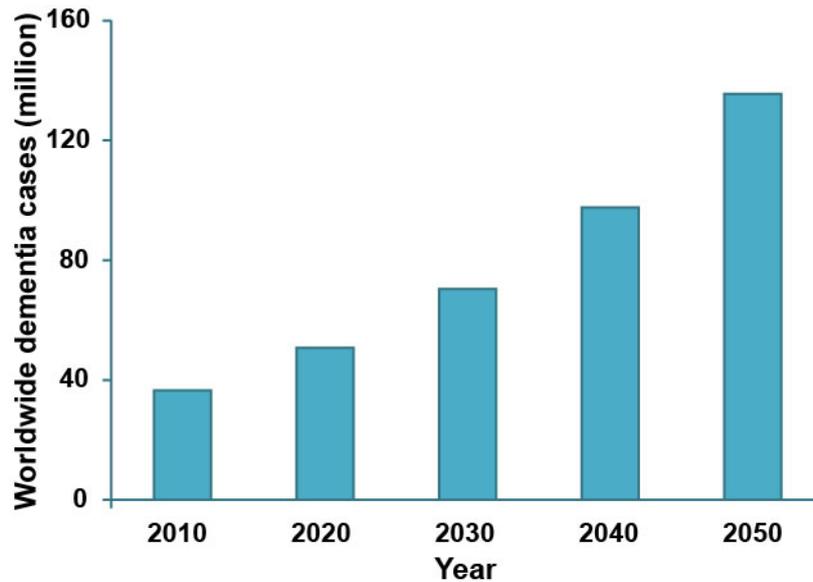


Figure 6.2-4 Estimated yearly cases of Alzheimer's disease. Source World Alzheimer Report (2016)

6.2.3.2 Disease pathology of Alzheimer's disease

There are two established pathological hallmarks of Alzheimer's disease shown in Figure 6.2-5 [29]. They are:

- β -amyloid ($A\beta$) plaques also known as senile plaques (SPs) are the extracellular deposits of beta-amyloid peptides. These $A\beta$ deposits are toxic to neurons as they cause loss of long term potentiation (LTP) and leads to neuronal death
- Neurofibrillary Tangles (NFTs) also known as neuritic plaques or neuropil threads are characterized as intracellular accumulation of the insoluble hyperphosphorylated paired helical filaments of tau proteins that are associated microtubules

$A\beta$ deposits are very specific to Alzheimer's disease and characterized to be the primary cause of Alzheimer's disease, but their deposition amount does not correlate well with increasing dementia. Accumulation of tau proteins which second pathological hallmark of Alzheimer's disease is also observed in other neurodegenerative diseases like FTD and prion disease. Hence, they were considered secondary to $A\beta$ deposits. However, their accumulation amount correlated well with increasing dementia and became the foundation of Braak and Braak staging system of Alzheimer's disease [30].

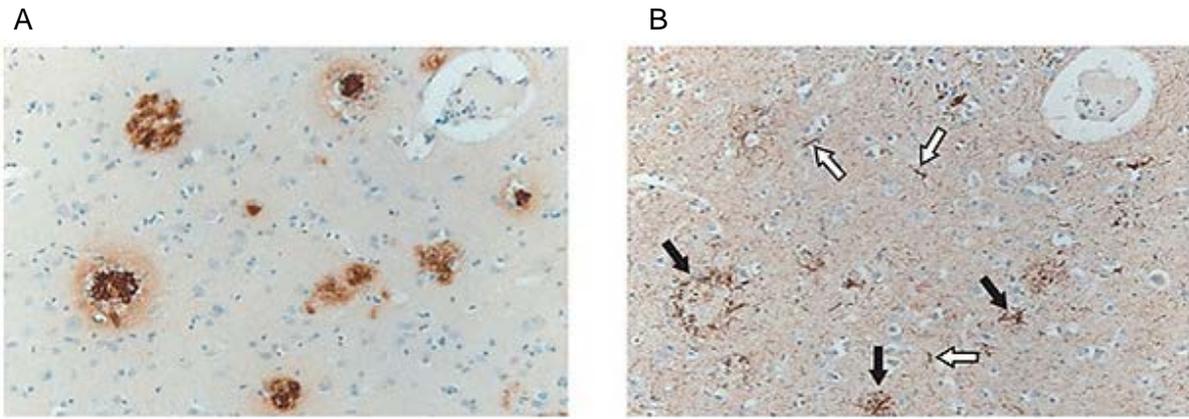


Figure 6.2-5 Amyloid plaques and tau neurofibrillary tangles in a human Alzheimer's disease brain.

The histopathology shows the A β plaques and neurofibrillary tangles in cortical tissue of an Alzheimer's disease suffering individual. A. Brown stains (spherical lesions) show amyloid- β plaques. B. Neurofibrillary tangles are shown by black arrows and neuropil threads shown by the white arrows. (Adapted from Nicoll et.al. [29])

6.2.3.3 Stages and Symptoms of Alzheimer's disease

As Alzheimer's disease is the most common form of dementia, its symptoms are also the same to start with. They include loss of memory and disorientation, having difficulty with communication and rational thinking, displaying poor judgement and inability to learn new skills [31]. Alzheimer's disease is a progressive neurodegenerative disorder and the disease progression occurring in various stages (Figure 6.2-6).

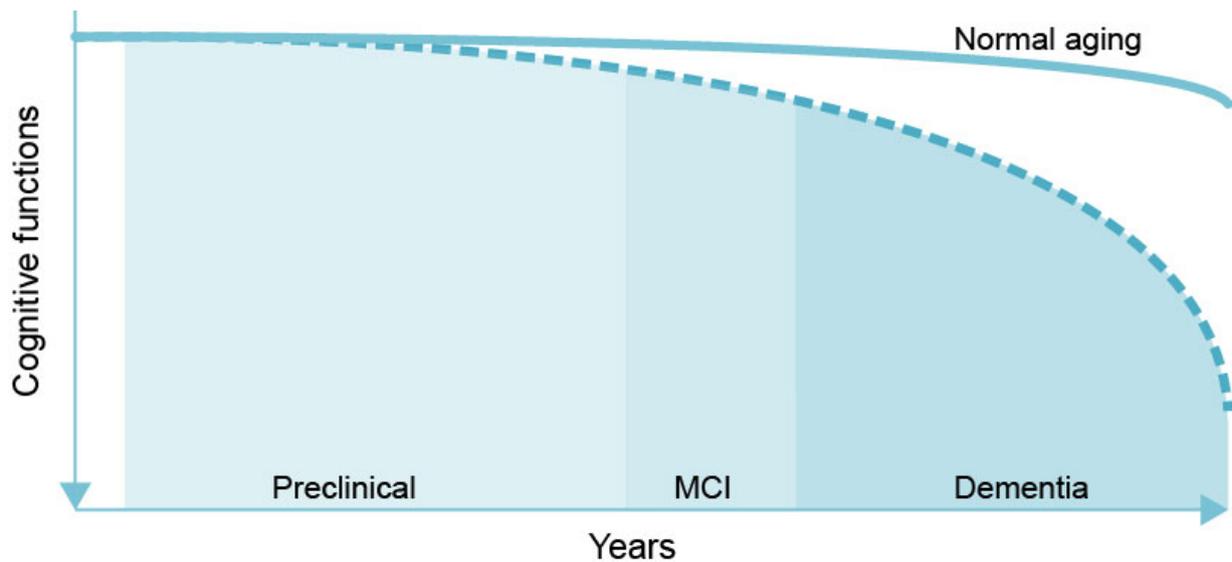


Figure 6.2-6 A model of progression of Alzheimer's disease. A hypothetical model showing the progression of a normal individual from healthy to Alzheimer's disease with the decline in cognition. The solid line representing aging shows a normal decline in cognitive functions compared to a demented individual (represented by dotted lines)

The main stages that characterize Alzheimer's disease are [32]:

- **Preclinical Alzheimer's disease:** There is a distinct lack of any clinical symptoms of Alzheimer's disease at this stage [33]. Although the changes in the pathology have already begun by the accumulation of insoluble proteins. This typically goes on for years without any visible symptoms. The advent of new diagnostic techniques is now helping in diagnosing Alzheimer's disease at this stage. The new Cerebrospinal fluid (CSF) biomarkers and advanced imaging techniques are showing some positive results
- **MCI due to Alzheimer's disease:** Individuals suffering from an MCI start to show the sign of cognitive decline [12]. Although not all individuals go to develop Alzheimer's disease, the ones that do show a consistent progression in the loss of memory and other cognitive tasks starts to get their daily life affected
- **Mild Alzheimer's disease:** This is one of the early stages where Alzheimer's disease is clinically diagnosed as an individual shows worsen symptoms in terms of memory, thinking, judgement, organization skills and behavior. They are still able to act independently, but their daily life activities started to get affected [34, 35]

- **Moderate Alzheimer's disease:** This is the stage where the individual started to show obvious and clear signs of dementia [36]. This stage can last for several years where an individual starts to forget important personal information and events, erratic sleeping patterns, increasing the risk of getting lost, changes in the personality and behavior [37]
- **Severe Alzheimer's disease:** The final stage of Alzheimer's disease where the individual their independence to do daily activities and requires constant care. At this stage, they are unable to respond to their environment, losing awareness of their surroundings, increasing episodes of forgetting family members with difficulty in communicating with them and a steep decline in performing physical activities [38]

The disease onset is progressive and happens gradually over the years. The average life span after the diagnosis of Alzheimer's disease is between 10 to 20 years [39, 40].

6.2.3.4 Diagnosis and treatment of Alzheimer's disease

Like all neurodegenerative diseases, there is no cure for Alzheimer's disease. The treatment can prolong the symptoms and may slow down the decline of cognitive abilities but they cannot stop them. The diagnosis is also improving with time. Doctors and clinicians are now looking at the detailed medical and family history of the individual suffering from Alzheimer's disease in order to get a bigger picture of the individual's development of the Alzheimer's disease symptoms [41]. This may also help them diagnose Alzheimer's disease at an earlier stage and help them slow down the disease progression [42]. Various other diagnostic tools include various laboratory tests for metabolic disorders, physical tests to check muscle strengths, cognitive and behavioral tests to check for mental abilities [43, 44]. The new diagnostic tools involve the imaging techniques like MRI [45], PET [46, 47], CT scans [48, 49] and the use of new Cerebrospinal fluid (CSF) biomarkers to look the levels of beta amyloid plaques and neurofibrillary tangles [50-54]. The role of small non-coding RNAs like miRNAs and piRNAs can also be used as potential diagnostic tools in the future [55-58].

6.3 Exosomes

6.3.1 Overview of Exosomes

Extracellular vesicles (EVs)[59] are released by all cells and historically were called "debris" as their functions were unknown. There are many types of extracellular vesicles that include microvesicles, ectosomes and exosomes. Microvesicles are groups of small vesicles that are budded off from the surface of cells through a budding mechanism. They may be released

containing surface receptors, which can attach to the surface of another cell or may be completely taken up by another cell. Ectosomes are multisized vesicles mostly larger than exosomes that are secreted directly from the plasma membrane into the extracellular space [60]. Exosomes are smaller vesicles compared to the microvesicles and are produced by multivesicular endosomes (Figure 6.3-1) [61] which carries different types of cellular materials [62]. They are also taken up by other cells.

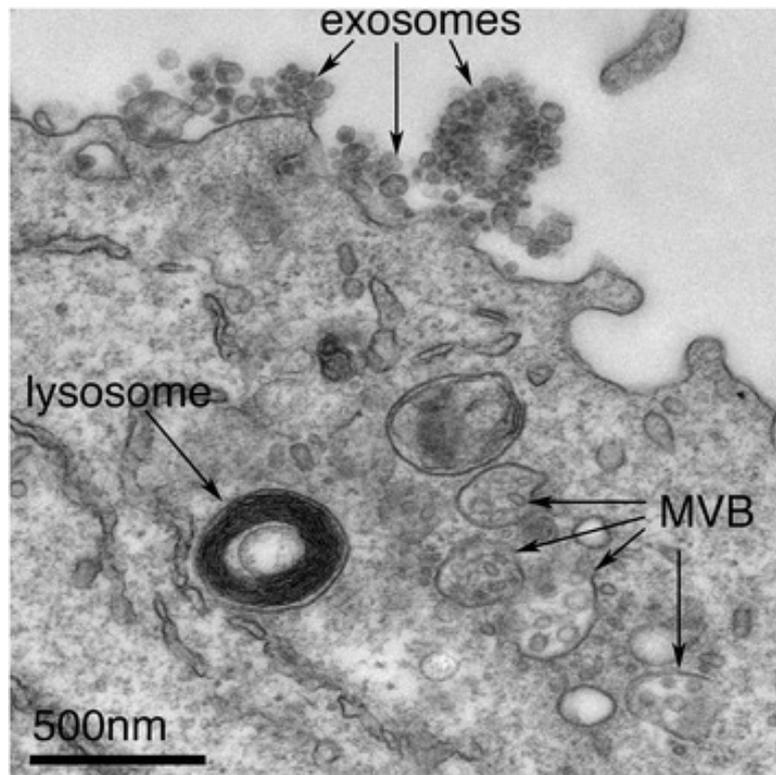


Figure 6.3-1 Exosome Electron micrograph. Exosomes secreted from Epstein–Barr virus-transformed B cell and multivesicular bodies (MVB). The MVBs content is either degraded in lysosomes or more exosomes are released after their fusion with the cellular membrane. Adapted from Edgar et. al. [61]

6.3.2 Biogenesis and functions of exosomes

Exosomes are normally 30-150nm sized vesicles, which are indicated that are derived from the multi-vesicular bodies. Exosomes are generated via endocytic pathways when cell internalizes receptors and deliver them to lysosomes for degradation by proteolysis [63]. The process begins (Figure 6.3-2) [64] when the cell surface receptors containing small peptides ubiquitin which marks them for degradation are fused inward into an endosome that squeezes into the interior of the endosome through invagination that results in a small intraluminal vesicle (ILVs) which goes to build many more of these ILVs and resulting in the formation of multi-vesicular bodies (MBVs). These MBVs primarily known to deliver the ILVs

to the lysosomes where these ILVs were degraded into sugars and amino acids to be used again by the cell. In the past couple of decades, another pathway is known where the MBVs instead of delivering the ILVs to the lysosomes for the degradation, actually travel to the cell surface where they fuse with the plasma membranes and the ILVs were expelled into the extracellular space which are now known as exosomes. The working mechanism of these pathways is still unclear [65]. These exosomes then can be targeted to other cells and tissues and the molecules that are contained within the exosomes are then be internalized by these targets to affect the signaling pathways or expression of the genes.

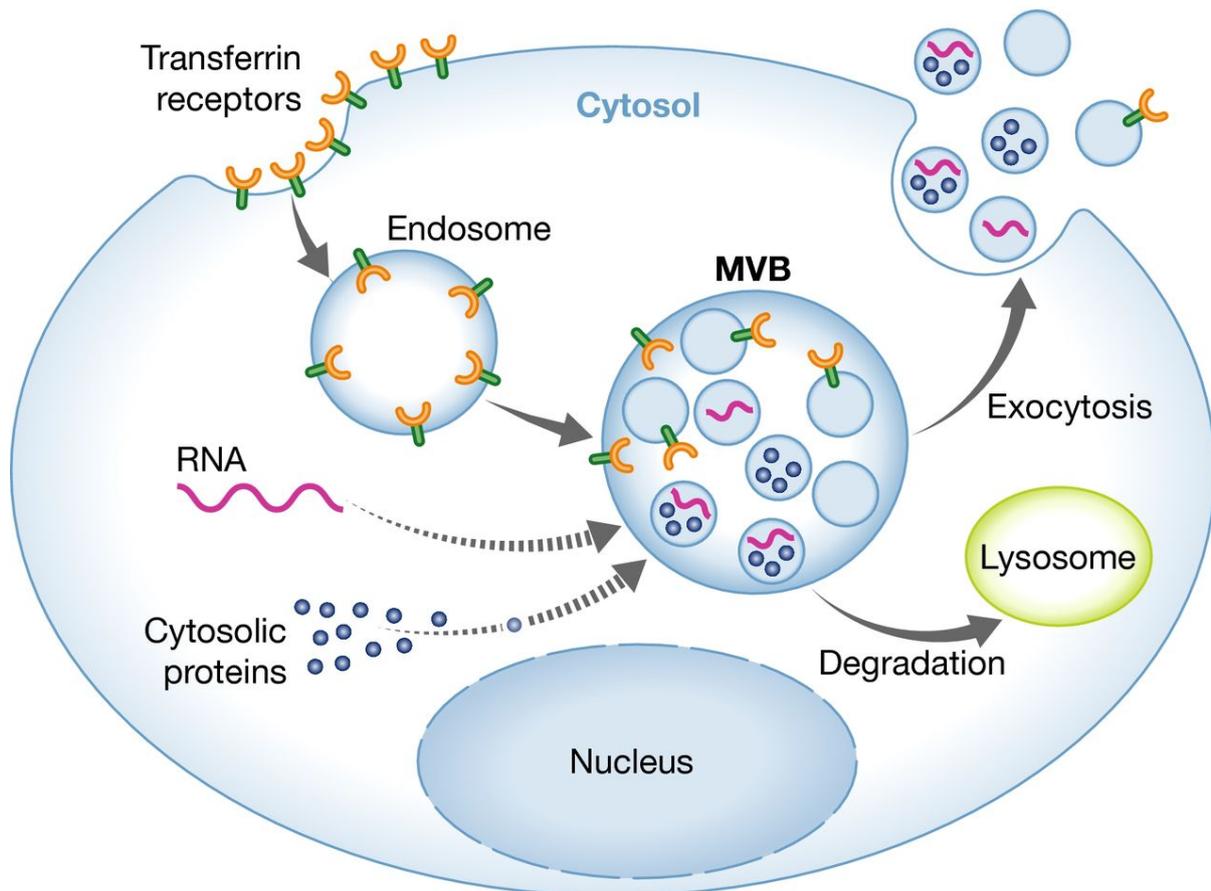


Figure 6.3-2 Schematic of exosomes biogenesis. The surface receptors buds to smallRNAs and proteins in cytoplasm ILVs after budding with MBVs. From there they can either release exosomes or get degraded in lysosomes. Adapted from Schorey et. al. [64]

Many biological functions have been identified for exosomes [63, 65]. They can act as proteins which are expelled out of the cells during some other cellular processes [66]. These also act as clearance mechanisms where unwanted proteins are secreted out of the cells. They have also been known to regulate immune response where their role in coagulation and inflammation are reported [67, 68]. They also function as messengers between cells, especially in neurons. They can also be selected where they are only captured and internalized by certain cell types [69].

6.3.3 Role of exosomes in Alzheimer's disease

The two pathological hallmarks of Alzheimer's disease are A β plaques and neurofibrillary tangles. The removal of unwanted materials from the cells through exosomes has been established and neurons use the exosomes to clear out the toxic beta amyloid (A β) proteins [70] (Figure 6.3-3). These toxic beta amyloid (A β) proteins accumulate in the endosomes and their ILVs are then transported out as exosomes after the fusion of their respective MVBs to the cell membrane into the extracellular space [71].

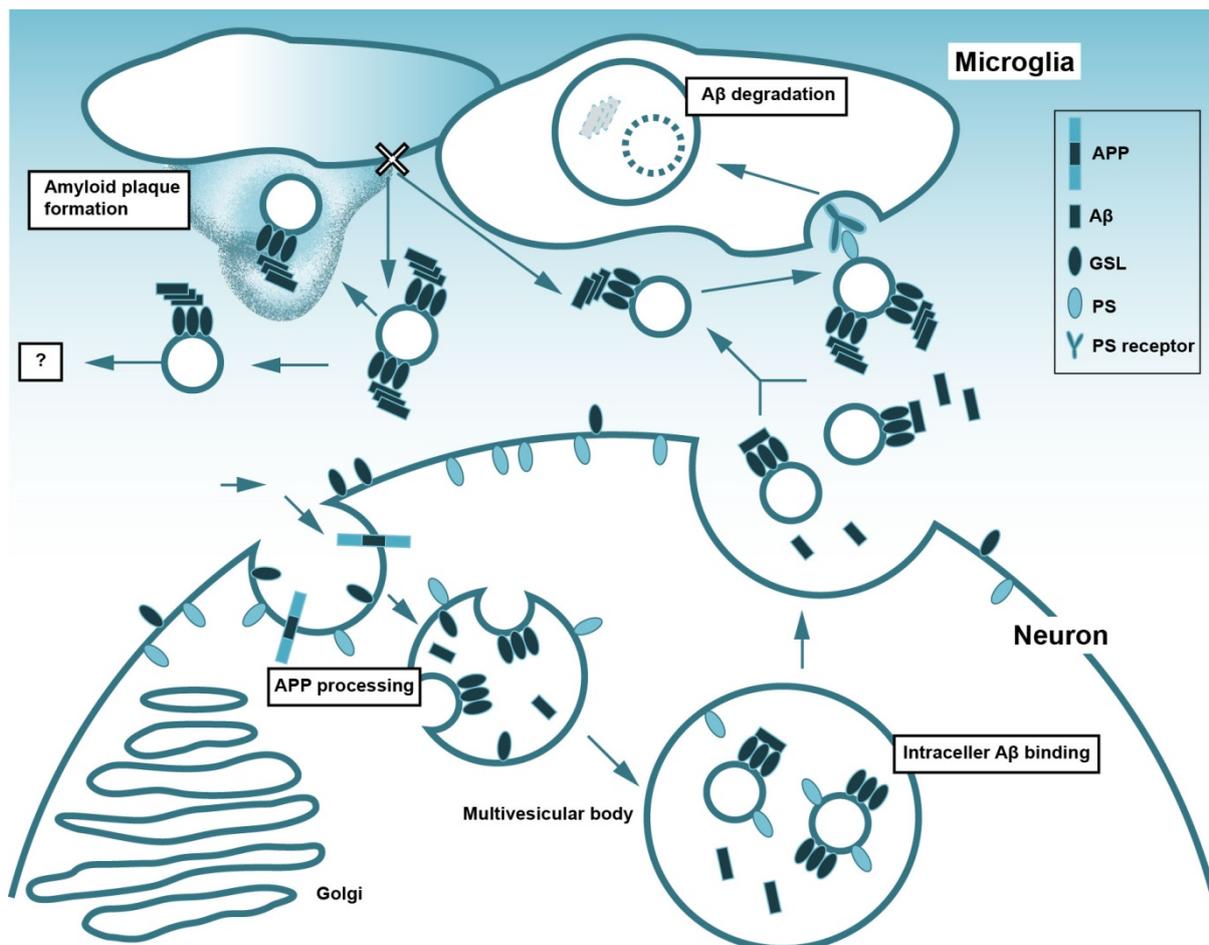


Figure 6.3-3 A β proteins clearance through exosomes. Amyloid- β precursor protein (APP) gets cleaved to form Amyloid- β (A β) proteins. They are then released into the extracellular space through exosomes. Adapted from Yuyama, K et.al. [71]

Second hallmark of Alzheimer's disease, the neurofibrillary tangles associated with hyperphosphorylated microtubule associated tau protein is also secreted through the use of exosomes [72-74]. In the clinics, increased levels of phosphorylated tau proteins were observed from the blood exosomes of Alzheimer's disease patients in comparison to the age-matched controls [75].

6.3.4 Exosomes as potential biomarkers for Alzheimer's disease

Exosomes can be excellent candidate biomarkers for early stage Alzheimer's disease. Exosomes extracted from the cerebrospinal fluid (CSF) contains neurons and non-neuronal markers like beta-amyloid (A β) proteins, tau proteins and pTAU proteins. By checking at the levels of these proteins from the CSF exosomes, it is possible to diagnose an individual with MCI or Alzheimer's disease [76, 77]. Exosomes also contains small non-coding RNAs like miRNAs and piRNAs which may also be used as potential biomarkers for the classification of Alzheimer's disease from their aged matched controls [78, 79].

6.4 Biomarkers

6.4.1 Definition of a biomarker

A biomarker or "biological marker" is anything that can be measured as an indicator of a biological process which can range from anything is going on in the body, whether it is a normal development of the body or response of an individual's body to a particular medicine [80]. These can range from simple measurements like blood pressure or cholesterol levels to complex measurements like the levels of tau proteins. Biomarkers can be used to provide information about an individual's risk of developing a particular disease, the prognosis of that disease and to predict response to a particular medication for that disease [81]. Individualized medicine critically relies on the development of biomarker [82, 83]. Especially for slowly processing brain diseases, it is important to:

- Develop markers that are rather stable over time in a healthy population
- Safe and easy to measure
- Cost efficient
- Inform about treatment efficacy

6.4.2 Types of biomarkers

Biomarkers fall into two major categories; Prognostic and Predictive [81, 84].

- Prognostic biomarkers can provide us with information regarding the progression of a disease. They are unaffected to the treatment plan and strategies of a particular patient suffering from the disease. These biomarkers are useful in the prognosis of the patient as to how aggressive the treatment plan should be in order to approach the treatment of the disease

- Predictive biomarkers can provide the prediction about the therapeutic outcome for a patient to receive certain medicine or treatment plan over another medicine or the treatment plan

These classifications can also be fuzzy where particular biomarkers can be both prognostic and predictive thus providing crucial evidence about what kind of medicine or treatment plan should be planned for a patient and how aggressive it should be implemented [85, 86].

6.4.3 Pros and cons of biomarkers

Biomarkers are definitely useful as they could lead to faster and accurate diagnosis of a neurodegenerative disorder. One of the biggest pros that biomarkers provide is the prognosis about the disease. This can also help reduce the use of animal studies and postmortem tissue studies [87]. The new biomarkers development has the additional advantages that they are non-invasive with the development of the blood based biomarkers. On the other hand, developing a cost efficient biomarker is still an open field as most of them require huge research infrastructure to develop [88]. Another drawback is that most of these biomarkers are not ready for use in the clinical community as they are mainly for research purposes.

6.4.4 Potential biomarkers for MCI and Alzheimer's disease

Both in MCI and Alzheimer's disease, the challenge is to find a good biomarker that is dealing with the brain, which is the most complex organ in the body[89]. Definitely, there cannot be a universal biomarker for these diseases and so developing a suitable list of biomarkers that can cover most of the disease symptoms will be very useful [90]. Using both neuroimaging biomarkers and cerebrospinal fluid biomarkers, early diagnosis of Alzheimer's disease is possible [91-93]. These biomarkers have shown that the levels of A β plaques arise very early in the Alzheimer's disease pathology and then plateau as the disease progresses. On the other hand, tau proteins come later and are affected by the levels of A β plaques. Using the use of these biomarkers, an accurate diagnosis of the disease state can be estimated along with the other clinical measurements like memory and cognitive tests. This provides a better understanding of the disease progression from the preclinical Alzheimer's disease stage to MCI and finally into the Alzheimer's disease [88, 94, 95].

6.5 Epigenetics and role of Non-coding RNAs

6.5.1 Introduction to Epigenetics

Epigenetics refers to the information related to the expression of the genes inherited stably that does not affect the primary nucleotide sequence of the DNA [96-98]. Epigenetics determines which genes are expressed by turning them on and off and the expression patterns are maintained through the cell division. Epigenetic dysregulation can lead to varieties of diseases by either suppressing or overexpressing certain genes. Epigenetics basically comprises all the chromatin and DNA modifications along with all the other regulatory elements that drive gene expression [99].

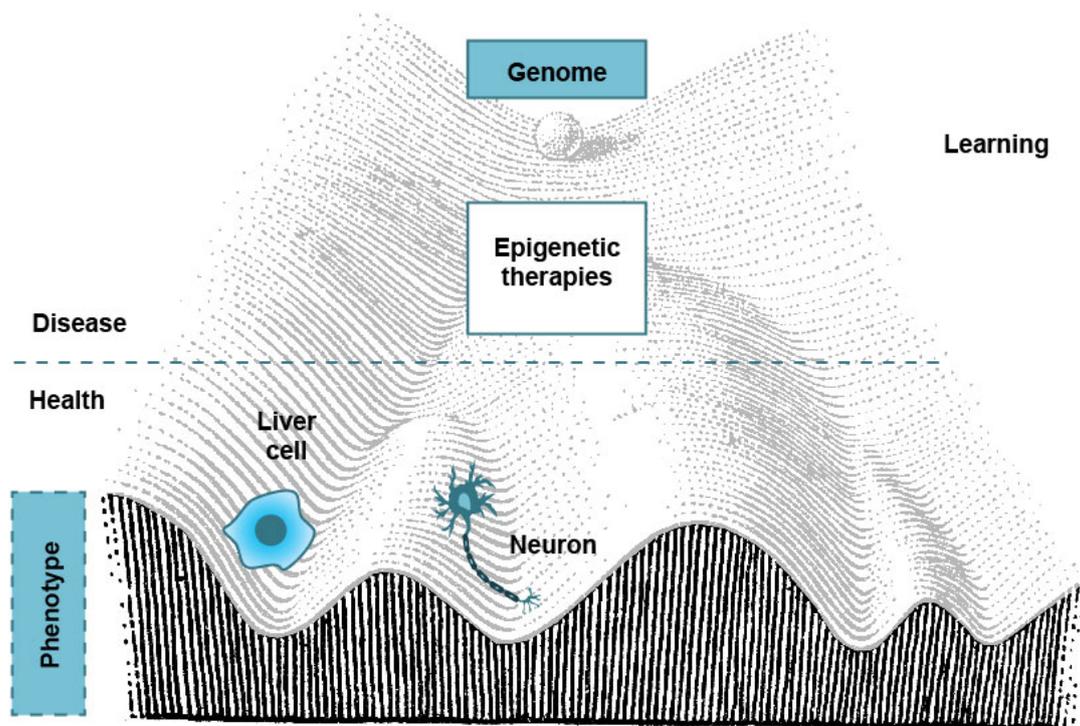


Figure 6.5-1 Epigenetics landscape in health and disease.

All the cell in our body has the same genome, for example, the liver cell has the same genome as the neurons, but clearly a different phenotype and the difference between them is mediated not to a large extent by epigenetic regulation [100, 101]. An interesting thing is that during learning, the neuron seems to take on a different epigenetic state and it again changes between health and disease. Since the sequence of the genome cannot be changed and epigenetics is reversible, there are increasing interest in epigenetic therapies for brain diseases that can bring back the neuron into homeostasis [102-104].

There are three key epigenetic mechanisms:

- Histone modifications
- DNA methylation
- The action of non-coding RNAs

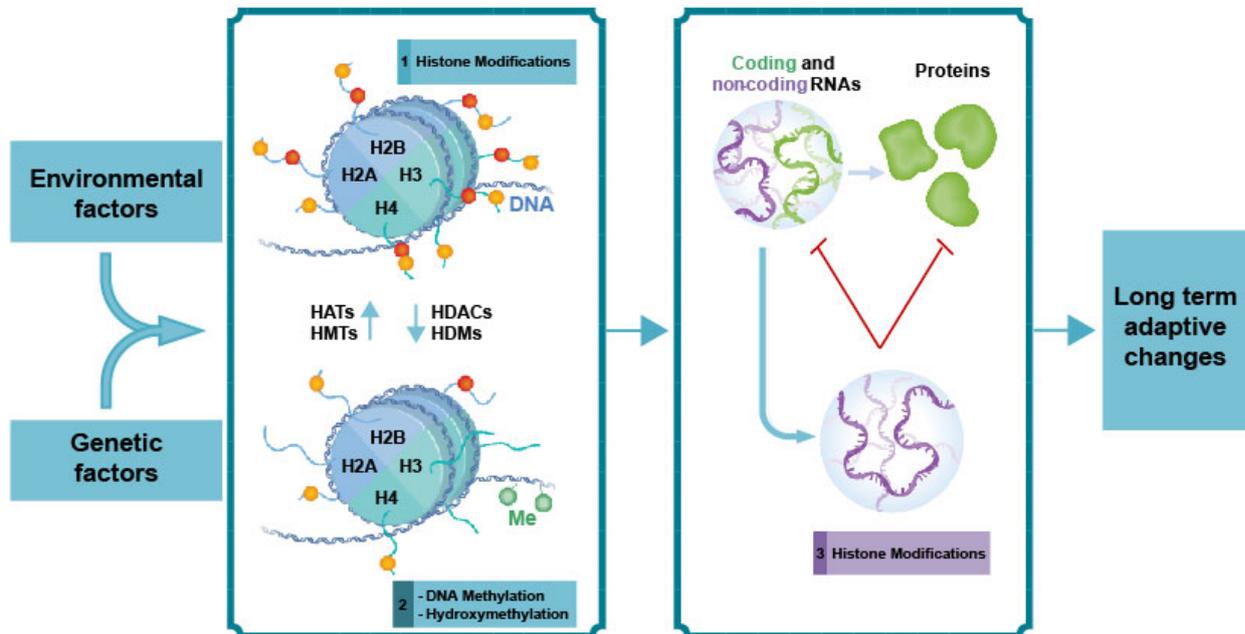


Figure 6.5-2 Key epigenetics mechanisms (Adapted from Fischer et.al. [105])

These factors are there to mediate the change from environmental and genetic factors into long term adaptive changes. Epigenetics provides the cell the molecular toolbox to drive the transient stimuli into long-term adaptive changes [105].

6.5.2 A general overview of small non-coding RNAs (sncRNAs)

The traditional central dogma of molecular biology dictates that the DNA is transcribed to RNA and then translated into proteins. Small RNAs have changed the landscape and modified this central dogma. Out of all the genome that is transcribed, only 2% of the transcripts actually translated into the proteins. The remaining 98% of the non-coding transcribed transcripts (ncRNAs) [106] are used for the infrastructural and regulatory functions in the human genome [107-109].

Infrastructural ncRNAs [110] include:

- Ribosomal RNAs (rRNAs): These are the most abundant RNAs constituting about 80% of total RNAs. These are synthesized in the nucleolus. They are the structural and catalytic component of the ribosomes

- Transfer RNAs (tRNAs): A ~80 nucleotide (nt) long sncRNA that transfers amino acids to ribosomes during protein synthesis. The tRNAs are amino acid specific with each molecule of tRNAs consists of an anticodon that are complementary to the codon on the mRNA molecule [111]
- Small nuclear RNAs (snRNAs): Performs splicing and intron removal from the primary transcript to form the mature mRNA transcript. It is also involved in the production of 3' ends of polyA deficient histone mRNAs [112]
- Small nucleolar RNAs (snoRNAs): Primarily involved in the processing of rRNA. These are encoded in introns and can be anywhere from 60 to 300nt in length [112]

Regulatory ncRNAs [109, 113-115] include:

- MicroRNAs (miRNAs): The miRNAs are about 16-27 nt in length that regulates gene expression post transcriptionally either by translational repression or degradation of mature mRNAs. They are highly conserved from plants to animals.
- Piwi-interacting RNAs (piRNAs): Longer than miRNAs, piRNAs are 26-32 nt long and mainly involved in epigenetic regulation of transcription, silencing of the transposable elements mainly in germ cell development and post-transcriptional gene silencing
- Small interfering RNAs (siRNAs): Similar to miRNAs and are about 19-25 nt in length which are involved in the degradation of the mRNA
- Long non-coding RNAs (lncRNAs): These noncoding RNAs ranges from several hundred nucleotides to several thousand nucleotides. They are involved in both transcriptional and post transcriptional regulation, chromatin remodeling, and genomic imprinting

Other ncRNAs include:

- Enhancer RNAs (eRNAs): These are a few hundred to about 1000 nucleotides long. They are involved in the transcriptional regulation by acting as transcription activators and can be used as the markers for enhancer identification [116, 117]
- Promoter associated RNAs (PARs): These are about 20 to 200 nt long. They are mostly involved in enhancing or blocking the transcription of neighboring genes and resides in the vicinity of the basal promoter and TSS [118, 119]

6.5.3 Biogenesis of microRNAs (miRNAs) and piRNAs

6.5.3.1 miRNAs biogenesis

Gene regulation is important as they make cells different from each other since the genome of each cell is the same. miRNAs are regulating a lot of these genes. The biogenesis starts with the key proteins AGO, Drosha and Dicer. PolII polymerase transcribes the miRNAs as a part of a much larger transcript, which happens in the nucleus (Figure 6.5-3). Then the Drosha cleaves the hairpin just one helical turn from the base of the hairpin. That releases the pre-miRNA hairpin which is then transported out of the nucleus by the exportin 5 complexes. In the cytoplasm, dicer cleaves the hairpin loop of the pre-miRNA to form a miRNA duplex and from there it is loaded onto a silencing complex containing AGO protein which is then used to target the mature mRNAs for silencing [120].

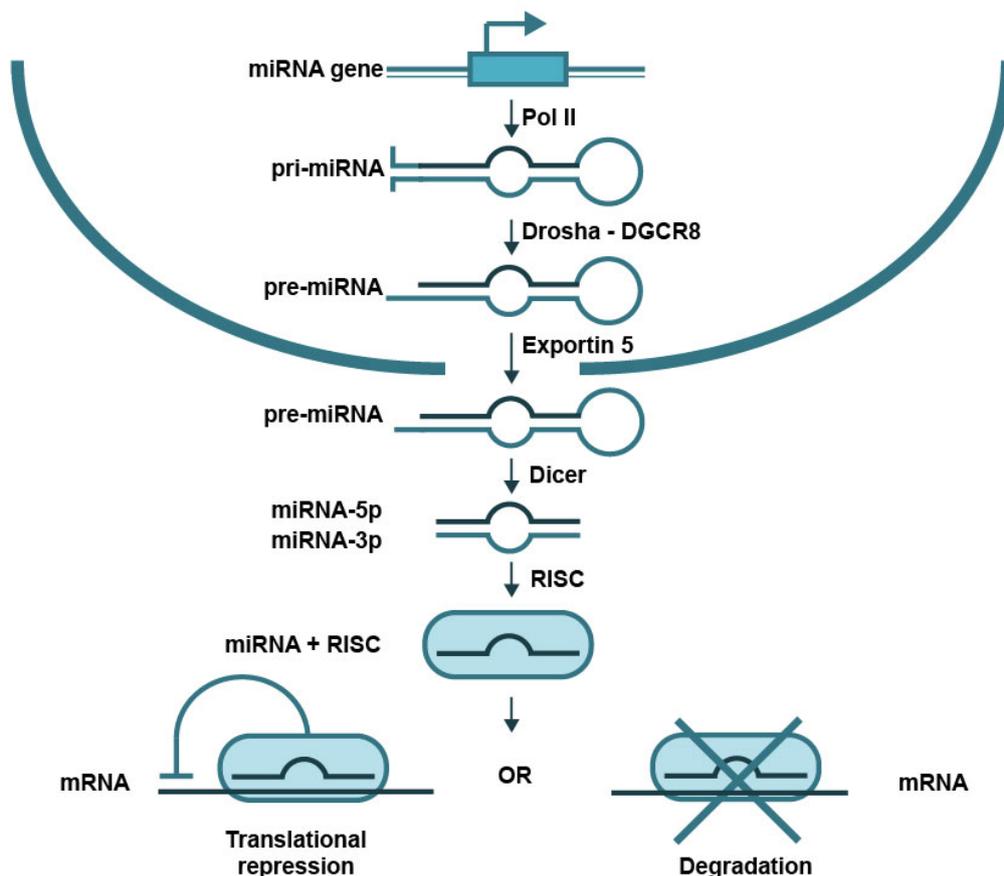


Figure 6.5-3 miRNA biogenesis (Adapted from Devaux et.al. [121])

6.5.3.2 piRNA biogenesis

piRNAs although best characterized in the *Drosophila* but recent developments in the field have [122-124] provided more understanding about them in the mammals. The biogenesis starts from the transcription of the transposon or repeat elements containing uni- or bi-directional promoters that are clustered or separated throughout the genome. This transcription produces piRNA precursors which are transported out of the nucleus. First, the longer precursors are cleaved to the length of 26-32 nucleotides in primary processing (Figure 6.5-4 A). From there they are loaded onto PIWI proteins in secondary processing which is also known as ping-pong amplification where piRNAs are amplified (Figure 6.5-4 B) and used for silencing [125].

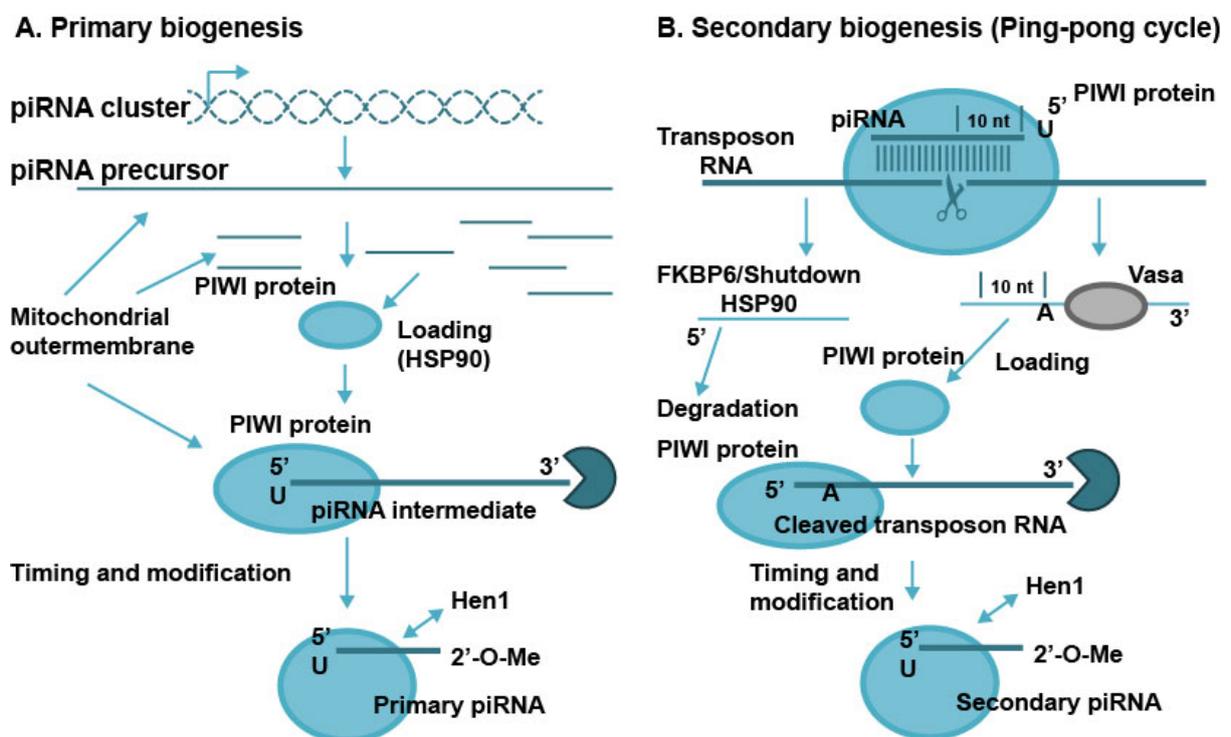


Figure 6.5-4 The biogenesis of piRNAs (Adapted from Watanabe & Lin et.al. (2014) [125])

6.5.4 Role of miRNAs and piRNAs in Alzheimer's disease

miRNAs and piRNAs although do not code for proteins themselves, but they play an important role in many biological processes by regulating the expression of the genes. The two hallmarks of Alzheimer's disease pathology are A β plaques and neurofibrillary tangles. The A β plaques are formed by the increased level of BACE1 protein and many miRNAs are found to be associated with the regulation of the BACE1 protein levels [126]. The second hallmark is the neurofibrillary tangles which are formed by the hyperphosphorylated tau proteins. miRNAs also affects the phosphorylation of these tau proteins [127]. The role of

piRNAs is still not well known. The reduction in the piRNAs expression in Alzheimer's disease is also linked to increase neuronal death as they drive transposable element dysregulation in tauopathy. piRNAs can also play a big role as a potential biomarker as the level of piRNAs are found to be upregulated in Alzheimer's disease patients compared to the healthy controls [122, 128].

6.6 Machine learning in biomarker development

6.6.1 Overview of machine learning

Machine learning [129] and big data are the two most powerful things in healthcare in today's world and revolutionized the field completely [130]. Machine learning is developing algorithms and code that can teach computer tasks over a period of time either from the gathered data or by interactively interacting with the environment. The digitization of the medical records has boosted the field by providing a huge amount of data which can be used to train and test the performance of the algorithms to better understand a disease mechanism [131, 132].

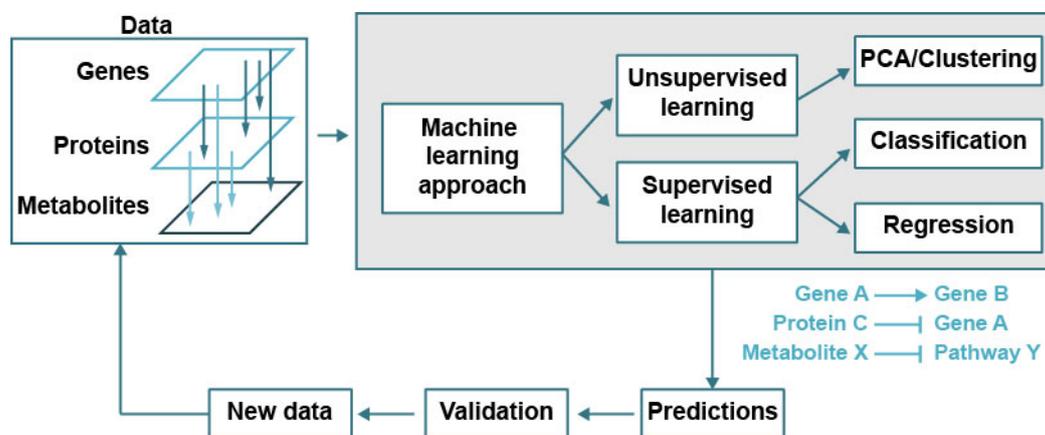


Figure 6.6-1 A typical machine learning application used in biological data analysis (Adapted from Camacho et.al. (2018) [133])

Here (Figure 6.6-1) different type of data is collected over time, including several samples for genes, proteins and metabolites and other data related to the correlation between them [133]. Depending on the type of data, whether it is labeled or unlabeled, a suitable machine learning algorithm will be chosen. For unsupervised learning, an example would be the dimensionality reduction algorithm PCA or clustering algorithm can be chosen. In case of supervised learning where the labels of the data are available and depending on the biological question, classification or prediction algorithm will be chosen. The machine

learning algorithm will output a model which will be validated on an untouched test dataset and depending on the performance of the test, the process will be either repeated to get better performance of the model are being used in the clinics and hospital for the purpose of diagnosis or prognosis of a disease [130, 134].

6.6.2 Types of learning

6.6.2.1 Supervised learning

There are mainly three different classes of machine learning algorithms, supervised, unsupervised and semi-supervised learning algorithms [135-137]. In supervised learning, the data is provided along with the labels where the data contains the dependent variables and the labels will be the independent variables. It can be used either for classification or prediction [138]. These algorithms typically require training data and training labels which the algorithm will learn and build a model and then measure the performance of the learning algorithm on the completely untouched test dataset and test labels [139, 140]. The main advantage is that these algorithms can directly be measured in terms of their performance and outputs a numerical measure on the accuracy and inaccuracy of these machine learning algorithms.

6.6.2.2 Unsupervised learning

On the other hand, unsupervised learning has no goals to predict or to classify the data, but mostly used to learn more about the data [141]. Basically trying to understand the underlying structure in the data or uncover some unusual patterns in the data. These algorithms do not require labeled data. The evaluations of these algorithms are also different in terms that there is no way of telling how accurate or inaccurate the algorithms are because there is no labeled data available. There are two kinds of evaluation that can be performed for these algorithms, indirect or qualitative evaluations. Qualitative evaluation is measured by looking at the results of the algorithms and seeing something relevant in the result that helps in the understanding of the data [142]. On the other hand, indirect evaluation is measured by providing the output of the unsupervised learning algorithm to some other algorithm and checking if the other algorithm has performed well or not.

6.6.2.3 Semi-supervised learning

Semi-supervised learning algorithms are the middle ground between supervised and unsupervised learning algorithms [136]. These are very useful when obtaining the labeled

data for supervised learning is expensive, but obtaining the unlabeled data is cheap. Semi-supervised learning data is a mix of supervised learning data, i.e. the labeled data and unsupervised learning data i.e. the unlabeled data. The algorithm first starts only with the labeled data and then used on the unlabeled data to find the tentative labels which can be then used to improve the performance of the model. This can be then repeated again and again until the algorithm converges [134]. At this point, the model is generated which is more accurate and used only a small amount of labeled data and the generated the labels for the unlabeled data.

6.6.3 Alternative categorization of learning algorithms

6.6.3.1 Generative learning

Generative learning algorithms use some characteristics of the population where it finds a distribution for each class and creates the model for each of them separately[143]. In a classification problem, the algorithm will try to find a boundary where it will choose a model that is more plausible than the other. These algorithms are mostly probabilistic [144] and can be implemented using both labeled and unlabeled data for estimating the overall distribution of the data [145, 146].

6.6.3.2 Discriminative learning

Discriminative learning algorithms focus mainly on finding the boundary between the two classes [147]. These boundaries can be linear or non-linear in nature. These algorithms do not penalize for the variance of the data. The variance can be very high or low and the algorithm will perform the same way if the boundary is found. It uses only the points that are closest to the boundary to make the decisions [148]. Basically, a lot of data can be ignored, including precious labeled data and focus on the important data which are used to identify the boundary. These algorithms are very powerful when there is plenty of training data but can only make use of the labeled data. These algorithms cannot be used for unsupervised learning [149].

6.6.4 Prior knowledge in learning

The machine learning algorithms cannot always provide better performance. In the absence of a huge amount of training data, the model cannot be created perfectly and the performance of the model would not be adequate. Incorporating some prior knowledge into the learning algorithm can improve the performance where the training data is limited and

the algorithm does not need to make deductions about the data from the training data itself [150]. The prior knowledge can be incorporated into a regularization function where many computationally efficient approaches are there to handle the knowledge [151-153].

6.6.5 Handling heterogeneous data

The data in today's research comes from various fields [154]. This introduces a new challenge as most of the data have different structure and different meaning. The heterogeneity of the data is a major challenge in this field [155]. This type of data can be anything starting from a free text to proper continuous data, images, codes, etc. [156]. The data are inaccurate and contains a lot of missing points. The data can be very dynamic and may evolve over time also [157].

6.6.6 Data munging and normalization

6.6.6.1 Data munging and wrangling

Data munging, which is also known as data wrangling is basically the process of preparing datasets for the data analysis [158]. A general rule of thumb in the data analysis is the 80/20 rule where the 80% of the time is spent in cleaning and formatting the data into the right format and the remaining 20% of the time is used for the actual data analysis. One important aspect is the use of standard data formats. There are many accepted standard available now, for example, CSV, TSV, XML, JSON and structured query language files. Another critical step is cleaning the data [159]. The obtained data can have both errors and artifacts that need to be removed or properly handled. Errors can be noise in the data that happened during the time of the data acquisition, which cannot be corrected in the data analysis step. On the other hand, artifacts are problems that are generated during processing and during the handling of the data systematically [160, 161].

6.6.6.2 Data normalization

In biological datasets, the data normally represent some biological phenomenon, but this is not the case most of the time. Along with the underlying biological or true variance in the data, there is also non-biological variances in the data that are the artifacts of the data and should be removed or normalized [162]. For example, in the microarray data, an artifact can be dye biases that are resulting from the overall spot intensity or location on the array [163]. In the RNAseq experiment, these artifacts can be the sequencing biases or GC content biases or unknown biases introduced during the library preparation [164]. The normalization

process is used to remove the non-biological sources of variation which allows comparing results from multiple samples. All the downstream analysis depends critically upon this step as using raw data or improper normalized data can provide misleading or completely wrong conclusions [165-167].

6.6.7 Random forest and other machine learning algorithms used in the analysis

With the advent of next generation sequencing (NGS), there is a huge influx of big data in the field of biology, which led to the development of new methods and techniques to process, analyze and visualize the data. Machine learning algorithms are one of the key classes of algorithms that are very useful with this kind of data. Both supervised and unsupervised methods are useful. The supervised methods can use the already existing labels for the biological data and can start to classify the data or perform prediction on new samples. One goal of these algorithms is to learn from the data and generate a model with high accuracy performance. This performance can further be increased by combining many learning models where the inherent noisy and unbiased models can create a new and better model with low variance. This is the main concept behind random forest (RF) algorithm [168, 169].

Random forest algorithm offers many advantages over many other machine learning algorithms like support vector machines (SVMs), k-nearest neighbors (KNNs), neural networks (NNs), logistic regression and many more [170]. RF does not overfit the data easily, has high accuracy, can be implemented in parallel for faster computing and provides variable importance information which can be extremely useful in narrowing down the big list of features that contain only a few informative features to a small list of useful and informative features [171, 172].

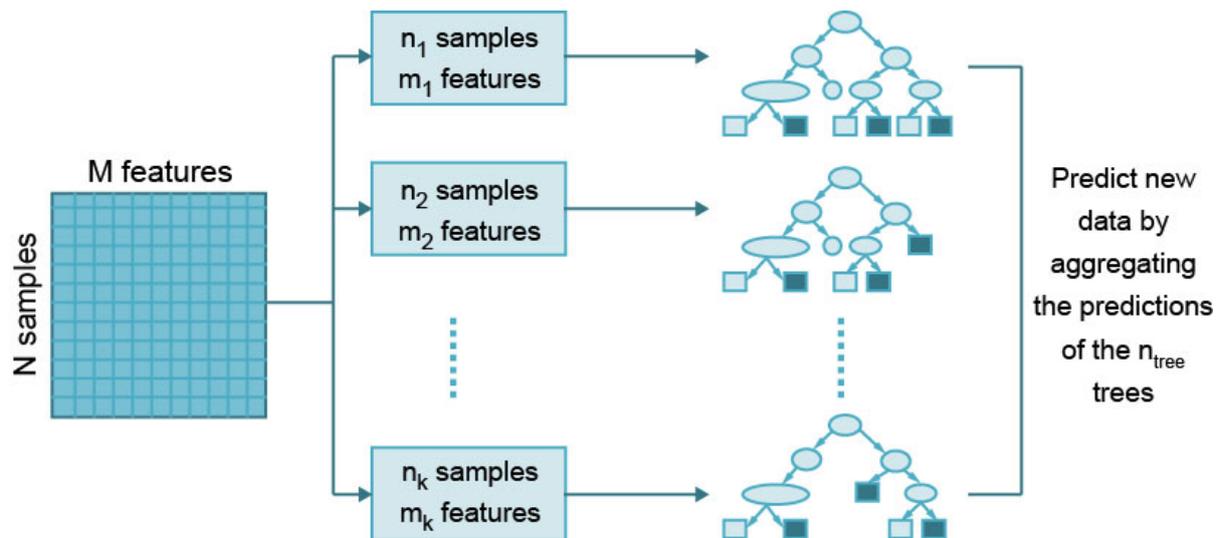


Figure 6.6-2 Schematic diagram of internal working of RF algorithm

RF is basically a bagging of de-correlated decision trees [173, 174]. The algorithms work on i iteration by selected B bootstrap samples of size s from the original training dataset

$$M = \langle (m_1, n_1), \dots, (m_k, n_k) \rangle$$

The random forest then grows the de-correlated tree T_i to these bootstrap sampled datasets (B). At the end of i iterations, ensembles of all de-correlated decision trees are returned. Prediction is done by calculating the average of all predictions of all decision trees. Classification, on the other hand, is done by taking the majority vote of all single tree predictions (Figure 6.6-2).

6.6.8 Feature selection

It is a common knowledge in the field of machine learning that the performance of a given machine learning algorithm heavily depends on the quality and quantity of the features provided to the algorithm. Finding and selecting a good feature is a challenging task [175]. Features contain information related to the target variables. In a classification problem, features should provide the information on the available labels and the classification model will be defined in terms of these features [176]. This leads to a misunderstanding as including more features does not provide the algorithm a better discriminative power [177]. In healthcare (clinics) field getting patient data is expensive. With the limited amount of samples and a huge number of parameters (features) that can be measured in these samples, developing a machine learning algorithm is not an easy task [178]. The algorithm's performance decreases as the amount of feature grow (Figure 6.6-3) [179].

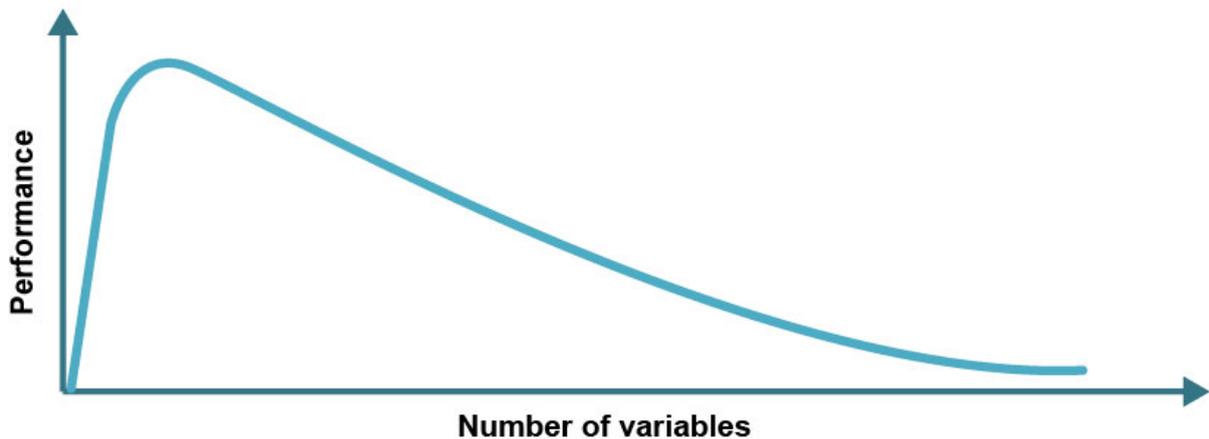


Figure 6.6-3 Relationship between a machine learning algorithm's performance and the number of features

This is due to the presence of irrelevant features which introduces noise in the data and redundant features that do not provide any additional information which degrades the performance of the learning algorithm. This curse of dimensionality is very problematic with limited training examples (clinical data, genomic and epigenomic datasets etc.) [179] and also limited computational resources as it increases the search space and ultimately this leads to overfitting. Thus, it becomes very important to perform feature reduction before training a learning algorithm. There are two types:

- Feature selection: is selecting a smaller subset of the features from a larger set of features
- Given a big set of features $F = \{x_1, x_2, \dots, x_N\}$ that contains both relevant, irrelevant and redundant features, a smaller set of relevant features $F' \subset F = \{x_1', x_2', \dots, x_{M'}\}$ that optimizes the learning algorithm for better performance
- Feature extraction: is the process of transforming or projecting the original set of features (N) into a higher dimensional plane with a reduced number of dimensions (M)

The final goal is to either improve or maintain the learning algorithm's performance by possibly simplifying the algorithm's complexity (Figure 6.6-4).

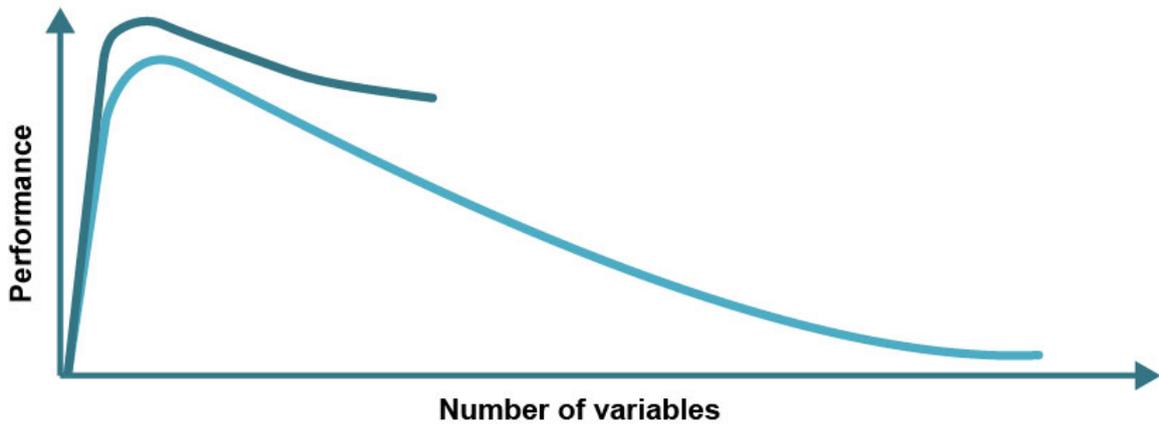


Figure 6.6-4 Modified relationship between a machine learning algorithm's performance and the number of features

6.6.9 Evaluation of learning models

A good machine learning algorithm not only uses fewer resources, but should also perform well during the evaluation phase. This is important as these algorithms are going to be used on unseen and unlabeled data points for classification and prediction. There are many choices for the performance evaluation of these learning algorithms. In supervised learning, there are classifications and regression evaluation metrics while unsupervised learning has their own metrics.

6.6.9.1 Classification metrics

Typical choices for performance evaluation for classification models [180] include:

Table 6-2 Common classification metrics [181-183]

Metrics	Description
Accuracy	<ul style="list-style-type: none"> • Most common and misused evaluation metric • It is basically the ratio of true predictions over total predictions • It is not suitable for imbalanced classes • The predictions or classifications are binary in nature as they assign each prediction or classification with either 0 or 1
Logarithmic Loss	<ul style="list-style-type: none"> • Also known as logloss which takes the uncertainty of predictions or classification into account • It is calculated by measuring the accuracy after penalizing the false classifications • The probability of each class has to be defined before the

	evaluation
Confusion Matrix	<ul style="list-style-type: none"> • Useful for multi-class classification problems • It is represented in the form of a table where rows represent actual class and column represents the predicted class
Precision-Recall	<ul style="list-style-type: none"> • Precision represents the exactness of the model and calculated as a percentage of predicted positives are actually positive • The Recall represents the completeness of the model and calculated as the total positives labeled by the model • It is useful with the imbalanced classes
Area Under ROC Curve [184]	<ul style="list-style-type: none"> • Is commonly known as AUROC (Area Under the Receiver Operating Characteristic curve) • Provides both aggregated accuracy of AUC and true positive rates vs. false positive rates (ROC) with the variation in classification threshold • It also provides a good visual summary of the performance of the model

6.6.9.2 Regression metrics

Table 6-3 Common regression metrics [185-189]

Metrics	Description
Mean Absolute Error	<ul style="list-style-type: none"> • Also known as MAE and is calculated by averaging the absolute difference of each prediction of actual values • It provides a sense of the magnitude of the error and it is also less sensitive to outliers • It has the same unit as the original data
Mean Squared Error	<ul style="list-style-type: none"> • Also known as MSE and is calculated by averaging the squares of the difference of each prediction to actual values • It is sensitive to large outliers • Can be used in the comparisons of various statistical models
R squared (r^2)	<ul style="list-style-type: none"> • Also known as the coefficient of determination and provides information on how good the predictions are compared to the actual values • It represents the amount of variation in the independent variable

that can be explained by the variance in the dependent variable

- The values lie between 0 and 1

6.6.9.3 Unsupervised learning metrics

Evaluation of an unsupervised learning algorithm is a difficult task compared to the supervised learning evaluation. It varies mostly subjective as the definition of success with those algorithms on the basis of the question being asked. One of the most common categories of unsupervised learning algorithms is clustering algorithms [190, 191]. There are mainly two types of cluster validity measurements:

- Extrinsic or external index: The clustering performance is measured indirectly on the basis of another learning algorithm (mostly supervised learning) where a good performance of supervised learning algorithms reflects a good clustering of the data points
- Intrinsic or internal index: The clustering performance is measured as to how well the clusters are by themselves without the use of any external input. It can be qualitative where the clustering is helpful in the understanding of the data

Apart from the above mentioned criteria [192, 193], unsupervised learning can also be evaluated using other evaluation metrics like Rand index, F-score and Mutual information [194].

6.7 Three dimensional organization of the genome

6.7.1 General features of the 3D genome organization

One of the most pressing questions in molecular biology was the organization of the DNA into the nucleus as the spatial organization and position of the genes and other regulatory elements drive the expression of these genes and any misregulation may result in a disease or disorder [195, 196]. Another question was how does the same genome result in so many different cell types in the body [197]? Biochemical elements like promoters, enhancers, and insulators are known for a long time now where promoters are normally the starting point of the transcription, enhancers controlling the rate of transcription and insulators controls which gene an enhancer can target [198]. At the smallest scale (Figure 6.7-1) [199] DNA is wrapped around the histone octamers also known as nucleosomes and at the highest scale, i.e. at the chromosome level scale each chromosome occupies a unique territory within the

nucleus [200, 201]. The intermediate steps in between are now being studied and the characterization of their functions is showing up.

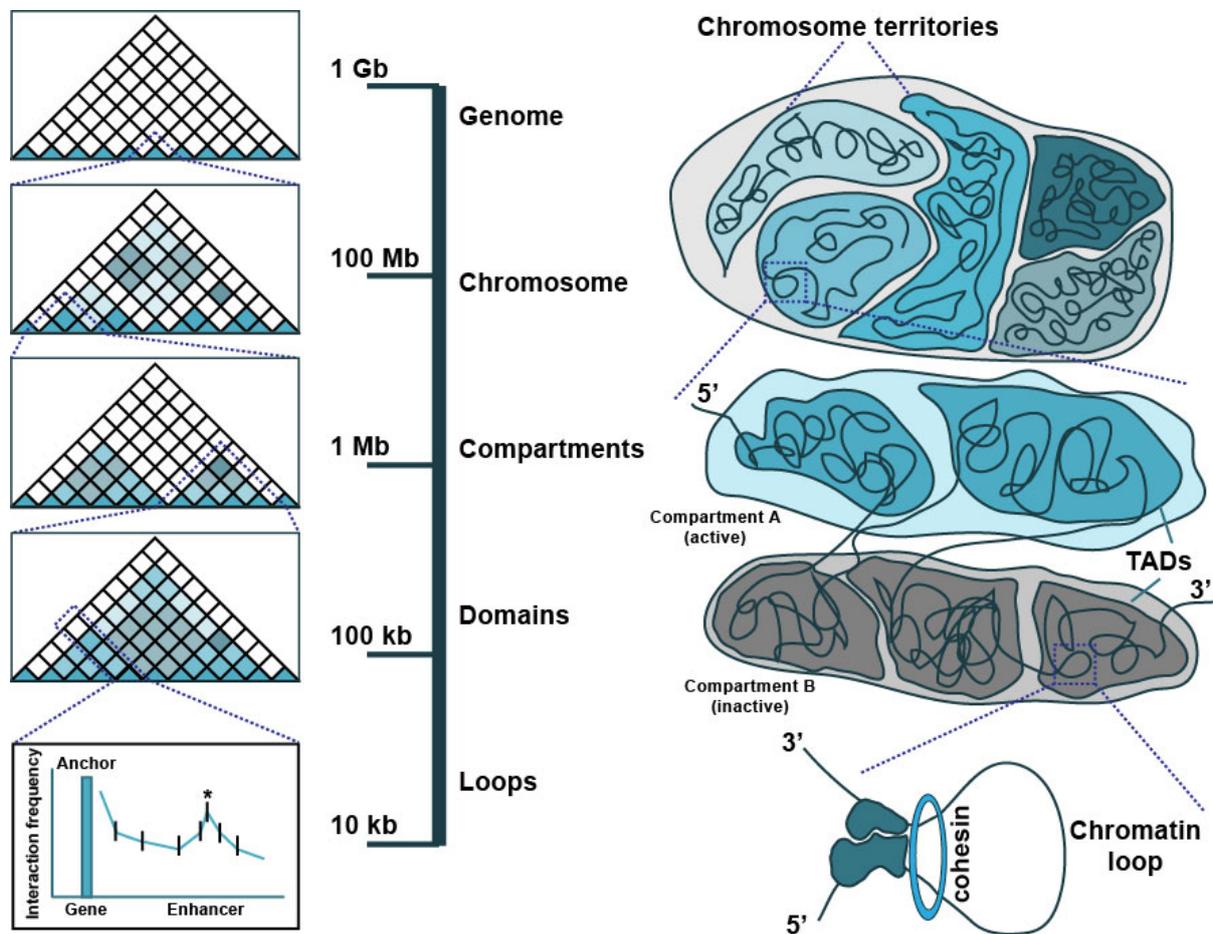


Figure 6.7-1 Common organizational structures in the genome adapted from Szalaj et.al (2018) [199]

6.7.1.1 Chromosome territories

The knowledge of chromosome territories (CTs) has been known more than a century. The chromatin in a non-dividing cell are known to occupy a certain place in the nucleus where chromosomes with high gene density tends to cluster together with an inclination towards the center of the nucleus while gene poor chromosomes [202] co-occupy the regions that are at the boundary of the nucleus.

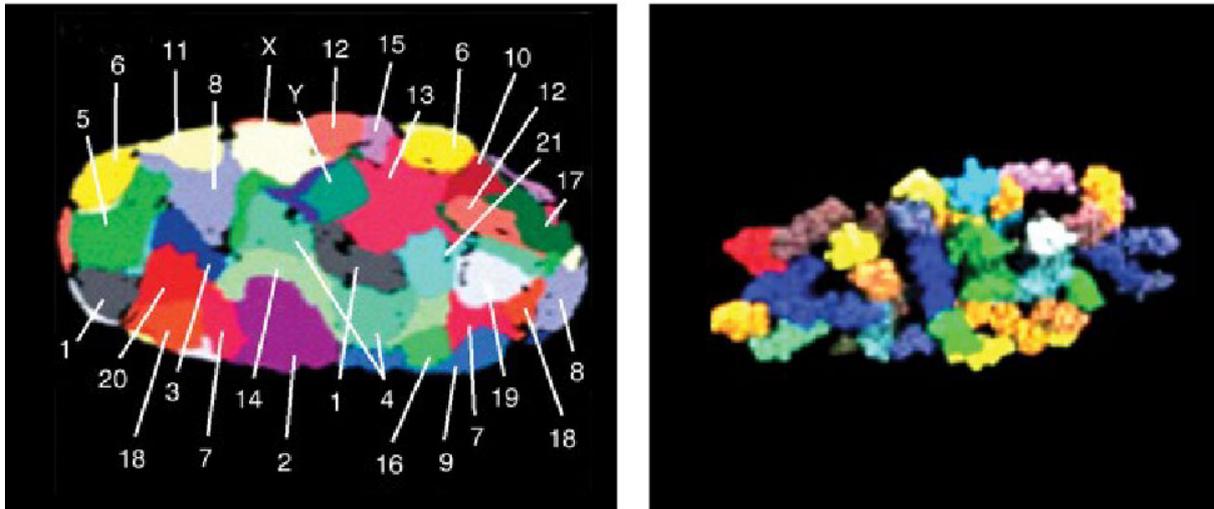


Figure 6.7-2 Chromosome territories (Adapted from Bolzer et.al. (2005) [203])

Figure 6.7-2 is portraying a spectral karyotyping where each chromosome is artificially labeled with different colors where they occupy their own territories [203]. They are around 50 – 250Mb in length. A common method of detection is Fluorescence In Situ Hybridization (FISH).

6.7.1.2 Compartments

The mammalian genomes have another level of organization in terms of active and inactive compartments [202, 204, 205]. The active compartment tends to interact with respective active compartments and an inactive compartment tends to interact with their respective inactive compartments [206]. The interactions between the active and inactive compartments are minimal. There is another observation that the active compartments have higher gene expressions compared to the inactive compartments where the gene expression is generally silenced [207]. However, the mechanism on how the compartments are formed or what constitutes an active compartment and what constitutes as inactive compartments is still unknown. Recent studies have shown that there are now at least 6 subcompartments [208].

6.7.1.3 TADs

Zooming into the specific compartment around 100 KB resolutions (Figure 6.7-1), there are different domains which are not alternating but more are domains. These domains are called topologically associating domains (TADs) [209, 210]. These were one of the biggest discoveries as a big black box in chromatin organization can be explained with TADs [211]. Genomic elements within the same TAD interact with each other, but they do not tend to interact with the genomic elements located in different TADs. These TADs are cell type

invariant. On average, TADs ranges from a few hundred KBs to several hundred MBs. The distributions of genomic elements have a specific pattern inside the TADs. Gene promoters, promoter associated histone marks such as H3K4me3 and insulator marks are tending to stay on the boundary of the toads while enhancer elements like H3K4me1 are spread all along the TAD. The boundary regions are especially enriched with insulator proteins like CTCF along with repeats and housekeeping genes [212, 213].

6.7.1.4 subTADs

Initially TADs are considered as cell type invariant and conserved across the species. This is still unclear as upto how much these domains are conserved. Although TADs are cell type invariant, the subdomains also known as subTADs are found to be cell type specific [214]. These subTADs are the domains where a specific regulatory mark like a promoter or enhancer can be co-regulated [215].

6.7.1.5 Chromatin Loops

The human genome is subpartitioned into approximately ten thousand loops [216]. Loops are basically two regions of chromatin that are linearly far apart but close to each other in 3D. On average, loops are approximately 200kb in length and do not overlap with each other (Figure 6.7-1). This is one of the mechanisms to control gene expression [217]. The TSS and enhancer loop each other which express a gene. Loops are primarily formed by the binding of CTCF proteins to their respective motif binding sites which occurs at the boundary of the loops [218]. Recent studies have shown that the CTCF motifs that are involved in the formation of loops should be pointing towards one another [219-221].

6.7.2 Tools to explore 3D genome organization

The three dimensional organization of the genome can be understood from different approaches. Mainly NSG bases techniques and imaging techniques [222] are used to visualize these chromatin structures and DNA editing tools like CRISPR-Cas9 for the functional understating of the genome [223, 224].

6.7.2.1 Imaging based tools

Imaging tools helps to view the chromatin structures and to study the organization of genomic elements. Among the various techniques that are used these days to quantify the distances between elements and view the structures, some are:

- Fluorescence in situ hybridization (FISH) [225]

- MB-FISH (Molecular Beacon-FISH) [226]
- Super-resolution dipole orientation mapping (SDOM) microscopy [227]
- Low-power super-resolution STED nanoscopy [228]
- Cryo-electron microscopy (cryo-EM) [229]

6.7.2.2 3C-based technologies

The emergence of 3C (Chromosome Conformation Capture) [230] based techniques has revolutionized the field of three dimensional genome organization. After the advent of the first 3C technique, many other variations have emerged. Some of them are:

- Chromosome conformation capture-on-Chip (4C) [231]
- Chromosome conformation capture carbon copy (5C) [232]
- 3C-chip [233]
- Combined 3C-ChIP-Cloning (6C) [234]
- Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) [235]
- Hi-C [236]
- Capture-C [237]

6.7.3 Role of 3D genome organization in diseases

In the healthy state of an individual, the three dimensional organization of the genome into proper domains and loops occur normally. Any disruptions in this organization results in the misregulation of the genomic elements causing disorder and diseases [238, 239]. In a healthy state the regulation of gene expression is controlled by restricting the enhancer promoter interactions to their domains. Disruptions in the boundaries of these domains (TADs) have shown to cause gene expression changes where an enhancer from one TAD can influence the expression of a gene in a different TAD. This can lead to various disorders like Schizophrenia, various types of cancer, etc. [240]. They are also linked to physical deformities like the F syndrome or Autosomal-dominant adult-onset demyelinating Leukodystrophy (ADLD). Developing a better understanding of these gene regulation can help understand the disease mechanism and provide the understanding on how they causes the pathogenic phenotypes [241, 242].

7. Objectives

The tight control of gene expression programs in the adult brain is critical for synaptic plasticity and memory formation. Epigenetic mechanisms [243, 244] are key processes that control gene expression at a systems level. There is emerging evidence that dynamic changes in histone modifications, DNA methylation/demethylation, and the expression of small noncoding RNAs provide brain cells a cellular and molecular toolkit by which environmental stimuli are transformed into long-term adaptive changes. In fact, epigenetic processes have been linked to synaptic plasticity and memory function and have emerged as promising novel therapeutic approaches to treat neuropsychiatric and neurodegenerative diseases [105]. These findings, however, are mechanistically not well understood. In almost over a decade, say from 2007, the new field of neuroepigenetics [245] has emerged, but the base publications in the field dated back to more than four decades ago [38]. During this time, although there has been tremendous development in technologies and discoveries of novel biomarker development, but there is a distinct lack of mechanistic insights.

An important question that has so far not been addressed is **how do neurons change their epigenome in a neuronal network?** With this broader question in mind which I may not find an answer during my Ph.D., I am hoping to solve a part of the question by working on two aims which are described below.

Aim1: To study the epigenetic regulation of information in a neuronal network like exosomal transfer of small noncoding RNAs that may lead to the development of the new category of biomarkers for neurodegenerative disease especially focusing on Alzheimer's disease

- The main hypothesis for this aim is that the small noncoding RNAome measured in the brain, CSF and blood may be used to detect biomarker signatures for neurodegenerative disorders
- First, use the high throughput next generation sequencing (NGS) techniques to quantify the expression of small noncoding RNAs in Alzheimer's disease patients and their respective control samples
- Then, develop targeted statistical and machine learning pipeline and tool kits to process the NGS data to identify informative and putative biomarker signatures
- Finally, test the performance of these small noncoding RNAs signatures in different replication cohorts for:
 - Disease diagnosis, i.e. to classify Alzheimer's disease patients from their respective controls with high precision and accuracy

- Disease onset, i.e. predict the conversion of some of the patients suffering from mild cognitive impairment to develop Alzheimer's disease in due course of time
- Perform the comparison of the performance of these small noncoding RNAs signatures to the existing clinical protein biomarkers

Aim2: To study the role of dynamic changes in higher order chromatin structure that control gene expression programs in neuronal plasticity

- Understand the three dimensional (3D) organization of the genome inside neuronal and non-neuronal cells and its impact on the regulatory network using Chromosome Conformation Capture (3C) [230, 246] based techniques like 5C [232, 247], Hi-C [236] and 3C-seq [248]
- The hypothesis is that the dynamic changes in higher order chromatin structure control gene expression [249] programs in synaptic plasticity, memory function, and neurodegenerative disorders
- First, in a pilot project, the epigenome organization and higher order chromatin structures in FACS sorted non-neuronal and neuronal nuclei from the hippocampal region (CA1) under the naïve condition will be studied in C57BL/6J wild type mice using 3C-seq technique
- Then, with the help of publicly available existing tools and if needed, develop new pipelines to identify long-range looping interactions and topologically associating domains (TADs) in neurons and non-neurons
- Perform comparisons in the looping patterns in neuronal/non-neuronal interactions and domains and also identify few important differences related to the deregulation of the genes that may be involved in neurodegenerative disorders like Alzheimer's disease

8. Materials and methods

8.1 Cerebrospinal fluid (CSF) from humans

8.1.1 CSF from Alzheimer's disease patients and control humans

Cerebrospinal fluid was collected by performing a lumbar puncture (spinal tap) on 82 control humans (nonAD), 42 Alzheimer's disease patients, 11 mciStable and 6 mciAD patients. The procedure was performed after the approval of Institutional review board (IRB) of Dept. of Medical Ethics and History of Medicine at University Medical Center Göttingen (IRB 02/05/09). The samples were collected at the University Medical Center Göttingen (Germany), Department of Psychiatry, Göttingen (Germany), University Department of Neurology at Universitätsklinikum Tübingen (Germany) and Paracelsus Elena clinic Kassel (Germany). The samples were collected from Jan 2012 and Mar 2013 and from Apr 2013 and Oct 2014. Standard operating protocols were applied during the collection and handling of the CSF from the individuals. 5 to 10 mL CSF was collected in 4 vials and protease inhibitors and preservative were added to the collected CSF. It was then centrifuged for 10 mins at 2000 x g at room temperature to get rid of extra membrane fragments or cell debris. The resulting supernatant was flash frozen and stored in a minus 80°C freezer for the extraction of RNA from the exosomes.

8.1.2 Extraction of CSF exosomes

The flash frozen supernatant was removed from the minus 80°C and centrifuged three times at 4°C. First centrifugation was performed for 10 mins at 3500 x g and the supernatant was then centrifuged again for 10 mins at 4500 x g and the resultant supernatant was then centrifuged again for 30 mins at 10000 x g. The supernatant was then further subjected to ultracentrifugation for 10 mins at 100000 x g resulting in an exosomal and exosomal-free fraction which was then resuspended in 0.2ml TRI Reagent® (RNA Isolation Reagent from Merck). The exosomal fraction was then stored at minus 80°C freezer prior to RNA isolation.

8.1.3 Total RNA isolation

The exosomal fraction was homogenized with 1ml of TRI Reagent® (RNA Isolation Reagent from Merck). 25 µl/ml of distilled H₂O [diethyl pyrocarbonate (DEPC) treated and 2 µl of glycogen was added to the mixture and incubated for 5 mins at room temperature followed by the addition of 0.2 ml of phenol/chloroform. The samples were shaken vigorously and re-

incubated for an additional 5 mins followed by 15 mins of centrifugation at 1200 x g at 4°C and clean transfer of phase separated aqueous RNA was carefully collected. Overnight precipitation of collected aqueous RNA was performed at minus 20°C with 0.5 ml isopropanol. The precipitated RNA was centrifuged at 4°C at 12000 x g for half hour and after discarding the supernatant, the isolated pellet was washed twice with 75% ethanol and resuspended in 10 µl of RNase free water after air drying the RNA pellet.

8.1.4 SmallRNA library preparation

Small RNA libraries were prepared from total RNA using the TruSeq®(Illumina) small RNA sample preparation kit according to the Illumina reference guide for TruSeq Small RNA Library Prep Kit. Briefly, the adapters are ligated to the RNA molecules. A cDNA library is generated by performing reverse transcription, amplification and gel purification of adapter ligated RNA molecules. SmallRNA sequencing was performed on the aliquoted cDNA libraries.

8.2 Processing of next generation sequencing data

8.2.1 Generation of raw fastq files

For the processing of the sequencing data, a customized in-house pipeline was used. First, the HiSeq™ 2000 Sequencing System performs image analysis and basecalling using the machine's internal software provided the primary data files in the *.bcl (binary base call) format. Illumina's bcl2fastq (v 1.8.4) [250] with default parameters was used to convert the base calls in the per-cycle BCL files to the per-read FASTQ format. Along with the conversion of base calls *.fastq reads, demultiplexing of the samples are also performed in parallel using the same Illumina's bcl2fastq (v 1.8.4) software with default parameters.

8.2.2 Quality control (QC) for raw sequencing data

Basic quality control for each generated *.fastq files was performed with illumine_qc.sh (v 1.2.0) [251] with default parameters which produced fastq_screen plots for an overview of sequencing quality. If the samples pass the quality control step, they are stored in the DZNE [252] long term storage file server and can be used for further analysis. Before the transfer to the long term storage from the local Illumina machine storage, the MD5 checksums of all the *.fastq files and the directory are generated using md5checker.py (v 0.3.2) with default parameters.

8.3 SmallRNA mapping pipeline (GJSrMap)

The smallRNAs mapping pipeline (GJSrMap) (v 2.5.0) [253] was used to align all the *.fastq files. All the following steps were performed on the GWDG's high performance computing cluster (HPCC) [254] using gwdu103.gwdg.de as the master node. The jobs were launched on to short or long queues with a maximum of 12 cores, maximum time constraint of 24 hours and maximum memory constraint of 16GB using FAT queue. The average run time of each job is approximately 13.58 mins, average memory used was 171.2 MB and average max threads were 26.75.

The mapping is performed in the iterative manner. First all the relevant files are downloaded and processed for the reference genome (Table 8-1). We trim low-quality ends from reads before adapter removal having Phred quality score [255] less than or equal to 28. We then trim the 3' adapters and filter out the reads with the minimum length of 15 nucleotides. We first map the reads to the reference genome created from microRNA and piRNA sequences. Reads greater than 33bp were mapped to the reference genome created by other small non-coding RNA. Remaining unmapped reads were then mapped to the human genome.

We used bowtie [256] for all the reads alignment o the reference genome. We allowed no mismatches for the first part of the mapping and one mismatch for the rest of the mapping. Read count distribution and visualization was generated using in-house python script. Final data, summary and QC is collected as shown in (Figure 8.3-2 A and B)

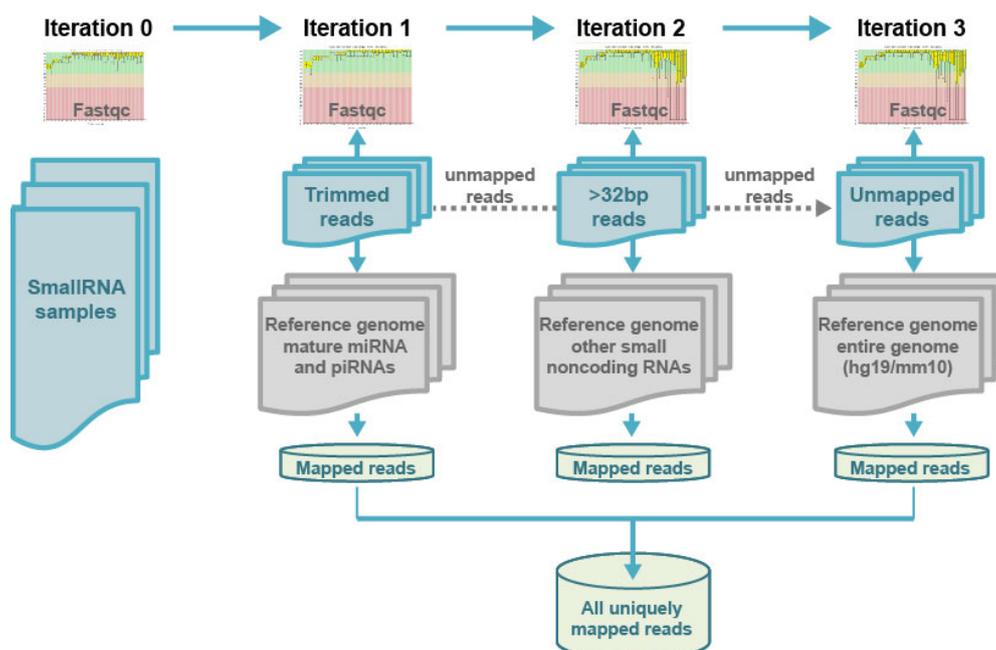


Figure 8.3-1 The smallRNA mapping pipeline (GJSrMap) [253]

8.3.1 Prepare custom reference genome

The smallRNAs reads are aligned to the custom reference genome made from the small noncoding RNA sequences and also to the reference human genome sequence. This is to avoid the problems of multiple mapped reads and to increase the sensitivity of the mapping.

Table 8-1 Description of small noncoding RNAs using in the mapping

smallRNAs	Description
miRNA	Mature microRNA
piRNA	Piwi-interacting RNA
rRNA	Ribosomal RNA
scRNA	Small cytoplasmic RNA
snRNA	Small nuclear RNA
snoRNA	Small nucleolar RNA
premiRNA	microRNA precursors
osncRNA	Other small noncoding RNA
- tRNA	- Transfer RNA
- Mt-tRNA	- Transfer RNA located in the mitochondrial genome
- misc_RNA	- Miscellaneous other RNA

8.3.1.1 Small noncoding RNA and full genome sequences source

Human precursor and mature miRNA sequences are downloaded from miRBase [257] (version.20) with genome-build-id: GRCh37.p5 and genome-build-accession: NCBI_Assembly:GCA_000001405.6.

Human piRNAs sequences are downloaded from piRNABank [258] (<http://pirnabank.ibab.ac.in/index.shtml>).

Human other small noncoding RNAs were downloaded from EnSEMBL release 75 - February 2014 for the GRCh37 version (ftp://ftp.ensembl.org/pub/grch37/current/fast/homo_sapiens/ncrna/Homo_sapiens.GRCh37.ncrna.fa.gz).

The assembly of the human genome (hg19, GRCh37 Genome Reference Consortium Human Reference 37 (GCA_000001405.1) in Feb. 2009) as well as repeat annotations and GenBank sequences were downloaded from <http://hgdownload.soe.ucsc.edu/goldenPath/hg19/bigZips> and random contigs were filtered out using the Unix command `grep` [259].

8.3.1.2 Pre-processing of the sequences

Alignment of small noncoding RNAs is always a challenging task due to their origin from multiple genomic locations. In order to avoid multi-mapped reads, a custom reference genome is created with unique sequences of every smallRNAs. Multiple smallRNAs with same sequences are collapsed into the single smallRNAs with a new unique ID that is a combination of the individual IDs separated by a “|” symbol. This solves the problem of the multiple mapped reads, as there won't be any read that needs to multi-mapped any more. It also makes mapping modular and fast. An example collapsed sequence would be:

```
>hsa-mir-6859-1|hsa-mir-6859-2|hsa-mir-6859-3|hsa-mir-6859-4  
TGTGGGAGAGGAACATGGGCTCAGGACAGCGGGTGTTCAGCTTGCCCTGACCCCATGTCGCCTCTGTAG
```

8.3.1.3 Custom reference genome indexes

The custom reference genome indexes are built for the iterative steps in the pipeline. There are three custom reference genomes; pimiRNAs (miRNAs + piRNAs) reference genome, sncRNAs (all other small noncoding RNAs other than miRNAs and piRNAs) and the full human genome. Bowtie-build command with default parameters of the Bowtie (v 1.1.2) [256] alignment tool is used to build the indexes for pimiRNAs and sncRNAs. The index of the human genome (hg19) is downloaded from the Pre-build index library of bowtie tools (ftp://ftp.ccb.jhu.edu/pub/data/bowtie_indexes/hg19.ebwt.zip)

8.3.2 Pre-processing and quality control before alignment

8.3.2.1 Quality control for FastQ files

At every iteration, from 0 to 3 of the iterative mapping pipeline, quality control of raw sequencing data is performed using FastQC (v 0.11.5) [260] using the default parameters and `--quite option` for the quite mode as the additional argument.

8.3.2.2 Low-quality read filtering

Due to the shorter length of small noncoding RNAs compared to the sequencing read length, the sequencing adapters are also sequences which can cause a problem in the mapping as the partial adapter sequences do not exist in the reference genome and hence can cause a very poor mapping. Before trimming the adapter sequences, low quality fastQ reads are removed to increase the quality of the mapping. FastQ reads having the Phred quality score of less than 28 encoded in ASCII33 are removed and not considered in the analysis using cutadapt (v 1.9.1) with the argument:

```
--trim-qualities 28 or -q 28
```

8.3.2.3 Adapter trimming and read size selection

3' adapter sequences are removed using the cutadapt (v 1.9.1) [261] and along with that two types of read size selection are also performed for the iterative mapping steps 1 and 2. For the first step, reads between 16 and 32 nucleotides are considered and for the second step, reads greater than 32 nucleotides are considered. Along with the adapter removal cutadapt also performs the size selection using the arguments:

- `--minimum-length 16` or `-m 16`: This option will remove any reads shorter than 16 nucleotides
- `--maximum-length 33` or `-M 33`: remove any reads greater than 33 nucleotides
- `--too-long-output <filename>`: This is the option that is set to save all the reads that are longer than 33 nucleotides in the same fastQ format

8.3.3 Iterative mapping of the filtered fastQ reads

This step is performed in three iterations followed by read quantification and the cleanup step.

8.3.3.1 Iteration 1: mapping to pimiRNA reference genome

The filtered fastQ reads between the size of 16 and 33 nucleotides are mapped to the custom pimiRNA reference genome. The mapping is performed using bowtie (v 1.1.2) with fine-tuned parameters for optimal mapping.

- `-v 0` is used is set for end-to-end hits with zero mismatches to keep the mapping very conservative for this iteration
- `-p/--threads 4` are the number of alignment threads for faster mapping

- `-S/--sam` is set for writing all the mapped reads into the Sequence Alignment Map (SAM) format
- `-a/--all` is set to get all alignments for every read
- `-m 1` is set to remove all alignments that have more than one reportable alignments
- `--un <unmapped_fastq>` to save all the unmapped reads for iteration 3 where these reads will be mapped again but to the entire human genome

The resulting SAM files are converted to BAM (Binary SAM) files and indexed using samtools (v 1.3.1) [262]. This iteration is run using the “fat” queue with four maximum numbers of cores.

8.3.3.2 Iteration 2: mapping to sncRNA reference genome

The fastQ reads longer than 32 nucleotides are mapped to the custom sncRNA reference genome. The mapping is performed using bowtie (v 1.1.2) with same parameters except

- `-v 1` is used is set for end-to-end hits with one possible mismatch for a little relaxed mapping strategy

The unmapped reads from the previous iteration 0 and current iteration 1 are then concatenated for the next step and saved to the file. The resulting SAM files are converted to BAM files and indexed using samtools (v 1.3.1). This iteration is also run using the “fat” queue with four maximum numbers of cores.

8.3.3.3 Iteration 3: mapping to full genome

The unmapped reads from both iteration 1 and iteration 2 are then mapped to the human genome. The mapping is performed using bowtie (v 1.1.2) with the same parameters as iteration 2. The resulting SAM file is converted to the BAM file. It is then concatenated with the BAM files from iteration 0 and iteration 1 along with the remaining unmapped fastQ reads in the iteration 3 and indexed using samtools (v 1.3.1). Low quality mapping reads are filtered using samtools with the option `mapq >= 30`. The final BAM file is then converted to the SAM file for the read counting step. This iteration 3 is run using the “fat” queue with four maximum numbers of cores.

8.3.4 Read counting and summary statistics

8.3.4.1 Read counting and summary quality control report

Read counting is performed on the SAM file using the AWK [263] unix tool and distributing the counts for each smallRNA into their respective classes. Along with the raw read counts

distribution, normalized counts are also generated by simple RPM (reads per million) where the counts are divided the total uniquely mapped reads for the sample and multiplying it by the million. This is sufficient as smallRNAs read length distribution is narrow and hence bias to GC content or read length do not occur in the normalization of these reads.

A summary quality control report is also generated apart from individual quality control reports for each iteration of the pipeline. MultiQC (v 0.9) [264] is used with default parameters for each iteration and also a combined QC for all the iterations. The report is provided in the form of plots and also HTML reports.

8.3.4.2 Summary mapping and smallRNA statistics

A customized python script is used to create the following mapping and smallRNA statistics plots.

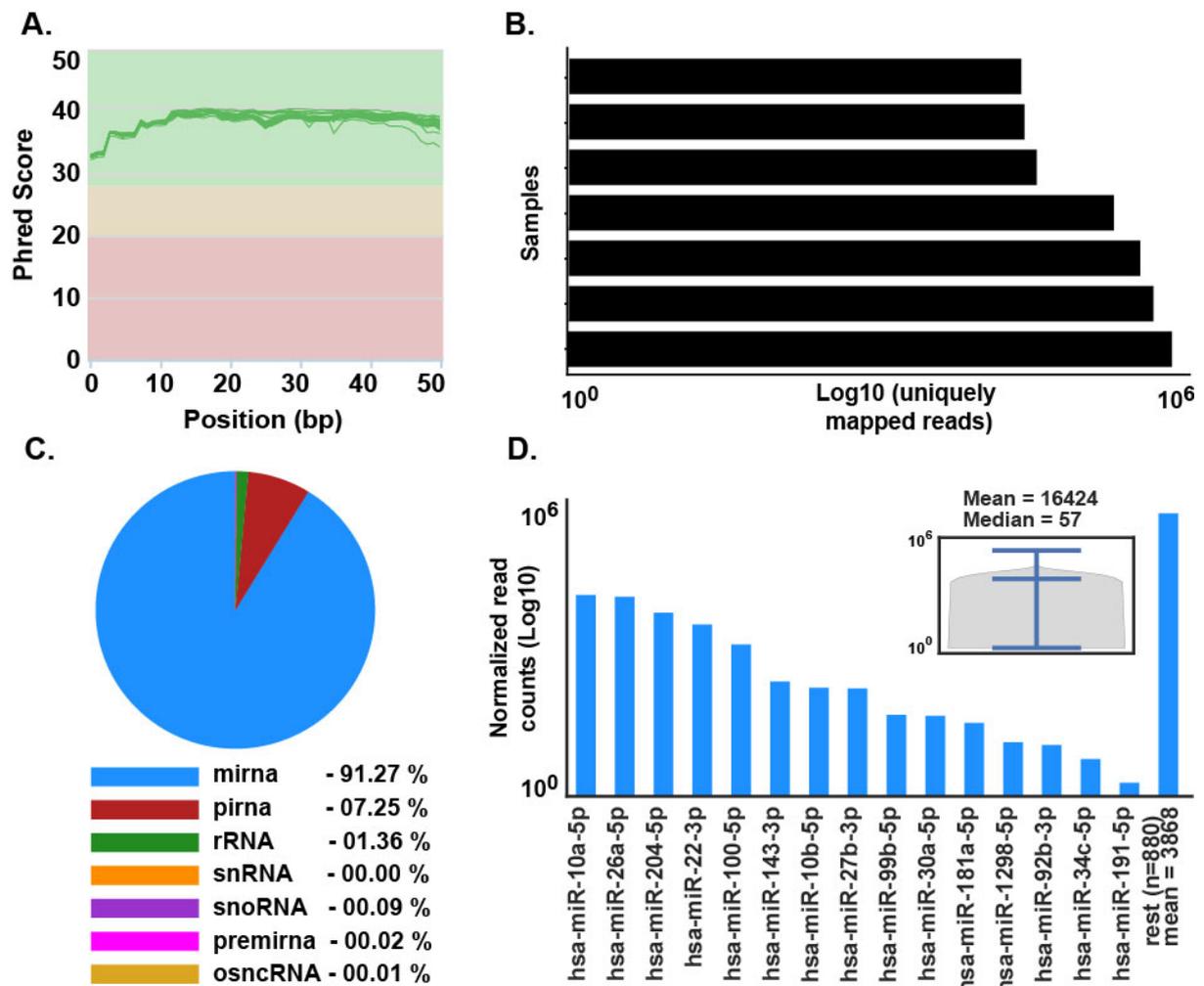


Figure 8.3-2 Sequencing quality control, mapping and smallrnaome distribution

Few plots are shown in Figure 8.3-2 Sequencing quality control, mapping and smallrnaome distribution. A. Shows the sequence quality information for 16 samples. X-axis represents the nucleotide position within the read and the Y-axis represents the quality scores (Phred score). The plot is divided into three parts according to the quality scores. The red part represents poor quality, orange part represents reasonable quality and the green part represents the good quality calls from the sequencer. The quality score from the Illumina is Q scores ($Q = -10\log_{10}P$) [255] which is an estimate of the probability of a particular base being called incorrectly. The score is averaged over the entire read length and deteriorates towards the end of the read. A Phred quality score of 30 tells us the probability of 1 incorrect base call for every 1000 base calls the sequencer makes. B. A barplot of uniquely mapped reads for each sample. On average, a quarter million reads were uniquely mapped to the small non-coding RNAome. One sample seems to have a failed sequencing. C. Representation of small non coding RNAs in serum samples. The majority of the detected small non coding are microRNAs covering 95% of entire small non coding RNAome. D. Top 10 expressed miRNAs (Blue). Last bar represents an aggregate summary of remaining expressed microRNAs. The inset violin plot shows the distribution of entire detected microRNAome. Similar plots are provided for all the other smallRNAs.

8.4 Data filtering and normalization

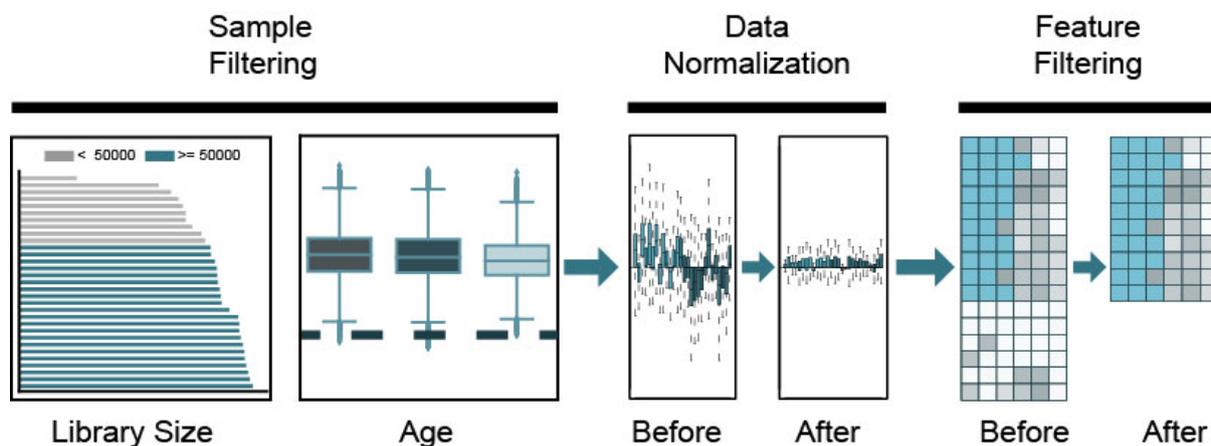


Figure 8.4-1 Sample filtering, feature normalization and filtering

All the samples with library size (total uniquely mapped reads) less than 50000 reads were removed from the analysis as they do not contribute valuable information to the downstream analysis. Samples that were less than 40 years were also removed from the analysis in order to maintain the uniform age distribution (Figure 8.4-1).

8.4.1 Normalization of smallRNAs

Various normalization strategies for the miRNA and piRNA normalization were tested as there was no standard normalization protocol available for the smallRNAs although many comparison references are available [265, 266].

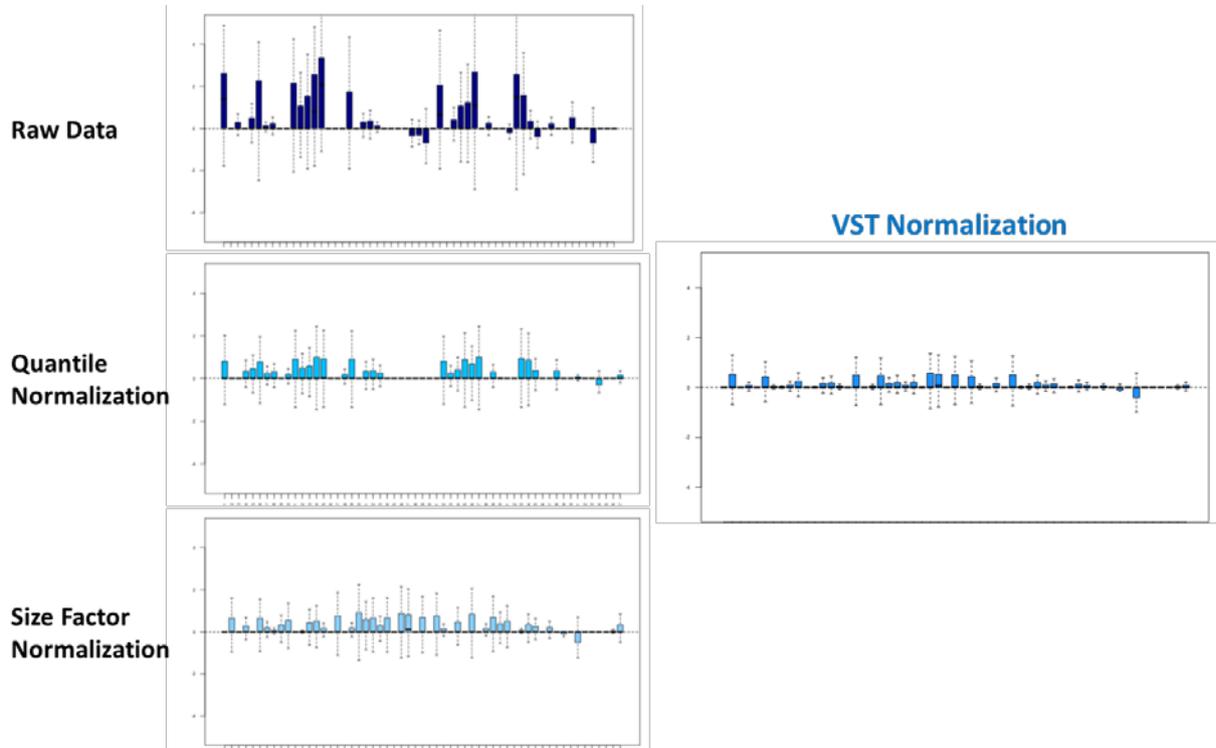


Figure 8.4-2 Various normalization strategies

The raw data after visualization was found to be pretty variable (Figure 8.4-2A).

Quantile normalization [267, 268] aligns the count distribution of samples. The quantile scaling factor for i th sample φ_i has the form of

$$\varphi_i^{quantile} = 10^{Log_{10} Q_i - \left(\frac{1}{N}\right) \sum_{j=1}^N \log_{10} Q_j}$$

Where N is the total number of samples, Q_i, Q_j are upper quantiles of i th and j th samples.

Figure 8.4-2 B shows the upper quantile normalization of the samples which is a lot less variation compared to the raw data.

Median normalization [269] implemented DESeq2 which is specifically implemented for the negative binomial distributed counts data has the form of

$$\varphi_i^{median} = Median_s \left(\frac{C_s^i}{\left(\prod_{j=1}^N C_s^j\right)^{\frac{1}{N}}} \right)$$

Where N is the total number of samples, C_s^i , C_s^j are the counts of smallRNA s for the i th and j th sample. Figure 8.4-2 C shows the median normalization of the samples which is similar to the quantile normalization and also has a lot less variation compared to the raw data.

Variance stabilization normalization [269-273] for the negative binomial distribution is given by

$$\varphi(x)^{VST} = \frac{\log[1 + 2x\alpha_0 + \alpha_1 + 2\sqrt{x\alpha_0(1 + x\alpha_0 + \alpha_1)}]}{\log[2]}$$

Where, α_0, α_1 are the two constants for dispersion α representing asymptotic dispersion and extra-Poisson factor respectively. The VST is a transformation $\varphi(x)^{VST}$ for a random variable X whose variance- mean relationship is $v = \mu + \mu^2\alpha_0 + \mu\alpha_1$ and variance (X) is equal to $v(E(X))$ which makes the $\varphi(x)^{VST}$ homoscedastic.

Figure 8.4-2 D shows the VSN of the samples which is a lot less variation compared to the raw data and other transformation and used for all the downstream analysis. The VSN counts were also corrected for cohort effects as the samples were collected from multiple centers along with the removal of the unwanted variances using the R (v 3.2.2)[274] package RUVSeq (v 1.14.0) [164].

8.4.2 Filtering of non-expressed smallRNAs

miRNAs and piRNAs having variance stabilized normalized (VSN) read counts of least 0.5 in the 95% of Alzheimer's disease and control samples separately were considered as expressed smallRNAs (Figure 8.4-1). We performed the filtering on miRNAs and piRNAs obtaining a set of 154 expressed miRNAs and 43 expressed piRNAs and were considered for further analysis.

8.5 Statistical and machine learning analysis

Due to the low number of samples and high number of features, various statistical and machine learning methods are applied to reduce the number of uninformative features and get a list of informative features that can be used as a potential signature for biomarkers.

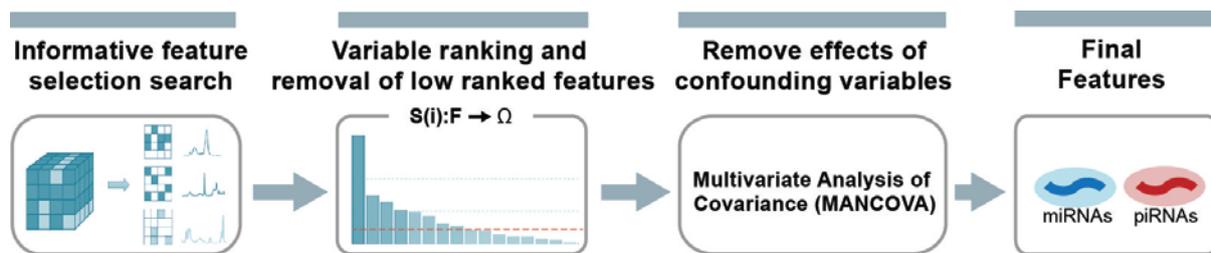


Figure 8.5-1 Feature selection

In order to obtain informative miRNAs (154) and piRNAs (43), “Measure of Relevance (MoR) procedure [275] on 61 samples of cohort 1 was applied. Considering high dimensionality of the features compared to a very small set of samples, the MoR procedure can reduce features set to small informative features set by evaluating distribution overlap (assuming that samples are independent), biological difference, dispersion parameters of the samples and weighing factor common to all features. The MoR^j value for j th smallRNA feature is given by:

$$MoR^j = \frac{1}{\sqrt{\frac{(n_1 + n_2)}{2}}} \cdot \frac{|p_1^j - p_2^j|}{\sqrt{\left(p^j(1 - p^j) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)}} \cdot \frac{\bar{U}_1^j - \bar{U}_2^j}{\sqrt{s_{U_{k_1}^j}^2 + s_{U_{k_2}^j}^2 - 2s_{U_{k_1}^j U_{k_2}^j}}}$$

Where, n_1, n_2 are two sub samples from N , p shows all smallRNA features and p^j represents j th feature, $\bar{U}_1^j, \bar{U}_2^j, s_{U_{k_1}^j}^2, s_{U_{k_2}^j}^2, s_{U_{k_1}^j U_{k_2}^j}$ represents means, variances and covariance of rank transformed values of j th smallRNA feature for two sub samples 1 and 2.

The absolute MoR values for each feature are then sorted in decreasing order and a suitable selection and evaluation criterion is applied to the information chain as described in [275]. We also applied reliability analysis (RiA) [276] to obtain a reduced set of informative features by applying MoR procedure (500 iterations) to a randomly selected subgroup of samples and features and features with a relative frequency higher than 0.25 during 500 iterations were selected. We chose features that were common between MoR and RiA as reliable features for further analysis.

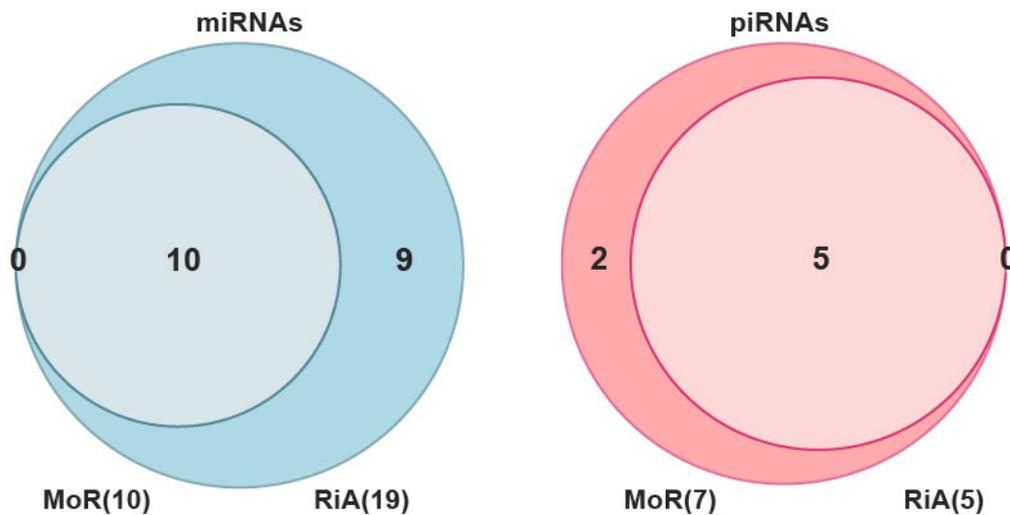


Figure 8.5-2 miRNAs and piRNAs

8.5.1 Variable ranking and removal of low ranked variables

Given the very low number of samples, the reliable features were further ranked [277] by sorting the features according to some scoring function $S(i)$ values given by [188, 278]:

$$S(i): F \rightarrow \Omega$$

Where, $S(i)$ is the scoring function computed from the values of training data with a set of features $F = (F_1, \dots, F_n)$ and Ω is the probability space which is the set of possible classifications $\{c_1, \dots, c_n\}$.

A combination of various machine learning algorithms were used and the mean score was calculated from the scoring functions which were used to rank the reliable features.

First, Ridge regression [279] which is also known as Tikhonov regularization was used. Here the loss function is a linear least squares function with l2 regularization. The linear least square function is calculated using singular value decomposition (SVD) on the input matrix $M(n, p)$ with the complexity of $O(np^2)$ where $n \geq p$. The ridge algorithm with default parameter given by

```
Ridge(alpha=0.5, copy_X=True, fit_intercept=True, max_iter=None,
       normalize=False, random_state=None, solver='auto', tol=0.001)
```

is used. The output are the coefficients which the Ridge algorithm minimized by applying a penalty on the size of the coefficients (the residual sum of squares). The minimization function is given by

$$\min_{\omega} \|M\omega - y\|_2^2 + \alpha \|\omega\|_2^2$$

Where, α is the regularization strength, y is the output variable and ω are the coefficients of the linear model.

Along with the Ridge, where l2 regularization parameters are set, Bayesian regression can estimate these parameters by fine tuning from the data. Bayesian Ridge [280-283] regression given by

```
BayesianRidge(alpha_1=1e-06, alpha_2=1e-06, compute_score=False,
copy_X=True, fit_intercept=True, lambda_1=1e-06, lambda_2=1e-06,
n_iter=300, normalize=False, tol=0.001, verbose=False)
```

basically, estimates the probabilistic model where the output y is assumed to be Gaussian around $M\omega$ is given by

$$p(y|M, \omega, \alpha) = N(y|M\omega, \alpha)$$

Where, α is treated as a random variable that will be estimated from the data.

The priors for ω is given by spherical Gaussian

$$p(\omega|\lambda) = N(\omega|0, \lambda^{-1}I_p)$$

Where λ is the estimated precision of the weights.

Then, a univariate linear regression tests [284] ($f_{regression}()$) was used for testing the effect of a single regressor, sequentially for many regressors. It is performed by first calculating the correlation between each regressor (i) and the target using

$$\frac{((M[:, i] - \text{mean}(M[:, i])) * (y - \text{mean}_y))}{(\text{std}(M[:, i]) * \text{std}(y))}$$

It is first converted to a F score and then to a p-value.

A random forest regressor (RF) [168, 169] given by

```
RandomForestRegressor(n_estimators=10, criterion='mse', max_depth=None,
min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0,
max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0,
```

```
min_impurity_split=None, bootstrap=True, oob_score=False, n_jobs=1,  
random_state=None, verbose=0, warm_start=False)
```

results in the meta estimator that fits many sub classification decision trees. The classifier uses the average of probabilities predicted by each sub decision trees to get the final classification for each class.

Then, LassoLarsCV [285], a cross-validated Lasso, using the LARS algorithm given by

```
LassoLars(alpha=1.0, fit_intercept=True, verbose=False, normalize=True,  
precompute='auto', max_iter=500, eps=2.220446049250313e-16, copy_X=True,  
fit_path=True, positive=False)
```

The optimization objective function to minimize is:

$$\min_{\omega} \frac{1}{2n_{samples}} \|M\omega - y\|_2^2 + \alpha \|\omega\|_1$$

Where, α is the regularization strength, y is the output variable and ω are the coefficients of the linear model. The lasso estimate thus solves the minimization of the least-squares penalty with $\alpha \|\omega\|_1$ added, where α is a constant and $\|\omega\|_1$ is the l_1 norm of the parameter vector.

In the end, mean rank is calculated by averaging the scores from each algorithm and features with mean rank lower than 0.30 are filtered out.

8.5.2 Multivariate analysis of covariance

We also filtered out miRNAs and piRNAs that were confounded by age and gender by applying multivariate analysis of covariance (MANCOVA) [286, 287] on the reliable features with significant values less than 0.05.

8.5.3 Model selection and performance

The performance of selected features is then evaluated in an independent test cohort. The average Error and average number of trees parameter were calculated using a 10 fold cross validation on CV data which was obtained from the training data.

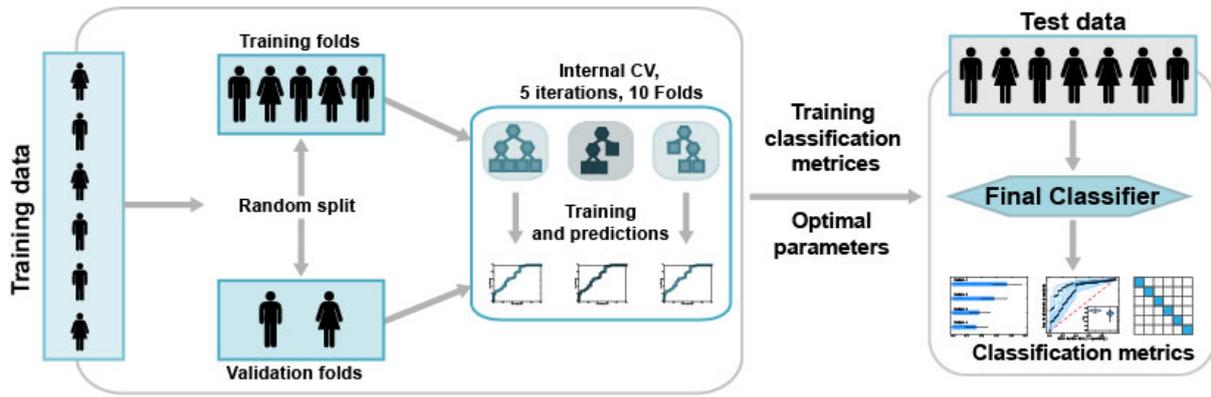


Figure 8.5-3 Optimal model selection and performance evaluation

The random forest model implemented in R (v 3.2.2) with the package randomforest (v 4.6.14) is trained with stratified sampling and class weights of 0.5 and 1.0 were used for the control class and the Alzheimer’s disease class to minimize the false negatives. Prediction on the training and test data is performed using the predict function from randomforest and roc function is used to get the test performance from the untouched test cohort data. The AUC values are plotted using the pROC [288] package with 500 stratified bootstrap [289] iterations to provide the confidence interval computed with Delong’s method [290] for the AUC values. Smoothing of the AUROC curve was performed by calculating the α and β coefficients of a linear regression line of the smoothed curve which is given by

$$\phi^{-1}(SE) = \alpha + \beta\phi^{-1}(SP)$$

Where, ϕ is the normal quantile function value of sensitivities (SE) and specificities (SP).

The variable importance is calculated from the importance function in randomforest package. The total decrease in node impurities is measured by taking the average node impurities on every split of the variable for all the trees generated in the random forest model. This node impurity is measured by Gini index (I_G) which is given by

$$I_G(p) = 1 - \sum_{i=1}^C p_i^2$$

Where, C is the number of classes in the total samples set with $i \in \{1, 2, \dots, C\}$, and p_i are the number of samples that belong to the particular class i in the total sample set. The Gini index for each variable is obtained and a Barplot of decreasing variable importance is plotted.

8.6 Handling of mice and nuclei preparation

8.6.1 Ordering and upkeep of mice

All mice used in the pilot experiment were adult male mice bought from JANVIER LABS. These wild type (WT) C57B16/J mice were pathogen free and were housed in the standard conditions with *ad libitum* access to food and water. The animal husbandry facility is located in the basement of the European Neuroscience Institute (ENI), Göttingen that follows and ensure ethical experimental practices in accordance with animal protection laws and were approved by the Veterinary Institute in Oldenburg.

8.6.2 Tissue collection

Animals were sacrificed by cervical dislocation and the whole brain was isolated in ice-cold Dulbecco's Phosphate Buffered Salt (DPBS, PAN-biotech GmbH) supplemented with EDTA-free protease inhibitor cocktail (Roche). The CA1 and ACC regions were isolated, snap frozen in liquid nitrogen, and stored at -80°C .

8.6.3 Nuclei isolation and sorting

In order to look at the interactions from neuronal and non-neuronal population of the cells for CA1 region, fluorescence-activated cell sorting (FACS) is used to sort the isolated nuclei. 10 mice were pooled in each group to extract a sufficient amount of chromatin. Frozen mouse tissue from 10 mice was homogenized with the help of a micro pestle in low sucrose buffer (0.32 M Sucrose, 10 mM HEPES pH 8.0, 5mM CaCl_2 , 3 mM $\text{Mg}(\text{CH}_3\text{COO})_2$, 0.1 mM EDTA, 1% Triton X-100) and crosslinked with 1% formaldehyde (Sigma-Aldrich) for 5min at room temperature. The reaction was quenched by incubating for 5 min with 125mM glycine. The nuclei were centrifuged and re-suspended in 3ml of low sucrose buffer with protease inhibitors (Roche Complete) using. Nuclei were further homogenized mechanical homogenizer (IKA T 10 basic Ultra-Turrax) for half a minute. The nuclei were purified through a sucrose cushion (10mM HEPES pH 8.0, 1M sucrose, 3 mM $\text{Mg}(\text{CH}_3\text{COO})_2$, 1mM DTT) by centrifugation (3,200 rcf for 10 min in Oak Ridge centrifuge tubes). 6ml of cushion was used per 1.5 ml of lysate. The nuclear pellet was re-suspended in PBS, and aggregated were cleared by filtering through 70 μm filter. The nuclei were stained with anti-NeuN Antibody, clone A60 (Trade name: Chemicon and catalogue number MAB377) diluted 1:500 in PBS-T (0.1% Tween 20 in PBS) with 5% BSA and 3% goat serum, incubated for 30 min at 4°C . The nuclei were washed 4 times with PBD-T and stained for 15min with anti-mouse Alexa 488 (Life Technologies) diluted 1:1000. The nuclei were washed once with PBS-T and

re-suspended in PBS-T with 5% BSA. Nuclei sorting was performed using the FACS machine (BD FACSAria III) which is a 13 color 15 parameter high speed digital benchtop cell sorter with four excitation lasers that are adjustable.

8.7 Chromosome Conformation Capture Sequencing (3C-Seq)

8.7.1 Preparation of 3C library

The sorted nuclei (NeuNpos and NeuNneg) from CA1 were then crosslinked with formaldehyde (1% final concentration) at room temperature (25°C approximately) for 10 mins. The crosslinking reaction was stopped by adding glycine (final concentration 125mM). The crosslinked nuclei were pelleted down and washed with 1X PBS. The crosslinked nuclei were then again pelleted, frozen in dry ice and stored at -80°C.

For the 3C library preparation, both samples were processed in parallel under similar experimental conditions. Briefly, the crosslinked nuclei were resuspended in 200 µL of NEBuffer DpnII (50 mM Bis-Tris-HCl, 100 mM NaCl, 10 mM MgCl₂, 1 mM DTT (pH 6 at 25°C)). The chromatin was then solubilized by adding SDS (0.1% final concentration) and incubating at 65°C for exactly 10 min. The SDS is quenched by Triton X-100 with 1% final concentration. The chromatin was next digested with 600 units of DpnII (NEB) restriction enzyme and incubated overnight at 37°C in a thermomixer (Eppendorf). DpnII was heat inactivated the following morning by incubating the digestion mixture at 65°C for 30 minutes. A ligation mixture (cocktail mix in Table 8-2 The ligation master mix) was prepared containing 30U of T4 DNA ligase (Invitrogen) to the samples and incubated for 4 hours at 16°C. The ligated samples were reverse crosslinked by adding 600 µg of Proteinase K and incubating them at 65°C overnight. DNA was purified using standard phenol-chloroform extraction followed by DNA precipitation using sodium acetate pH 5.2 (to a final concentration of 0.3M) and 2.5 volume of ice-cold absolute ethanol. The mixture was incubated at minus 80°C for 1 hour and then centrifuged at 12000 x g for 20 minutes at 4°C. The pellet was then resuspended in TE buffer and passed through Amicon-Ultra 30K column (0.5 mL) and washed with TE buffer pH 8 for 3 times. The washed DNA which contains the 3C library was collected by inverting the column on a fresh collection tube placed by spinning at 1000 x g for 2 minutes. The NeuN-3C library was then subjected to 3C QC step.

Table 8-2 The ligation master mix

1 reaction (µl) per 15ml tube

10X ligation buffer	447
10mg/ml BSA	48
100mM ATP	48
Water	4023
T4 DNA ligase (ligase separately added into individual tubes)	30

8.7.2 Quantification of 3C library

Successful preparation of 3C library was validated using 3C qPCR with standard conditions. Due to the very low yield of 3C library, QC of only one sample (NeuNeg) could be performed. 3C primers were designed (Table 8-3) on a gene desert region of the mouse genome (mm9) in Chr 5 near the DpnII cut site to probe for a short range interaction. The qPCR analysis showed the positive amplification curve (Fig 7.3.2). The qPCR product was run on an EtBr stained 2% agarose gel. A DNA band of 384 bp was detected which conform to the expected amplicon size showing the successful preparation of the NeuN-3C library. One pair of primers (primer 25 and 26) probing an interaction 4.1 kb apart were used for 3C qPCR using 3 μ L of NeuN-3C library with expected Amplicon size 384 bp. The location of the primers and restriction enzymes around the primers and between them are shown in Figure 8.7-1.

Table 8-3 3C primers details

#	Name of the Primers	Chr5 Start	Chr5 End	Nucleotide sequence (5'-3')	Tm (°C)	GC (%)	bp
25	Mouse_MMC9_DpnII_Fwd1	133376673	133376693	CAGGACCTGATTGCTAAACTG	53.4	47.6	21
26	Mouse_MMC9_DpnII_Rev1	133384528	133384549	TGTATACCCGCACACAATGAAT	54.5	40.9	22

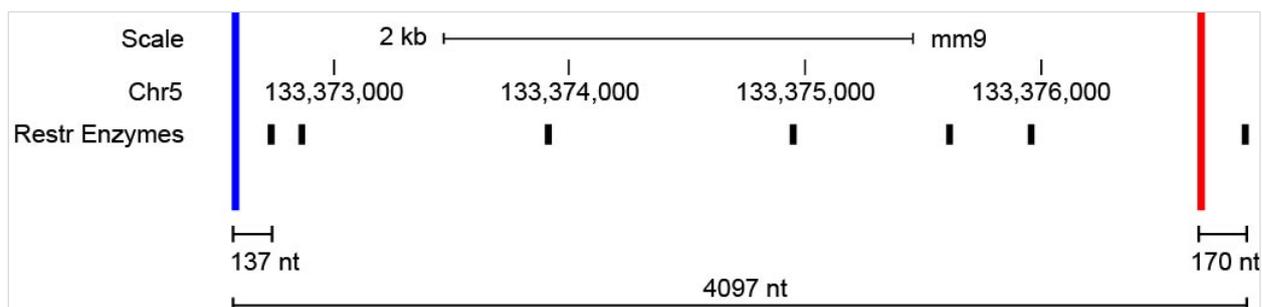


Figure 8.7-1 Genomic location of primers and restriction enzymes

After the primer design and checking for an interaction, 3C reactions were performed. 3 μ L of the NeuN-3C DpnII library were used per reaction (Table 8-4)

Table 8-4 Reaction usage

	Components	1X (μ l)
1	KAPA SYBR mix (2X)	10.0
2	Forward 3C primer (10uM)	1.0
3	Reverse 3C primer (10uM)	1.0
4	Template (3C library)	3.0
5	DNase and RNase free water	5.0
	Total volume (per reaction)	20.0

Then PCR amplification was performed using the parameters (Table 8-5). The fluorescence signal measured in relative fluorescence units (RFU) and the number of PCR cycles are plotted in a PCR amplification curve (Figure 8.7-2). The background fluorescence was plotted as green line.

Table 8-5 PCR tubes

#	Step	Temperature ($^{\circ}$ C)	Time	Number of cycle(s)
1	Initial denaturation	98	3 mins	1
2	Denaturation	98	15 seconds	1
3	Annealing and Elongation	60	30 seconds	54

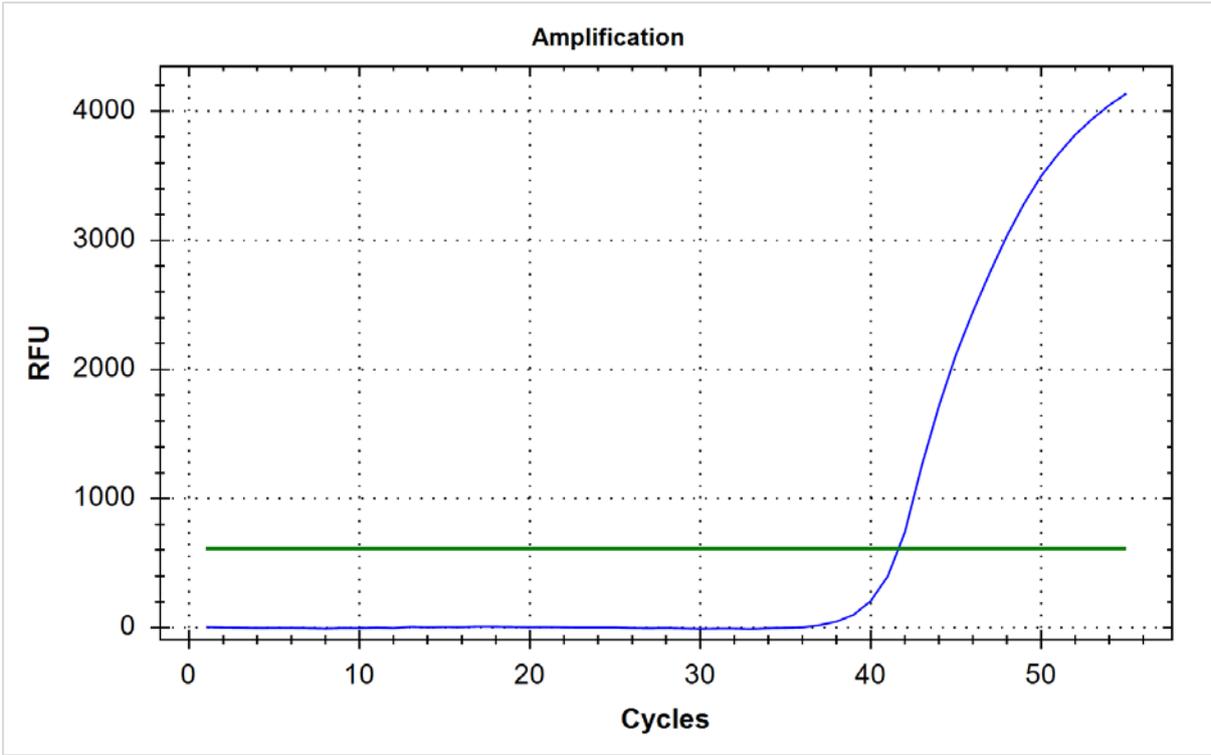


Figure 8.7-2 PCR amplification curve

Table 8-6 Standard curve range

Well	Fluor	Target	Sample name	Cq
C03	SYBR	25-26 (nearby)	250516 NeuNneg DpnII-3C library (0.6ng)	41.63

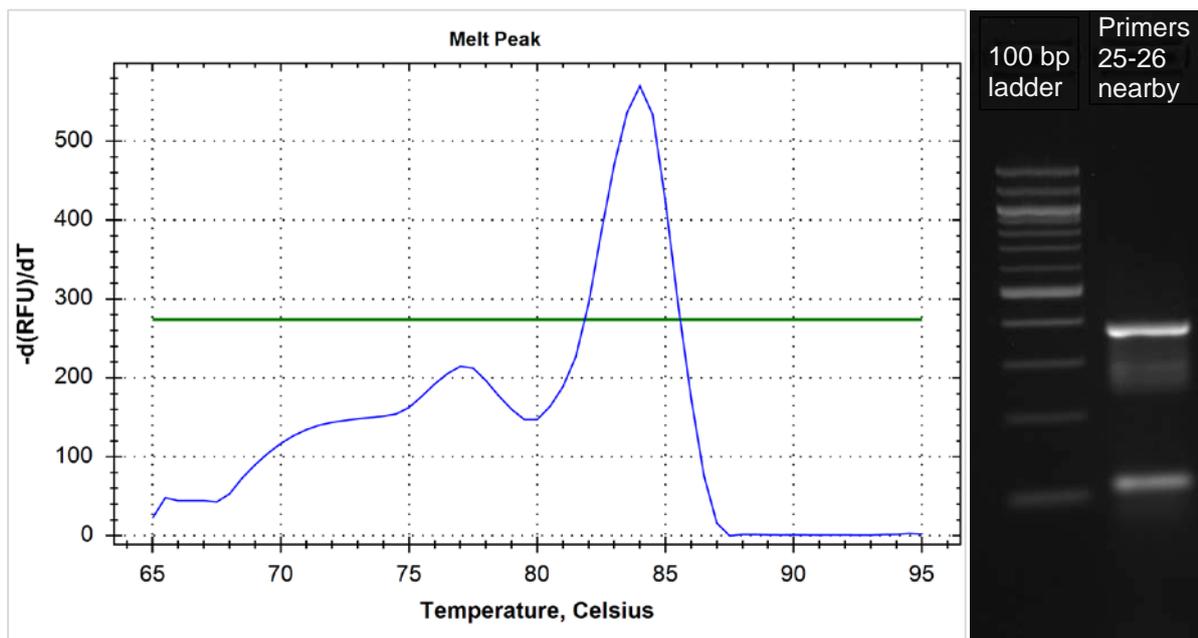


Figure 8.7-3 Melting curve and 3C interaction on a gel

The C_q value of the standard curve is calculated and shown in Table 8-6

The melt curve analysis program (Figure 8.7-3):

- 65°C for 5 seconds
- 95°C for 0.5°C/cycle
- 12°C for ∞

The negative rate of change of fluorescence units compared to the temperature (°C) is plotted in Figure 8.7-3 A showing a single peak. The qPCR product was run on a 2% agarose gel for the confirmation of the short range interaction (3C product) of the expected size from the 3C library from Figure 8.7-3 B.

8.8 Sequencing of 3C-seq library

8.8.1 Preparation of 3C library

The sorted nuclei were first homogenized by Dounce homogenizer to lyse almost all cells. The quality of homogenization was checked under the microscope. It is then crosslinked with 1% formaldehyde at room temperature (25°C approximately) for 10 mins. The crosslinked chromatin was digested with DpnII (GATC) restriction enzyme overnight at 37°C. DpnII was inactivated the next morning and a ligation mixture was added and samples were kept for 5 hours for DNA ligation. The samples were heated at 65°C with proteinase K treatment

overnight to reverse crosslink the ligated samples. DNA was precipitated and washed thoroughly.

8.8.2 Quantification of 3C library and library preparation

Short and long looping interactions were quantified using PCR with standard conditions (50 cycles). Due to the very low yield of genomic DNA, only one QC of (NeuN minus) could be performed. Small distance primers were designed at the DpnII site to check for short range interactions. The PCR product was run on 2% agarose gel with the water control and the band intensities were quantified with the standard imaging software. The titration curve of PCR product vs. input DNA should be plateaued for a successful 3C library preparation. The library preparation was performed with NEBNext® Ultra[®] DNA Library Prep Kit for Illumina (cat. no. E7370L). For PCR primers primer with index 2 for samples 3C CA1 pos and index 4 for 3C CA1neg was used. Finally, the sequencing was performed on Illumina HiSeq2500.

8.9 Processing of 3C-seq sequencing data

8.9.1 Generation of raw fastQ files and quality control

The generation and quality control of the raw fastQ files for the 3C-seq sequencing data is performed in the same way as mentioned in the section 8.2.1 above for the generation of raw fastQ files and section 8.2.2 above for the quality control reporting of the raw fastQ files. The quality control statistics were presented in Table 9-5, Figure 9.6-1 and Figure 9.6-2. Additional file conversions and quality controls are described in the subsequent sections with the subsequent analysis.

8.9.2 Iterative mapping to the reference genome

The 3C-seq paired end data is unique in the sense that the paired end mode cannot be used even when the reads are paired end as they do belong to the same fragment. The two ends of the read are in-fact comes from two different genomic locations. In order to optimize the mapping, they are mapped like single end reads and later the information is combined to get the chromatin looping interaction. Another, important point about the 3C-seq data is that the DNA fragments that goes into the sequencing contain a restriction enzyme junction. With the four-base cutter DpnII, an average fragment size is around 300-500 bp. The fragments are then sequenced with PE 2x150 bp paired end sequencing. With these parameters, around a quarter of the reads contains a restriction enzyme junction as only a part of the fragment is sequenced.

The mapping is performed iteratively starting with the first 25 bp of the reads from the 5' end are mapped to the mouse reference genome. The assembly of the mouse genome (mm9, NCBI Build 37 in July 2007) as well as repeat annotations and GenBank sequences were downloaded from <http://hgdownload.soe.ucsc.edu/goldenPath/mm9/bigZips> and random contigs were filtered out using Unix command `grep [259]`.

The mapping for each read pair is individually performed using bowtie (v 1.1.2) [256] with fine-tuned parameters for optimal mapping. The first iteration is conservative in nature, allowing no mismatches during the mapping of the reads. The parameters are:

- `-v 0` is used is set for end-to-end hits with zero mismatches to keep the mapping very conservative for this iteration
- `-p/--threads 8` are the number of alignment threads for faster mapping
- `-S/--sam` is set for writing all the mapped reads into the Sequence Alignment Map (SAM) format
- `-a/--all` is set to get all alignments for every read
- `-m 1` is set to remove all alignments that have more than one reportable alignment
- `--un <unmapped_fastq>` to save all the unmapped reads for next iteration where these reads will be mapped again but with a longer read length

The resulting SAM files are converted to BAM (Binary SAM) files using samtools (v 1.3.1) [262] with the command:

```
samtools view -S -b input.sam > output.bam
```

The resulting BAM file is indexed using the command:

```
samtools index output.bam
```

In the next and all subsequent iterations until the read length is reached or all the reads are mapped uniquely, the reads are trimmed with 5 bp longer than the previous iteration and mapped with one mismatch using bowtie (v 1.1.2) with same parameters except

- `-v 1` is used is set for end-to-end hits with one possible mismatch for a little relaxed mapping strategy

The individual mapped BAM files from every iteration was combined into one BAM file and indexed using samtools. All the unmapped reads were discarded in this case. All the mapping iterations were performed on the GWDG's high performance computing cluster (HPCC) [254] using gwdu103.gwdg.de as the master node. The jobs were launched on to

“fat” queue with a maximum of 16 cores, a maximum time constraint of 48 hours and maximum memory constraint of 32GB.

8.9.3 Reads assignment to restriction enzyme fragments and filtering

The list of restriction enzyme fragment is created by assigning all the fragments that are separated by the restriction enzyme DpnII (GATC) which is a four-base cutter. For DpnII, there will be approximately 6 million restriction enzyme fragments. After the mapping is finished, the reads are also assigned to these genomic restriction enzyme fragments. This is important as it will be used to filter out the various 3C-seq library preparation and sequencing related artifacts are shown in Figure 8.9-1. Following the assignment of all the mapped reads to the restriction enzyme fragments, all the reads that do not fall into the unique valid pairs are filtered out and only the unique valid pairs are considered for the downstream analysis. The results of the mapping and all the artifacts are presented in Table 9-7.

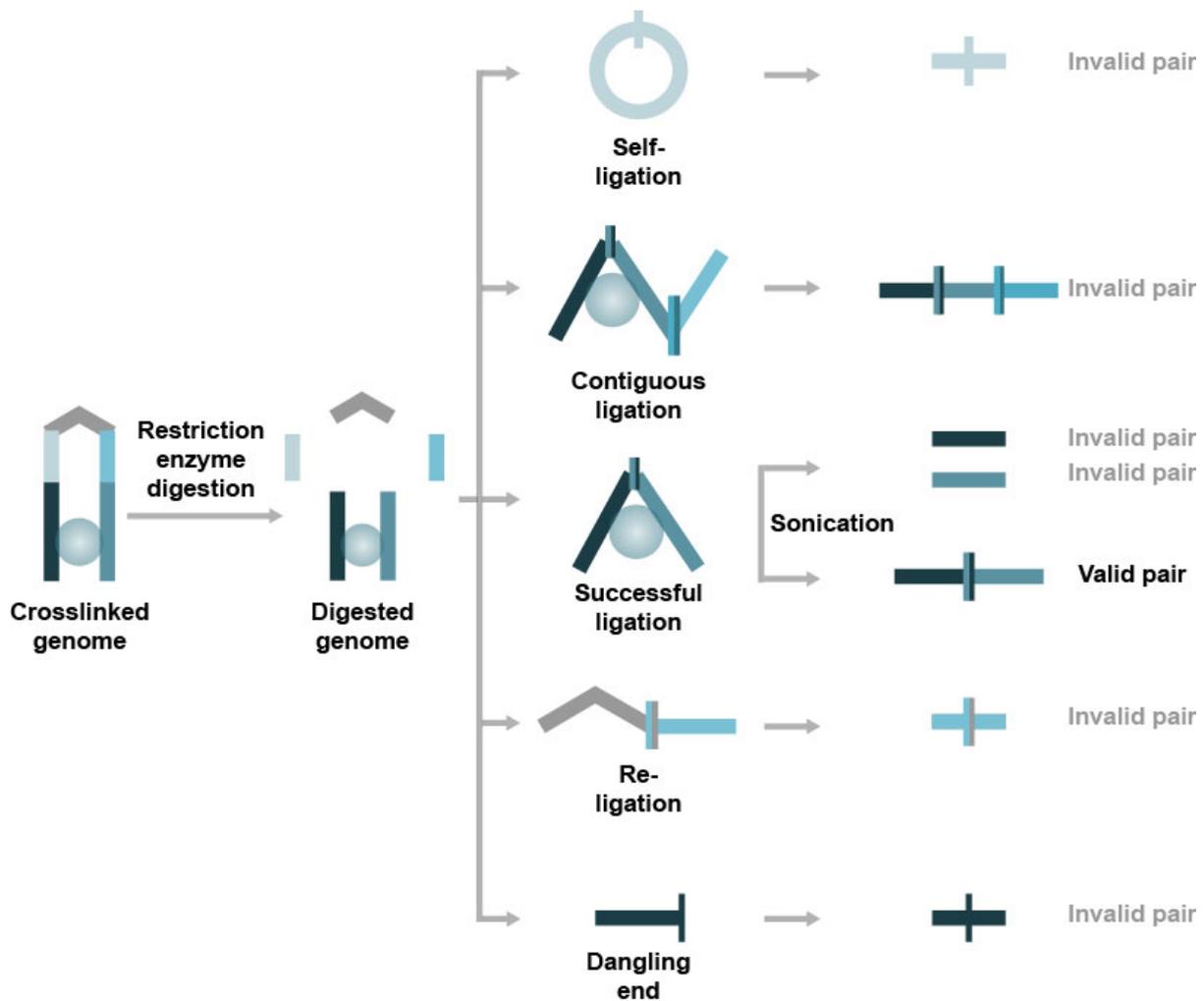


Figure 8.9-1 3C-seq mapping artifacts (Adapted from Fides Lay's HiC workshop [291])

8.9.4 Assign reads to genomic bins and perform iterative correction

After the identification of the unique valid pairs, the reads are binned to a particular resolution for further analysis. As mentioned in the previous section, there are approximately 6 million genomic fragments identified by 3C-seq. This makes approximately 36 trillion possible pairwise interactions. Even with the advanced next generation sequencing, working with the fragment level resolution requires a very high depth of sequencing, which need a high amount of input material and can be very expensive. Due to the lack of enough reads, a lot of fragments and their corresponding pairwise looping interaction cannot be identified. To get an abstract idea, instead of working with fragment level resolution, the data were binned to a certain resolution, for example 1 MB bins or 500 KB bins [236]. This will also level out outliers and reduce noise. All the restriction fragments whose center falls into the bin is assigned to the bin. The 3C-seq neuronal and non-neuronal data are binned to 100 kb, 200 kb, 500 kb, and 1 Mb sized bins. Due to an insufficient number of unique valid pairs, 100 kb

bins and 200 kb bins do not provide enough information to call topologically associating domains or get any conclusion from the binned data. In this pilot study, 500 KB and 1 MB data are used. After binning is performed, low quality bins that do not convey any information are filtered out (marked as NaNs) and not be used in further downstream calculations. After the filtering at the bin level, Iterative correction of the matrix is performed. This removes various known bias like variable length of fragments and GC content [292, 293] and unknown bias [294]. The iterative correction and normalization is performed by applying Sinkhorn–Knopp balancing algorithm [295] to the binned data [296, 297]. The normalized matrix $\Psi_{normalized}$ being doubled stochastic with both the sum of rows and column being 1 is given by

$$\Psi_{normalized_{i,j}} = \Psi_{raw} \Psi_{diagonal_{i,i}} \Psi_{diagonal_{j,j}}$$

Where, Ψ_{raw} is the initial raw count matrix binned at a particular resolution (1 Mb or 500 kb), $\Psi_{diagonal}$ is the bias matrix and i, j are the bins of the matrix. These matrices are all symmetric matrix and can be visualized as square heatmaps (Figure 9.7-1) or triangular heatmaps (Figure 9.7-5).

8.9.5 Defining topological regions and domains

Within the various levels of chromatin organization, at a certain level of abstractions lies the few 100s KB structures that have a specific looping interaction profile. These domains are called topologically associating domains or TADs [209, 210, 292] (6.7.1.3 above). Both the insulation score [298] and directionality index [209] methods are used to identify TADs for the matrices binned at 500 kb resolution.

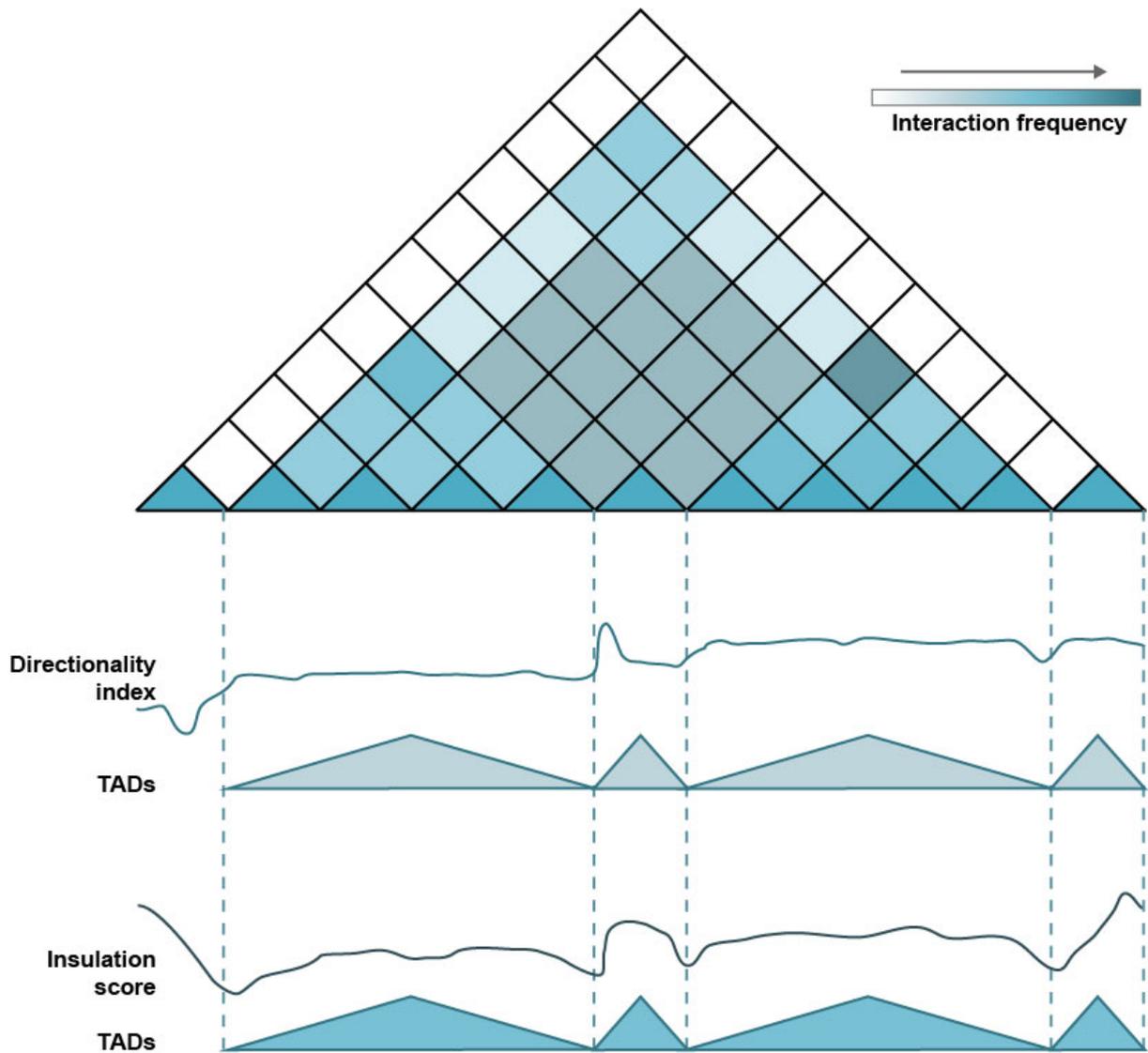


Figure 8.9-2 Schematic to identify topologically associating domains (TADs) by two competing methods

TADs used in the study for the neuronal and non-neuronal populations are called with insulation delta span which is the insulation delta window of size 2 Mb and insulation square of size 1.5 Mb for the 500 kb resolution TADs identification.

9. Results

9.1 CSF exosomes and small non-coding-RNAome

9.1.1 Analysis of isolated RNA from CSF exosomes and exosomal free CSF fraction

We started the analysis with the aim to confirm that small non-coding RNAs are enriched within CSF exosomes. To this end, we obtained CSF from 36 individuals and isolated exosomes, which were confirmed via immunoblot analysis of the marker proteins Flotilin-2 and CD-63, electron microscopy and via analysis of particle size (Figure 9.1-1 A, B, C). Next, we analyzed isolated RNA from these exosomes and from the exosomal free CSF fraction via a bioanalyzer microfluidics device. The corresponding electropherograms show that a significant amount of RNA with a particularly high peak of a small RNA species is detectable in the exosomal CSF fraction. In contrast, comparatively little RNA was obtained from exosomal free CSF. Treating RNA samples obtained from CSF exosomes with DNAase did not affect RNA integrity while treatment with RNAase eliminated the small RNA peak (Figure 9.1-1 D, E, F, G). This data confirms that the human CSF contains exosomes that carry small RNAs.

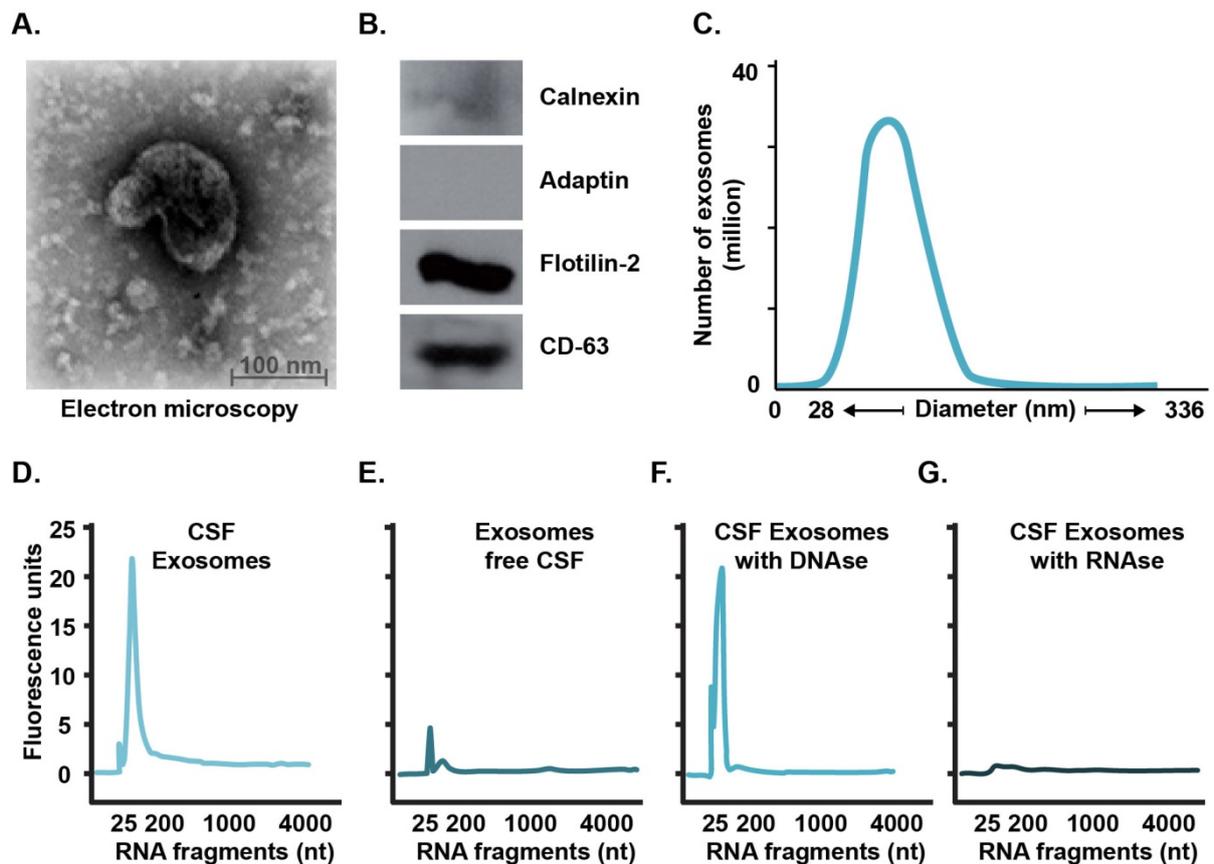


Figure 9.1-1 Profile of exosomal RNA from human CSF A. Exosomes isolated from human CSF was analyzed via EM. B. Fragment size using a nanosight instrument and via immunoblot for exosomal marker proteins. C. Electropherogram showing the profile of RNA isolated from exosomes. D. Electropherogram showing the profile of RNA isolated from exosome free CSF. E. Electropherogram showing the profile of RNA isolated from exosomes treated with DNAase (F) and RNAase (G).

9.1.2 miRNAs and piRNAs in CSF exosome small non-coding-RNAome

The advent of the next generation sequencing (NGS) technologies and the decreasing cost of sequencing, enabled the scientific community to study small RNAs in detail. Analysis of small RNAs is challenging due to the short and repeatable sequences. There are many tools to analysis small RNA data [299]. However, there were many shortcomings associated with these existing tools and to get a global and complete picture of the small RNAs I developed the small RNA mapping and analysis pipeline: GJSrMap (<https://github.com/gauravj49/gjsrmap>). The pipeline is available as a standalone installation (cloned from github) or through the webtool which will be accessible in the future from the

webpage (<http://www.fischerlab.uni-goettingen.de/tools/gjsrmap>). Details and inner working of the pipeline is mentioned in the methods and material section 8.3 above.

We analyzed the small RNA content of CSF exosomes obtained from two independent cohorts of Alzheimer's disease patients and individuals that did not suffer from any neurodegenerative disorder and performed small RNA-sequencing according to DZNE RNAome database SOPs using Illumina reagents on an in-house Illumina HiSeq 2000. All the analysis of the data is performed in-house. Majority of CSF smallRNAs are miRNAs and piRNAs (Figure 9.1-2 A). Amongst the 5 highest expressed miRNAs in CSF exosomes were miR-10a-5p, miR-22-3p and miR-204-5p that have been previously linked to memory function and/or neurological diseases [300-304]. Recent studies also show the role of piRNAs in neurological diseases like Alzheimer's disease [122] and cancer [305]. The 5 top expressed piRNAs showed comparable expression levels to miRNAs (Figure 9.1-2 B). Compared to the genomic piRNAome, the CSF piRNAome has a completely different profile. A majority of genomic piRNAs are coming from distal intergenic regions and only a fraction comes from 1st exon. On the other hand, the majority of piRNAs in the CSF piRNAome comes from 1st exon and a small fraction come from the distal intergenic regions (Figure 9.1-2 A). This is quite interesting as a recent study [306] shown that the small non-coding RNAs derived from the first exons of protein coding genes are better at distinguishing between patients compared to healthy individuals. This suggests that piRNAs may have a potential role in classifying and predicting the disease pathology.

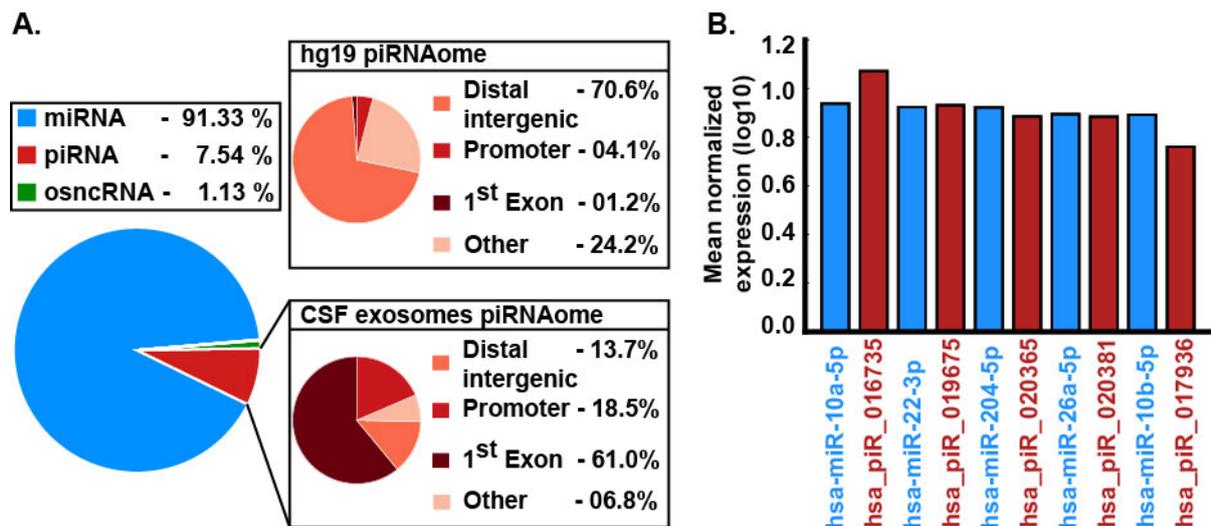


Figure 9.1-2 miRNAome and piRNAome in CSF exosomes

A. Pie chart (Left) representing small non coding RNAs in human CSF exosomes. Here miRNAs and piRNAs make up almost entire small non coding RNAome. Pie chart (top right) showing genomic annotation of human piRNAome. Here, the majority of piRNAs lies in the distal intergenic region. Pie chart (bottom right) shows genomic annotation of expressed

piRNAome in the human CSF exosomes. Here, the majority of piRNAs lies in the 1st exon of genes. B. Top 5 expressed miRNAs (Blue) and piRNAs (Red) in human CSF exosomes. Here y-axis represents the mean normalized expression and x-axis represents miRNAs\piRNAs.

9.1.3 miRNAs arm expression in CSF exosomes and brain tissues

Mature microRNAs are derived from both 5' and 3' ends of their pre-microRNAs which are known as 5-p and 3-p arms. The 5-p and the 3-p arm of a given miRNA are often expressed at different levels. We need to quantify the expression of both arms to define if the 5-p, 3-p or both arms of a given miRNA are biologically active since the precise biological function of many miRNAs are still unclear. The highest expressed arm is often considered to be the active one. It was previously suggested that the inactive arm of a given microRNA is sorted into exosomes as a cellular clearance mechanism, while at the same time there is evidence that transport of molecules from one cell to another via exosomes serves important biological functions.

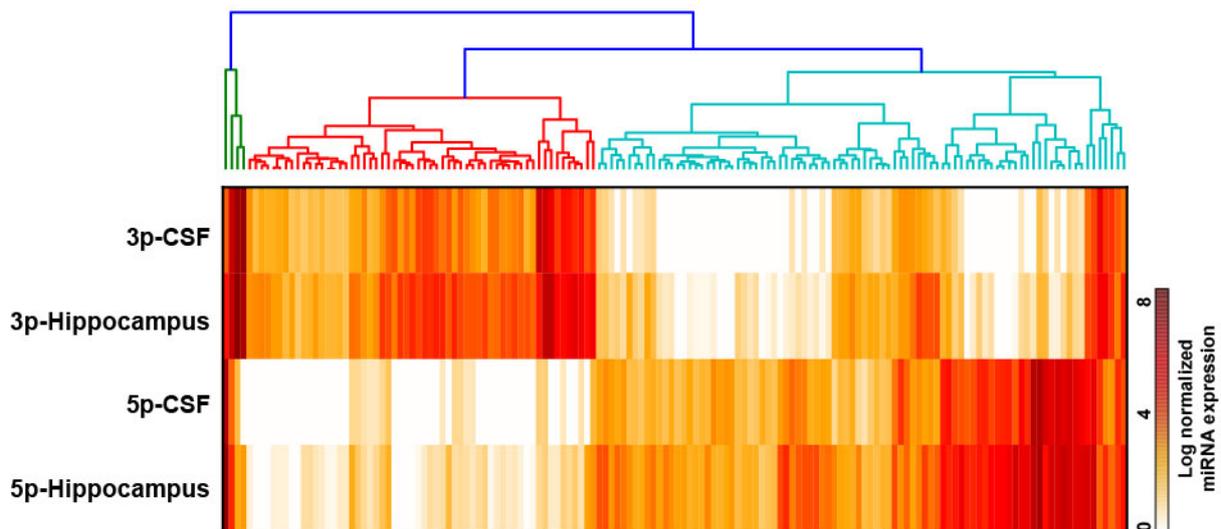


Figure 9.1-3 Heat map showing expression values of the 3-p and 5-p arms of all miRNAs detected in human CSF and in the human cortex (Brodmann Area 9)

Thus, we performed small RNA-sequencing from postmortem human prefrontal cortex (Brodmann area 9) and compared the expression pattern of the 5-p and 3-p arms for all miRNAs detected in CSF to the corresponding expression pattern observed in the human brain. The pattern of the 3-p vs. 5-p arms of microRNAs detected in CSF was similar to the pattern observed in postmortem human brain tissue and confirmed that for the majority of the miRNAs only one arm was highly expressed (Figure 9.1-3). This finding supports the view that the small non-coding-RNA content of CSF exosomes – at least in part – resembles the small non-coding-RNA content of the parent cells. The data have to be interpreted with

care, since the parental cell for CSF detected exosomes is not precisely known and likely includes not only various brain regions but also other cells [307].

9.1.4 Cellular small non-coding-RNAome and exosomal small non-coding-RNAome

After looking into the expression of mature microRNA arms and profiling the small non-coding-RNAome in exosome, the follow-up analysis is to study the relationship between cellular small non-coding-RNAome and exosomal small non-coding-RNAome which may lead to the understanding of cellular small non-coding-RNAome from the profile of exosomal small non-coding-RNAome.

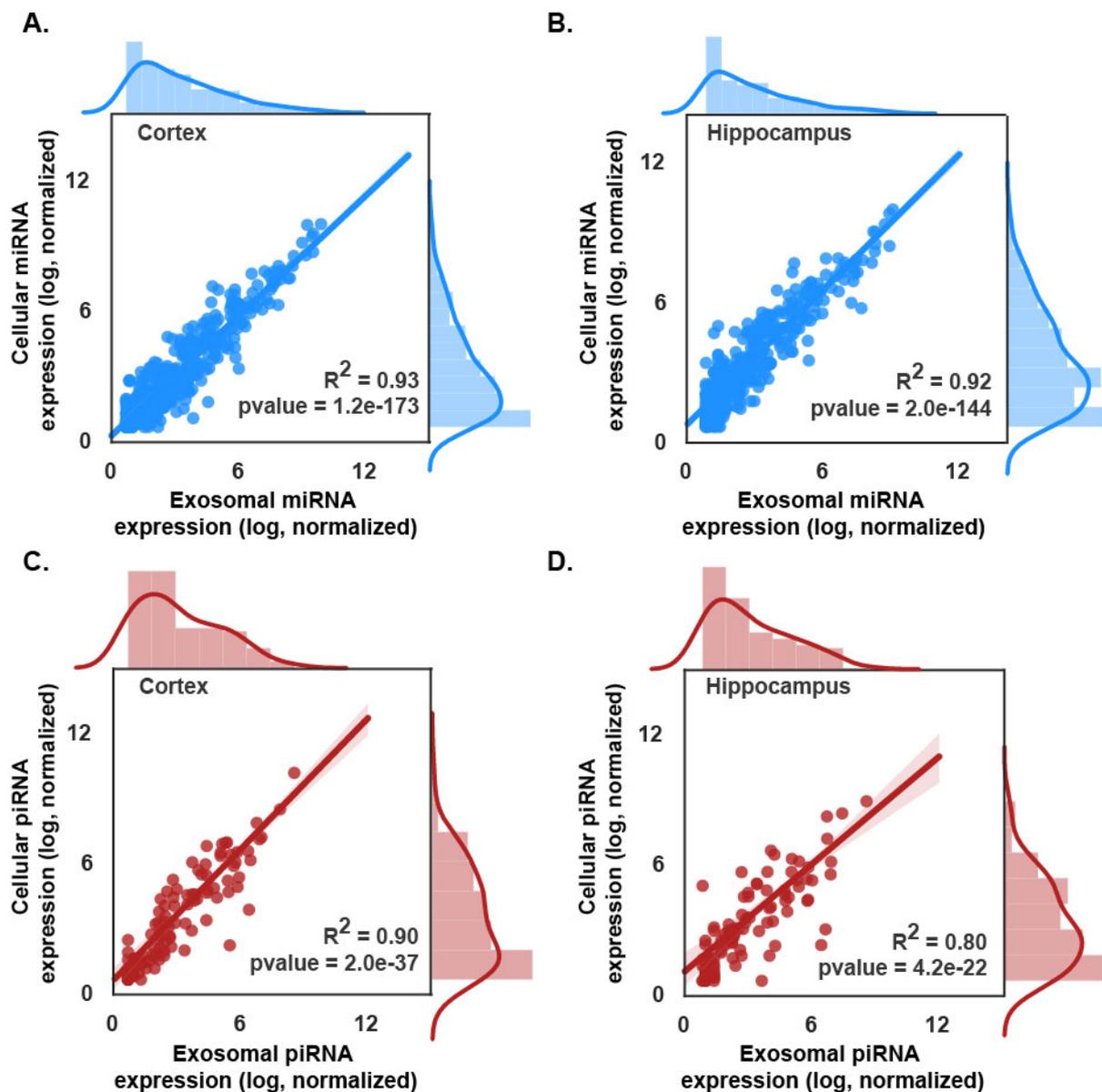


Figure 9.1-4 Pearson correlation between miRNA and piRNA expression values of hippocampal and cortical neurons against the miRNA and piRNA expression in the exosomes released from these cells. Here x-axis represents the normalized expression of exosomal small RNA and the y-axis represents the normalized expression of cellular small RNA. The Pearson correlation coefficient and significance level are mentioned in the plot.

To address the questions if the small non-coding-RNA content of exosomes released from neurons allows making conclusions about the corresponding cellular small non-coding-RNAome, we decided to test the correlation of miRNA and piRNA expression in primary cortical and hippocampal neurons and in their corresponding exosomes. Exosomes were isolated from the media supernatant of the corresponding cells. Subsequently the exosomal and the cellular RNA was prepared and subjected to smallRNA sequencing. For both, miRNAs and piRNAs, we detected a highly significant correlation between cellular and exosomal fractions (Figure 9.1-4).

9.2 A CSF miRNAs/piRNAs signature to diagnose Alzheimer’s disease patients along with the available clinical markers

9.2.1 Demographic information for the samples included in the biomarker signature analysis

The study was performed on two independent cohorts of Alzheimer’s disease patients and controls. Cohort 1 consisted of 23 Alzheimer’s disease patients and 38 control individuals after removing outliers and age-matched control individuals that did not suffer from any neurodegenerative disorder (Table 9-1).

Table 9-1 showing demographic information for the samples included in the biomarker signature analysis. Abbreviations: AD, Alzheimer’s disease; F, female; M, male.

Group	Gender				Age	
	Control		AD		Control	AD
Cohort 1 (Signature identification)						
Cohort 2 (Signature testing)						

Samples were collected at the University Medical Center Göttingen (Germany) between Jan 2012 and Mar 2013. Cohort 2 consisted of 19 Alzheimer's disease cases and 44 control individuals after removing outliers and age-matched control individuals from which CSF samples were collected at the University Medical Center's in Göttingen, Tübingen, and Kassel (Germany) between Apr 2013 and Oct 2014.

9.2.2 Performance of clinical signatures on replication cohort

The hallmarks of the Alzheimer's disease are the presence of Amyloid- β plaques and tau tangles in the brain. The traditional CSF clinical biomarkers to identify Alzheimer's disease correspond to A β 1-42, total-Tau (tau) and phospho-Tau (pTAU) [308]. Various studies [309-315] have shown that Alzheimer's disease patients have higher levels of pTAU (Figure 9.2-1 A).

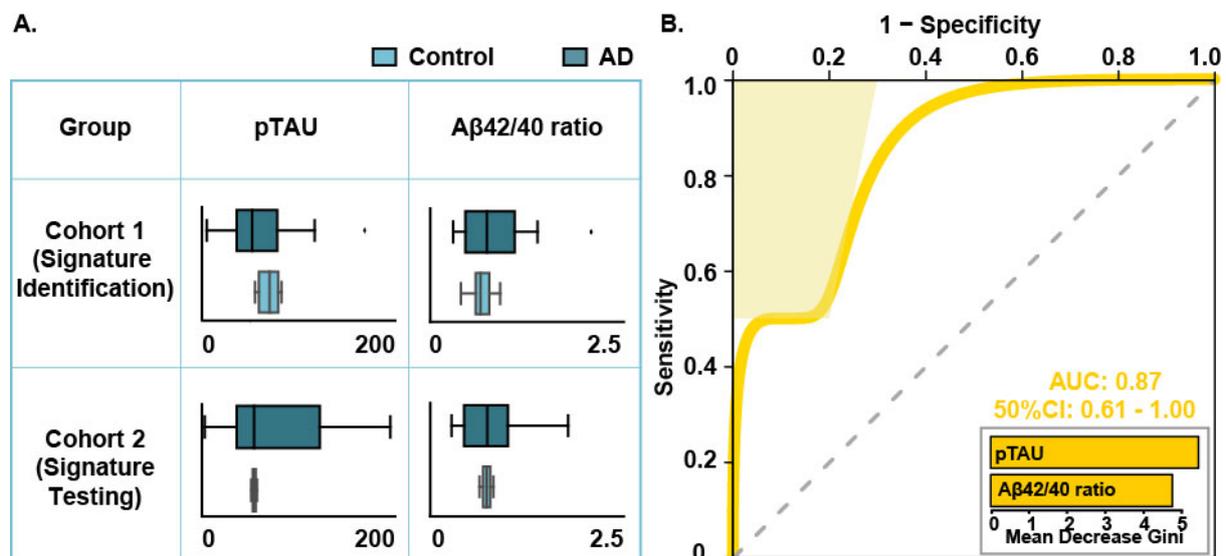


Figure 9.2-1 showing summary statistics about the clinical markers used in the study. Abbreviations: Alzheimer's disease, Alzheimer's disease; A β 40 and A β 42, amyloid beta peptide 40 and 42 amino acids; pTAU is phosphorylated TAU. B. Receiver operating characteristic (ROC) plot was obtained during the performance testing of the clinical markers on replication cohort 2 with mean test AUC of 0.87. The training was done on cohort 1 with a tenfold cross validation. An inset plot showing the variable importance which explains how good a variable in classifying the data. Here x-axis represents the mean decrease in Gini and the y-axis shows each variable which are ordered top-to-bottom as most- to least-important. In this case, pTAU is most informative.

A random forest machine learning classification approach with a 10 fold cross-validation protocol was used to assess the combined performance of both clinical markers (pTAU and

A β 42/40 ratio). Clinical marker pTAU was found to be more informative than the A β 42/40 ratio (Inset plot Figure 9.2-1 B). The performance of this model was then tested on cohort 2 which allowed us to distinguish Alzheimer's disease patients from controls with an AUC of 0.87 (Figure 9.2-1 B).

9.2.3 Identification of small non-coding RNAs signature

Small RNA sequencing data obtained from cohort 1 was used to perform an iterative feature selection protocol (section 8.5.1 above and section 8.5.2 above). Data filtering and normalization of the data (section 8.4 above) was performed first given the small number of samples and a large number of miRNAs and piRNAs in the data. The measure of relevance procedure [275] was then used to trim down the number of miRNAs and piRNAs and to keep the informative variables (Figure 9.2-2).

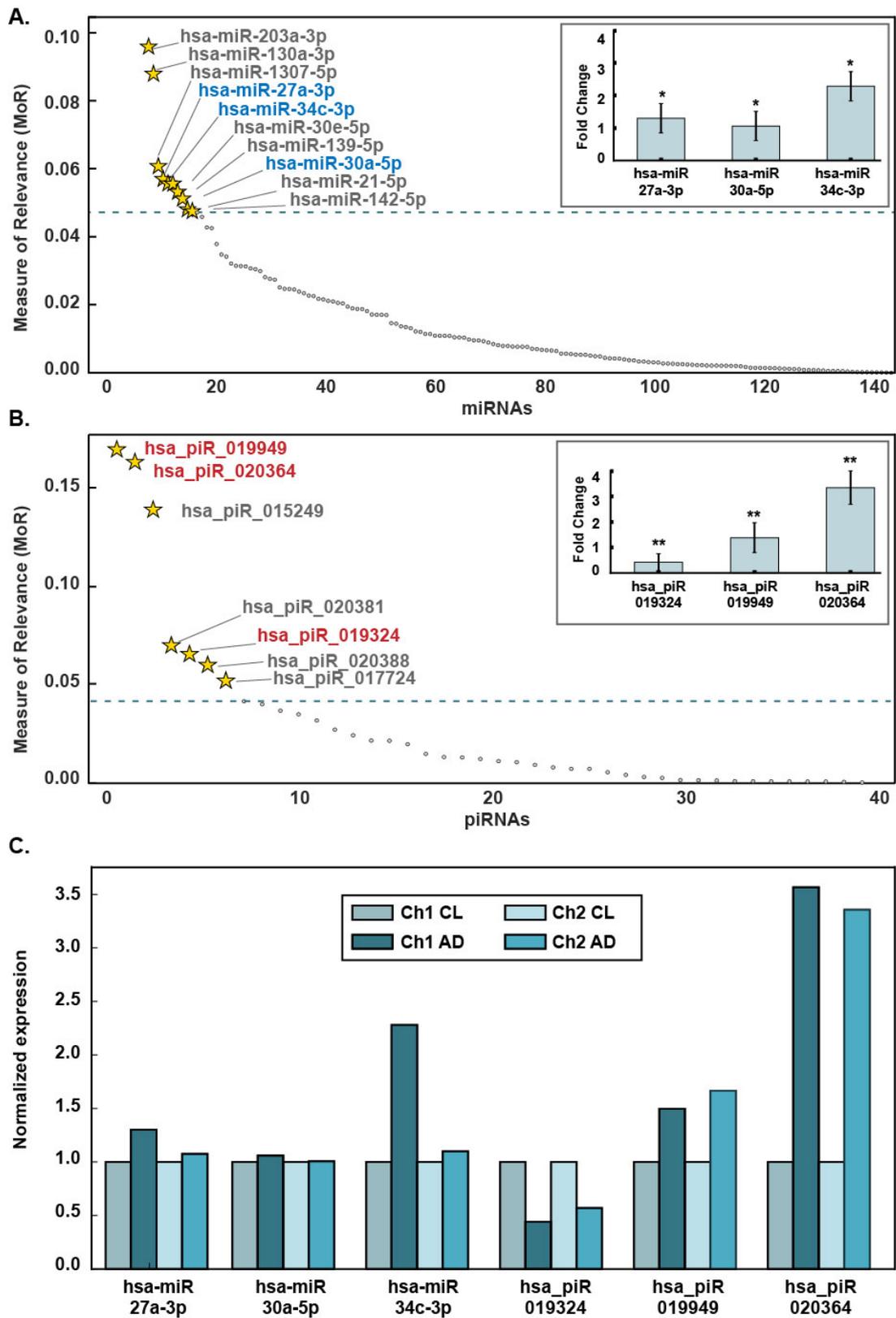


Figure 9.2-2 Identification and quantification of small non-coding RNAs signature

A. Showing the Measure of Relevance (MoR) for miRNAs (left) and piRNAs (right) signature for identifying differences between Alzheimer's disease and control samples in cohort 1. The dotted red line represents critical more value. The miRNAs and piRNAs that are above dotted red line are considered informative. miRNAs marked in blue and piRNAs marked in red are part of the final signature. Inset showing fold change (log) of most informative signature and their significance level (p-value ≤ 0.1) with Significance codes: 0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 after performing MANCOVA by age and gender as confounding factors. B Barplots showing the expression of the putative signature in training cohort 1 and replication cohort 2. The expression of control samples was scaled to one and the expression of Alzheimer's disease samples were scaled appropriately. Dark color shows control samples and light shade of the same color represents the Alzheimer's disease samples. Here, the signature follows the same trend in both training and replication cohort. Five signatures are upregulated in Alzheimer's disease compared to controls while one signature is downregulated.

In addition, a reliability test [276] was performed to ensure the effectiveness of the informative variables that can discriminate between the samples reliably. MiRNAs and piRNAs that were detected in both of the above mentioned approaches and were not affected by age or gender were considered as the candidate signature for further analysis. The set of 'informative and "relevant' miRNAs and piRNAs was further reduced after filtering out low ranking miRNAs and piRNAs measured using machine learning feature selection methods (Figure 9.2-2 C). In the end, three miRNAs, namely hsa-miR-27a-3p, hsa-miR-30a-5p, miR-34c-3p and three piRNAs, namely hsa-piR-019949, hsa-piR-020364 and hsa-piR-019324 were identified and treated as potential small non-coding RNAs signature.

9.2.4 The role of clinical and small non-coding RNA signatures

At present clinical markers are used extensively in hospitals and clinics for the diagnosis of Alzheimer's disease. These clinical markers have been shown to have a high sensitivity rates (95%) but poor specificity rates (36%) [311]. Nonetheless, these clinical biomarkers are now used as a common diagnostic tool for the diagnosis of Alzheimer's disease [316].

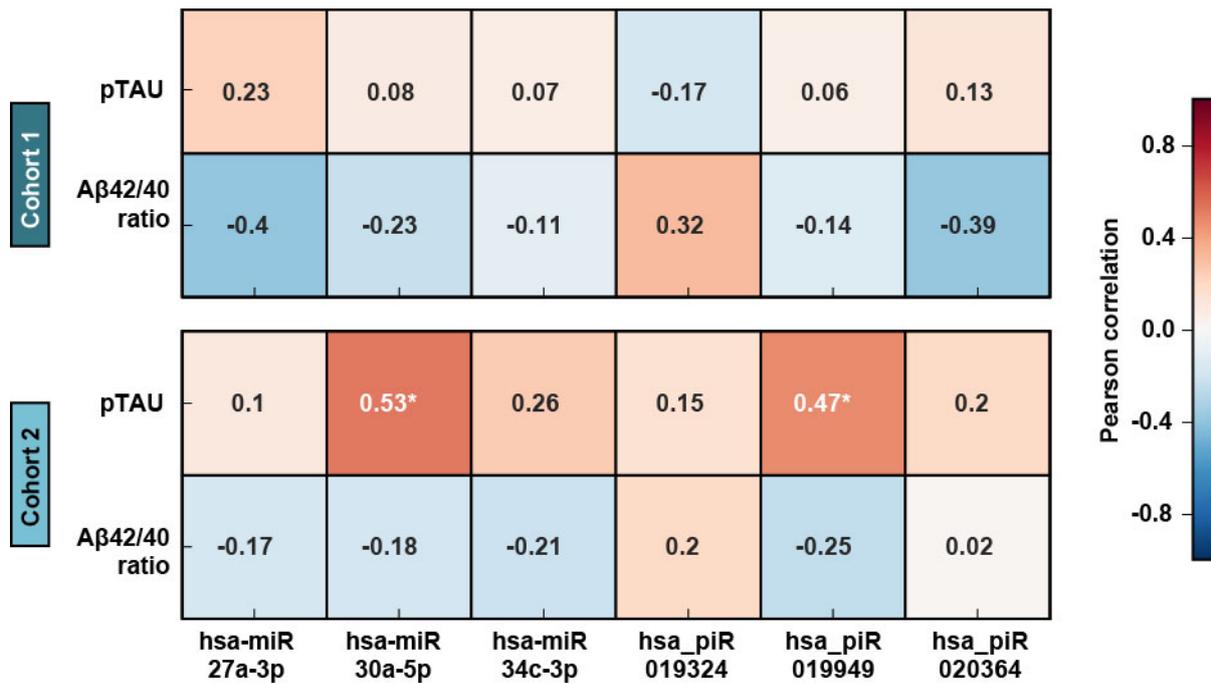


Figure 9.2-3 Heatmaps showing Pearson correlation coefficient between the normalized expression of putative biomarker signature and levels of clinical markers in cohort 1 (top) and in the replication cohort 2 (bottom). The Pearson correlation coefficients and a two-sample paired t-test were calculated and the significance is marked as (*) with p-value < 0.05.

Due to the complex pathologies of neurodegenerative diseases and the fact that multiple clinical phenotypes can be linked to just one disease pathology and vice versa, it is important to know whether the same information and resulting diagnosis is provided by the clinical markers and small non-coding RNA signatures or they cover different aspects of disease pathology and whether the combination of both can provide a clearer picture. For this, a correlation analysis of clinical and small non-coding RNA signatures is performed. Almost all clinical and small non-coding RNA signatures are not correlated (Figure 9.2-3).

9.2.5 The performance of small non-coding RNAs signatures on replication cohort

The performance of the small non-coding RNAs signatures was performed in three different stages. First, the performance of only miRNAs signature was performed using a random forest machine learning classification approach with a 10 fold cross-validation protocol. The miRNAs alone signature performed very close to a random classifier with the poor AUC of 0.56 (A). The most informative feature was hsa-miR-30a-5p (Figure 9.2-4 A inset). In comparison to miRNAs, little is known about the role of piRNAs that were however, expressed at similar levels in the CSF exosomes (Figure 9.2-2 C). The performance of

piRNAs signature was far better in comparison to the miRNAs and were able to distinguish Alzheimer's disease patients from controls with an AUC of 0.82 (Figure 9.2-4 B). The most informative feature was hsa_piR_019949 (Figure 9.2-4 B inset). The performance of combined miRNAs and piRNAs signature provides a slight improvement on the diagnosis and were able to distinguish Alzheimer's disease patients from controls with an AUC of 0.83 (Figure 9.2-4 C). The top two most informative features were hsa_piR_019949 and hsa-miR30a-5p (Figure 9.2-4 C inset).

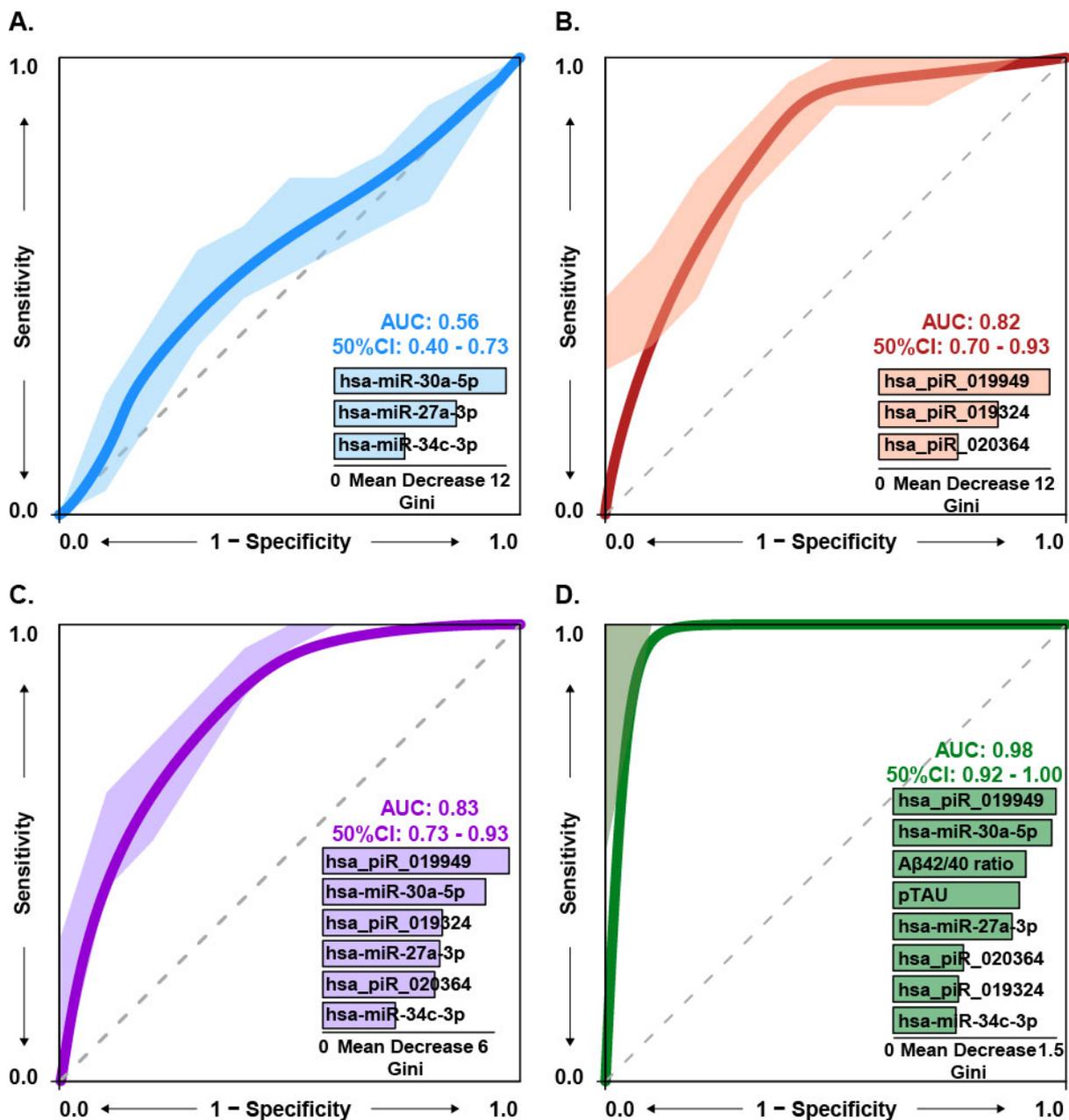


Figure 9.2-4 Performance of signature on replication cohort. A. ROCs showing the performance of miRNA signature on separate replication cohort (cohort 2). The training was done on cohort 1 with tenfold cross validation. Mean AUC of 0.56. Barplot (right) shows the

variable importance explaining how good a variable is in classifying the data. Here x-axis represents the mean decrease in Gini and the y-axis shows each variable which are ordered top-to-bottom as most to least important. In this case, hsa-miR-30a-5p is most informative. B. ROCs showing the performance of piRNA signature on separate replication cohort (cohort 2). Mean AUC of 0.82 was obtained. Barplot (right) show the variable importance explaining how good a variable is in classifying the data. Here x-axis represents mean decrease Gini value and the y-axis shows each variable which are ordered top-to-bottom as most- to least-important. In this case, hsa_piR_019949 is most informative. C. ROCs showing the performance of combined miRNA and piRNA signature on separate replication cohort (cohort 2). Interestingly the mean AUC of 0.83 was obtained suggest that the combined signature performed way better than the miRNA signature and marginally better than the piRNA signature. An inset plot showing the variable importance which explains how good a variable in classifying the data. Here x-axis represents the mean decrease in Gini and the y-axis shows each variable which are ordered top-to-bottom as most- to least-important. In this case, hsa_piR_019949 is most informative. D. ROCs showing the performance of combined clinical markers and smallRNA signature on separate replication cohort (cohort 2). The mean AUC of 0.98 was obtained showing that the combined signature performed massively better than the individual clinical markers or the smallRNAs signature.

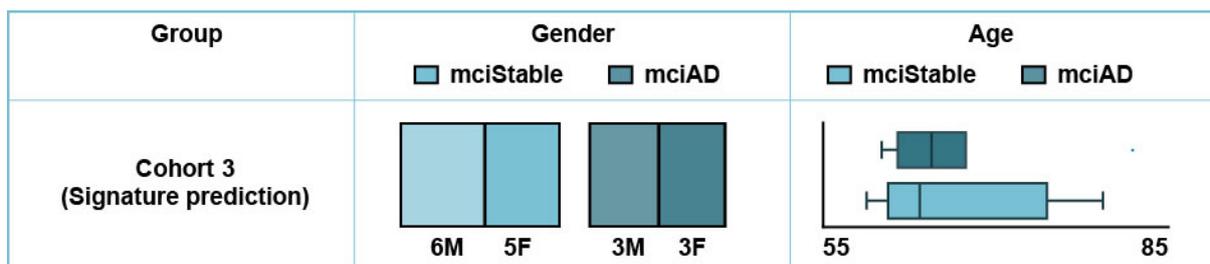
Apart from the similar performance, the miRNAs/piRNAs signature also does not correlate significantly with the A β and phospho-Tau levels (Figure 9.1-4) suggesting that the small non-coding RNAs signatures provide an additional layer of support to the current diagnosis of Alzheimer's disease patients. In order to cover all the aspects that both clinical markers and smallRNAs signature provide, the performance of both clinical markers and smallRNA signature was evaluated and had a massive improvement in the diagnosis where they were able to distinguish Alzheimer's disease patients from controls with an AUC of 0.98 (Figure 9.2-4 D). The top three most informative features were hsa_piR_019949, hsa-miR-30a-5p and A β 42/40 ratio (Figure 9.2-4 D inset).

9.3 CSF smallRNAs signature along with the available clinical markers as a promising biomarker to predict the progression from mild cognitive impairment (mci) to Alzheimer’s disease

9.3.1 Demographic information for the samples included in the biomarker signature analysis

The prediction cohort used in this study is cohort 3 where the samples were collected Erlangen (Germany). Cohort 3 consisted of 17 individuals with MCI of which 10 years later, 6 individuals had converted to Alzheimer’s disease (mciAD) while 11 did not develop Alzheimer’s disease (mciStable) (Table 9-2).

Table 9-2 Showing demographic information for the samples included in the biomarker signature analysis. Abbreviations: mciAD, mild cognitive impairment patients that went on to develop Alzheimer’s disease; mciStable, mild cognitive impairment patients that did not develop Alzheimer’s disease; F, female; M, male.



9.3.2 The role of small non-coding RNA signatures in predicting the onset of disease progression

The expression profile of many smallRNAs signatures of the MCI individuals (Figure 9.3-1) does not follow the same direction for their respective counterparts showing there are some differences that are unique to MCI individuals.

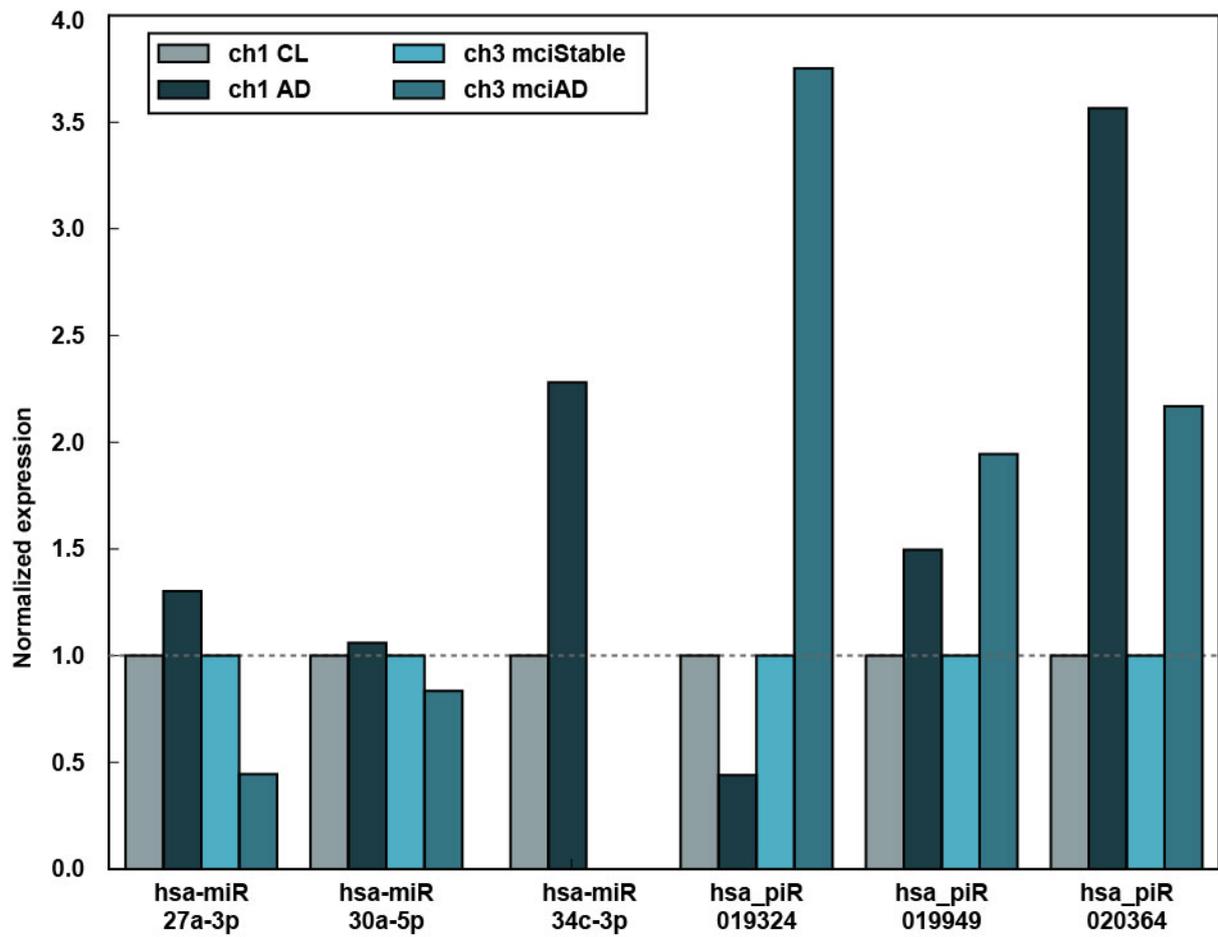


Figure 9.3-1 A. Barplot (left) showing the expression of the putative signature in training cohort 1 and predictive replication cohort 3. The expression of mciStable samples was scaled to one and the expression of mciAD samples was scaled appropriately. Dark color shows control/mciStable samples and light shade of the same color represents the Alzheimer's disease/mciAD samples. Three signatures are downregulated in mciAD compared to mciStable and remaining three signatures are unregulated.

9.3.3 The performance of small non-coding RNAs signatures on predictive replication cohort 3 for the predicting the onset of disease progression

Similar to disease diagnosis, the performance of the small non-coding RNAs signatures was performed in three different stages. First, the performance of only miRNAs signature was performed using a random forest machine learning classification approach with a 10 fold cross-validation protocol. The miRNAs alone signature performed very close to a random classifier with the AUC of 0.70 (Figure 9.3-2 A). The performance of combined miRNAs and piRNAs on the other hand performed poorly with an AUC of 0.62 (Figure 9.3-2 C). The performance of piRNAs signature is excellent with an AUC of 0.86 (Figure 9.3-2 B).

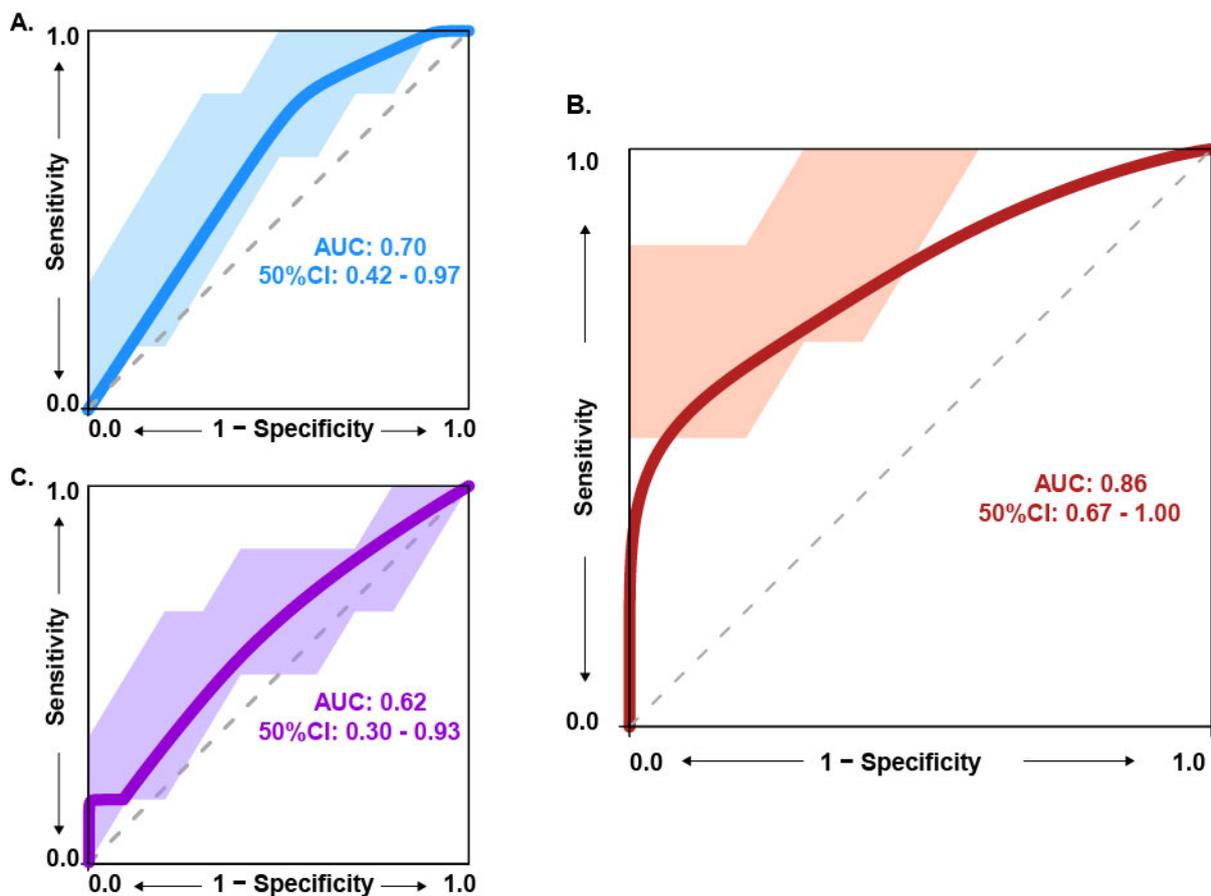


Figure 9.3-2 Performance of signature on predictive replication cohort

A. ROCs (top left) showing the performance of miRNA signature on separate predictive replication cohort 3. Mean AUC of 0.70 was obtained suggesting decent predictive ability. B. ROCs showing the performance of only piRNA signature on separate predictive replication cohort 3. The training was done on cohort 1 with a tenfold cross validation. Mean AUC of 0.86 suggesting excellent predictive ability. C. ROC shows the performance of combined miRNA and piRNA signature on separate predictive replication cohort 3. The mean AUC of 0.62 was obtained suggesting a poor performance of smallRNAs signature.

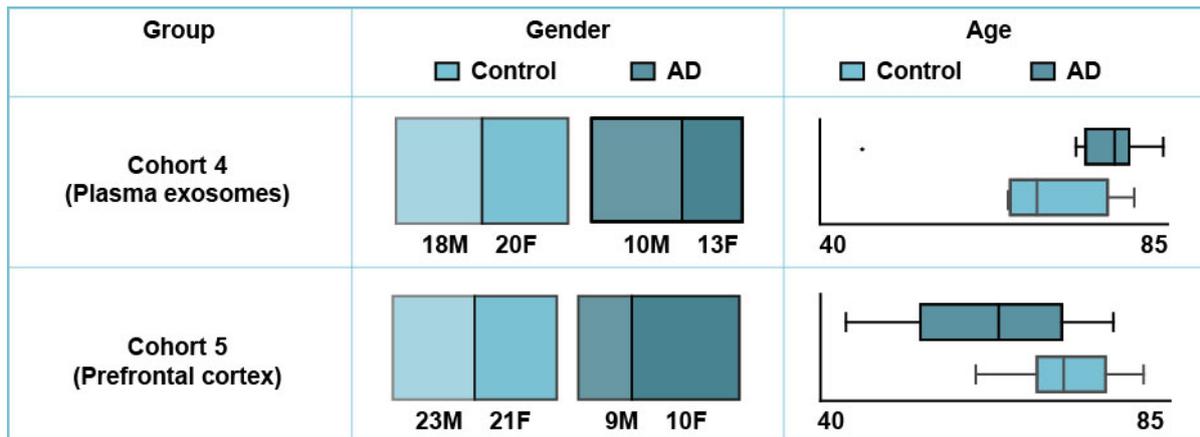
9.4 CSF smallRNAs signature in blood and brain

9.4.1 Demographic information for the samples included in the biomarker signature analysis

The samples from the blood plasma exosomes cohort (cohort 4) were collected in Göttingen (Germany). Cohort 4 consisted of 10 Alzheimer's disease patients and 12 control individuals after removing outliers and age-matched control individuals that did not suffer from any

neurodegenerative disorder (Table 9-3). The samples for cohort 5 are the postmortem brain tissue samples from the Prefrontal Cortex region. These samples are taken from the published study [317]. The samples were collected and processed in Leuven, Belgium. Cohort 5 consisted of 6 Alzheimer's disease and 6 control individuals.

Table 9-3 showing demographic information for the samples included in the biomarker signature analysis. Abbreviations: PFC, Prefrontal Cortex; AD, Alzheimer's disease; F, female; M, male



9.4.2 Role of CSF exosomes small non-coding RNA signatures in blood and brain

The expression profile of many CSF exosomes smallRNAs signature in the blood plasma exosomes and prefrontal cortex region of the brain. The entire smallRNAs signatures obtained from the CSF exosomes are expressed in the prefrontal cortex of the brain. However, few piRNAs were not expressed in the plasma exosomes (Figure 9.4-1).

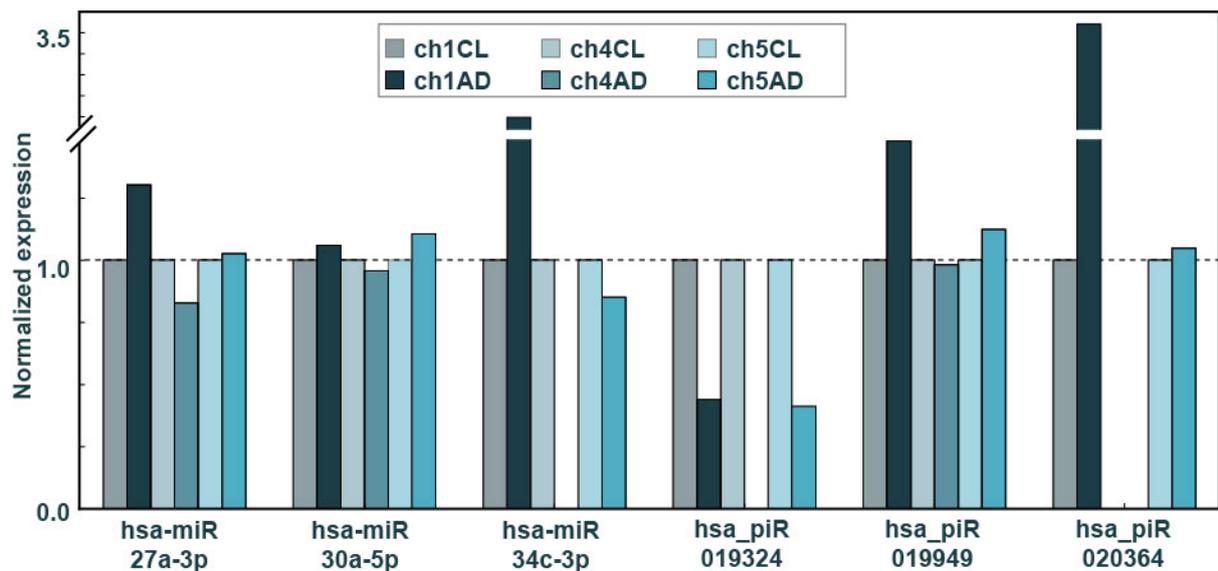


Figure 9.4-1 Barplots showing the expression of the putative signature in training cohort1, plasma exosomes cohort 4 and PFC brain cohort 5. The expression of control samples was scaled to one and the expression of Alzheimer's disease samples was scaled appropriately. Dark color shows control samples and light shade of the same color represents the Alzheimer's disease samples.

9.4.3 Performance of CSF exosomes small non-coding RNAs signatures in plasma exosomes blood samples (cohort 4) and from the post mortem brain tissue samples from the Prefrontal Cortex region (cohort 5)

Similar to disease diagnosis, the performance of the small non-coding RNAs signatures was performed at three different stages. First, the performance of only miRNAs signature was performed using a random forest machine learning classification approach with a 10 fold cross-validation protocol. The miRNAs alone signature performed very well in classifying Alzheimer's disease against the controls with an AUC of 0.84 (Figure 9.4-2 A) in cohort 4. The same miRNAs alone signature also performed well in cohort 5 with an AUC of 0.70 (Fig. Figure 9.4-2 D). Then the performance of piRNAs signature was assessed in both cohorts. The piRNAs only signature performance was not good at classifying Alzheimer's disease against the controls with an AUC of 0.67 (Figure 9.4-2 B) in cohort 4. On the other hand, the same piRNAs alone signature performed extremely well in cohort 5 with an AUC of 0.97 (Figure 9.4-2 E).

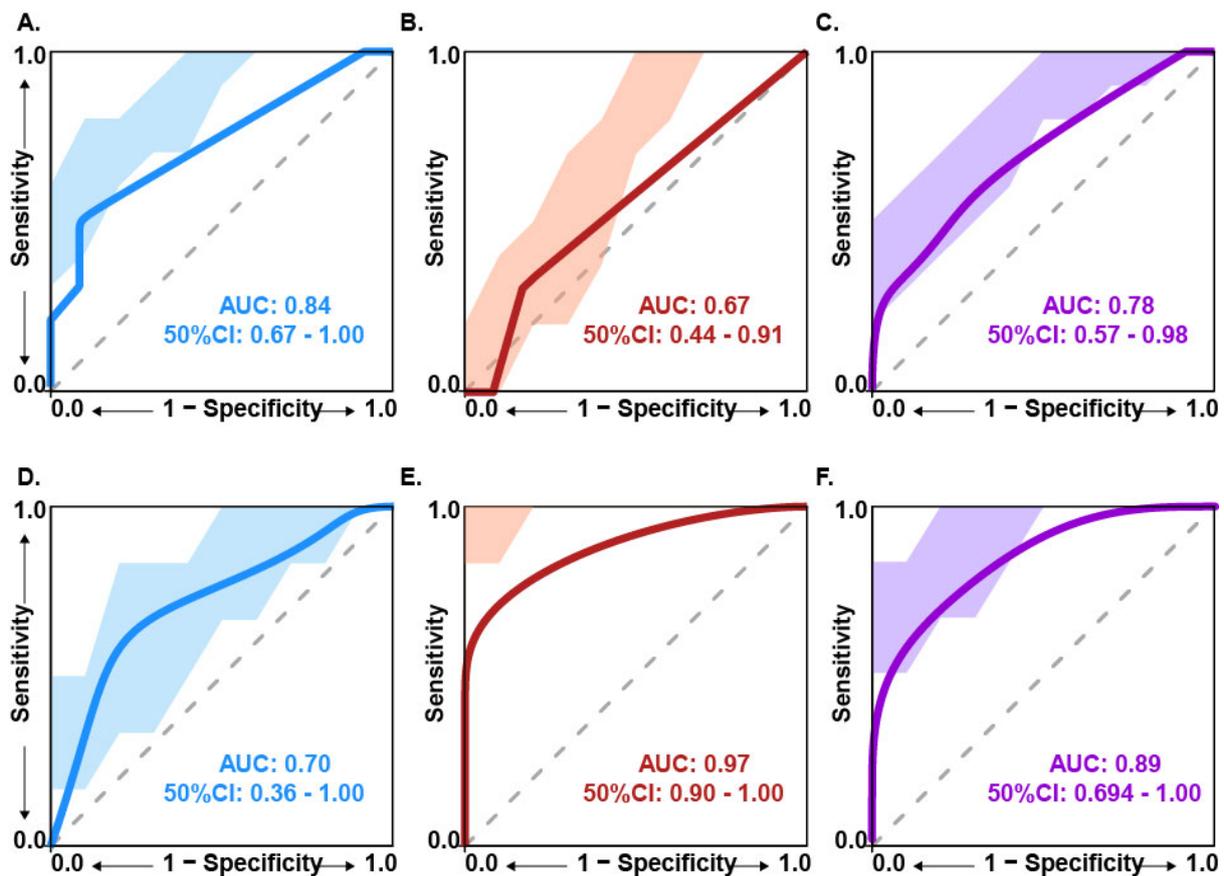


Figure 9.4-2 Performance of signature in blood and brain samples

A. ROCs showing the performance of miRNA signature on plasma exosomes cohort 4. Mean AUC of 0.83 was obtained suggesting good classification ability. B. ROCs showing the performance of only piRNA signature on plasma exosomes cohort 4. The training was done on cohort 1 with a tenfold cross validation. Mean AUC of 0.67 suggesting some classification ability. C. ROCs shows the performance of combined miRNA and piRNA signature on plasma exosomes cohort 4. The mean AUC of 0.78 was obtained suggesting that the combined signature also has good classification ability on the blood data. D. ROCs showing the performance of miRNA signature on PFC brain cohort 5. Mean AUC of 0.70 was obtained suggesting classification ability. E. ROCs showing the performance of only piRNA signature on PFC brain cohort 5. The training was done on cohort 1 with a tenfold cross validation. Mean AUC of 0.97 suggesting extremely good classification ability. F. ROCs shows the performance of combined miRNA and piRNA signature on PFC brain cohort 5. The mean AUC of 0.89 was obtained suggesting that the combined signature has very good classification ability on the PFC brain data.

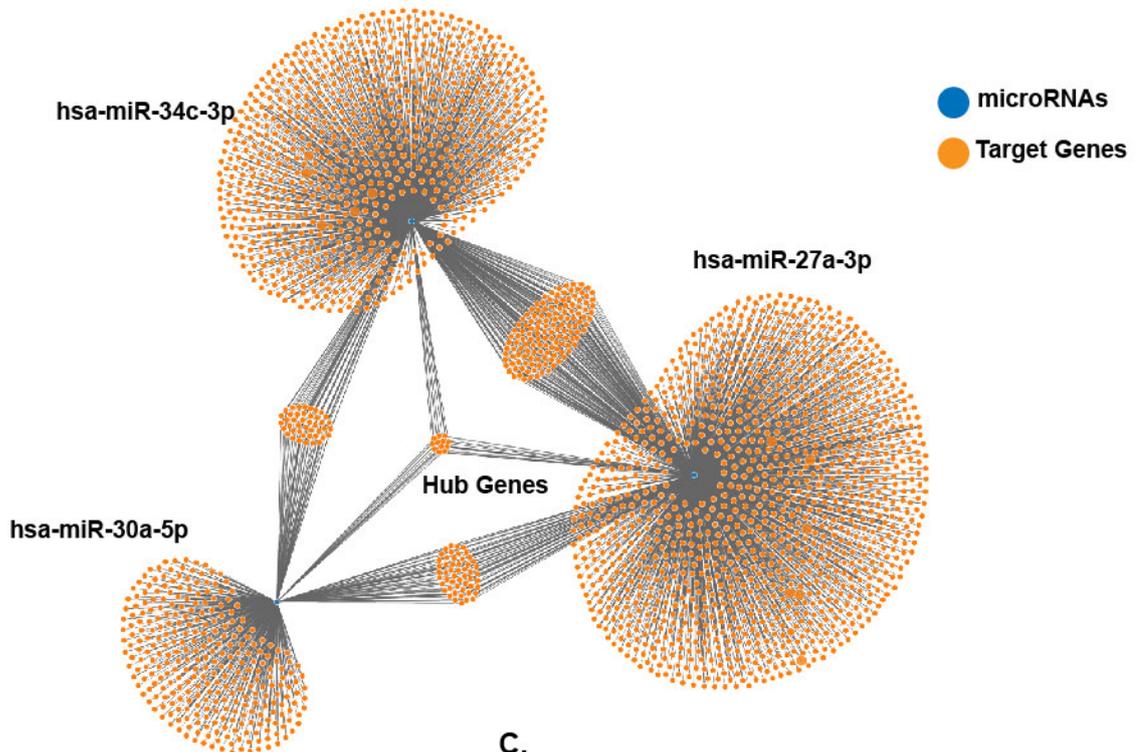
The performance of combined miRNAs and piRNAs, on the other hand, performed well in the diagnosis of Alzheimer's disease. Here the smallRNAs signature performed with an AUC of 0.78 (Figure 9.4-2 C). On the other hand, the same signature performed even better in

classifying Alzheimer's disease and controls in the brain region with an AUC of 0.89 (Figure 9.4-2 F). This result suggests that we can use the non-invasive techniques with these smallRNAs signatures to perform a good diagnosis between Alzheimer's disease and controls.

9.5 Functionally enriched piRNAs and miRNAs gene target pathways

9.5.1 Pathway analysis of miRNAs signature functional targets

A.



B.

IL1-mediated signaling events
HIV-1 Nef: Negative effector of Fas and TNF-alpha
RXR and RAR heterodimerization with other nuclear receptor
Role of Calcineurin-dependent NFAT signaling in lymphocytes
IGF1 pathway
Calcium signaling in the CD4+ TCR pathway
mTOR signaling pathway
Cellular roles of Anthrax toxin
CD40/CD40L signaling
HIF-1-alpha transcription factor network

C.

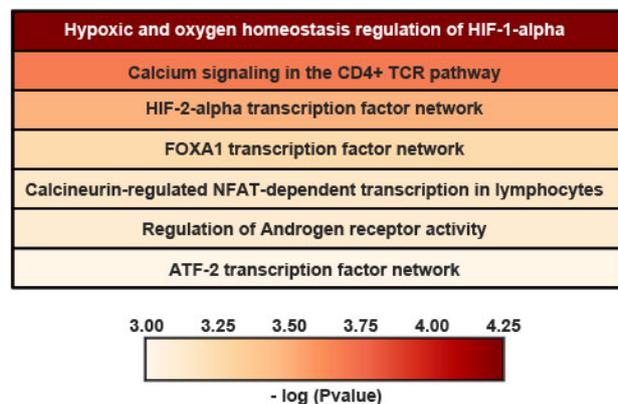


Figure 9.5-1 Signature miRNAs-gene targets regulatory network

A. Signature miRNAs are represented by blue circles and the orange circles represent the gene targets of the corresponding miRNAs. The grey edge connecting two miRNAs represents that there is a regulatory link between the two entities. The network is generated by the webtool miRwalk (v 3.0) [318, 319]. The most common targets of all miRNAs are called as hub genes. B. Pathways that are enriched in all genes-targets of the three miRNAs signature. C. Pathways that are enriched in only hub gene-targets of the three miRNAs signature.

In order to look at the involvement of the miRNAs target genes in the relevant biological pathways that shows the disease characteristics, pathway analysis of the confirmed target genes from the three miRNAs is performed. As one miRNA can regulate several genes and one gene can be regulated by several miRNAs [320], it forms a complex network of functionally enriched gene targets of the miRNAs. The miRNAs-gene target network for the signature miRNAs is shown in Figure 9.5-1 A. The analysis of all the confirmed targets of the three signature miRNAs reveals the association of the target genes with pathways that are highly relevant to Alzheimer's disease pathogenesis. Some of the major players that are well known are pathways linked to inflammatory processes [321-330] various signalling pathways like IGF1[331-336] and mTOR signaling [337, 338] and to HIF1alpha related hypoxia [339-342] (Figure 9.5-1 B). After investigating all the targets of the miRNAs signature, a subset of gene targets that are commonly regulated by all the three signatures also termed "hub genes" were investigated (Figure 9.5-1 A). Coincidentally, very similar and common pathways were identified by these hub genes (Figure 9.5-1 C) and mostly involved HIF1 α mediated hypoxia, several pathways that are related to inflammatory processes and unique pathways like regulation of androgen receptor activity [343, 344] that are shown to have a role in Alzheimer's disease.

9.5.2 Functional annotation of piRNAs signature targets

The other smallRNAs in the signature are the three piRNAs and till date, unlike miRNAs, the functions of piRNAs in the brain are still not well studied. One important role or function of these piRNAs is the silencing of the transposon elements [345-347] along with maintaining the integrity of the genome [348]. The piRNAs are conserved poorly evolutionary among the various species [349]. In the human genome, a majority of the piRNAs (~70%) comes from the distal intergenic regions (Figure 9.1-2 A top right pie chart). However, in comparison to the globally expressed piRNAs, the piRNAs that are expressed in the CSF exosomes, a larger number of the expressed piRNAs comes from the first exons (Figure 9.1-2 A bottom right pie chart). It has also been reported recently that the smallRNAs coming from the first exon are predictive of disease [306]. This is also true for the three piRNAs in the signature that are also mainly coming from the first exon and the targets of the genes, for example neurexin 1 (NRXN1) [350-352] have been reported to be involved in Alzheimer's disease (Table 9-4).

Table 9-4 Detailed Annotation of CSF piRNAs signature with their Alzheimer's disease associated alterations

piRNA	Annotation	Symbol	Alzheimer's disease Association
hsa_piR_019324 chr5:71851020-71851050 +	Distal Intergenic	CARTPT	[353]
hsa_piR_019949 chr11:122559947-122559977 +	Exon (uc021qrq.1/uc021qrq.1, exon 1 of 1)	UBASH3B	[354]
hsa_piR_020364 chr10:72972395-72972420 -	Distal Intergenic	PLA2G12B	[355]
hsa_piR_020364 chr14:102317092-102317117 +	Promoter (<=1kb)	ZNF839	[356]
hsa_piR_020364 chr17:8227039-8227064 -	3' UTR	LINC00324	[357]
hsa_piR_020364 chr2:50936204-50936229 -	Intron (uc021vhg.1/9378, intron 21 of 22)	NRXN1	[358]
hsa_piR_020364 chr6:26554122-26554147 +	Exon (uc021ynp.1/uc021ynp.1, exon 1 of 1)	HMGN4	[359, 360]
hsa_piR_020364 chr6:27177263-27177288 -	Exon (uc021ypd.1/uc021ypd.1, exon 1 of 1)	HIST1H2BK	[361]
hsa_piR_020364 chr6:27237619-27237644 -	Exon (uc021ypk.1/uc021ypk.1, exon 1 of 1)	POM121L2	[362]
hsa_piR_020364 chr6:27293892-27293917 +	Exon (uc021ypq.1/uc021ypq.1, exon 1 of 1)	VN1R10P	[363]

9.6 3C-seq Quality control and mapping statistics

9.6.1 3C-seq library quality

The sequencing is performed on Illumina HiSeq2500. As the project was the pilot project and the initial input was low, the quality of the data was not of high resolution. The data is still of very good quality.

Table 9-5 General sequencing statistics

Sample	Duplicates (%)	GC content (%)	Average sequence length (bp)	Initial number of reads
NeuNneg_R1	78.76	43	112	144279949
NeuNneg_R2	75.21	43	113	144279949
NeuNpos_R1	90.79	42	93	215566744
NeuNpos_R2	87.19	43	94	215566744

The average number of sequencing reads for the NeuNneg samples are around 144 million reads and 215 million reads for the NeuNpos samples. The overall %GC of all bases in all sequences were found to be similar in both samples with nearly all of them to be 43%. Although the initial number of reads is higher in the NeuNpos sample compared to the NeuNneg samples, the average sequence length of the NeuNneg samples were higher than NeuNpos samples (Table 9-5). The overall quality of the samples for both NeuNneg and NeuNpos samples is found to be good with average quality per read to be around Phred score of 35 (Figure 9.6-1 A). The x-axis represents the length of sequencing reads and the y-axis shows the Phred quality score. The Phred quality [255] Q_{Phred} is given by

$$Q_{Phred} = -10 \log_{10} P_r$$

Where P_r is the probability of incorrect base call.

Table 9-6 Relationship of base call accuracy and Phred quality score

Phred quality score (Q_{Phred})	Probability (P_r)	Base call accuracy (%)
10	10^{-1}	90.000
20	10^{-2}	99.000
30	10^{-3}	99.900
40	10^{-4}	99.990
50	10^{-5}	99.999

There is however a small percentage of reads with poor quality, but they are still above the recommended a good quality threshold of a Phred score of 30 (Table 9-6). The level of duplicates is also normal for a 3C-seq library [364] with 15% of library have an enrichment artifact peak around 10000 duplication level (Figure 9.6-1 B). The plot shows the number of duplicates on the axis and y-axis representing the percentage of library containing those duplicate numbers of reads. The high duplicate number of reads at the end of the plot (Figure 9.6-1 B) showed in solid lines shows the high enrichment or PCR amplification of certain interactions. Dotted line representing de-duplicated set of reads, however, disappears the peak representing that the library does not contain any contaminant or any high sequencing based technical duplication.

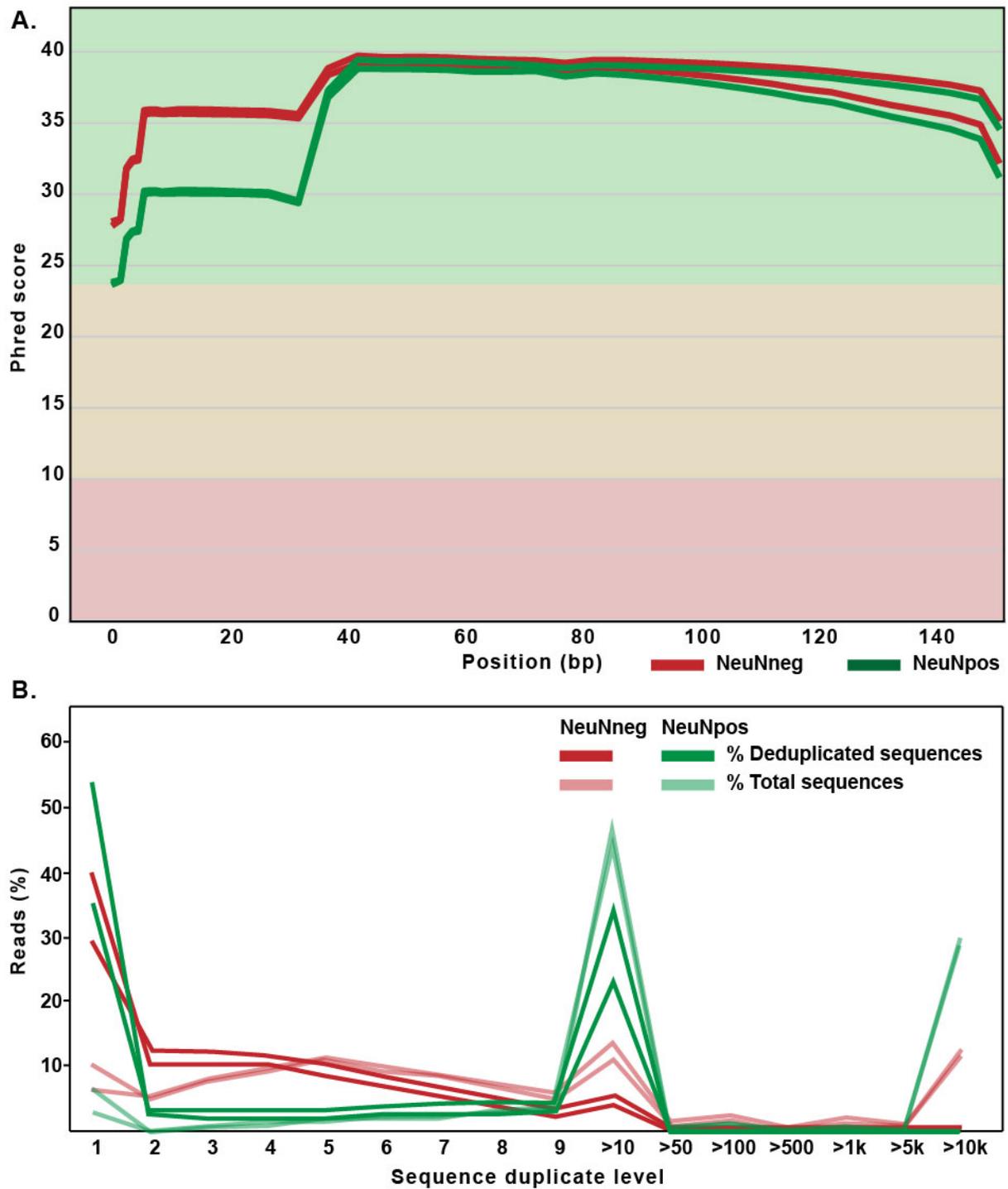


Figure 9.6-1 Quality score distribution over all sequences and duplicate sequences

The distributions of over-represented sequences are less in non-neuronal samples compared to the neuronal samples (Figure 9.6-2). It might be due to some biological sequences or just PCR amplification artifacts. These are however taken care of in the normalization step.

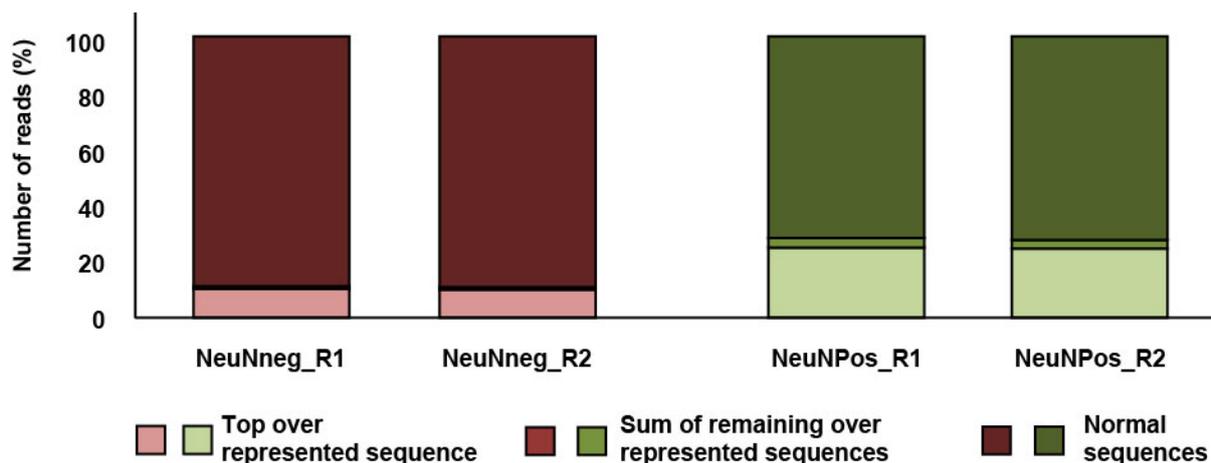


Figure 9.6-2 Overrepresented Sequences

9.6.2 3C-seq iterative mapping and iterative correction

After the through quality check for the 3C-seq raw data, the reads were mapped to the mouse genome (mm10). Due to the use of the four cutter DpnII, average fragment sizes are around 300-500 bp and with 101bp sequencing paired-end reads, only 20% of the reads are a valid read pair that contains the restriction enzyme junction. The reads were trimmed to contain the largest potential ligation junction. The reads are then mapped iteratively starting with the length of 25bp from 5' end and increasing 5 bp for every subsequent iteration till the length of the read (section 8.9.2 above). Approximately 52% reads were mapped to the neuronal samples and 67% reads were mapped for non-neuronal samples (Table 9-7). After all the filtering of various types, approximately 12.5 M and 1.8 M reads were found to be uniquely valid pairs of neuronal and non-neuronal samples. The iterative correction was then performed to reveal the relative interaction probabilities both cis (intra) and trans (inter) chromosomal interactions.

Table 9-7 Iterative mapping results

General HiC terms		NeuNneg	NeuNPos
Mapping	Side 1 mapped	96027256	111813998
	Side 2 mapped	95512770	111460268
Total Reads	Reads without unused chromosomes	102270189	120752492
	Reads removed unused chromosomes	0	0
DS reads	Total DS reads	89269837	102521774

	DS + SS	102270189	120752492
	SS reads removed	13000352	18230718
Reads removed	Same fragment	70549394	81622867
	Self-circles	3870368	2706058
	Dangling ends	66665741	78902539
	Extra dangling ends	9082595	7152212
	Error	13285	14270
Valid pairs	Total valid pairs	9637848	7152212
	Duplicates removed	7852492	13746695
	Unique valid pairs	1785356	12567075

9.7 Three dimensional landscape of neuronal and non-neuronal populations

9.7.1 Chromatin interaction patterns in the two types of CA1 cell population

Another way to understand the neurodegenerative disorders is to look at how the various neuronal and non-neuronal cells interact. To explore the three dimensional architecture of the neuronal and non-neuronal cells, 3C-seq was performed. The two libraries of approximately 144M and 215 M (Table 9-5) reads were generated for non-neuronal and neuronal samples. The samples were iteratively mapped (Table 9-7), corrected and normalized using iterative correction (balancing) [296] and final contact matrices were produced. As expected, the inter-chromosomal interactions are higher compared to the intra-chromosomal interactions (Figure 9.7-1). Even with higher number of reads in the non-neuronal population (Figure 9.7-1 Left), the numbers of looping interactions are higher than the neuronal population (Figure 9.7-1 Right). The alternative interaction pattern also reflects the compartments in the two populations [365, 366].

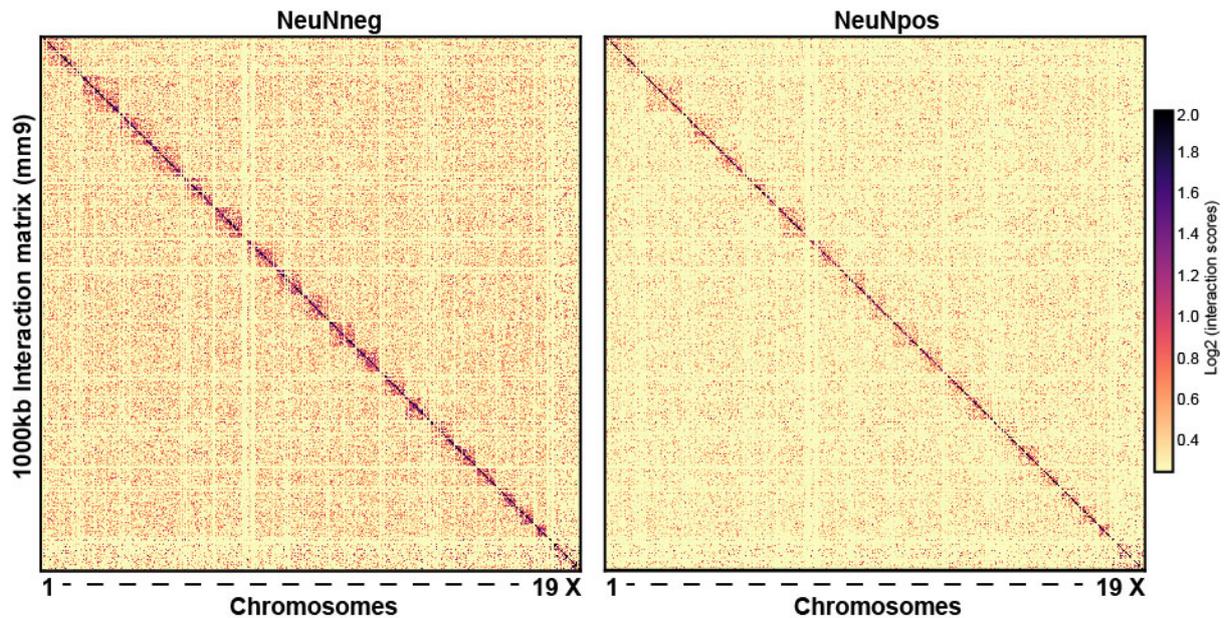


Figure 9.7-1 Chromatin interaction contact maps for non-neuronal and neuronal population

9.7.2 TADs in neuronal and non-neuronal population

To understand the higher order chromatin structure in the neuronal and non-neuronal populations, Topologically Associating Domains (TADs) were identified for all the chromosomes using the insulation score [367] and the directionality index [209]. However, TADs calculated with the insulation scores were considered due to better similarities with the published data [366, 368-370]. The number of TADs in the neuronal population is marginally higher than the non-neuronal population (Figure 9.7-2 (inset)). This may not affect the overall interaction pattern, but it may affect certain genes which may be specific to those populations. This will be further explained in the section 9.7.4 below. The TADs from the mouse mm9 cortex are taken from the original published study [209] with 40kb resolution.

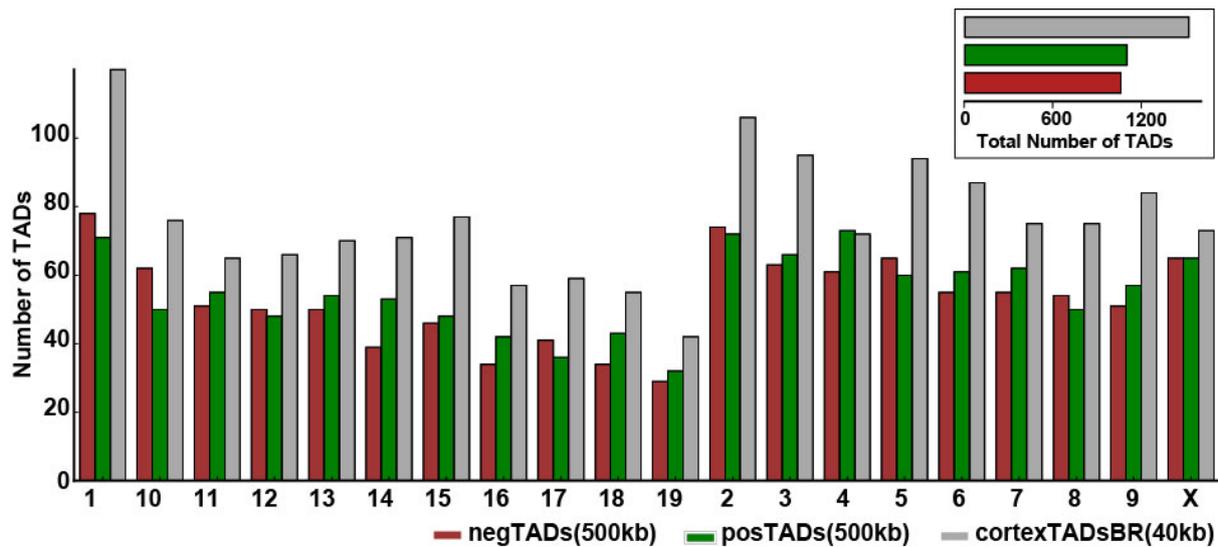


Figure 9.7-2 Number of Topologically Associating Domains (TADs) for neuronal and non-neuronal populations

9.7.3 Local interaction patterns in chromosome 9 (mm9) for neuronal and non-neuronal populations

The chromatin interactions tend to stay within the same TAD and inter-TAD interactions are not that common. This can be seen in both neuronal and non-neuronal population. However, the density of neuronal inter-tad interactions is less than the non-neuronal population (Figure 9.7-4). This is an interesting finding as that means there is more long range looping in the non-neuronal population compared to the neuronal population. The result, however, should be considered with precaution given the low sequencing depth.

The difference in interaction pattern can be observed clearly from the Figure 9.7-5. There are many places, for example, the region on chromosome 9 with genomic location chr9:10,000,000-40,000,000 and chr9:40,000,000-55,000,000 where the genes that are the boundary of the TADs has a lower expression in the neuronal population compared to the non-neuronal population (Figure 9.7-5). This may be due to the miRNAs that are present in the region or genes may be regulated by H3K27m3 peaks (denoted by the Cbellum H3K27me3 [371, 372]) Track. On the other hand, the genomic region in the neuronal population chr9: 90,000,000-92,000,000 (marked by red arrow in Figure 9.7-5) has higher neuronal interactions compared to the non-neurons and a distinct reduction in the number of H3K27m3 peaks in the region. This might suggest that the high interaction pattern may be between the TSS of the expressed genes and the H3K4me1 marks. The targets of miRNAs that are involved in neurodegeneration from example with the Alzheimer's disease (miR-30a-5p [373], let-7i-5p [374], 181a-5p [375] etc.) are also present in the region (for example: Sik2

(SIK2 salt inducible kinase 2)) with links to neurodegeneration [376-378]. A thorough analysis needs to be performed in the next phase of the study.

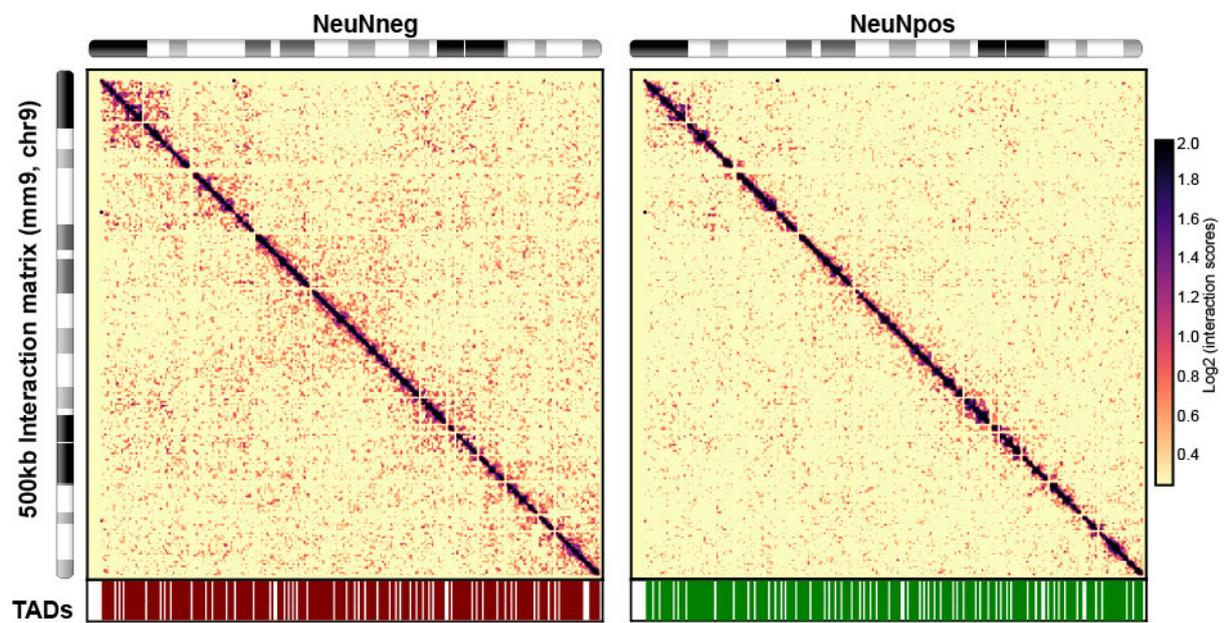


Figure 9.7-3 Figure 9.7-4 Local interaction patterns in chromosome 9 (mm9) for neuronal and non-neuronal populations

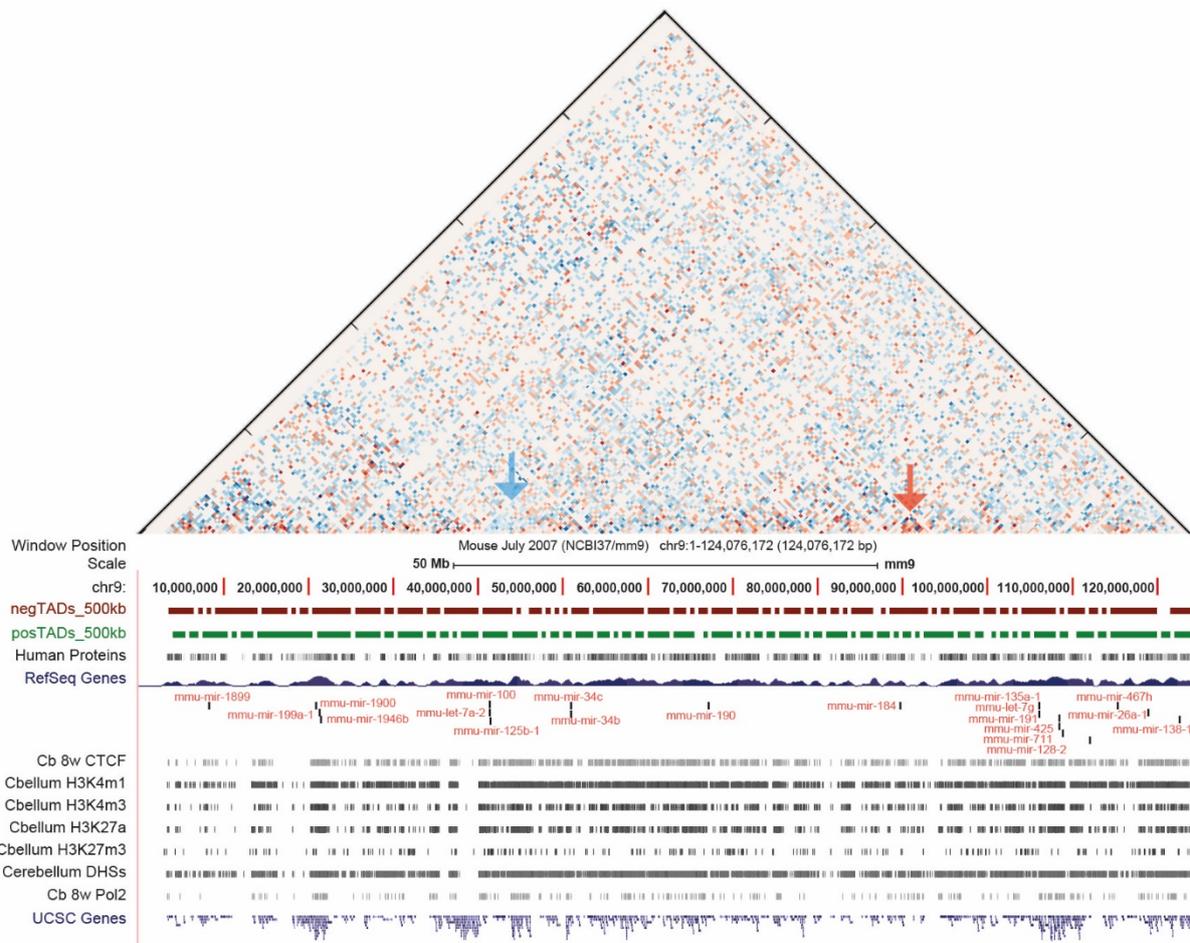


Figure 9.7-5 CA1 interactions looping profiles in neuronal and non-neuronal population

9.7.4 Interactions pattern near the BACE1 gene

One of the pathological hallmarks of Alzheimer's disease (Alzheimer's disease) is the formation of the A β plaques. BACE1 plays the key role in initiating the formation of these A β plaques [379-382]. The interaction profile of the region containing the BACE1 genes becomes crucial to understand. For this, a 5 MB region looping interaction profile is generated (Figure 9.7-6). The same neuronal over non-neuronal interaction values are plotted as the heatmap. As mentioned in section 9.7.2 above that there is a marginal decrease in the number of TADs in the neuronal and the non-neuronal population. The TAD in the genomic location of the BACE1 gene is missing the non-neuronal population, while the TAD in the neuronal population is expanded to have more interactions. However, the TAD left to the BACE1 gene in the non-neuronal population shows a higher number of looping interactions (chr9:44,500,000–45,000,000) as it is more bluish. Also, the bins containing the BACE1 gene have higher interactions in the neuronal samples compared to the non-neuronal samples. The gene density (Figure 9.7-6: Refseq genes) and density of enhancers (H3k4me1) marks are also higher near the BACE1 gene resulting in an increase in the

looping interaction of TSS in the neuronal samples. The absence of TADs around the BACE1 gene in the non-neuronal population, and, an increase in the looping interactions of the TSS and the enhancers of the BACE1 genes in the neuronal population suggesting the role of 3D genome organization in the Alzheimer's disease. At this point, the results should be very carefully examined as this is the pilot study that resulted in the low number of mapped reads to perform the analysis at a higher resolution.

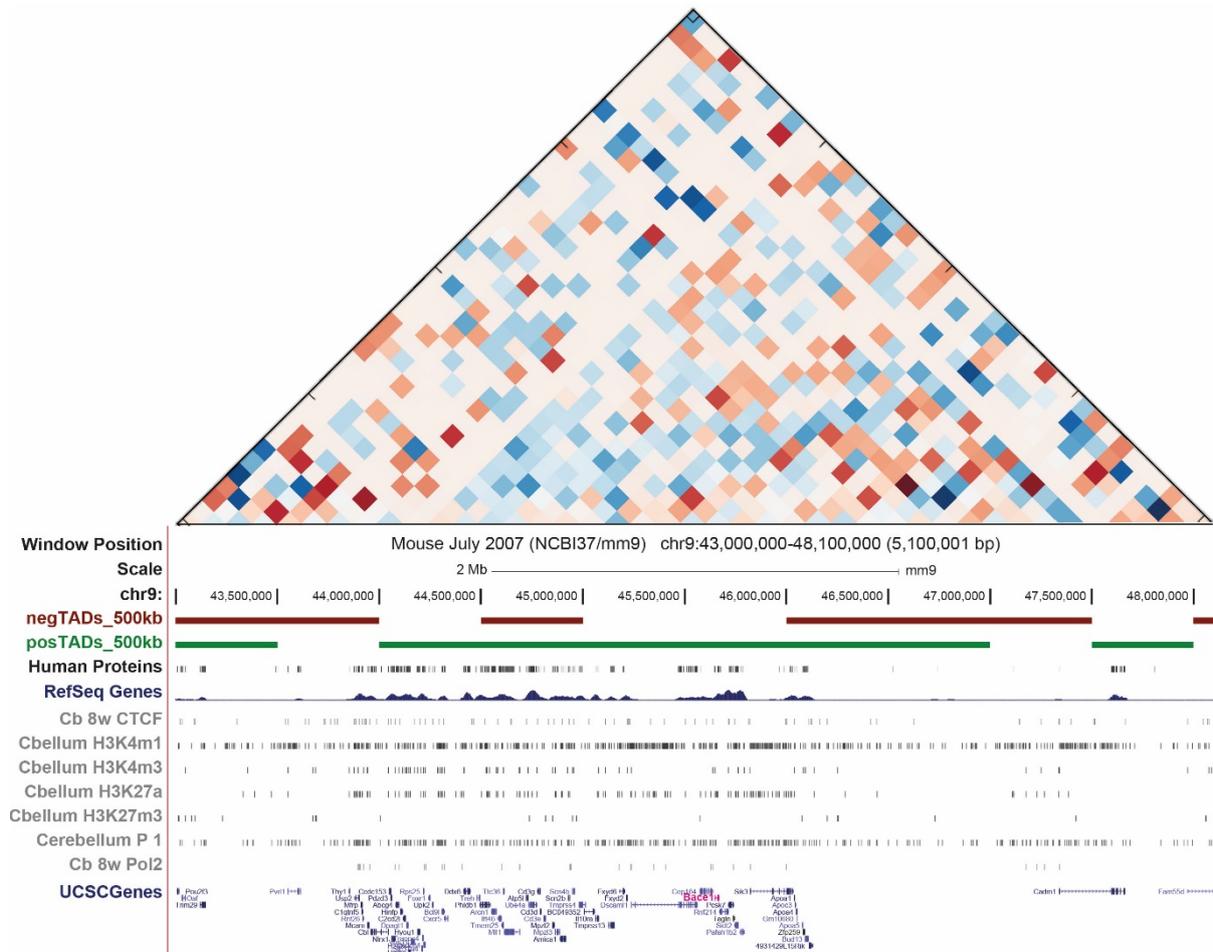


Figure 9.7-6 The interaction pattern in the vicinity of the BACE1 gene

10. Discussion

The work reported in this thesis provides an outlook into the fast progressing neurodegenerative disorders especially the Alzheimer's disease. It shows the need for the development of not only the sensitive, accurate, cheap and non-invasive biomarkers that can help in the diagnosis, prognosis, and prediction of the onset of a disease, but also focuses on the development of the statistical and computational pipelines and toolkits to provide a way to replicate the results and perform new analysis without introducing new technical biases. It also sheds a light on how the chromatin is organized in the genome and basic differences in the looping interactions between the neuronal and non-neuronal cell populations by the use of the latest techniques like 3C-seq.

The discussion is divided into two parts, one for each aim similar to methods and results sections to provide a clear understanding and interpretation of the work performed in line with the current published research. The conclusion finally provides a bird's eye view of the study that wraps up the two aims and provides a bigger picture to the current state of the biological question that is investigated in this study.

10.1 miRNAs and piRNAs as biomarker for Alzheimer's disease

Neurodegenerative disorders, for example, Alzheimer's disease and mild cognitive impairment that are identified by cognitive decline over a period of time and ultimately resulting in dementia, in general, diminishes the quality of life and affects not just the individual suffering from it but everyone around them [383-387]. At present, the goal in the research community is to develop biomarkers for the diagnosis [388-390] and prognosis [391-394] of these neurodegenerative disorders as the underlying disease mechanism that may help find good and effective biomarkers remains poorly understood. There has been an increasing number of reports [395-397] on circulating miRNAs and piRNAs as potential biomarkers for neurodegenerative disorders, providing much needed abilities for early diagnosis, disease distinction, disease prognosis and possible therapeutic benefits. During the course of my studies, I investigated a small number of miRNAs and piRNAs as a putative signature that shows both diagnostic potential (Figure 9.2-4) as well as the disease progression from MCI to Alzheimer's disease (Figure 9.3-2).

I focused my study to identify the potential biomarker signatures for the Alzheimer's disease that uses the expression level of small noncoding RNAs (for diagnosis and prediction of the

disease) by putting significant effort into the characterization of the small noncoding RNAome in humans. In order to identify a fairly non-invasive biomarker signature, the first step is to check the content of small noncoding RNAs from the CSF exosomes as it has been reported recently that CSF exosomes contain proteins and small noncoding RNAs that can be used as biomarker signature [58, 69, 398] (Section 6.4 above) for these neurodegenerative disorders. After successfully establishing that CSF exosomes contain the small noncoding RNAs (Section 8.1 above), another challenge was to check their expression. With the advent of NGS technologies, this was made possible experimentally. However, there were many bioinformatics challenges to quantify and summarize the expression of these small noncoding RNAs. These small noncoding RNAs often originates from multiple locations in the genome. This leads to the problem of uniquely aligning a particular small noncoding RNA that has multiple genomic origins to the reference genome. Traditional mapping software uses the common strategy of mapping (aligning) the small noncoding RNA reads to a reference genome, which normally leads to poor alignment (mapping) percentage as they either discard the multiple aligned small noncoding RNAs or randomly choose one of the mapping (normally the first alignment found). To address such issues, I successfully developed a pipeline (Section 8.3 above) that takes care of the multiple genomic origins of these small noncoding RNAs in the genome and provides the correct expression for these small noncoding RNA molecules.

From there, identification of informative small noncoding RNAs among the thousands of small noncoding RNAs was the challenge. I applied various statistical and machine learning approaches (Section 8.4 above and 8.5 above) to achieve this task where in the end, I identified a signature containing three miRNAs and three piRNAs. I further evaluated their diagnostic performance (Section 9.2.5 above) and the disease progression performance from MCI stable to MCI Alzheimer's disease (Section 9.3.3 above). I also compared the diagnostic (Alzheimer's disease vs. CL) performance of these small noncoding RNAs signature to the existing protein based clinical biomarkers specifically the levels of pTAU and A β 42/A β 40 ratio for the same human samples (Section 9.2.2 above). In the end, the small noncoding RNAs signature alone had good diagnostic and predictive abilities (Figure 9.2-4 C and Figure 9.3-2 C). Moreover, the diagnostic ability of these signatures increased tremendously by the use of the combination of both the protein based clinical biomarkers and the small noncoding RNAs signature (Figure 9.2-4 D).

Another important task is to find non-invasive biomarkers which can diagnose brain conditions as accurately as possible. The hypothesis here is that whether the expression of the small noncoding RNAs present in the bodily fluids such as blood, saliva or urine can also

be used as a biomarker signature. In order to check this hypothesis, I looked at the expression level of the particular small noncoding RNAs that I obtained as a biomarker signature (Section 9.2.5 above) in plasma exosomes. Almost all the small noncoding RNAs signature was expressed in plasma exosomes. Moreover, the diagnostic ability of the small noncoding RNAs signature in plasma exosomes was also good (Figure 9.4-2 A, B, C). Finally, I also wanted to check that the expression level of the small noncoding RNAs signature obtained from CSF exosomes reflects some (not all) changes in the brain. To this end, I checked the expression level of the small noncoding RNAs signature in brain tissue. All six small noncoding RNAs signature were expressed in the brain tissue. Furthermore, the diagnostic ability of these small noncoding RNAs signature to classify between Alzheimer's disease and controls was excellent (Figure 9.4-2 D, E, F). This suggests that the small noncoding RNAs can be targeted as biomarker candidates for the Alzheimer's disease diagnosis.

Finally, in order to get a better understanding of the functions of these small noncoding RNAs (miRNAs and piRNAs); I analyzed the role their targets play in the regulatory machinery of the disease pathway. I characterized these small noncoding RNAs from their target genes to find functionally similar co-expressing smallRNAs-gene target pairs that can be used as potential therapeutic drug targets for the Alzheimer's disease. For all small noncoding RNAs (miRNAs and piRNAs) signature, I report plausible functional targets that have a role in the Alzheimer's disease pathogenesis (Figure 9.5-1).

10.1.1 Characterization of CSF exosomes

Most neurodegenerative diseases including Alzheimer's disease can be characterized by the aggregation of certain proteins in the brain and can be only be confirmed after the post-mortem analysis of the brain tissue. Targeting the pathways that lead to the aggregation of these proteins are reported to have therapeutic potential [399-401]. Detection of these aggregated proteins in the brain is also possible by the use of neuroimaging techniques [402] which led to the rapid advances in the field of neuroimaging and the development of the neuroimaging markers [403-405]. However, the accuracy and sensitivity of these neuroimaging markers are still poor [406]. Identification of protein and small noncoding RNAs based markers [407, 408] offer a complementary approach for the characterization and early detection of the neurodegenerative disorders. These markers can be measured from bodily fluids like Cerebrospinal fluid (CSF), peripheral blood, saliva, breastmilk or urine and can be used in the clinics as non-invasive biomarkers.

Recent studies have shown that small noncoding RNAs especially miRNAs and piRNAs are present in the bodily fluids as free smallRNAs or within the extracellular vesicles (EVs) like exosomes [409-413]. One of the major challenges that other researchers and I have faced is the lack of standard operating protocols in the collection, extraction, and handling of bodily fluids for RNA isolation. The lack of standardization affects the downstream analysis as it can introduce major biases. Few studies [414-416] tried to address this issue, but still, there is an urgent need for standardization of such protocols for multi cohort studies. Furthermore, development and open access to the manufacturing standards, clinical certification, and standard data analysis protocols should be encouraged.

One drawback of CSF based biomarkers is the complicated and invasive collection procedure of CSF (through a lumbar puncture (Section 8.1.1 above)). On the other hand, a fairly practical and non-invasive collection procedure of peripheral blood makes the blood based biomarkers a promising candidate for the diagnosis of Alzheimer's disease. There are few studies reported recently on the development of the non-invasive biomarkers from blood plasma [417-419] or blood serum [420, 421]. However, a major caveat in the development of the non-invasive blood based biomarkers is the concentration of the Alzheimer's disease pathology related proteins in the brain that is reflected at a far lower amount in the blood than compared to CSF. For example, the concentration of TAU proteins in CSF is ~250 pg/ml [422], while the concentration of TAU measured in blood plasma is ~5 pg/ml [423]. This makes the quantification of TAU proteins in blood plasma difficult as the TAU proteins need to be isolated from blood plasma containing highly abundant proteins such as albumin that has a concentration of ~50 mg/ml [424].

SmallRNAs from CSF offers more and detailed information about the brain than the smallRNAs coming from blood plasma or serum as the extracellular matrix of the brain is in direct contact with the CSF [425]. Many smallRNAs detected in CSF exosomes are also highly expressed in the brain, suggesting the possible communication link between the brain and distal organs by the use of exosomes [426]. The cellular and exosomal small noncoding RNAome profile is found to be highly correlated as reported in the section 9.1.4 above and shown in the

Figure 9.1-4. I also investigated the profile of mature miRNAs in whole cells and exosomes. Mature miRNAs are derived from the precursor miRNAs. Moreover, there is no mechanism known yet that reports which mature miRNAs arm is functional [427, 428]. One hypothesis is that after the stem loop of premature miRNAs is cleaved by the Dicer, two mature strands 5p and 3p arms are produced. Depending on the stability of either or both arm may be

functional and the unstable arm is degraded [429, 430]. It has been reported previously [431, 432] that the inactive strand of the miRNAs are cleared by the use of exosomes and hence most miRNAs found in the exosomes are for degradation. However, as shown in the Figure 9.1-3 there is no such pattern observed. Both the cellular miRNAs and exosomal miRNAs showed the same expression pattern, whether it be the active or inactive arm. This relationship also strengthens the idea that the small noncoding RNAs in CSF exosomes in some part reflects the small noncoding RNAs snapshot of the cell of origin which may reside in the brain.

10.1.2 Characterization of NGS analysis and small noncoding RNAs

Profiling of small noncoding RNAs using the next generation sequencing technologies is a challenging task. This is due to the duplication or multiple origins of several small noncoding RNAs within the genome. Many published tools [302, 303, 433-439] for the profiling of the small noncoding RNAs either completely failed to address or poorly address the multi-origin of the small noncoding RNAs. Briefly, their mapping strategy included mapping the small noncoding RNA reads to the full reference genome and then annotating the mapped reads with the annotation database or file. During this process, they either discard the multiple mapped reads completely or use only the first or a randomly mapped location strategy. This leads to a significant drop in assessing the expression of some of highly expressed small noncoding RNAs. For example, one of the highest expressed miRNA in the brain hsa-miR-181a-5p [303, 440-443] is either completely missing or reported extremely low read counts using the current approach. However, there are other tools [444-447] that offer a different strategy quantified miRNAs correctly, but do not quantify other small noncoding RNAs such as piRNAs, snRNAs or snoRNAs.

Thus, I developed my own mapping strategy. I used the custom reference genome created by the sequences of the small noncoding RNAs and mapped the small noncoding RNA reads to this custom reference genome with iterative mapping strategy shown in Figure 8.3-1. The mapping pipeline (GJSrMap) [253] described in detail in section 8.3 above, has several advantages over the existing pipelines and available tools. One of the biggest advantages is the modularity of the pipeline. All of its sections are customizable from the choice of alignment software (section 8.3.3 above) to the building of the custom reference genome (section 8.3.1.3 above) and the annotation of the smallRNAs (section 8.3.4.1 above). By default, the pipeline is optimized to run a high performance cluster computing (HPCC) system, but it can easily be customized to run on a local server or computer. The

pipeline also provides the logs and statistics for every iteration related to the mapping of the reads.

A big part of the pipeline is the quality control steps taken to ensure the clean and high accuracy during the mapping and read counting process. First, the fastQ files are checked for the quality of sequencing. Then, before the alignment, adapters are trimmed and the reads with low quality scores are filtered out. Afterward, alignment of reads to the customized reference genome is performed iteratively. Final BAM files are filtered to remove the low quality mapping and sorted for the counting. After read counting is performed, the reads are assigned to each small noncoding RNAs class such as miRNAs and piRNAs. Both raw and normalized reads (Counts per million (CPM) [448] as default) are provided for the assessment of the mapping. Future versions of the pipeline will include the quantile, median and variance stabilized normalization (default) methods (section 8.4.1 above).

The pipeline also provides a detailed summary of mapping quality, library size distribution, distribution of small noncoding RNA classes found in the sequencing reads and the distribution of individual small noncoding RNAs within each class (Figure 8.3-2). Samples with low uniquely mapped read counts are recommended to be removed from the further downstream analysis as do not offer any information. The pipeline can distinguish between the real useful biological products and degradation products and provides additional information on the distribution of the special infrastructural small RNA classes such as rRNAs, pre-mature miRNAs, and snoRNAs (section 6.5.2 above). If they are present in high percentages, then the RNA isolation or the library preparation part needs to be tested.

In the CSF data, miRNAs distribution is the most highest as expected, but surprisingly second most abundant class of small RNAs are the piRNAs (Figure 9.1-2). All the top five miRNAs in the human CSF samples are reported to play a role in neurodegeneration. MiR-10-5p is reported to be involved with Sporadic Amyotrophic Lateral Sclerosis, Parkinson disease and Alzheimer disease [301, 449]. MiR-22-3p is reported to regulate cell proliferation [450] and inhibits cell apoptosis [451]. MiR-204-5p is reported as a biomarker candidate in Frontotemporal Dementia (FTD) [452] and as a therapeutic target in endometrial carcinoma [453]. MiR-26a-5p is reported to regulate the expression of inducible nitric oxide synthase in human osteoarthritis chondrocytes [454] and play an active role in tumorigenesis [455] growth and in multiple sclerosis [456]. MiR-10b-5p is involved in Huntington's disease [457] and breast cancer [458]. However, not much known about the roles of piRNAs and only recently a few studies reported the role of piRNAs in Alzheimer's

disease [122, 459] in cancer [460-462] in cardiac regeneration [463] and retinal degeneration [464].

In the human genome (Figure 9.1-2), 70.6% of piRNAs comes from the distal intergenic regions, 4.1% from the promoters of the genes and only about 1.2% comes from the 1st exon of a gene. Compared to piRNAs from the CSF exosomes, I found that 61% of piRNAs originate from 1st exon of a gene, 18.5% from the promoters of the genes and only 13.5% from the distal intergenic regions of the genome. A recent study [306] in cancer research reported that the smallRNAs from the 1st exons can distinguish between cancer and healthy individuals. The ones near the TSS are also found to be conserved between tissues. The three piRNAs from the biomarker signature has origin from the protein coding genes that are associated with Alzheimer's disease (Table 9-4). This shows that there is a great need to study these classes of small noncoding RNAs apart from miRNAs to get a complete picture of the regulation of the genes that are involved in the disease pathology.

10.1.3 Role of clinically established CSF biomarkers for Alzheimer's disease diagnosis

At the core of Alzheimer's disease pathology are the key molecular hallmarks amyloid-beta (A β) peptides and phosphorylated TAU, which are the key components of insoluble plaques and tangles respectively. Both have been closely correlated with the Alzheimer's disease progression. One of the major functions that are disrupted [465, 466] in the pathogenesis of Alzheimer's disease is the interaction of various signalling pathways believe to regulate the phosphorylation of TAU [467-470]. The highly soluble TAU protein when gets hyper-phosphorylated particularly mediated by CDK5 [471, 472], dissociates from its microtubules in the axons and form paired helical filaments (PHF) [473, 474] which are insoluble aggregates and believed to cause axonal transport impairment [475]. This in turn is believed to cause neuronal dysfunction [476] and cognitive decline in Alzheimer's disease [477, 478].

One of the major updates to the *Alzheimer's Disease Diagnostic Guidelines* [479] that were included in 2011 by the National Institutes of Health and the Alzheimer's Association is to recognize the potential use of the brain and CSF biomarkers. The use of positron emission tomography (PET) scans [480] and cerebrospinal fluid (CSF) analysis to check the elevated levels of TAU [481, 482] and/or decreased levels of amyloid-beta (A β) [483, 484] in CSF for the amyloid plaques and neurofibrillary tangles build up is suggested but not implemented in the clinics yet [12, 479, 485, 486].

The role of A β and TAU from CSF as biomarkers [487-489] has now been validated in many population studies [490, 491] and cohorts [492]. Recently, many studies [493-495] have shown that A β 42/40 ratio has better accuracy in the clinical diagnosis of Alzheimer's disease compared to the levels of A β 40 or A β 42 alone. It has also been shown that the A β 42/40 ratio is highly correlated to the levels of pTAU in the Alzheimer's disease patients [496, 497]. In line with the current research that is reported, I also found the levels of pTAU and A β 42/40 ratio elevated in the Alzheimer's disease patients compared to the controls (Figure 9.2-1). In terms of classification abilities of these protein based clinical biomarkers, I also see a good diagnostic performance with an AUC of 0.87 (Figure 9.2-1) for A β 42/40 ratio and pTAU in distinguishing the Alzheimer's disease patients from controls. An interesting observation here is that when compared to each other, pTAU is found to be more informative than A β 42/40 ratio (Figure 9.2-1 inset barplot).

10.1.4 The CSF miRNA/piRNA signature as biomarker for Alzheimer's disease diagnosis

One important aspect of the clinically established CSF protein biomarkers (A β 40, A β 42 TAU, pTAU and A β 42/40 ratio) [498] along with positron emission tomography (PET) scans is that they are biased towards the clinical hallmark pathology of Alzheimer's disease i.e. the build-up of insoluble amyloid plaques and neurofibrillary tangles in the brain. However, other methodologies such as exploring epigenetics mechanisms that might explain alternative part of Alzheimer's disease pathogenesis and might provide other methods for disease diagnosis, prognosis and therapeutic targets [499] should be investigated. Small noncoding RNAs like miRNAs and piRNAs are present in bio fluids like CSF and blood and are shown to have both diagnostic and prognostic abilities as a biomarker for Alzheimer's disease.

Out of all the small noncoding RNAs, miRNAs are the most studied ones. Their abundance and stability mark their key features to be a good biomarker candidate [500]. Many studies have reported miRNAs as novel and non-invasive biomarkers for Alzheimer's disease [501-506] by the use of various approaches. One array based study reported miRNAs in brain and CSF from Alzheimer's disease patients and non-demented controls and could identify 60 differentially expressed miRNAs in CSF [502]. Another recent study investigated eight selected miRNAs in CSF from Alzheimer's disease patients and reported lower miR-146a levels in Alzheimer's disease [507]. In a targeted approach miR-let-7b was found to be increased in the CSF from Alzheimer's disease patients [508]. Another study employed the array based nano-string technology to study miRNAs level in CSF from Alzheimer's disease patients and identified miR-27a-3p to be reduced in the CSF from Alzheimer's disease

patients [509], while miR-100, miR-103 and miR-375 levels in CSF were found to discriminate Alzheimer's disease patients from non-demented individuals [276]. Another group of scientists reported that the loss of miR-9, miR-137 miR-181c and miR-29a/b-1 increases the levels of the rate limiting enzyme Serine palmitoyltransferase (SPT) resulting in the increased levels of A β [510]. It has also been shown that the loss of miR-29a/b-1 cluster resulted in elevated levels of BACE1 gene and consequently levels of A β in sporadic Alzheimer's disease [511].

Recently, another class of small noncoding RNAs, piRNAs has gained traction. Although, many studies already reported their various functions such as post transcriptional regulation of protein coding genes and stability of mature mRNAs [512], chromatin stability [513, 514], regulation of epigenetic mechanisms [515, 516] and most importantly suppression of transposons [347, 517, 518]. First reports have only been recently published that piRNAs also play a role in the pathogenesis of various disease [122, 459, 519-522]. Some reported piRNAs with their respective target genes that are shown to play a role in Alzheimer's disease are piR-38240 (cytochrome c), piR-34393 (karyopherin subunit alpha 6), and piR-40666 (RAB11A) having an inverse gene expression relationship [122].

In this study, I developed a specialized mapping pipeline (section 8.3 above) and used various statistical and machine learning algorithms (section 8.5 above) to obtain a set of miRNAs and piRNAs that can be useful in early diagnosis, disease distinction, disease prognosis and possible therapeutic benefits in Alzheimer's disease. Identification of a few most informative miRNAs and piRNAs from the entire small noncoding RNAome with a limited number of samples was a challenging task. First, I applied the Measure of Relevance (MoR) procedure [275] and a reliability test [276] to obtain a significantly reduced set of informative miRNAs and piRNAs that are expressed in at least 95% of the samples (section 8.5.1 above). After the removal of the uninformative smallRNAs, a more sophisticated variable ranking is calculated using various machine learning algorithms that provide a mean ranked score for each informative signature (section 8.5.1 above). I chose a threshold of 0.30 after the discussion with experts in machine learning and filtered out the miRNAs and piRNAs that were found below this threshold. The miRNAs and piRNAs are also checked if they are confounded by age and gender by using MANCOVA analysis (section 8.5.2 above).

After taking all the measures to ensure the effectiveness of the miRNAs and piRNAs, I got three miRNAs, namely hsa-miR-27a-3p, hsa-miR-30a-5p, miR-34c-3p and three piRNAs, namely hsa-piR-019949, hsa-piR-020364 and hsa-piR-019324 as the putative biomarker signature. In order to avoid redundancy of the information provided by these small

noncoding RNAs signature and protein based clinical biomarkers used in the study (pTAU and A β 42/40 ratio), I performed the correlation analysis (section 9.2.4 above Table 9-1). I found that they do not correlate which means that they provide information on a different aspects of the Alzheimer's disease pathology. This is also very promising as in the future a combination of both types of markers can provide a bigger picture rather than an incomplete picture by looking at just the individual marker's diagnosis and prognosis. From here, I investigated the diagnostic abilities of the miRNAs and piRNAs signature. I developed a random forest model by training the model on cohort 1 data (Table 9-1 row 1) containing 23 Alzheimer's disease and 38 control samples. I then used a 10 fold cross-validated approach to test the performance of the model (details of the entire procedure is mentioned in section Model selection and performance) on an independent replication cohort 2 (Table 9-1 row 2). Surprisingly, miRNAs alone signature performed very close to a random model (Figure 9.2-4 A), while piRNAs signature performed with a high classification accuracy (Figure 9.2-4 B). When both miRNAs and piRNAs signature was checked together, a marginal (1%) increase in performance to the piRNAs is observed (Figure 9.2-4 C). As the protein based clinical biomarkers and the small noncoding RNAs signature provide different information (Section 9.2.4 above), I looked at the combined power of both types of signature. The combined signature performed significantly better to the point of almost perfect classification of Alzheimer's disease and controls (Figure 9.2-4 D) with an AUC of 0.98. This is very promising as it can be a very good diagnostic marker, although it needs to be tested on a larger cohort with more patients and age matched controls.

10.1.5 The predictive power of CSF miRNAs/piRNAs signature from MCI to Alzheimer's disease

While cost-efficient and reliable biomarkers for correct diagnosis of Alzheimer's disease are important, an equally pressing the issue is the detection of markers that could predict the conversion of patients suffering from mild cognitive impairment (MCI) to Alzheimer's disease [523]. To this end, I investigated the predictive abilities of the small noncoding RNAs signature in a separate predictive replication cohort 3 (Table 9-2) with 17 MCI individuals of which 10 years later, 6 individuals progressed to develop Alzheimer's disease (mciAD) while 11 did not develop Alzheimer's disease (mciStable).

When I evaluated the predictive abilities of the small noncoding RNAs signature, miRNAs alone signature performed well with an AUC of 0.70 (Figure 9.3-2A). The piRNAs signature showed an excellent predictive ability (Figure 9.3-2 B) with an AUC of 0.86. The combined miRNAs and piRNAs signature did not demonstrate predictive abilities with an AUC of 0.62

(Figure 9.3-2 C). Nonetheless, this shows the excellent diagnostic and predictive abilities of the piRNAs signature in Alzheimer's disease pathogenesis which has not been reported earlier. This data suggest that the CSF exosomal miRNAs and mostly piRNAs signature can be used to diagnose Alzheimer's disease patients and might also be suitable to predict MCI patients that are likely to convert to Alzheimer's disease with careful consideration.

10.1.6 Diagnostic properties of CSF miRNA/piRNA signature for plasma and brain tissue samples

So far, I have discussed the diagnostic and predictive abilities of the two types of CSF biomarkers: proteins and small noncoding RNAs based markers. These markers are reported to be most investigated by the research community and included as the recommended biomarkers in the updated guidelines for clinical diagnostic criteria for Alzheimer's disease by the National Institutes of Health and the Alzheimer's Association [479]. However, obtaining CSF is still invasive as compared to other bodily fluids like saliva, urine and blood. There has been a significant increase in the number of studies published recently reporting the development and potential therapeutic use of blood based biomarkers for Alzheimer's disease [506, 524-527]. Besides the obvious ease of obtaining blood compared to CSF, another advantage these blood based biomarkers provide is that they can be obtained in higher volume and are more economic than obtaining CSF. However, this also means that the expression of the proteins and smallRNAs that are measured from blood based techniques will be diluted and specificity to the brain might not be correlated. The technological recent advancements in the field have led to better and more accurate detection of these proteins and smallRNAs from blood that can be used as biomarkers for MCI and Alzheimer's disease [78, 528-532].

To test the diagnostic abilities of the current miRNAs and piRNAs signature found in my study, I evaluated them in an independent blood plasma exosomes cohort 4 with 10 Alzheimer's disease and 12 control individuals. The miRNAs signature displayed high diagnostic abilities (Figure 9.4-2 A) classifying the Alzheimer's disease individuals from controls with an AUC of 0.84. On the other hand, piRNAs displayed some diagnostic abilities with an AUC of 0.67 (Figure 9.4-2 B). The combined signature also displayed high diagnostic abilities in classifying the Alzheimer's disease individuals from controls with an AUC of 0.78 (Figure 9.4-2 C). The ability of small noncoding RNAs signature to classify Alzheimer's disease individuals from controls in plasma exosomes makes them a potential candidate for biomarker signature.

It has been reported that the cerebrospinal fluid (CSF) is in contact with the brain and reflects the events that happened inside the brain [533-535]. This makes the small noncoding RNAs derived from the CSF exosomes highly informative as they can be used to understand the processes happening inside the brain [536]. One can also argue that the small RNA signature from these CSF exosomes can also closely reflect the brain events that can be used as markers to identify certain disease symptoms [532, 537-539]. The exosomes (that end up in CSF and blood plasma/serum) may also originate from other cells, including diseased neurons and can inform us about the cellular status inside the brain. The hypothesis is whether the miRNAs and piRNAs signature I identified from CSF exosomes can demonstrate a diagnostic ability to classify Alzheimer's disease individuals and controls in the brain tissue. I downloaded a publicly available dataset and used the samples published in that study [317] to look at the diagnostic abilities of the small noncoding RNA signature. The miRNAs signature shows decent diagnostic ability with an AUC of 0.70 (Figure 9.4-2 D). In comparison to the miRNAs, piRNAs show excellent diagnostic ability with the AUC of 0.97 in classifying individuals with Alzheimer's disease from controls (Figure 9.4-2 E). The combined signature also displayed excellent diagnostic ability to classify individuals with Alzheimer's disease from controls (Figure 9.4-2 E).

However, there is still no clear understanding on the origin of small RNAs in the CSF exosomes from the brain. The small RNAs in the CSF exosomes cannot be assumed to come from the neurons exclusively. These small RNAs can very well come from other non-neuronal cells, which may alter our understanding of the results. There is also the issue of small cohort sizes that only provides a small snapshot from which Alzheimer's disease associated neurodegeneration can be studied. Certainly, a large cohort study is needed to confirm the findings I evaluated and reported here. In the meantime, the small noncoding RNA signature (three miRNAs, namely hsa-miR-27a-3p, hsa-miR-30a-5p, miR-34c-3p and three piRNAs, namely hsa-piR-019949, hsa-piR-020364 and hsa-piR-019324) may be used as a diagnostic tool for Alzheimer's disease classification and an early measure for disease prediction from MCI to Alzheimer's disease with considerable precautions.

10.1.7 Functional characterization of CSF miRNAs/piRNAs signature

The target gene predictions from the miRNAs and piRNAs signature identified many genes that play a key role in Alzheimer's disease pathogenesis related pathways. Some of them include the heterodimerization of Retinoid X receptors (RXRs)/Retinoic Acid Receptors (RARs) [540] with other receptors and selective targeting of those receptors may help in the

A β removal and better cognition; the mTOR signalling pathways which may play an important part in the development of the Alzheimer's disease [541], the influence of the CD40 and its ligand (CD40L) in the Alzheimer's disease pathogenesis [542-544] inflammation related pathways like IL1-mediated signalling events [325, 326, 328, 545]; IGF1 pathway [332, 336]; and hypoxia related pathway [339-342] showing to play a role in Alzheimer's disease pathology. I also looked at the set of genes that were the common targets for the three miRNAs signature, which I termed as "hub genes" to see if there is a core set of pathways that is involved in the disease pathology. I found a very similar set of pathways like the hypoxia (HIF-1a and HIF-2a regulated) pathways and calcium signalling in the CD4+ TCR pathways [439] playing the role in disease pathology. Since, piRNAs are new to this field and very little is known about the target of these piRNAs like we know about the miRNAs, a general overview about the piRNAs targets was checked. I found that a majority of piRNAs genomewide lies in the distal intergenic regions (Figure 9.1-2 A). On the other hand, contrary to the genomewide consensus, the majority of the CSF exosomes piRNAs lies in the 1st exons of their targets (Figure 9.1-2 A). The three piRNAs in signature also lies in the 1st exons. This is an important finding as it is been shown recently that the smallRNAs that comes from the 1st exon of their targets plays an important role in disease prediction [306]. Some of the genes the piRNAs are originating from for example NRXN1 [350-352] and UBASH3B [546, 547] are also linked with Alzheimer's disease. Thus the results described in section 9.5 above support our hypothesis and cement our belief that the miRNAs and piRNAs signature could possibly reflect the interactions of these gene regulatory pathways that might play a key role in Alzheimer's disease pathogenesis.

10.2 Three dimensional architecture of neuronal and non-neuronal cells

In the first aim, I looked at the epigenetic regulation of gene expression and mainly the role of small noncoding RNAs involved in the regulatory mechanism. Another way to understand the regulation of the gene expression is to study the role of chromatin conformation in looping of the genomic elements that are involved in this process of regulation. This field of study is quite recent and new mechanisms are constantly being discovered that elucidate the role of three dimensional organization of the chromatin involved in controlling the gene expression. The hypothesis here is that by studying the chromosomal conformation differences between the neuronal and non-neuronal cells, especially at the locations of regulatory elements that are crucial in the neurodegenerative disease pathways may open up new research frontiers to understand the disease mechanism. For this, we performed a

pilot experiment to look at the conformational differences in the chromatin architecture of neurons and non-neurons in the CA1 region of wildtype mouse.

10.2.1 3C-seq library quality assessment, reads alignment and balancing

The overall quality of the samples was good with average read quality about 35 (Phred score) for all the samples (Figure 9.6-1). One point to note is that the subset of reads are always going to have variable read quality with few good quality subsets and others with poor quality. It is however only considered when the majority of the subset has poor quality as they are supposed to be only a very small percentage of total sequences. Low quality reads have a small Phred score (Table 9-5) which, if present in large number in the sequencing library can be costly in the price of the experiment and may lead to an inaccurate understanding of the results. The high quality reads were then aligned to the mouse genome (mm9) by using the iterative mapping procedure which led to the mapping of 52% and 67% in the neuronal and non-neuronal populations respectively (Table 9-6). The 3C-seq technical artifacts were filtered out and iterative correction and normalization (balancing) were performed to enhance the possibility of detecting the looping interactions with high probabilities.

10.2.2 The conformational differences and identification of TADs in the neuronal and non-neuronal CA1 populations

I then looked at the genomewide interaction patterns of neurons and non-neurons in the CA1 cells. The initial libraries had approximately 215 and 144 million reads for neuronal and non-neuronal population. After iterative mapping, iterative correction and balancing of the interaction matrix were performed, I was left with approximately 12.5 and 1.7 million valid unique interaction pairs for the neuronal and non-neuronal population. As the experiment was a pilot and the quality of the libraries was good, I went ahead to perform the basic analysis to look at the differences between the two population cells. The results, however, should be very carefully and cautiously interpreted as they may or may not represent the full picture. However, given the quality of the data, I can still look at the basic interaction patterns. My first observation was that even with the higher number of initial reads as well as raw unique valid pairs, after balancing, there was a small increase in the looping interactions in the non-neuronal population compared to the neuronal population when visualized and inspected through the genomewide chromatin interactions heatmaps (Figure 9.7-1). This should, however, be cautiously interpreted as quantification is required both in the biological setup and computational setup. I then identified the topologically associating

domains (TADs) (6.7.1.3 above) from the neuronal and non-neuronal population and compared them to the published results (Figure 9.7-1). The number of TADs were also marginally higher in the neuronal population (1098) compared to the non-neuronal populations (1057). This number, however, might change when the TADs calling is performed at a higher resolution (current resolution is 500 kb genomic bins). It may be speculated that this may not affect the overall interaction pattern, but may be responsible for the differences in the interaction patterns that may be involved in disease mechanism.

10.2.3 Interaction pattern differences on a single chromosome (chr9) and near a single gene (BACE1)

In the end, I wanted to check the instances of the looping interactions that were different between neurons and non-neurons. I looked into the particular chromosome, chr 9 which has a very important gene (BACE1) that is involved in the Alzheimer's disease pathology [379, 381, 382, 548, 549]. However, due to the low number of uniquely valid pairs, I performed the analysis by the visual inspection of the data which still provided some striking evidences in the looping patterns of neurons and non-neurons near the BACE1 gene. The data and results should be cautiously interpreted. When I looked at the general interaction pattern on chromosome 9, there were few genomic locations with lower number of interactions and few genomic locations have a higher number of interactions in neurons compared to non-neurons (section 9.7.3 above). For one of the examples, there was a reduction in the gene expression pattern (Figure 9.7-5) that has a miRNAs cluster and also an increase peak density for H3K27m3 peaks which is a repressive mark. This calls for a compressive analysis for the next part of the study in the future. The absence of 500 Kb binned resolution TAD in non-neurons may play a role in the looping interaction pattern of the BACE1 gene which right now is a speculation. A high resolution TADs map should confirm the absence of this particular TAD in the BACE1 gene vicinity. The specific interaction profile of the BACE1 gene was crucial to understand and even with the low resolution data, I was able to identify that there was a decrease in the interaction pattern of the BACE1 genes in non-neurons compared to neurons. Along with that, the enhancer's density was also higher for the neurons near the BACE1 gene compared to the non-neurons (Figure 9.7-6). All the results in this pilot study suggest that there is a difference in the conformation of the cells from neuronal and non-neuronal population with a distinctive interaction pattern but it should be interpreted carefully.

11. Conclusions

The epigenetic regulation of gene expression plays a major role in memory formation and synaptic plasticity. The epigenetic mechanisms involve the regulation of proper gene expression by DNA methylation and demethylation, histone modifications and the role of noncoding RNAs along with the genetic and environmental factors. Deregulation of these epigenetic mechanisms [428, 550-553] along with the disruption of the three dimensional architecture of the genome results in a crisscrossing of the complex interconnection of long range looping interactions [238, 554] leading to the disease pathogenesis. Understanding these mechanisms becomes critical in fast progressing neurodegenerative disorders like Alzheimer's disease, Parkinson's disease and Amyotrophic Lateral Sclerosis (ALS). The need to develop biomarkers for the diagnosis, prognosis, and prediction of the onset of Alzheimer's disease - which can lead to personalized therapeutic tools - is very important. The key objectives of my study were to identify putative small noncoding RNAs biomarkers by the use of machine learning and statistical methods and get a basic understanding of the underlying mechanism that may be probable causes of Alzheimer's disease by looking into the three dimensional topology of the chromatin in the neuronal and non-neuronal samples. Those were the two main goals of my study and a brief summary of both aims is presented below.

The first aim was to study the epigenetic regulation of information in a neuronal network like exosomal transfer of small noncoding RNAs, which leads to the development of putative biomarkers for disease diagnosis, prognosis, and prediction. The data suggest that exosomes released from neurons mainly reflect the small noncoding RNA composition of the parental cells. This data is in line with other studies performed, for example on cancerous cells [555-558]. The analysis of CSF exosomes at least in the part that reveals the pathological small noncoding RNAome changes occurring in the brain as the top expressed miRNAs and piRNAs found in the CSF exosomes are linked to brain function [373, 509, 522, 559, 560]. This data also suggest that the analysis of miRNAs and piRNAs expression alone is suitable for the diagnosis and prognosis of Alzheimer's disease in an independent replication cohort during the biomarker development phase but not sufficiently high for use in clinical practice as of now. However, the combination of both the existing clinical protein markers like pTAU and A β 42/A β 40 ratio and small noncoding RNAs signature (three miRNAs, namely hsa-miR-27a-3p, hsa-miR-30a-5p, miR-34c-3p and three piRNAs, namely hsa-piR-019949, hsa-piR-020364 and hsa-piR-019324) provides very high accuracy and may be used in the clinics for the disease diagnosis and prognosis compared to the use of

individual markers. Taking into account that small noncoding RNAs are extremely stable [561-563] and are rather unaffected by repeated freezing and thawing cycles our findings suggest that both miRNAs and piRNAs [564] should be considered as the additional biomarker for diagnosis and patient stratification. It is also reported that the changes in these small noncoding RNAs represent a combination of environmental and genetic risk factors and not simply A β or Tau load providing additional and valuable information for the onset of Mild Cognitive Impairment (MCI) and Alzheimer's disease. Finally, the study reported that the analysis of these small noncoding RNAs from CSF exosomes potentially informs about many molecular and biological processes that might fail to classify patients when assessed individually. This may be the primary reason for the excellent performance of the reported piRNAs signature and a good performance for miRNAs signature in predicting the conversion of MCI patients from mciStable to mciAD.

Through this study, I also provide various statistical and machine learning toolkits and pipelines with open source access to the code. This not only ensures that the results of the study can be replicated, but also, at the same time, make sure that the same pipelines can be used on a larger cohort so that the data is analyzed in a standard way which will avoid various biases that may introduce during data analysis and interpretation of the results. This is very important, as before, making any major deduction of the usability of these small noncoding RNAs signatures as biomarkers from this study, the data should be systematically tested in a study with bigger population size and conducted over a longer period of time in the clinical setup.

The second aim was to study the role of dynamic changes in higher order chromatin structure that control gene expression programs in neuronal plasticity. This was a pilot study with the goal to get a glimpse of the three dimensional organization of the genome inside the neuronal and non-neuronal population of the cells and its impact on the genetic and epigenetic regulatory network by the use of Chromosome Conformation Capture (3C) based technique (3C-seq). The study was successfully performed with the 3C-seq technique in the FACS sorted neuronal and non-neuronal nuclei from CA1 (the specific sub region of the dorsal hippocampus) in the mouse. The 3C-seq experiment was successful, but due to the low input RNA concentration, the library could not be sequenced deep enough to get a fragment level resolution data for the analysis. Even with the low resolution data (500 kb bins), the study reported that there is a difference in the number of topologically associating domains in the neuronal and non-neuronal population of the cells. The study reports many regions where there are distinct interaction pattern in the neuronal and non-neuronal cells for example between the TSS of a gene and an enhancer mark (H3k4me1); suggesting their

involvement in the regulation of a gene that might be a part of the disease pathway. One such finding reported a distinct looping interaction pattern of the BACE1 gene with the neighboring enhancers present in neuronal cells compared to the corresponding non-neuronal cells. The data and results, however, should be interpreted by carefully as the data cannot be binned at a higher resolution for more concrete TAD calling and assessment of the looping interactions within those TADs. However, it does provide an overview of how the looping interaction pattern is present in the two sub populations of the cells in the CA1 region. In the future, I want to look at the role of the selected small noncoding RNAs found the aim 1 in different neuronal and non-neuronal populations. I also want to see the effects of these small noncoding RNAs on the proteome if we knock out some of them to look at the gain or loss of function.

In the end, my Ph.D. work will significantly contribute to better understanding on how the epigenetic code controls neuronal plasticity and memory function which will help not only to better understand the brain, but may also help to identify novel therapeutic avenues to treat cognitive dysfunction and complex brain diseases.

12. References

- [1] M.-T. Heemels, Neurodegenerative diseases, *Nature*, 539 (2016) 179.
- [2] T.A. Yacoubian, Chapter 1 - Neurodegenerative Disorders: Why Do We Need New Therapies?, in: A. Adejare (Ed.) *Drug Discovery Approaches for the Treatment of Neurodegenerative Disorders*, (Academic Press, 2017), pp. 1-16.
- [3] A. Abeliovich, et al., Defects in trafficking bridge Parkinson's disease pathology and genetics, *Nature*, 539 (2016) 207.
- [4] C. Butler, et al., Neurological syndromes which can be mistaken for psychiatric conditions, *Journal of neurology, neurosurgery, and psychiatry*, 76 Suppl 1 (2005) i31-i38.
- [5] B.R. Rund, Is schizophrenia a neurodegenerative disorder?, *Nord J Psychiatry*, 63 (2009) 196-201.
- [6] A.M. Gorman, Neuronal cell death in neurodegenerative diseases: recurring themes around protein handling, *Journal of cellular and molecular medicine*, 12 (2008) 2263-2280.
- [7] F.M. Elahi, et al., A clinicopathological approach to the diagnosis of dementia, *Nature reviews. Neurology*, 13 (2017) 457-476.
- [8] Factsheet 470LP - What is mild cognitive impairment (MCI)?
- [9] E.R. Edwards, et al., Neuropsychiatric symptoms in mild cognitive impairment: differences by subtype and progression to dementia, *International journal of geriatric psychiatry*, 24 (2009) 716-722.
- [10] S.Y. Pandya, et al., Predictors of Reversion from Mild Cognitive Impairment to Normal Cognition, *Dementia and geriatric cognitive disorders*, 43 (2017) 204-214.
- [11] M.D.C. Díaz-Mardomingo, et al., Problems in Classifying Mild Cognitive Impairment (MCI): One or Multiple Syndromes?, *Brain sciences*, 7 (2017).
- [12] M.S. Albert, et al., The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 7 (2011) 270-279.
- [13] I. Contador, et al., Cause of death in mild cognitive impairment: a prospective study (NEDICES), *European journal of neurology*, 21 (2014) 253-e259.
- [14] What is mild cognitive impairment (MCI)?
- [15] Mayo Clinic Staff, Mild cognitive impairment (MCI) <https://www.mayoclinic.org/diseases-conditions/mild-cognitive-impairment/symptoms-causes/syc-20354578>, (2011).
- [16] F. Portet, et al., Mild cognitive impairment (MCI) in medical practice: a critical review of the concept and new diagnostic procedure. Report of the MCI Working Group of the European Consortium on Alzheimer's Disease, *Journal of neurology, neurosurgery, and psychiatry*, 77 (2006) 714-718.

- [17] R.C. Petersen, et al., Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology, *Neurology*, 90 (2018) 126-135.
- [18] A. Burns, et al., *Dementia*, *BMJ*, 338 (2009) b75.
- [19] World Health Organization, *Dementia: a public health priority* (World Health Organization, 2012).
- [20] World Health Organization, *Dementia - A public health priority*, First WHO Ministerial Conference on the Global Action Against Dementia, (2015).
- [21] J.K. Harrison, et al., Outcomes measures in a decade of dementia and mild cognitive impairment trials, *Alzheimer's research & therapy*, 8 (2016) 48.
- [22] L.O. Killin, et al., Environmental risk factors for dementia: a systematic review, *BMC geriatrics*, 16 (2016) 175.
- [23] R.W. Aslam, et al., A systematic review of the diagnostic accuracy of automated tests for cognitive impairment, *International journal of geriatric psychiatry*, 33 (2018) 561-575.
- [24] J. Cerejeira, et al., Behavioral and psychological symptoms of dementia, *Frontiers in neurology*, 3 (2012) 73.
- [25] Alzheimer's Association, 2017 Alzheimer's disease facts and figures, *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 13 (2017) 325-373.
- [26] R. Barber, et al., White matter lesions on magnetic resonance imaging in dementia with Lewy bodies, Alzheimer's disease, vascular dementia, and normal aging, *Journal of neurology, neurosurgery, and psychiatry*, 67 (1999) 66-72.
- [27] C.A. Taylor, et al., Deaths from Alzheimer's Disease—United States, 1999–2014, *MMWR. Morbidity and mortality weekly report*, 66 (2017) 521.
- [28] Alzheimer's Association, 2016 Alzheimer's disease facts and figures, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 12 (2016) 459-509.
- [29] J.A.R. Nicoll, et al., Neuropathology of human Alzheimer disease after immunization with amyloid- β peptide: a case report, *Nature Medicine*, 9 (2003) 448.
- [30] H. Braak, et al., Neuropathological staging of Alzheimer-related changes, *Acta Neuropathologica*, 82 (1991) 239-259.
- [31] H. Hampel, et al., Revolution of Alzheimer Precision Neurology. Passageway of Systems Biology and Neurophysiology, *Journal of Alzheimer's disease : JAD*, 64 (2018) S47-S105.
- [32] C.R. Jack, Jr., et al., An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease, *Annals of neurology*, 71 (2012) 765-775.
- [33] B. Dubois, et al., Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 12 (2016) 292-323.
- [34] J.T. Shelton, et al., Improving Prospective Memory in Healthy Older Adults and Individuals with Very Mild Alzheimer's Disease, *Journal of the American Geriatrics Society*, 64 (2016) 1307-1312.

- [35] A.M. Berardi, et al., Sustained attention in mild Alzheimer's disease, *Developmental neuropsychology*, 28 (2005) 507-537.
- [36] M. Laisney, et al., Cognitive and affective Theory of Mind in mild to moderate Alzheimer's disease, *Journal of neuropsychology*, 7 (2013) 107-120.
- [37] R. Rosa, et al., Caregivers' resilience in mild and moderate Alzheimer's disease, *Aging & mental health*, (2018) 1-9.
- [38] F.A. Schmitt, et al., A systematic review of assessment and treatment of moderate to severe Alzheimer's disease, *Primary care companion to the Journal of clinical psychiatry*, 8 (2006) 158-159.
- [39] B.T. Hyman, et al., National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 8 (2012) 1-13.
- [40] G. Šimić, et al., Monoaminergic neuropathology in Alzheimer's disease, *Progress in neurobiology*, 151 (2017) 101-138.
- [41] T. Voisin, et al., Diagnosis and treatment of patients with severe Alzheimer's disease, *Drugs & aging*, 26 (2009) 135-144.
- [42] L.W. Chu, Alzheimer's disease: early diagnosis and treatment, *Hong Kong medical journal = Xianggang yi xue za zhi*, 18 (2012) 228-237.
- [43] P.N. Tariot, et al., Current treatment for Alzheimer disease and future prospects, *Alzheimer disease and associated disorders*, 17 Suppl 4 (2003) S105-113.
- [44] J.L. Cummings, Treatment of Alzheimer's disease: the role of symptomatic agents in an era of disease-modifying therapies, *Reviews in neurological diseases*, 4 (2007) 57-62.
- [45] R. Casanova, et al., High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization, *Frontiers in neuroinformatics*, 5 (2011) 22.
- [46] R. Ossenkoppele, et al., Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease, *Brain : a journal of neurology*, 139 (2016) 1551-1567.
- [47] J.F. Burke, et al., Assessment of mild dementia with amyloid and dopamine terminal positron emission tomography, *Brain : a journal of neurology*, 134 (2011) 1647-1657.
- [48] Health Quality Ontario, The appropriate use of neuroimaging in the diagnostic work-up of dementia: an evidence-based analysis, *Ontario health technology assessment series*, 14 (2014) 1-64.
- [49] K.A. Johnson, et al., Brain imaging in Alzheimer disease, *Cold Spring Harbor perspectives in medicine*, 2 (2012) a006213.
- [50] E. Cavedo, et al., The Road Ahead to Cure Alzheimer's Disease: Development of Biological Markers and Neuroimaging Methods for Prevention Trials Across all Stages and Target Populations, *The journal of prevention of Alzheimer's disease*, 1 (2014) 181-202.
- [51] J.B. Toledo, et al., Alzheimer's disease cerebrospinal fluid biomarker in cognitively normal subjects, *Brain : a journal of neurology*, 138 (2015) 2701-2715.
- [52] J.L. Molinuevo, et al., Current state of Alzheimer's fluid biomarkers, *Acta Neuropathol*, 136 (2018) 821-853.

- [53] M.W. Weiner, et al., The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 9 (2013) e111-194.
- [54] A. Atri, Effective pharmacological management of Alzheimer's disease, *The American journal of managed care*, 17 Suppl 13 (2011) S346-355.
- [55] P.M. McKeever, et al., MicroRNA Expression Levels Are Altered in the Cerebrospinal Fluid of Patients with Young-Onset Alzheimer's Disease, *Molecular neurobiology*, 55 (2018) 8826-8841.
- [56] D.A. Bennett, et al., Epigenomics of Alzheimer's disease, *Translational research : the journal of laboratory and clinical medicine*, 165 (2015) 200-220.
- [57] A. Sierksma, et al., Deregulation of neuronal miRNAs induced by amyloid- β or TAU pathology, *Molecular neurodegeneration*, 13 (2018) 54.
- [58] J.J. Chen, et al., Potential Roles of Exosomal MicroRNAs as Diagnostic Biomarkers and Therapeutic Application in Alzheimer's Disease, *Neural plasticity*, 2017 (2017) 7027380.
- [59] T. Croese, et al., Extracellular vesicles in neurodegenerative diseases, *Molecular aspects of medicine*, 60 (2018) 52-61.
- [60] S.L.N. Maas, et al., Extracellular Vesicles: Unique Intercellular Delivery Vehicles, *Trends in cell biology*, 27 (2017) 172-188.
- [61] J.R. Edgar, Q&A: What are exosomes, exactly?, *BMC Biology*, 14 (2016) 46.
- [62] D. Xu, et al., The role of exosomes and microRNAs in senescence and aging, *Advanced drug delivery reviews*, 65 (2013) 368-375.
- [63] C. Arenaccio, et al., The Multifaceted Functions of Exosomes in Health and Disease: An Overview, *Advances in experimental medicine and biology*, 998 (2017) 3-19.
- [64] J.S. Schorey, et al., Exosomes and other extracellular vesicles in host-pathogen interactions, *EMBO reports*, 16 (2015) 24.
- [65] M. Tkach, et al., Why the need and how to approach the functional diversity of extracellular vesicles, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 373 (2018).
- [66] M. Moras, et al., From Erythroblasts to Mature Red Blood Cells: Organelle Clearance in Mammals, *Frontiers in Physiology*, 8 (2017) 1076.
- [67] J.S. Schorey, et al., Exosomes and other extracellular vesicles in host-pathogen interactions, *EMBO Rep*, 16 (2015) 24-43.
- [68] H.R. M, et al., Exosomes: From Garbage Bins to Promising Therapeutic Targets, *International journal of molecular sciences*, 18 (2017).
- [69] T. Xiao, et al., The role of exosomes in the pathogenesis of Alzheimer' disease, *Transl Neurodegener*2017), pp. 3.
- [70] T. Malm, et al., Exosomes in Alzheimer's disease, *Neurochemistry international*, 97 (2016) 193-199.
- [71] K. Yuyama, et al., Exosomes as Carriers of Alzheimer's Amyloid- β , *Frontiers in Neuroscience*, 11 (2017) 229.

- [72] X. Chai, et al., Constitutive secretion of tau protein by an unconventional mechanism, *Neurobiology of disease*, 48 (2012) 356-366.
- [73] S. Saman, et al., Exosome-associated tau is secreted in tauopathy models and is selectively phosphorylated in cerebrospinal fluid in early Alzheimer disease, *The Journal of biological chemistry*, 287 (2012) 3842-3849.
- [74] D. Simon, et al., Proteostasis of tau. Tau overexpression results in its secretion via membrane vesicles, *FEBS letters*, 586 (2012) 47-54.
- [75] M.S. Fiandaca, et al., Identification of preclinical Alzheimer's disease by a profile of pathogenic proteins in neurally derived blood exosomes: A case-control study, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 11 (2015) 600-607 e601.
- [76] J. Kocki, et al., Dysregulation of Amyloid-beta Protein Precursor, beta-Secretase, Presenilin 1 and 2 Genes in the Rat Selectively Vulnerable CA1 Subfield of Hippocampus Following Transient Global Brain Ischemia, *Journal of Alzheimer's disease : JAD*, 47 (2015) 1047-1056.
- [77] R. Pluta, et al., Sporadic Alzheimer's disease begins as episodes of brain ischemia and ischemically dysregulated Alzheimer's disease genes, *Molecular neurobiology*, 48 (2013) 500-515.
- [78] L. Cheng, et al., Prognostic serum miRNA biomarkers associated with Alzheimer's disease shows concordance with neuropsychological and neuroimaging assessment, *Molecular psychiatry*, 20 (2015) 1188-1196.
- [79] J. Satoh, et al., MicroRNA-Seq Data Analysis Pipeline to Identify Blood Biomarkers for Alzheimer's Disease from Public Data, *Biomarker insights*, 10 (2015) 21-31.
- [80] R.M. Califf, Biomarker definitions and their applications, *Experimental biology and medicine (Maywood, N.J.)*, 243 (2018) 213-221.
- [81] J.K. Aronson, et al., Biomarkers-A General Review, *Current protocols in pharmacology*, 76 (2017) 9 23 21-29 23 17.
- [82] X. Zhao, et al., Biomarkers in Pharmaceutical Research, *Clinical chemistry*, 61 (2015) 1343-1353.
- [83] J.L. Allinson, Clinical biomarker validation, *Bioanalysis*, 10 (2018) 957-968.
- [84] J.M. Sauer, et al., Preclinical biomarker qualification, *Experimental biology and medicine (Maywood, N.J.)*, 243 (2018) 222-227.
- [85] J. Nielsen, Systems Biology of Metabolism: A Driver for Developing Personalized and Precision Medicine, *Cell metabolism*, 25 (2017) 572-579.
- [86] D.F. Hayes, Biomarker validation and testing, *Molecular oncology*, 9 (2015) 960-966.
- [87] W.C. Albrich, et al., Pros and cons of using biomarkers versus clinical decisions in start and stop decisions for antibiotics in the critical care setting, *Intensive care medicine*, 41 (2015) 1739-1751.
- [88] W. Chen, et al., On the assessment of the added value of new predictive biomarkers, *BMC medical research methodology*, 13 (2013) 98.
- [89] V. Mantzavinos, et al., Biomarkers for Alzheimer's Disease Diagnosis, *Current Alzheimer research*, 14 (2017) 1149-1154.

- [90] A. Jeromin, et al., Biomarkers in Neurodegenerative Diseases, *Advances in neurobiology*, 15 (2017) 491-528.
- [91] B. Olsson, et al., CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis, *The Lancet. Neurology*, 15 (2016) 673-684.
- [92] R.A. Hickman, et al., Alzheimer Disease and Its Growing Epidemic: Risk Factors, Biomarkers, and the Urgent Need for Therapeutics, *Neurologic clinics*, 34 (2016) 941-953.
- [93] W. Bao, et al., PET Imaging for Early Detection of Alzheimer's Disease: From Pathologic to Physiologic Biomarkers, *PET clinics*, 12 (2017) 329-350.
- [94] C.C. Tan, et al., Biomarkers for preclinical Alzheimer's disease, *Journal of Alzheimer's disease : JAD*, 42 (2014) 1051-1069.
- [95] S.E. Counts, et al., Biomarkers for the Early Detection and Progression of Alzheimer's Disease, *Neurotherapeutics : the journal of the American Society for Experimental NeuroTherapeutics*, 14 (2017) 35-53.
- [96] J. Gayon, From Mendel to epigenetics: History of genetics, *Comptes rendus biologies*, 339 (2016) 225-230.
- [97] A. Ganesan, Epigenetics: the first 25 centuries, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 373 (2018).
- [98] C. Deans, et al., What do you mean, "epigenetic"?, *Genetics*, 199 (2015) 887-896.
- [99] A. Nicoglou, et al., Epigenetics: A way to bridge the gap between biological fields, *Studies in history and philosophy of biological and biomedical sciences*, 66 (2017) 73-82.
- [100] C. Waddington, *The strategy of the genes: a discussion of some aspects of theoretical biology.*, (London: Allen & Unwin 1957).
- [101] C.H. Waddington, *The strategy of the genes* (Routledge, 2014).
- [102] V. Prachayasittikul, et al., Exploring the epigenetic drug discovery landscape, *Expert opinion on drug discovery*, 12 (2017) 345-362.
- [103] L. Adwan, et al., Epigenetics: a novel therapeutic approach for the treatment of Alzheimer's disease, *Pharmacology & therapeutics*, 139 (2013) 41-50.
- [104] I.M. Fingerman, et al., NCBI Epigenomics: what's new for 2013, *Nucleic acids research*, 41 (2013) D221-225.
- [105] A. Fischer, Epigenetic memory: the Lamarckian brain, *The EMBO journal*, (2014) e201387637.
- [106] S. Hombach, et al., Non-coding RNAs: Classification, Biology and Functioning, *Advances in experimental medicine and biology*, 937 (2016) 3-17.
- [107] J.W. Wei, et al., Non-coding RNAs as regulators in epigenetics (Review), *Oncology reports*, 37 (2017) 3-9.
- [108] J.V. Philippe, et al., Probing small non-coding RNAs structures, *Methods in molecular biology* (Clifton, N.J.), 1296 (2015) 119-136.
- [109] B. Huang, et al., Regulatory non-coding RNAs: revolutionizing the RNA world, *Molecular biology reports*, 41 (2014) 3915-3923.

- [110] C.J. Chen, et al., Small RNAs derived from structural non-coding RNAs, *Methods* (San Diego, Calif.), 63 (2013) 76-84.
- [111] L. Zhu, et al., tRNA-derived small non-coding RNAs in human disease, *Cancer letters*, 419 (2018) 1-7.
- [112] A.G. Matera, et al., Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs, *Nature reviews. Molecular cell biology*, 8 (2007) 209-220.
- [113] M.D. Shortridge, et al., Structure based approaches for targeting non-coding RNAs with small molecules, *Current opinion in structural biology*, 30 (2015) 79-88.
- [114] H. Kobayashi, et al., RISC assembly: Coordination between small RNAs and Argonaute proteins, *Biochimica et biophysica acta*, 1859 (2016) 71-81.
- [115] I. Dugaard, et al., Biogenesis and Function of Ago-Associated RNAs, *Trends in genetics : TIG*, 33 (2017) 208-219.
- [116] Y. Shibayama, et al., Visualization of Enhancer-Derived Noncoding RNA, *Methods in molecular biology* (Clifton, N.J.), 1468 (2017) 19-32.
- [117] F. Liu, Enhancer-derived RNA: A Primer, *Genomics Proteomics Bioinformatics*, 15 (2017) 196-200.
- [118] B.X. Yan, et al., Promoter-associated RNAs and promoter-targeted RNAs, *Cellular and molecular life sciences : CMLS*, 69 (2012) 2833-2842.
- [119] S. Napoli, Targeting Promoter-Associated RNAs by siRNAs, *Methods in molecular biology* (Clifton, N.J.), 1543 (2017) 209-219.
- [120] V.N. Kim, MicroRNA biogenesis: coordinated cropping and dicing, *Nature reviews. Molecular cell biology*, 6 (2005) 376-385.
- [121] Y. Devaux, et al., MicroRNAs: new biomarkers and therapeutic targets after cardiac arrest?, *Critical care* (London, England), 19 (2015) 54.
- [122] J. Roy, et al., Small RNA sequencing revealed dysregulated piRNAs in Alzheimer's disease and their probable role in pathogenesis, *Molecular bioSystems*, 13 (2017) 565-576.
- [123] G. Singh, et al., Genome-wide profiling of the PIWI-interacting RNA-mRNA regulatory networks in epithelial ovarian cancers, *PLoS one*, 13 (2018) e0190485.
- [124] W. Sun, et al., Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies, *Nature neuroscience*, 21 (2018) 1038-1048.
- [125] T. Watanabe, et al., Posttranscriptional regulation of gene expression by Piwi proteins and piRNAs, *Molecular cell*, 56 (2014) 18-27.
- [126] M. Basavaraju, et al., Alzheimer's disease: presence and role of microRNAs, *Biomolecular concepts*, 7 (2016) 241-252.
- [127] C. Ballatore, et al., Tau-mediated neurodegeneration in Alzheimer's disease and related disorders, *Nature reviews. Neuroscience*, 8 (2007) 663-672.
- [128] M.L. Idda, et al., Noncoding RNAs in Alzheimer's disease, *Wiley interdisciplinary reviews. RNA*, 9 (2018).

- [129] Y. Bastanlar, et al., Introduction to machine learning, *Methods in molecular biology* (Clifton, N.J.), 1107 (2014) 105-128.
- [130] H.O. Alanazi, et al., A Critical Review for Developing Accurate and Dynamic Predictive Models Using Machine Learning Methods in Medicine and Health Care, *Journal of medical systems*, 41 (2017) 69.
- [131] A.M. Darcy, et al., Machine Learning and the Profession of Medicine, *Jama*, 315 (2016) 551-552.
- [132] R.C. Deo, et al., Learning About Machine Learning: The Promise and Pitfalls of Big Data and the Electronic Health Record, *Circulation. Cardiovascular quality and outcomes*, 9 (2016) 618-620.
- [133] D.M. Camacho, et al., Next-Generation Machine Learning for Biological Networks, *Cell*, 173 (2018) 1581-1592.
- [134] M.G. Shafto, et al., Teacher and learner: Supervised and unsupervised learning in communities, *The Behavioral and brain sciences*, 38 (2015) e64.
- [135] J.C. Ang, et al., Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection, *IEEE/ACM transactions on computational biology and bioinformatics*, 13 (2016) 971-989.
- [136] G. Huang, et al., Semi-supervised and unsupervised extreme learning machines, *IEEE transactions on cybernetics*, 44 (2014) 2405-2417.
- [137] B.C. Love, Comparing supervised and unsupervised category learning, *Psychonomic bulletin & review*, 9 (2002) 829-835.
- [138] M.W. Libbrecht, et al., Machine learning applications in genetics and genomics, *Nature reviews. Genetics*, 16 (2015) 321-332.
- [139] G. James, et al., *An introduction to statistical learning* (Springer, 2013).
- [140] A.C. Rencher, *Methods of multivariate analysis* (John Wiley & Sons, 2003).
- [141] C. Santolin, et al., Unsupervised statistical learning in newly hatched chicks, *Current biology : CB*, 26 (2016) R1218-R1220.
- [142] G. Liu, et al., Unsupervised Learning in PET Radiomics, *IEEE Nuclear Science Symposium conference record. Nuclear Science Symposium*, 2017 (2017).
- [143] G. Alessi, Generative strategies and teaching for generalization, *The Analysis of verbal behavior*, 5 (1987) 15-27.
- [144] A. Testolin, et al., Probabilistic Models and Generative Neural Networks: Towards an Unified Framework for Modeling Normal and Impaired Neurocognitive Functions, *Frontiers in computational neuroscience*, 10 (2016) 73.
- [145] V.R. Bejjanki, et al., Learning and inference using complex generative models in a spatial localization task, *Journal of vision*, 16 (2016) 9.
- [146] A. Kandler, et al., Generative inference for cultural evolution, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 373 (2018).
- [147] W. Schultz, Neuronal Reward and Decision Signals: From Theories to Data, *Physiological reviews*, 95 (2015) 853-951.

- [148] L. Zhou, et al., Learning Discriminative Bayesian Networks from High-Dimensional Continuous Neuroimaging Data, *IEEE transactions on pattern analysis and machine intelligence*, 38 (2016) 2269-2283.
- [149] B. Lei, et al., Discriminative Learning for Alzheimer's Disease Diagnosis via Canonical Correlation Analysis and Multimodal Fusion, *Frontiers in aging neuroscience*, 8 (2016) 77.
- [150] D. Ferranti, et al., The value of prior knowledge in machine learning of complex network systems, *Bioinformatics*, 33 (2017) 3610-3618.
- [151] M. Diligenti, et al., Integrating Prior Knowledge into Deep Learning, 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)2017), pp. 920-923.
- [152] X. Fang, et al., Perturbation of old knowledge precedes integration of new knowledge, *Neuropsychologia*, 99 (2017) 270-278.
- [153] D. Yurovsky, et al., The role of partial knowledge in statistical word learning, *Psychonomic bulletin & review*, 21 (2014) 1-22.
- [154] L.H. Wooley JC, On the Nature of Biological Data, National Research Council (US) Committee on Frontiers at the Interface of Computing and Biology, (2005).
- [155] E. Baralis, et al., Exploring Heterogeneous Biological Data Sources, 2008 19th International Workshop on Database and Expert Systems Applications2008), pp. 647-651.
- [156] A. Gough, et al., Biologically Relevant Heterogeneity: Metrics and Practical Insights, *SLAS discovery : advancing life sciences R & D*, 22 (2017) 213-237.
- [157] A. Nazabal, et al., Handling incomplete heterogeneous data using VAEs, arXiv preprint arXiv:1807.03653, (2018).
- [158] D. Pyle, *Data Preparation for Data Mining* (Morgan Kaufmann Publishers Inc., 1999).
- [159] S.M. Ali, et al., From conceptual design to performance optimization of ETL workflows: current state of research and open problems, *The VLDB Journal*, 26 (2017) 777-801.
- [160] S. Kandel, et al., Research directions in data wrangling: Visualizations and transformations for usable and credible data, *Information Visualization*, 10 (2011) 271-288.
- [161] S. Ahuja, et al., Using Machine Learning to Accelerate Data Wrangling, 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)2016), pp. 343-349.
- [162] F. Abbas-Aghababazadeh, et al., Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing, *PloS one*, 13 (2018) e0206312.
- [163] T. Park, et al., Evaluation of normalization methods for microarray data, *BMC Bioinformatics*, 4 (2003) 33.
- [164] D. Risso, et al., Normalization of RNA-seq data using factor analysis of control genes or samples, *Nature biotechnology*, 32 (2014) 896-902.
- [165] K.P. McCormick, et al., Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments, *Silence*, 2 (2011) 2.
- [166] M. Carandini, et al., Normalization as a canonical neural computation, *Nature Reviews Neuroscience*, 14 (2013) 152.

- [167] Z. Hu, et al., SeqVISTA: a graphical tool for sequence feature visualization and comparison, *BMC Bioinformatics*, 4 (2003) 1.
- [168] L. Breiman, Random Forests, *Machine Learning*, 45 (2001) 5-32.
- [169] L. Breiman, et al., Package randomForest, Software available at: <http://stat-www.berkeley.edu/users/breiman/RandomForests>, (2011).
- [170] R. Couronné, et al., Random forest versus logistic regression: a large-scale benchmark experiment, *BMC Bioinformatics*, 19 (2018) 270.
- [171] Y.V. Karpievitch, et al., An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++, *PloS one*, 4 (2009) e7087.
- [172] A. Sarica, et al., Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease: A Systematic Review, *Frontiers in aging neuroscience*, 9 (2017) 329.
- [173] J. Friedman, et al., *The elements of statistical learning* (Springer series in statistics New York, 2001).
- [174] J. Yao, et al., Checking account activity and credit default risk of enterprises: An application of statistical learning methods, *arXiv preprint arXiv:1707.00757*, (2017).
- [175] J. Cai, et al., Feature selection in machine learning: A new perspective, *Neurocomputing*, 300 (2018) 70-79.
- [176] L. Wang, et al., Feature selection methods for big data bioinformatics: A survey from the search perspective, *Methods (San Diego, Calif.)*, 111 (2016) 21-31.
- [177] J. Krawczuk, et al., The feature selection bias problem in relation to high-dimensional gene data, *Artificial intelligence in medicine*, 66 (2016) 63-71.
- [178] M. De Marco, et al., Machine-learning Support to Individual Diagnosis of Mild Cognitive Impairment Using Multimodal MRI and Cognitive Assessments, *Alzheimer disease and associated disorders*, 31 (2017) 278-286.
- [179] L.B. Holder, et al., Machine learning for epigenetics and future medical applications, *Epigenetics*, 12 (2017) 505-514.
- [180] M. Hossin, et al., A review on evaluation metrics for data classification evaluations, *International Journal of Data Mining & Knowledge Management Process*, 5 (2015) 1.
- [181] A. Tharwat, Classification assessment methods, *Applied Computing and Informatics*, (2018).
- [182] R.J. Urbanowicz, et al., Benchmarking relief-based feature selection methods for bioinformatics data mining, *Journal of biomedical informatics*, 85 (2018) 168-188.
- [183] B.S. Cole, et al., Analysis of Gene-Gene Interactions, *Current protocols in human genetics*, 95 (2017) 1 14 11-11 14 10.
- [184] T. Fawcett, An introduction to ROC analysis, *Pattern Recognition Letters*, 27 (2006) 861-874.
- [185] R.S. Olson, et al., PMLB: a large benchmark suite for machine learning evaluation and comparison, *BioData mining*, 10 (2017) 36.

- [186] J.D. Kelleher, et al., *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies* (The MIT Press, 2015).
- [187] L. Mosley, A balanced approach to the multi-class imbalance problem, *Graduate Theses and Dissertations*. 13537. , (2013).
- [188] I. Guyon, et al., Design of the 2015 ChaLearn AutoML challenge, *2015 International Joint Conference on Neural Networks (IJCNN)2015*, pp. 1-8.
- [189] G.W. BRIER, VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY, *Monthly Weather Review*, 78 (1950) 1-3.
- [190] C. Lopez, et al., An unsupervised machine learning method for discovering patient clusters based on genetic signatures, *Journal of biomedical informatics*, 85 (2018) 30-39.
- [191] H. Jia, et al., A New Distance Metric for Unsupervised Learning of Categorical Data, *IEEE transactions on neural networks and learning systems*, 27 (2016) 1065-1079.
- [192] S. Yong, et al., Feature Selection With L2,1-2 Regularization, *IEEE transactions on neural networks and learning systems*, 29 (2018) 4967-4982.
- [193] J. Peltonen, et al., Improved learning of Riemannian metrics for exploratory analysis, *Neural networks : the official journal of the International Neural Network Society*, 17 (2004) 1087-1100.
- [194] S. Park, et al., hc-OTU: A Fast and Accurate Method for Clustering Operational Taxonomic Units Based on Homopolymer Compaction, *IEEE/ACM transactions on computational biology and bioinformatics*, 15 (2018) 441-451.
- [195] Y.B. Schwartz, et al., Three-Dimensional Genome Organization and Function in *Drosophila*, *Genetics*, 205 (2017) 5-24.
- [196] J. Achinger-Kawecka, et al., Disruption of the 3D cancer genome blueprint, *Epigenomics*, 9 (2017) 47-55.
- [197] B. Bonev, et al., Organization and function of the 3D genome, *Nature reviews. Genetics*, 17 (2016) 661-678.
- [198] L. Koch, Genome organization: 3D genome architecture--of loops and globules, *Nature reviews. Genetics*, 15 (2014) 780.
- [199] P. Szalaj, et al., Three-dimensional organization and dynamics of the genome, *Cell Biol Toxicol*, 34 (2018) 381-404.
- [200] J. Dekker, et al., The 3D Genome as Moderator of Chromosomal Communication, *Cell*, 164 (2016) 1110-1121.
- [201] K. Baumann, Genome organization: A vision of 3D chromatin organization, *Nature reviews. Molecular cell biology*, 18 (2017) 532.
- [202] V. Ramani, et al., Understanding Spatial Genome Organization: Methods and Insights, *Genomics, Proteomics & Bioinformatics*, 14 (2016) 7-20.
- [203] A. Bolzer, et al., Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes, *PLoS biology*, 3 (2005) e157.
- [204] K. Sengupta, Genome 3D-architecture: Its plasticity in relation to function, *Journal of biosciences*, 43 (2018) 417-419.

- [205] B.A. Bouwman, et al., Getting the genome in shape: the formation of loops, domains and compartments, *Genome Biol*, 16 (2015) 154.
- [206] J.H. Gibcus, et al., The hierarchy of the 3D genome, *Molecular cell*, 49 (2013) 773-782.
- [207] S.V. Ulianov, et al., Nuclear compartments, genome folding, and enhancer-promoter communication, *International review of cell and molecular biology*, 315 (2015) 183-244.
- [208] S.S. Roy, et al., Insights about genome function from spatial organization of the genome, *Human genomics*, 12 (2018) 8.
- [209] J.R. Dixon, et al., Topological domains in mammalian genomes identified by analysis of chromatin interactions, *Nature*, 485 (2012) 376-380.
- [210] E.P. Nora, et al., Spatial partitioning of the regulatory landscape of the X-inactivation centre, *Nature*, 485 (2012) 381-385.
- [211] C. Weinreb, et al., Identification of hierarchical chromatin domains, *Bioinformatics*, 32 (2016) 1601-1609.
- [212] A.S. Hansen, et al., Recent evidence that TADs and chromatin loops are dynamic structures, *Nucleus (Austin, Tex.)*, 9 (2018) 20-32.
- [213] A. Gonzalez-Sandoval, et al., On TADs and LADs: Spatial Control Over Gene Expression, *Trends in genetics : TIG*, 32 (2016) 485-495.
- [214] B.J. Matthews, et al., Insulation of gene expression by CTCF and cohesin-based subTAD (ubiquitous intra-TAD) loop structures in mouse liver, *bioRxiv*, (2018) 271023.
- [215] W. Yu, et al., Identifying topologically associating domains and subdomains by Gaussian Mixture model And Proportion test, *Nature communications*, 8 (2017) 535.
- [216] J. Nuebler, et al., Chromatin organization by an interplay of loop extrusion and compartmental segregation, *Proceedings of the National Academy of Sciences of the United States of America*, 115 (2018) E6697-E6706.
- [217] S. Kadauke, et al., Chromatin loops in gene regulation, *Biochimica et biophysica acta*, 1789 (2009) 17-25.
- [218] B.J. Matthews, et al., Computational prediction of CTCF/cohesin-based intra-TAD loops that insulate chromatin contacts and gene expression in mouse liver, *eLife*, 7 (2018).
- [219] M. Oti, et al., CTCF-mediated chromatin loops enclose inducible gene regulatory domains, *BMC genomics*, 17 (2016) 252.
- [220] K.P. Eagen, et al., Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map, *Proceedings of the National Academy of Sciences of the United States of America*, 114 (2017) 8764-8769.
- [221] A.S. Hansen, et al., CTCF and cohesin regulate chromatin loop stability with distinct dynamics, *eLife*, 6 (2017).
- [222] T. Ma, et al., Developing novel methods to image and visualize 3D genomes, *Cell Biol Toxicol*, 34 (2018) 367-380.
- [223] Y. Guo, et al., CRISPR-mediated deletion of prostate cancer risk-associated CTCF loop anchors identifies repressive chromatin loops, *Genome Biol*, 19 (2018) 160.

- [224] Y. Guo, et al., CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function, *Cell*, 162 (2015) 900-910.
- [225] P.R. Langer-Safer, et al., Immunological method for mapping genes on Drosophila polytene chromosomes, *Proceedings of the National Academy of Sciences of the United States of America*, 79 (1982) 4381-4385.
- [226] Y. Ni, et al., Super-resolution imaging of a 2.5 kb non-repetitive DNA in situ in the nuclear genome using molecular beacon probes, *eLife*, 6 (2017).
- [227] K. Zhanghao, et al., Super-resolution dipole orientation mapping via polarization demodulation, *Light: Science & Applications*, 5 (2016) e16166.
- [228] G. Vicidomini, et al., Sharper low-power STED nanoscopy by time gating, *Nature methods*, 8 (2011) 571-573.
- [229] P.T. Ho, et al., Rapid increase of near atomic resolution virus capsid structures determined by cryo-electron microscopy, *Journal of structural biology*, 201 (2018) 1-4.
- [230] J. Dekker, et al., Capturing chromosome conformation, *Science (New York, N.Y.)*, 295 (2002) 1306-1311.
- [231] M. Simonis, et al., Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture–on-chip (4C), *Nature Genetics*, 38 (2006) 1348-1354.
- [232] J. Dostie, et al., Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements, *Genome research*, 16 (2006) 1299-1309.
- [233] J.Q. Ling, et al., CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1, *Science (New York, N.Y.)*, 312 (2006) 269-272.
- [234] V.K. Tiwari, et al., A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations, *Genome research*, 18 (2008) 1171-1179.
- [235] M.J. Fullwood, et al., An oestrogen-receptor-alpha-bound human chromatin interactome, *Nature*, 462 (2009) 58-64.
- [236] E. Lieberman-Aiden, et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome, *Science (New York, N.Y.)*, 326 (2009) 289-293.
- [237] J.R. Hughes, et al., Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment, *Nat Genet*, 46 (2014) 205-212.
- [238] D.G. Lupianez, et al., Breaking TADs: How Alterations of Chromatin Domains Result in Disease, *Trends in genetics : TIG*, 32 (2016) 225-237.
- [239] V.B. Kaiser, et al., When TADs go bad: chromatin structure and nuclear organisation in human disease, *F1000Research*, 6 (2017).
- [240] A. Mishra, et al., Three-dimensional genome architecture and emerging technologies: looping in disease, *Genome medicine*, 9 (2017) 87.
- [241] S. Remeseiro, et al., Cohesin in development and disease, *Development (Cambridge, England)*, 140 (2013) 3715-3718.

- [242] Z. Hu, et al., Enhancers and chromatin structures: regulatory hubs in gene expression and diseases, *Bioscience reports*, 37 (2017).
- [243] E. Ballestar, *Epigenetic contributions in autoimmune disease* (Springer Science & Business Media, 2011).
- [244] M.A. Barbara, et al., An Introduction to Epigenetics, *Neonatal Netw*, 36 (2017) 124-128.
- [245] J.D. Sweatt, The emerging field of neuroepigenetics, *Neuron*, 80 (2013) 624-632.
- [246] N.F. Cope, et al., Chromosome conformation capture, *Cold Spring Harbor protocols*, 2009 (2009) pdb prot5137.
- [247] J. Dostie, et al., Chromosome conformation capture carbon copy technology, *Current protocols in molecular biology*, Chapter 21 (2007) Unit 21 14.
- [248] R. Stadhouders, et al., Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions, *Nature protocols*, 8 (2013) 509-524.
- [249] A. Sanyal, et al., The long-range interaction landscape of gene promoters, *Nature*, 489 (2012) 109-113.
- [250] Illumina, *bcl2fastq Conversion Software v1.8.4* (2013).
- [251] P. Briggs, *genomics-bcftbx: A set of utility programs and scripts plus a Python library developed to support NGS and genomics-related bioinformatics within the Bioinformatics Core Facility (BCF) of the Faculty of Biology, Medicine and Health (FBMH) at the University of Manchester (UoM)*. <https://genomics-bcftbx.readthedocs.io/en/latest/>.
- [252] DZNE, *The German Center for Neurodegenerative Diseases*.
- [253] G. Jain, *GJSrMap: The smallRNA mapping pipeline*, (2018).
- [254] GWDG HPCC, *GWDG Scientific Compute Cluster, HIGH PERFORMANCE COMPUTING*.
- [255] Illumina, https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf, (2011).
- [256] B. Langmead, et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, *Genome Biol*, 10 (2009) R25.
- [257] A. Kozomara, et al., miRBase: integrating microRNA annotation and deep-sequencing data, *Nucleic acids research*, 39 (2011) D152-157.
- [258] S. Sai Lakshmi, et al., piRNABank: a web resource on classified and clustered Piwi-interacting RNAs, *Nucleic acids research*, 36 (2008) D173-177.
- [259] K. Thompson, et al., *grep is a command-line utility for searching plain-text data sets for lines that match a regular expression. g/re/p (globally search a regular expression and print)*, AT&T Bell Laboratories, (1974).
- [260] S. Andrews, *FastQC: a quality control tool for high throughput sequence data*, (2010).
- [261] M. Martin, *Cutadapt removes adapter sequences from high-throughput sequencing reads*, *EMBnet. journal*, 17 (2011) 10-12.

- [262] H. Li, et al., The Sequence Alignment/Map format and SAMtools, *Bioinformatics*, 25 (2009) 2078-2079.
- [263] D. Dougherty, et al., *sed & awk: UNIX Power Tools* (" O'Reilly Media, Inc.", 1997).
- [264] P. Ewels, et al., MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics*, 32 (2016) 3047-3048.
- [265] E. Maza, In Papyro Comparison of TMM (edgeR), RLE (DESeq2), and MRN Normalization Methods for a Simple Two-Conditions-Without-Replicates RNA-Seq Experimental Design, *Frontiers in genetics*, 7 (2016) 164.
- [266] J. Zypych-Walczak, et al., The Impact of Normalization Methods on RNA-Seq Data Analysis, *BioMed Research International*, 2015 (2015) 10.
- [267] B.M. Bolstad, et al., A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, *Bioinformatics*, 19 (2003) 185-193.
- [268] J.H. Bullard, et al., Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments, *BMC Bioinformatics*, 11 (2010) 94.
- [269] S. Anders, et al., Differential expression analysis for sequence count data, *Genome Biol*, 11 (2010) R106.
- [270] S. Ander, Variance Stabilizing Transformation (VST) in DESeq2, *Bioconductor* (2010).
- [271] W. Huber, et al., Parameter estimation for the calibration and variance stabilization of microarray data, *Statistical applications in genetics and molecular biology*, 2 (2003) Article3.
- [272] B.P. Durbin, et al., A variance-stabilizing transformation for gene-expression microarray data, *Bioinformatics*, 18 (2002) S105-S110.
- [273] R. Tibshirani, Variance Stabilization and the Bootstrap, *Biometrika*, 75 (1988) 433-444.
- [274] B.D. Ripley, The R project in statistical computing, *MSOR Connections. The newsletter of the LTSN Maths, Stats & OR Network*, 1 (2001) 23-25.
- [275] A. Yassouridis, et al., A new way of identifying biomarkers in biomedical basic-research studies, *PloS one*, 7 (2012) e35741.
- [276] J. Denk, et al., MicroRNA Profiling of CSF Reveals Potential Biomarkers to Detect Alzheimer's Disease, *PloS one*, 10 (2015) e0126423.
- [277] R. Kohavi, et al., Wrappers for feature subset selection, *Artificial Intelligence*, 97 (1997) 273-324.
- [278] D. Koller, et al., Toward optimal feature selection, *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*, (Morgan Kaufmann Publishers Inc., Bari, Italy, 1996), pp. 284-292.
- [279] R.M. Rifkin, et al., Notes on regularized least squares, (2007).
- [280] R.M. Neal, *Bayesian Learning for Neural Networks* (Springer-Verlag, 1996).
- [281] C.M. Bishop, *Neural Networks for Pattern Recognition* (Oxford University Press, Inc., 1995).
- [282] D.J.C. MacKay, Bayesian Interpolation, *Neural Computation*, 4 (1992) 415-447.

- [283] C.M. Bishop, et al., Bayesian Regression and Classification, *Advances in Learning Theory: Methods, Models and Applications*, (IOS Press, 2003), pp. 267-285.
- [284] F. Pedregosa, et al., Scikit-learn: Machine Learning in Python, *J. Mach. Learn. Res.*, 12 (2011) 2825-2830.
- [285] A. Abraham, et al., Machine learning for neuroimaging with scikit-learn, *Frontiers in neuroinformatics*, 8 (2014) 14.
- [286] C.J. Huberty, et al., Multivariate analysis of variance and covariance, *Handbook of applied multivariate statistics and mathematical modeling*, (Elsevier, 2000), pp. 183-208.
- [287] F.B. Bryant, et al., Principal-components analysis and exploratory and confirmatory factor analysis, *Reading and understanding multivariate statistics.*, (American Psychological Association, Washington, DC, US, 1995), pp. 99-136.
- [288] X. Robin, et al., pROC: an open-source package for R and S+ to analyze and compare ROC curves, *BMC Bioinformatics*, 12 (2011) 77.
- [289] J. Carpenter, et al., Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, *Statistics in Medicine*, 19 (2000) 1141-1164.
- [290] E.R. DeLong, et al., Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, *Biometrics*, 44 (1988) 837-845.
- [291] F. Lay, Workshop 10: Hi-C QCB Fellow (2017).
- [292] T. Sexton, et al., Three-dimensional folding and functional organization principles of the *Drosophila* genome, *Cell*, 148 (2012) 458-472.
- [293] M. Hu, et al., Bayesian inference of spatial organizations of chromosomes, *PLoS computational biology*, 9 (2013) e1002893.
- [294] N. Naumova, et al., Organization of the mitotic chromosome, *Science (New York, N.Y.)*, 342 (2013) 948-953.
- [295] R. Sinkhorn, et al., Concerning nonnegative matrices and doubly stochastic matrices, *Pacific J. Math.*, 21 (1967) 343-348.
- [296] M. Imakaev, et al., Iterative correction of Hi-C data reveals hallmarks of chromosome organization, *Nature methods*, 9 (2012) 999-1003.
- [297] A. Cournac, et al., Normalization of a chromosomal contact map, *BMC genomics*, 13 (2012) 436.
- [298] B.R. Lajoie, et al., The Hitchhiker's guide to Hi-C analysis: practical guidelines, *Methods (San Diego, Calif.)*, 72 (2015) 65-75.
- [299] M.R. Friedlander, et al., Discovering microRNAs from deep sequencing data using miRDeep, *Nature biotechnology*, 26 (2008) 407-415.
- [300] P.H. Cheng, et al., miR-196a ameliorates phenotypes of Huntington disease in cell, transgenic mouse, and induced pluripotent stem cell models, *American journal of human genetics*, 93 (2013) 306-312.
- [301] Y. Gui, et al., Altered microRNA profiles in cerebrospinal fluid exosome in Parkinson disease and Alzheimer disease, *Oncotarget*, 6 (2015) 37043-37053.

- [302] C.P. Mohammed, et al., miR-204 downregulates EphB2 in aging mouse hippocampal neurons, *Aging cell*, 15 (2016) 380-388.
- [303] R. Saba, et al., Dopamine-regulated microRNA MiR-181a controls GluA2 surface expression in hippocampal neurons, *Molecular and cellular biology*, 32 (2012) 619-632.
- [304] B. Stepniak, et al., Accumulated common variants in the broader fragile X gene family modulate autistic phenotypes, *EMBO molecular medicine*, 7 (2015) 1565-1579.
- [305] P. Krishnan, et al., Piwi-interacting RNAs and PIWI genes as novel prognostic markers for breast cancer, *Oncotarget*, 7 (2016) 37944-37956.
- [306] A. Zovoillis, et al., The expression level of small non-coding RNAs derived from the first exon of protein-coding genes is predictive of cancer status, *EMBO Rep*, 15 (2014) 402-410.
- [307] A. Tietje, et al., Cerebrospinal fluid extracellular vesicles undergo age dependent declines and contain known and novel non-coding RNAs, *PloS one*, 9 (2014) e113116.
- [308] N. Sheikh-Bahaei, et al., Current Role for Biomarkers in Clinical Diagnosis of Alzheimer Disease and Frontotemporal Dementia, *Current treatment options in neurology*, 19 (2017) 46.
- [309] N. Andreasen, et al., CSF markers for Alzheimer's disease: total tau, phospho-tau and Abeta42, *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry*, 4 (2003) 147-155.
- [310] N. Andreasen, et al., Cerebrospinal fluid levels of total-tau, phospho-tau and A beta 42 predicts development of Alzheimer's disease in patients with mild cognitive impairment, *Acta neurologica Scandinavica. Supplementum*, 179 (2003) 47-51.
- [311] G. De Meyer, et al., Diagnosis-independent Alzheimer disease biomarker signature in cognitively normal elderly people, *Archives of neurology*, 67 (2010) 949-956.
- [312] D. Galasko, et al., High cerebrospinal fluid tau and low amyloid beta42 levels in the clinical diagnosis of Alzheimer disease and relation to apolipoprotein E genotype, *Archives of neurology*, 55 (1998) 937-945.
- [313] C. Hock, et al., Increased CSF levels of nerve growth factor in patients with Alzheimer's disease, *Neurology*, 54 (2000) 2009-2011.
- [314] J. Marksteiner, et al., Cerebrospinal fluid biomarkers for diagnosis of Alzheimer's disease: beta-amyloid(1-42), tau, phospho-tau-181 and total protein, *Drugs of today (Barcelona, Spain : 1998)*, 43 (2007) 423-431.
- [315] P.D. Mehta, et al., Increased amyloid beta protein levels in children and adolescents with Down syndrome, *Journal of the neurological sciences*, 254 (2007) 22-27.
- [316] B. Dubois, et al., Alzheimer disease: from biomarkers to diagnosis, *Revue neurologique*, 169 (2013) 744-751.
- [317] P. Lau, et al., Alteration of the microRNA network during the progression of Alzheimer's disease, *EMBO molecular medicine*, 5 (2013) 1613-1634.
- [318] H. Dweep, et al., miRWalk database for miRNA-target interactions, *Methods in molecular biology (Clifton, N.J.)*, 1182 (2014) 289-305.
- [319] H. Dweep, et al., miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes, *Journal of biomedical informatics*, 44 (2011) 839-847.

- [320] Y. Hashimoto, et al., Multiple-to-multiple relationships between microRNAs and target genes in gastric cancer, *PLoS one*, 8 (2013) e62589.
- [321] S.T. Ferreira, et al., Inflammation, defective insulin signaling, and neuronal dysfunction in Alzheimer's disease, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 10 (2014) S76-83.
- [322] T. Liu, et al., Transcriptional signaling pathways inversely regulated in Alzheimer's disease and glioblastoma multiform, *Scientific reports*, 3 (2013) 3467.
- [323] S. Ogishima, et al., A map of Alzheimer's disease-signaling pathways: a hope for drug target discovery, *Clinical pharmacology and therapeutics*, 93 (2013) 399-401.
- [324] W.S. Griffin, et al., Interleukin-1 expression in different plaque types in Alzheimer's disease: significance in plaque evolution, *Journal of neuropathology and experimental neurology*, 54 (1995) 276-281.
- [325] R.E. Mrak, et al., Interleukin-1 and the immunogenetics of Alzheimer disease, *Journal of neuropathology and experimental neurology*, 59 (2000) 471-476.
- [326] R.E. Mrak, et al., Interleukin-1, neuroinflammation, and Alzheimer's disease, *Neurobiology of aging*, 22 (2001) 903-908.
- [327] S.S. Shafiq, et al., The role of interleukin-1 in neuroinflammation and Alzheimer disease: an evolving perspective, *Journal of neuroinflammation*, 5 (2008) 7.
- [328] S.S. Shafiq, et al., Sustained hippocampal IL-1 beta overexpression mediates chronic neuroinflammation and ameliorates Alzheimer plaque pathology, *The Journal of clinical investigation*, 117 (2007) 1595-1604.
- [329] J.G. Cui, et al., Differential regulation of interleukin-1 receptor-associated kinase-1 (IRAK-1) and IRAK-2 by microRNA-146a and NF-kappaB in stressed human astroglial cells and in Alzheimer disease, *The Journal of biological chemistry*, 285 (2010) 38951-38960.
- [330] L.M. Grimaldi, et al., Association of early-onset Alzheimer's disease with an interleukin-1 α gene polymorphism, *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 47 (2000) 361-365.
- [331] C.W. Cotman, et al., Exercise: a behavioral intervention to enhance brain health and plasticity, *Trends in neurosciences*, 25 (2002) 295-301.
- [332] L. Gasparini, et al., Potential roles of insulin and IGF-1 in Alzheimer's disease, *Trends in neurosciences*, 26 (2003) 404-406.
- [333] C. George, et al., The Alzheimer's disease transcriptome mimics the neuroprotective signature of IGF-1 receptor-deficient neurons, *Brain : a journal of neurology*, 140 (2017) 2012-2027.
- [334] M.L. Giuffrida, et al., Beta-amyloid monomer and insulin/IGF-1 signaling in Alzheimer's disease, *Molecular neurobiology*, 46 (2012) 605-613.
- [335] E. Steen, et al., Impaired insulin and insulin-like growth factor expression and signaling mechanisms in Alzheimer's disease--is this type 3 diabetes?, *Journal of Alzheimer's disease : JAD*, 7 (2005) 63-80.
- [336] A.J. Westwood, et al., Insulin-like growth factor-1 and risk of Alzheimer dementia and brain atrophy, *Neurology*, 82 (2014) 1613-1619.

- [337] C. Wang, et al., Targeting the mTOR signaling network for Alzheimer's disease therapy, *Molecular neurobiology*, 49 (2014) 120-135.
- [338] S. Wang, et al., Dysregulated mTOR-dependent signaling in neurodegeneration or carcinogenesis: implication for Alzheimer's disease and brain tumors, *Journal of Alzheimer's disease : JAD*, 37 (2013) 495-505.
- [339] B.S. Ashok, et al., Hypoxia-inducible factors as neuroprotective agent in Alzheimer's disease, *Clinical and experimental pharmacology & physiology*, 44 (2017) 327-334.
- [340] M. Guglielmotto, et al., The up-regulation of BACE1 mediated by hypoxia and ischemic injury: role of oxidative stress and HIF1alpha, *J Neurochem*, 108 (2009) 1045-1056.
- [341] O.O. Ogunshola, et al., Contribution of hypoxia to Alzheimer's disease: is HIF-1alpha a mediator of neurodegeneration?, *Cellular and molecular life sciences : CMLS*, 66 (2009) 3555-3563.
- [342] X. Zhang, et al., Hypoxia-inducible factor 1alpha (HIF-1alpha)-mediated hypoxia increases BACE1 expression and beta-amyloid generation, *The Journal of biological chemistry*, 282 (2007) 10873-10880.
- [343] R. Ferrari, et al., Androgen receptor gene and sex-specific Alzheimer's disease, *Neurobiology of aging*, 34 (2013) 2077 e2019-2020.
- [344] J. Raber, Potential Role of Androgens and Androgen Receptors in Alzheimer's Disease, in: R.W. Richter, et al. (Eds.) *Alzheimer's Disease: A Physician's Guide to Practical Management*, (Humana Press, Totowa, NJ, 2004), pp. 259-271.
- [345] E.S. Kelleher, et al., Analysis of piRNA-mediated silencing of active TEs in *Drosophila melanogaster* suggests limits on the evolution of host genome defense, *Molecular biology and evolution*, 30 (2013) 1816-1829.
- [346] S. Luo, et al., Silencing of Transposable Elements by piRNAs in *Drosophila*: An Evolutionary Perspective, *Genomics Proteomics Bioinformatics*, 15 (2017) 164-176.
- [347] S.R. Mani, et al., Untangling the web: the diverse functions of the PIWI/piRNA pathway, *Molecular reproduction and development*, 80 (2013) 632-664.
- [348] C. Ernst, et al., The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity, *Nature communications*, 8 (2017) 1411.
- [349] J. Lu, et al., Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*, *Genome research*, 20 (2010) 212-227.
- [350] N. Bot, et al., Processing of the synaptic cell adhesion molecule neurexin-3beta by Alzheimer disease alpha- and gamma-secretases, *The Journal of biological chemistry*, 286 (2011) 2762-2773.
- [351] C. Lowther, et al., Molecular characterization of NRXN1 deletions from 19,263 clinical microarray cases identifies exons important for neurodevelopmental disease expression, *Genetics in medicine : official journal of the American College of Medical Genetics*, 19 (2017) 53-61.
- [352] I.A. Sindi, et al., Role for the neurexin-neurologin complex in Alzheimer's disease, *Neurobiology of aging*, 35 (2014) 746-756.

- [353] P. Mao, Potential Antidepressant Role of Neurotransmitter CART: Implications for Mental Disorders, *Depression research and treatment*, 2011 (2011) 762139.
- [354] W.J. Hendriks, et al., Protein tyrosine phosphatase variants in human hereditary disorders and disease susceptibilities, *Biochimica et biophysica acta*, 1832 (2013) 1673-1696.
- [355] E.L. Schaeffer, et al., Phospholipase A2 activation as a therapeutic approach for cognitive enhancement in early-stage Alzheimer disease, *Psychopharmacology*, 202 (2009) 37-51.
- [356] L. Zhang, et al., Potential hippocampal genes and pathways involved in Alzheimer's disease: a bioinformatic analysis, *Genetics and molecular research : GMR*, 14 (2015) 7218-7232.
- [357] D. Hadley, et al., The impact of the metabotropic glutamate receptor and other gene family interaction networks on autism, *Nature communications*, 5 (2014) 4074.
- [358] A. Martinez-Mir, et al., Genetic study of neurexin and neuroligin genes in Alzheimer's disease, *Journal of Alzheimer's disease : JAD*, 35 (2013) 403-412.
- [359] Y. Birger, et al., HMGN4, a newly discovered nucleosome-binding protein encoded by an intronless gene, *DNA and cell biology*, 20 (2001) 257-264.
- [360] H. Shirakawa, et al., NBP-45, a novel nucleosomal binding protein with a tissue-specific and developmentally regulated expression, *The Journal of biological chemistry*, 275 (2000) 6368-6374.
- [361] R.R. Nayak, et al., Stress-induced changes in gene interactions in human cells, *Nucleic acids research*, 42 (2014) 1757-1771.
- [362] V. Stepanov, et al., M15 - Overlapping Pattern of Genetic Associations of Alzheimer's Disease And Schizophrenia Mediated By Cognitive Endophenotypes Genes, *European Neuropsychopharmacology*, 27 (2017) S376.
- [363] R.J. Strawbridge, et al., Genome-wide analysis of self-reported risk-taking behaviour and cross-disorder genetic correlations in the UK Biobank cohort, *Translational psychiatry*, 8 (2018) 39.
- [364] M.R. Paul, et al., Condensin Depletion Causes Genome Decompaction Without Altering the Level of Global Gene Expression in *Saccharomyces cerevisiae*, *Genetics*, 210 (2018) 331-344.
- [365] B. Bonev, et al., Multiscale 3D Genome Rewiring during Mouse Neural Development, *Cell*, 171 (2017) 557-572 e524.
- [366] S.S. Rao, et al., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping, *Cell*, 159 (2014) 1665-1680.
- [367] E. Crane, et al., Condensin-driven remodelling of X chromosome topology during dosage compensation, *Nature*, 523 (2015) 240-244.
- [368] J. Fraser, et al., Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation, *Molecular systems biology*, 11 (2015) 852.
- [369] J.E. Phillips-Cremins, et al., Architectural protein subclasses shape 3D organization of genomes during lineage commitment, *Cell*, 153 (2013) 1281-1295.

- [370] B.D. Pope, et al., Topologically associating domains are stable units of replication-timing regulation, *Nature*, 515 (2014) 402-405.
- [371] G. Robertson, et al., Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing, *Nature methods*, 4 (2007) 651-657.
- [372] M.D. Young, et al., ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity, *Nucleic acids research*, 39 (2011) 7415-7427.
- [373] N. Croce, et al., NPY modulates miR-30a-5p and BDNF in opposite direction in an in vitro model of Alzheimer disease: a possible role in neuroprotection?, *Molecular and cellular biochemistry*, 376 (2013) 189-195.
- [374] S.S. Sorensen, et al., miRNA expression profiles in cerebrospinal fluid and blood of patients with Alzheimer's disease and other types of dementia - an exploratory study, *Transl Neurodegener*, 5 (2016) 6.
- [375] S.F. Zhang, et al., miR-181a involves in the hippocampus-dependent memory formation via targeting PRKAA1, *Scientific reports*, 7 (2017) 8480.
- [376] J. Ozanne, et al., The clinically approved drugs dasatinib and bosutinib induce anti-inflammatory macrophages by inhibiting the salt-inducible kinases, *The Biochemical journal*, 465 (2015) 271-279.
- [377] S. Popov, et al., Lack of salt-inducible kinase 2 (SIK2) prevents the development of cardiac hypertrophy in response to chronic high-salt intake, *PloS one*, 9 (2014) e95771.
- [378] F.C. Yang, et al., Reversible acetylation regulates salt-inducible kinase (SIK2) and its function in autophagy, *The Journal of biological chemistry*, 288 (2013) 6227-6237.
- [379] S.L. Cole, et al., The Alzheimer's disease beta-secretase enzyme, BACE1, *Molecular neurodegeneration*, 2 (2007) 22.
- [380] Y. Luo, et al., Mice deficient in BACE1, the Alzheimer's beta-secretase, have normal phenotype and abolished beta-amyloid generation, *Nature neuroscience*, 4 (2001) 231-232.
- [381] R. Vassar, BACE1 inhibitor drugs in clinical trials for Alzheimer's disease, *Alzheimer's research & therapy*, 6 (2014) 89.
- [382] R. Vassar, et al., The beta-secretase enzyme BACE in health and Alzheimer's disease: regulation, cell biology, function, and therapeutic potential, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29 (2009) 12787-12794.
- [383] T. Erkinjuntti, et al., The effect of different diagnostic criteria on the prevalence of dementia, *The New England journal of medicine*, 337 (1997) 1667-1674.
- [384] E. Howe, Improving the quality of life in patients with Alzheimer's disease, *Psychiatry (Edgmont (Pa. : Township))*, 5 (2008) 51-56.
- [385] K. Inouye, et al., Quality of life of elderly with Alzheimer's disease: a comparative study between the patients and the caregivers report, *Revista latino-americana de enfermagem*, 18 (2010) 26-32.
- [386] K. Kahle-Wroblewski, et al., Assessing quality of life in Alzheimer's disease: Implications for clinical trials, *Alzheimer's & dementia (Amsterdam, Netherlands)*, 6 (2017) 82-90.
- [387] M.W. Riepe, et al., Quality of life as an outcome in Alzheimer's disease and other dementias--obstacles and goals, *BMC neurology*, 9 (2009) 47.

- [388] M.C. Biagioni, et al., Using biomarkers to improve detection of Alzheimer's disease, *Neurodegenerative disease management*, 1 (2011) 127-139.
- [389] C. Humpel, Identifying and validating biomarkers for Alzheimer's disease, *Trends in biotechnology*, 29 (2011) 26-32.
- [390] H.M. Snyder, et al., Developing novel blood-based biomarkers for Alzheimer's disease, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 10 (2014) 109-114.
- [391] J.L. Cummings, Biomarkers in Alzheimer's disease drug development, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 7 (2011) e13-44.
- [392] A.D. Henriques, et al., Fluid and imaging biomarkers for Alzheimer's disease: Where we stand and where to head to, *Experimental gerontology*, 107 (2018) 169-177.
- [393] S.K. Herukka, et al., Recommendations for cerebrospinal fluid Alzheimer's disease biomarkers in the diagnostic evaluation of mild cognitive impairment, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 13 (2017) 285-295.
- [394] I.O. Korolev, et al., Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification, *PloS one*, 11 (2016) e0138866.
- [395] G. Ferrero, et al., Small non-coding RNA profiling in human biofluids and surrogate tissues from healthy individuals: description of the diverse and most represented species, *Oncotarget*, 9 (2018) 3097-3111.
- [396] P. Krishnan, et al., The Challenges and Opportunities in the Clinical Application of Noncoding RNAs: The Road Map for miRNAs and piRNAs in Cancer Diagnostics and Prognostics, *International Journal of Genomics*, 2018 (2018) 18.
- [397] J.P. Lopez, et al., Biomarker discovery: quantification of microRNAs and other small non-coding RNAs using next generation sequencing, *BMC medical genomics*, 8 (2015) 35.
- [398] K.M. Kanninen, et al., Exosomes as new diagnostic tools in CNS diseases, *Biochimica et biophysica acta*, 1862 (2016) 403-410.
- [399] S. Mondragon-Rodriguez, et al., Amyloid Beta and tau proteins as therapeutic targets for Alzheimer's disease treatment: rethinking the current strategy, *International journal of Alzheimer's disease*, 2012 (2012) 630182.
- [400] C.J. Roberts, Therapeutic protein aggregation: mechanisms, design, and control, *Trends in biotechnology*, 32 (2014) 372-380.
- [401] P. Sweeney, et al., Protein misfolding in neurodegenerative diseases: implications and strategies, *Transl Neurodegener*, 6 (2017) 6.
- [402] S.L. Risacher, et al., Neuroimaging and other biomarkers for Alzheimer's disease: the changing landscape of early detection, *Annual review of clinical psychology*, 9 (2013) 621-648.
- [403] K. Herholz, et al., Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET, *NeuroImage*, 17 (2002) 302-316.
- [404] C.R. Jack, Jr., et al., Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade, *The Lancet. Neurology*, 9 (2010) 119-128.

- [405] L.M. Shaw, et al., Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects, *Annals of neurology*, 65 (2009) 403-413.
- [406] T.K. Khan, et al., Alzheimer's Disease Cerebrospinal Fluid and Neuroimaging Biomarkers: Diagnostic Accuracy and Relationship to Drug Efficacy, *Journal of Alzheimer's disease : JAD*, 46 (2015) 817-836.
- [407] H. Dong, et al., Serum MicroRNA Profiles Serve as Novel Biomarkers for the Diagnosis of Alzheimer's Disease, *Disease markers*, 2015 (2015) 625659.
- [408] N. Mattsson, et al., CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment, *Jama*, 302 (2009) 385-393.
- [409] M.L. Alvarez, et al., Comparison of protein, microRNA, and mRNA yields using different methods of urinary exosome isolation for the discovery of kidney disease biomarkers, *Kidney international*, 82 (2012) 1024-1032.
- [410] J.H. Bahn, et al., The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva, *Clinical chemistry*, 61 (2015) 221-230.
- [411] A. Drusco, et al., A differentially expressed set of microRNAs in cerebro-spinal fluid (CSF) can diagnose CNS malignancies, *Oncotarget*, 6 (2015) 20829-20839.
- [412] D. Koppers-Lalic, et al., Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes, *Cell reports*, 8 (2014) 1649-1658.
- [413] Q. Zhou, et al., Immune-related microRNAs are abundant in breast milk exosomes, *International journal of biological sciences*, 8 (2012) 118-123.
- [414] M. Eldh, et al., Importance of RNA isolation methods for analysis of exosomal RNA: evaluation of different methods, *Molecular immunology*, 50 (2012) 278-286.
- [415] D. Enderle, et al., Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method, *PloS one*, 10 (2015) e0136133.
- [416] K.W. Witwer, et al., Standardization of sample collection, isolation and analysis methods in extracellular vesicle research, *Journal of extracellular vesicles*, 2 (2013).
- [417] D. Galasko, et al., Biomarkers for Alzheimer's disease in plasma, serum and blood - conceptual and practical problems, *Alzheimer's research & therapy*, 5 (2013) 10.
- [418] M.C. Irizarry, Biomarkers of Alzheimer disease in plasma, *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics*, 1 (2004) 226-234.
- [419] H. Sohma, et al., Plasma Biomarkers in Alzheimer's Disease, *Update on Dementia*, (IntechOpen, 2016).
- [420] S.E. O'Bryant, et al., A blood-based screening tool for Alzheimer's disease that spans serum and plasma: findings from TARC and ADNI, *PloS one*, 6 (2011) e28092.
- [421] M. Zabel, et al., Assessing candidate serum biomarkers for Alzheimer's disease: a longitudinal study, *Journal of Alzheimer's disease : JAD*, 30 (2012) 311-321.
- [422] M. Sjogren, et al., Tau and Abeta42 in cerebrospinal fluid from healthy adults 21-93 years of age: establishment of reference values, *Clinical chemistry*, 47 (2001) 1776-1781.
- [423] H. Zetterberg, et al., Plasma tau levels in Alzheimer's disease, *Alzheimer's research & therapy*, 5 (2013) 9.

- [424] K. Blennow, et al., Protein analysis in cerebrospinal fluid. III. Relation to blood-cerebrospinal fluid barrier function for formulas for quantitative determination of intrathecal IgG production, *Eur Neurol*, 33 (1993) 134-142.
- [425] L.K. Page, Cerebrospinal fluid and extracellular fluid: their relationship to pressure and duration of canine hydrocephalus, *Child's Nervous System*, 1 (1985) 12-17.
- [426] L. Alvarez-Erviti, et al., Delivery of siRNA to the mouse brain by systemic injection of targeted exosomes, *Nature biotechnology*, 29 (2011) 341-345.
- [427] S. Griffiths-Jones, et al., miRBase: microRNA sequences, targets and gene nomenclature, *Nucleic acids research*, 34 (2006) D140-144.
- [428] H.I. Suzuki, et al., Emerging complexity of microRNA generation cascades, *Journal of biochemistry*, 149 (2011) 15-25.
- [429] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, *Cell*, 116 (2004) 281-297.
- [430] D.S. Schwarz, et al., Asymmetry in the assembly of the RNAi enzyme complex, *Cell*, 115 (2003) 199-208.
- [431] G. Desdin-Mico, et al., Role of exosomes in the protection of cellular homeostasis, *Cell adhesion & migration*, 11 (2017) 127-134.
- [432] M. Ragusa, et al., Asymmetric RNA Distribution among Cells and Their Secreted Exosomes: Biomedical Meaning and Considerations on Diagnostic Applications, *Frontiers in molecular biosciences*, 4 (2017) 66.
- [433] N.W. Breakfield, et al., High-resolution experimental and computational profiling of tissue-specific known and novel miRNAs in Arabidopsis, *Genome research*, 22 (2012) 163-176.
- [434] V. Capece, et al., Oasis: online analysis of small RNA deep sequencing data, *Bioinformatics*, 31 (2015) 2205-2207.
- [435] S. Cho, et al., MiRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting, *Nucleic acids research*, 41 (2013) D252-257.
- [436] M. Fasold, et al., DARIO: a ncRNA detection and analysis tool for next-generation sequencing experiments, *Nucleic acids research*, 39 (2011) W112-117.
- [437] J. Kim, et al., MAGI: a Node.js web service for fast microRNA-Seq analysis in a GPU infrastructure, *Bioinformatics*, 30 (2014) 2826-2827.
- [438] G.Z. Luo, et al., ISRNA: an integrative online toolkit for short reads from high-throughput sequencing data, *Bioinformatics*, 30 (2014) 434-436.
- [439] E.J. Song, et al., Restricted CD4+ T cell receptor repertoire impairs cognitive function via alteration of Th2 cytokine levels, *Neurogenesis (Austin, Tex.)*, 4 (2017) e1256856.
- [440] G. Chen, et al., Hsa-miR-181a-5p expression and effects on cell proliferation in gastric cancer, *Asian Pacific journal of cancer prevention : APJCP*, 14 (2013) 3871-3875.
- [441] W. Gao, et al., Deregulated expression of miR-21, miR-143 and miR-181a in non small cell lung cancer is related to clinicopathologic characteristics or patient prognosis, *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*, 64 (2010) 399-408.

- [442] Y. Li, et al., miR-181a-5p Inhibits Cancer Cell Migration and Angiogenesis via Downregulation of Matrix Metalloproteinase-14, *Cancer research*, 75 (2015) 2674-2685.
- [443] J.M. Moon, et al., Inhibition of microRNA-181 reduces forebrain ischemia-induced neuronal loss, *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 33 (2013) 1976-1982.
- [444] G. Giurato, et al., iMir: an integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-Seq, *BMC Bioinformatics*, 14 (2013) 362.
- [445] M. Hackenberg, et al., miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments, *Nucleic acids research*, 39 (2011) W132-138.
- [446] S.D. Mackowiak, Identification of novel and known miRNAs in deep-sequencing data with miRDeep2, *Current protocols in bioinformatics*, Chapter 12 (2011) Unit 12 10.
- [447] B.C. Richardson, et al., Epigenetics in 2013. DNA methylation and miRNA: key roles in systemic autoimmunity, *Nature reviews. Rheumatology*, 10 (2014) 72-74.
- [448] M.D. Robinson, et al., edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics*, 26 (2010) 139-140.
- [449] R. Waller, et al., Small RNA Sequencing of Sporadic Amyotrophic Lateral Sclerosis Cerebrospinal Fluid Reveals Differentially Expressed miRNAs Related to Neural and Glial Activity, *Front Neurosci*, 11 (2017) 731.
- [450] J. Chen, et al., Berberine upregulates miR-22-3p to suppress hepatocellular carcinoma cell proliferation by targeting Sp1, *American journal of translational research*, 8 (2016) 4932-4941.
- [451] K.T. Lv, et al., MiR-22-3p Regulates Cell Proliferation and Inhibits Cell Apoptosis through Targeting the eIF4EBP3 Gene in Human Cervical Squamous Carcinoma Cells, *International journal of medical sciences*, 15 (2018) 142-152.
- [452] R. Schneider, et al., Downregulation of exosomal miR-204-5p and miR-632 as a biomarker for FTD: a GENFI study, *Journal of neurology, neurosurgery, and psychiatry*, 89 (2018) 851-858.
- [453] W. Bao, et al., A TrkB-STAT3-miR-204-5p regulatory circuitry controls proliferation and invasion of endometrial carcinoma cells, *Molecular cancer*, 12 (2013) 155.
- [454] Z. Rasheed, et al., MicroRNA-26a-5p regulates the expression of inducible nitric oxide synthase via activation of NF-kappaB pathway in human osteoarthritis chondrocytes, *Archives of biochemistry and biophysics*, 594 (2016) 61-67.
- [455] J. Gao, et al., The role of miR-26 in tumors and normal tissues (Review), *Oncology letters*, 2 (2011) 1019-1023.
- [456] N. Potenza, et al., Human miR-26a-5p regulates the glutamate transporter SLC1A1 (EAAT3) expression. Relevance in multiple sclerosis, *Biochimica et biophysica acta. Molecular basis of disease*, 1864 (2018) 317-323.
- [457] A.G. Hoss, et al., miR-10b-5p expression in Huntington's disease brain relates to age of onset and the extent of striatal involvement, *BMC medical genomics*, 8 (2015) 10.

- [458] J. Zhang, et al., MicroRNA-10b expression in breast cancer and its clinical association, *PLoS one*, 13 (2018) e0192509.
- [459] W. Qiu, et al., Transcriptome-wide piRNA profiling in human brains of Alzheimer's disease, *Neurobiology of aging*, 57 (2017) 170-177.
- [460] C.B. Assumpcao, et al., The role of piRNA and its potential clinical implications in cancer, *Epigenomics*, 7 (2015) 975-984.
- [461] K.W. Ng, et al., Piwi-interacting RNAs in cancer: emerging functions and clinical utility, *Molecular cancer*, 15 (2016) 5.
- [462] W. Weng, et al., Novel evidence for a PIWI-interacting RNA (piRNA) as an oncogenic mediator of disease progression, and a potential prognostic biomarker in colorectal cancer, *Molecular cancer*, 17 (2018) 16.
- [463] S. Vella, et al., PIWI-interacting RNA (piRNA) signatures in human cardiac progenitor cells, *The international journal of biochemistry & cell biology*, 76 (2016) 1-11.
- [464] S. Sivagurunathan, et al., &em>In silico&em> analysis of piRNAs in retina reveals potential targets in intracellular transport and retinal degeneration, *bioRxiv*, (2018) 305144.
- [465] G.S. Bloom, Amyloid-beta and tau: the trigger and bullet in Alzheimer disease pathogenesis, *JAMA neurology*, 71 (2014) 505-508.
- [466] M. Rapoport, et al., Tau is essential to beta -amyloid-induced neurotoxicity, *Proceedings of the National Academy of Sciences of the United States of America*, 99 (2002) 6364-6369.
- [467] J. Busciglio, et al., beta-amyloid fibrils induce tau phosphorylation and loss of microtubule binding, *Neuron*, 14 (1995) 879-888.
- [468] S.M. Greenberg, et al., Secreted beta-APP stimulates MAP kinase and phosphorylation of tau in neurons, *Neurobiology of aging*, 16 (1995) 403-407; discussion 407-408.
- [469] M. Ingelsson, et al., Early Abeta accumulation and progressive synaptic loss, gliosis, and tangle formation in AD brain, *Neurology*, 62 (2004) 925-931.
- [470] T. Tokutake, et al., Hyperphosphorylation of Tau induced by naturally secreted amyloid-beta at nanomolar concentrations is modulated by insulin-dependent Akt-GSK3beta signaling pathway, *The Journal of biological chemistry*, 287 (2012) 35222-35233.
- [471] A. Alvarez, et al., Inhibition of tau phosphorylating protein kinase cdk5 prevents beta-amyloid-induced neuronal death, *FEBS letters*, 459 (1999) 421-426.
- [472] V. Shukla, et al., Deregulated Cdk5 activity is involved in inducing Alzheimer's disease, *Archives of medical research*, 43 (2012) 655-662.
- [473] J.P. Brion, et al., Neurofibrillary tangles and tau phosphorylation, *Biochemical Society symposium*, (2001) 81-88.
- [474] K. Iqbal, et al., Mechanisms of neurofibrillary degeneration and the formation of neurofibrillary tangles, *Journal of neural transmission. Supplementum*, 53 (1998) 169-180.
- [475] M. Goedert, Tau protein and the neurofibrillary pathology of Alzheimer's disease, *Annals of the New York Academy of Sciences*, 777 (1996) 121-131.

- [476] T.L. Spires-Jones, et al., The intersection of amyloid beta and tau at synapses in Alzheimer's disease, *Neuron*, 82 (2014) 756-771.
- [477] P.V. Arriagada, et al., Distribution of Alzheimer-type pathologic changes in nondemented elderly individuals matches the pattern in Alzheimer's disease, *Neurology*, 42 (1992) 1681-1688.
- [478] P. Giannakopoulos, et al., Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease, *Neurology*, 60 (2003) 1495-1500.
- [479] G.M. McKhann, et al., The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 7 (2011) 263-269.
- [480] J.Q. Trojanowski, et al., Update on the biomarker core of the Alzheimer's Disease Neuroimaging Initiative subjects, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 6 (2010) 230-238.
- [481] A.M. Fagan, et al., Cerebrospinal fluid tau and ptau(181) increase with cortical amyloid deposition in cognitively normal individuals: implications for future clinical trials of Alzheimer's disease, *EMBO molecular medicine*, 1 (2009) 371-380.
- [482] A.M. Fagan, et al., Cerebrospinal fluid tau/beta-amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults, *Archives of neurology*, 64 (2007) 343-349.
- [483] B.J. Cummings, et al., Image analysis of beta-amyloid load in Alzheimer's disease and relation to dementia severity, *Lancet (London, England)*, 346 (1995) 1524-1528.
- [484] J. Naslund, et al., Correlation between elevated levels of amyloid beta-peptide in the brain and cognitive decline, *Jama*, 283 (2000) 1571-1577.
- [485] C.R. Jack, Jr., et al., Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 7 (2011) 257-262.
- [486] E.M. Reiman, et al., Brain imaging and fluid biomarker analysis in young adults at genetic risk for autosomal dominant Alzheimer's disease in the presenilin 1 E280A kindred: a case-control study, *The Lancet. Neurology*, 11 (2012) 1048-1056.
- [487] K. Blennow, et al., The past and the future of Alzheimer's disease CSF biomarkers-a journey toward validated biochemical tests covering the whole spectrum of molecular events, *Front Neurosci*, 9 (2015) 345.
- [488] J.L. Molinuevo, et al., The clinical use of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: a consensus paper from the Alzheimer's Biomarkers Standardization Initiative, *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 10 (2014) 808-817.
- [489] H. Struyfs, et al., Validation of the AD-CSF-index in autopsy-confirmed Alzheimer's disease patients and healthy controls, *Journal of Alzheimer's disease : JAD*, 41 (2014) 903-909.

- [490] M.E. Murray, et al., Clinicopathologic and 11C-Pittsburgh compound B implications of Thal amyloid phase across the Alzheimer's disease spectrum, *Brain : a journal of neurology*, 138 (2015) 1370-1381.
- [491] S. Villeneuve, et al., Existing Pittsburgh Compound-B positron emission tomography thresholds are too high: statistical and pathological evaluation, *Brain : a journal of neurology*, 138 (2015) 2020-2033.
- [492] D.F. Wong, et al., An in vivo evaluation of cerebral cortical amyloid with [18F]flutemetamol using positron emission tomography compared with parietal biopsy samples in living normal pressure hydrocephalus patients, *Molecular imaging and biology : MIB : the official publication of the Academy of Molecular Imaging*, 15 (2013) 230-237.
- [493] O. Hansson, et al., Evaluation of plasma A β (40) and A β (42) as predictors of conversion to Alzheimer's disease in patients with mild cognitive impairment, *Neurobiology of aging*, 31 (2010) 357-367.
- [494] P. Lewczuk, et al., Amyloid-beta 42/40 cerebrospinal fluid concentration ratio in the diagnostics of Alzheimer's disease: validation of two novel assays, *Journal of Alzheimer's disease : JAD*, 43 (2015) 183-191.
- [495] S. Slaets, et al., Cerebrospinal fluid A β 1-40 improves differential dementia diagnosis in patients with intermediate P-tau181P levels, *Journal of Alzheimer's disease : JAD*, 36 (2013) 759-767.
- [496] E. Beaufils, et al., Confirmation of the amyloidogenic process in posterior cortical atrophy: value of the A β 42/A β 40 ratio, *Journal of Alzheimer's disease : JAD*, 33 (2013) 775-780.
- [497] J. Wiltfang, et al., Amyloid β peptide ratio 42/40 but not A β 42 correlates with phospho-Tau in patients with low- and high-CSF A β 40 load, *Journal of Neurochemistry*, 101 (2007) 1053-1059.
- [498] R. Ghidoni, et al., Cerebrospinal Fluid Biomarkers for Alzheimer's Disease: The Present and the Future, *Neurodegenerative Diseases*, 8 (2011) 413-420.
- [499] J.F. Quinn, Biomarkers for Alzheimer's disease: showing the way or leading us astray?, *Journal of Alzheimer's disease : JAD*, 33 Suppl 1 (2013) S371-376.
- [500] J.F. Quinn, et al., Extracellular RNAs: development as biomarkers of human disease, *Journal of extracellular vesicles*, 4 (2015) 27495.
- [501] K. Burgos, et al., Profiles of extracellular miRNA in cerebrospinal fluid and serum from patients with Alzheimer's and Parkinson's diseases correlate with disease status and features of pathology, *PloS one*, 9 (2014) e94839.
- [502] J.P. Cogswell, et al., Identification of miRNA changes in Alzheimer's disease brain and CSF yields putative biomarkers and insights into disease pathways, *Journal of Alzheimer's disease : JAD*, 14 (2008) 27-41.
- [503] V. Dorval, et al., Circulating microRNAs in Alzheimer's disease: the search for novel biomarkers, *Frontiers in molecular neuroscience*, 6 (2013) 24.
- [504] H. Geekiyanage, et al., Blood serum miRNA: non-invasive biomarkers for Alzheimer's disease, *Experimental neurology*, 235 (2012) 491-496.

- [505] T. Kiko, et al., MicroRNAs in plasma and cerebrospinal fluid as potential markers for Alzheimer's disease, *Journal of Alzheimer's disease : JAD*, 39 (2014) 253-259.
- [506] P. Kumar, et al., Circulating miRNA biomarkers for Alzheimer's disease, *PLoS one*, 8 (2013) e69807.
- [507] M. Muller, et al., MicroRNAs in Alzheimer's disease: differential expression in hippocampus and cell-free cerebrospinal fluid, *Neurobiology of aging*, 35 (2014) 152-158.
- [508] S.M. Lehmann, et al., An unconventional role for miRNA: let-7 activates Toll-like receptor 7 and causes neurodegeneration, *Nature neuroscience*, 15 (2012) 827-835.
- [509] C. Sala Frigerio, et al., Reduced expression of hsa-miR-27a-3p in CSF of patients with Alzheimer disease, *Neurology*, 81 (2013) 2103-2106.
- [510] H. Geekiyanage, et al., MicroRNA-137/181c regulates serine palmitoyltransferase and in turn amyloid beta, novel targets in sporadic Alzheimer's disease, *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31 (2011) 14820-14830.
- [511] S.S. Hebert, et al., Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression, *Proceedings of the National Academy of Sciences of the United States of America*, 105 (2008) 6415-6420.
- [512] S.T. Grivna, et al., A novel class of small RNAs in mouse spermatogenic cells, *Genes & development*, 20 (2006) 1709-1714.
- [513] B. Czech, et al., One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing, *Trends in biochemical sciences*, 41 (2016) 324-337.
- [514] G. Stefani, et al., Small non-coding RNAs in animal development, *Nature reviews. Molecular cell biology*, 9 (2008) 219-230.
- [515] A. Akkouche, et al., Maternally deposited germline piRNAs silence the transposon in somatic cells, *EMBO Rep*, 14 (2013) 458-464.
- [516] R.J. Ross, et al., PIWI proteins and PIWI-interacting RNAs in the soma, *Nature*, 505 (2014) 353-359.
- [517] A.A. Aravin, et al., The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race, *Science (New York, N.Y.)*, 318 (2007) 761-764.
- [518] A.A. Aravin, et al., Developmentally regulated piRNA clusters implicate MILI in transposon control, *Science (New York, N.Y.)*, 316 (2007) 744-747.
- [519] A. Dharap, et al., Altered expression of PIWI RNA in the rat brain after transient focal ischemia, *Stroke*, 42 (2011) 1105-1109.
- [520] Y. Hong, et al., Systematic characterization of seminal plasma piRNAs as molecular biomarkers for male infertility, *Scientific reports*, 6 (2016) 24229.
- [521] K.S. Rajan, et al., miRNA and piRNA mediated Akt pathway in heart: antisense expands to survive, *The international journal of biochemistry & cell biology*, 55 (2014) 153-156.
- [522] L. Zuo, et al., piRNAs and Their Functions in the Brain, *International journal of human genetics*, 16 (2016) 53-60.

- [523] L. Frolich, et al., Incremental value of biomarker combinations to predict progression of mild cognitive impairment to Alzheimer's dementia, *Alzheimer's research & therapy*, 9 (2017) 84.
- [524] J.L. Hsu, et al., Plasma biomarkers are associated with agitation and regional brain atrophy in Alzheimer's disease, *Scientific reports*, 7 (2017) 5035.
- [525] M. Kayano, et al., Plasma microRNA biomarker detection for mild cognitive impairment using differential correlation analysis, *Biomarker research*, 4 (2016) 22.
- [526] P. Leidinger, et al., A blood based 12-miRNA signature of Alzheimer disease patients, *Genome Biol*, 14 (2013) R78.
- [527] M. Thambisetty, et al., Blood-based biomarkers of Alzheimer's disease: challenging but feasible, *Biomarkers in medicine*, 4 (2010) 65-79.
- [528] R. Guo, et al., A 9-microRNA Signature in Serum Serves as a Noninvasive Biomarker in Early Diagnosis of Alzheimer's Disease, *Journal of Alzheimer's disease : JAD*, 60 (2017) 1365-1377.
- [529] N. Hara, et al., Serum microRNA miR-501-3p as a potential biomarker related to the progression of Alzheimer's disease, *Acta neuropathologica communications*, 5 (2017) 10.
- [530] L.H. Jia, et al., Downregulated serum miR-223 serves as biomarker in Alzheimer's disease, *Cell biochemistry and function*, 34 (2016) 233-237.
- [531] S. Nagaraj, et al., Profile of 6 microRNA in blood plasma distinguish early stage Alzheimer's disease patients from non-demented subjects, *Oncotarget*, 8 (2017) 16122-16143.
- [532] K.S. Sheinerman, et al., Plasma microRNA biomarkers for detection of mild cognitive impairment, *Aging*, 4 (2012) 590-605.
- [533] E.C. de Lange, Utility of CSF in translational neuroscience, *Journal of pharmacokinetics and pharmacodynamics*, 40 (2013) 315-326.
- [534] J.H. Lin, CSF as a surrogate for assessing CNS exposure: an industrial perspective, *Current drug metabolism*, 9 (2008) 46-59.
- [535] M. Pawlowski, et al., Cerebrospinal Fluid Biomarkers in Alzheimer's Disease-From Brain Starch to Bench and Bedside, *Diagnostics (Basel, Switzerland)*, 7 (2017).
- [536] J.H. Growdon, Biomarkers of Alzheimer disease, *Archives of neurology*, 56 (1999) 281-283.
- [537] A.S. Chen-Plotkin, Unbiased approaches to biomarker discovery in neurodegenerative diseases, *Neuron*, 84 (2014) 594-607.
- [538] K.S. Sheinerman, et al., Circulating brain-enriched microRNAs as novel biomarkers for detection and differentiation of neurodegenerative diseases, *Alzheimer's research & therapy*, 9 (2017) 89.
- [539] K.S. Sheinerman, et al., Circulating cell-free microRNA as biomarkers for screening, diagnosis and monitoring of neurodegenerative diseases and other neurologic pathologies, *Frontiers in cellular neuroscience*, 7 (2013) 150.
- [540] K.N. Nam, et al., RXR controlled regulatory networks identified in mouse brain counteract deleterious effects of Abeta oligomers, *Scientific reports*, 6 (2016) 24048.

- [541] J.J. Pei, et al., mTOR-dependent signalling in Alzheimer's disease, *J Cell Mol Med*, 12 (2008) 2525-2532.
- [542] B. Giunta, et al., Impact of the CD40-CD40L dyad in Alzheimer's disease, *CNS & neurological disorders drug targets*, 9 (2010) 149-155.
- [543] J. Tan, et al., CD40-CD40L interaction in Alzheimer's disease, *Current opinion in pharmacology*, 2 (2002) 445-451.
- [544] C.H. Volmar, et al., CD40/CD40L interaction induces Abeta production and increases gamma-secretase activity independently of tumor necrosis factor receptor associated factor (TRAF) signaling, *Experimental cell research*, 315 (2009) 2265-2274.
- [545] M. Kitazawa, et al., Blocking IL-1 signaling rescues cognition, attenuates tau pathology, and restores neuronal beta-catenin pathway function in an Alzheimer's disease model, *Journal of immunology (Baltimore, Md. : 1950)*, 187 (2011) 6539-6549.
- [546] O. Franzen, et al., Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases, *Science (New York, N.Y.)*, 353 (2016) 827-830.
- [547] J.L. Richens, et al., Rationalising the role of Keratin 9 as a biomarker for Alzheimer's disease, *Scientific reports*, 6 (2016) 22962.
- [548] B. Hitt, et al., beta-Site amyloid precursor protein (APP)-cleaving enzyme 1 (BACE1)-deficient mice exhibit a close homolog of L1 (CHL1) loss-of-function phenotype involving axon guidance defects, *The Journal of biological chemistry*, 287 (2012) 38408-38425.
- [549] M. Sathya, et al., BACE1 in Alzheimer's disease, *Clinica chimica acta; international journal of clinical chemistry*, 414 (2012) 171-178.
- [550] R. Cacabelos, Epigenomic networking in drug development: from pathogenic mechanisms to pharmacogenomics, *Drug development research*, 75 (2014) 348-365.
- [551] A. Fuso, The 'golden age' of DNA methylation in neurodegenerative diseases, *Clinical chemistry and laboratory medicine*, 51 (2013) 523-534.
- [552] J. Landgrave-Gomez, et al., Epigenetic mechanisms in neurological and neurodegenerative diseases, *Frontiers in cellular neuroscience*, 9 (2015) 58.
- [553] H. Suzuki, et al., Relationship Between Noncoding RNA Dysregulation and Epigenetic Mechanisms in Cancer, *Advances in experimental medicine and biology*, 927 (2016) 109-135.
- [554] D.G. Lupianez, et al., Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions, *Cell*, 161 (2015) 1012-1025.
- [555] S.A. Bellingham, et al., Small RNA deep sequencing reveals a distinct miRNA signature released in exosomes from prion-infected neuronal cells, *Nucleic acids research*, 40 (2012) 10937-10949.
- [556] C. Kilchert, et al., The regulation and functions of the nuclear RNA exosome complex, *Nature reviews. Molecular cell biology*, 17 (2016) 227-239.
- [557] G. Rabinowits, et al., Exosomal microRNA: a diagnostic marker for lung cancer, *Clinical lung cancer*, 10 (2009) 42-46.
- [558] D.D. Taylor, et al., MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer, *Gynecologic oncology*, 110 (2008) 13-21.

- [559] M. Schulze, et al., Sporadic Parkinson's disease derived neuronal cells show disease-specific mRNA and small RNA signatures with abundant deregulation of piRNAs, *Acta neuropathologica communications*, 6 (2018) 58.
- [560] A. Zovoilis, et al., microRNA-34c is a novel target to treat dementias, *The EMBO journal*, 30 (2011) 4299-4308.
- [561] M.A. Cortez, et al., MicroRNAs in body fluids--the mix of hormones and biomarkers, *Nature reviews. Clinical oncology*, 8 (2011) 467-477.
- [562] S. Kumar, et al., Are circulating microRNAs peripheral biomarkers for Alzheimer's disease?, *Biochimica et biophysica acta*, 1862 (2016) 1617-1627.
- [563] P.S. Mitchell, et al., Circulating microRNAs as stable blood-based markers for cancer detection, *Proceedings of the National Academy of Sciences of the United States of America*, 105 (2008) 10513-10518.
- [564] X. Yang, et al., Detection of stably expressed piRNAs in human blood, *International journal of clinical and experimental medicine*, 8 (2015) 13353-13358.