Multimodal analysis of neuroimaging and transcriptomic data in genetic frontotemporal dementia

Virgilio Kmetzsch



SORBONNE UNIVERSITÉ

ED 130 - INFORMATIQUE, TÉLÉCOMMUNICATIONS ET ÉLECTRONIQUE DE PARIS (EDITE)

Équipes de recherche Inria ARAMIS & DYLISS
Institut du Cerveau - Paris Brain Institute (ICM)

Multimodal analysis of neuroimaging and transcriptomic data in genetic frontotemporal dementia

---

Virgilio KMETZSCH

Thèse de doctorat d'informatique

Encadrée par Emmanuelle BECKER et Olivier COLLIOT

Présentée et soutenue publiquement le 26 septembre 2022

Composition du jury :

- **Anaïs BAUDOT** (Directrice de recherche), *rapportrice*
- **Blaise HANCZAR** (Professeur), *rapporteur*
- **Alessandra CARBONE** (Professeure), *examinatrice*
- **Yann LE CUNFF** (Maître de conférences), *examinateur*
- **Emmanuelle BECKER** (Maîtresse de conférences), *encadrante de thèse*
- **Olivier COLLIOT** (Directeur de recherche), *directeur de thèse*

Abstract

Multimodal analysis of neuroimaging and transcriptomic data in genetic frontotemporal dementia

Frontotemporal dementia (FTD) represents the second most common type of dementia in adults under the age of 65. Currently, there are no treatments that can cure this condition. In this context, it is essential that biomarkers capable of assessing disease progression are identified. Asymptomatic individuals who carry a genetic mutation that causes FTD are the ideal population for research.

This thesis has two objectives. First, to analyze the expression patterns of microRNAs taken from blood samples of patients, asymptomatic individuals who have certain genetic mutations causing FTD, and controls, to identify whether the expressions of some microRNAs correlate with mutation status and disease progression. Second, this work aims at proposing methods for integrating cross-sectional data from microRNAs and neuroimaging to estimate disease progression.

We conducted three studies. Initially, we focused on plasma samples from *C9orf72* expansion carriers. We identified four microRNAs whose expressions correlated with the clinical status of the participants. This study suggested that some microRNAs may be dysregulated even before the onset of symptoms. Next, we tested all microRNA signatures identified in the literature as potential biomarkers of FTD or amyotrophic lateral sclerosis (ALS), in two groups of individuals (with *C9orf72* expansion or with *GRN* mutation). The results of this study showed that microRNAs previously identified in sporadic or mixed cohorts may be useful to follow the progression of *C9orf72*-associated disease, but not of *GRN*-associated disease. Finally, in our third work, we proposed a new approach, using a supervised multimodal variational autoencoder, that estimates a disease progression score from cross-sectional microRNA expression and neuroimaging datasets with small sample sizes.

The work conducted in this interdisciplinary thesis showed that it is possible to use non-invasive biomarkers, such as circulating microRNAs and magnetic resonance imaging, to assess the progression of rare neurodegenerative diseases such as FTD and ALS. The main challenge in the future is to gather larger cohorts, with longitudinal data, to precisely assess the accuracy of these biomarkers in estimating disease progression of individual patients.

Résumé

Analyse multimodale des données de neuroimagerie et transcriptomiques dans la démence frontotemporale génétique

La démence frontotemporale (DFT) représente le deuxième type de démence le plus fréquent chez les adultes de moins de 65 ans. Il n'existe aucun traitement capable de guérir cette maladie. Dans ce contexte, il est essentiel d'identifier des biomarqueurs capables d'évaluer la progression de la maladie. Les personnes asymptomatiques porteuses d'une mutation génétique constituent la population idéale pour cette recherche.

Cette thèse a deux objectifs. Premièrement, analyser les profils d'expression des microARNs circulants prélevés dans le plasma sanguin de participants, afin d'identifier si l'expression de certains microARNs est corrélée au statut mutationnel et à la progression de la maladie. Deuxièmement, proposer des méthodes pour intégrer des données transversales de type microARN et de neuroimagerie pour estimer la progression de la maladie.

Nous avons mené trois études. D'abord, nous avons analysé des échantillons de plasma provenant de porteurs d'une expansion dans le gène *C9orf72*. Nous avons identifié quatre microARNs dont l'expression était corrélée avec l'état clinique. Cette étude a suggéré que certains microARNs peuvent être dérégulés avant même l'apparition des symptômes. Ensuite, nous avons testé toutes les signatures de microARNs identifiées dans la littérature comme biomarqueurs potentiels de la DFT ou de la sclérose latérale amyotrophique (SLA), dans deux cohortes indépendantes, avec une mutation dans le gène *C9orf72* ou *GRN*. Les résultats de cette étude ont montré que les microARNs précédemment identifiés dans des cohortes sporadiques ou mixtes peuvent être utiles pour suivre la progression de la maladie associée au gène *C9orf72*, mais pas celle de la maladie associée au gène *GRN*. Enfin, dans notre troisième étude, nous avons proposé une nouvelle méthode, utilisant un autoencodeur variationnel multimodal supervisé, qui estime à partir d'échantillons de petite taille un score de progression de la maladie en fonction de données transversales d'expression de microARNs et de neuroimagerie.

Les travaux menés dans cette thèse interdisciplinaire ont montré qu'il est possible d'utiliser des biomarqueurs non invasifs, tels que les microARNs circulants et l'imagerie par résonance magnétique, pour évaluer la progression de maladies neurodégénératives rares telles que la DFT et la SLA. Le principal défi à l'avenir est de rassembler des cohortes plus importantes, avec des données longitudinales, pour mieux déterminer la précision de ces biomarqueurs dans l'estimation de la progression de la maladie à l'échelle du patient.

Acknowledgements

I would like to express my sincere gratitude to both my supervisors Emmanuelle Becker and Olivier Colliot. First, for proposing such a fascinating subject and for recruiting me to carry out this thesis. Second, for their invaluable guidance, feedback and support during this work. Last, for their understanding, flexibility and kindness during a period of personal difficulties. It has been an honor and a pleasure working with them.

My gratitude goes also to Anaïs Baudot, Blaise Hanczar, Alessandra Carbone, and Yann Le Cunff, for their interest in my work and for having accepted to be part of the jury.

I wish to express my appreciation to Inria for funding this work, in particular I thank everyone involved in the Inria Project Lab Neuromarkers. I also extend my gratitude to the French government, for attracting international students and valuing their contribution. Through my present and future work, I sincerely hope I can give back to the French society.

Next, I would like to thank Dario Saracino and Isabelle Le Ber for their thoughtful suggestions and discussions, and their inestimable collaboration throughout this research project. I am also grateful to Vincent Anquetil and Morwena Latouche, whose work of collecting and preparing the plasma samples for our studies was crucial.

Then, I want to thank my colleagues at the Aramis team. It is a bit sad that the Covid-19 pandemic has negatively impacted our opportunities to be together (no team retreats, no conferences abroad, and no workshops on a péniche for those who started their PhD after 2019...). I am thankful and I will always remember the support of my colleagues, their good mood, and thoughtful debates during the pause-cafés and télépauses.

I am particularly grateful to my mother and my sister in Brazil. The last few years have been challenging, and finishing this work would not have been possible without their love, support and encouragement. Finally, I owe my deepest gratitude to Laura, for her love, for sharing this journey in France with me, and for always inspiring me to pursue my dreams.

Scientific Production

JOURNAL PAPERS

- **Kmetzsch, V.**, Anquetil, V., Saracino, D., Rinaldi, D., Camuzat, A., Gareau, T., Jornea, L., Forlani, S., Couratier, P., Wallon, D., Pasquier, F., Robil, N., de la Grange, P., Moszer, I., Le Ber, I., Colliot, O., Becker, E., PREV-DEMALS study group, "Plasma microRNA signature in presymptomatic and symptomatic subjects with C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis", *Journal of Neurology, Neurosurgery & Psychiatry*, 92(5):485-493 (2021). doi: 10.1136/jnnp-2020-324647 – hal-03046771. **(Chapter 4 of the dissertation)**.

- Muñoz-Ramírez, V., **Kmetzsch, V.**, Forbes, F., Meoni, S., Moro, E., Dojat, M., "Subtle anomaly detection: Application to brain MRI analysis of de novo Parkinsonian patients", *Artificial Intelligence in Medicine*, 125, 102251 (2022). doi: 10.1016/j.artmed.2022.102251

SUBMITTED JOURNAL PAPERS

- **Kmetzsch, V.**, Latouche, M., Saracino, D., Rinaldi, D., Camuzat, A., Gareau, T., the French research network on FTD/ALS, Le Ber, I., Colliot, O., Becker, E, "Circulating microRNA signatures as potential biomarkers in genetic frontotemporal dementia and amyotrophic lateral sclerosis". Submitted to *Molecular Psychiatry*. **(Chapter 5 of the dissertation)**.

- **Kmetzsch, V.**, Becker, E., Saracino, D., Rinaldi, D., Camuzat, A., Le Ber, I., Colliot, O., "Disease progression score estimation from multimodal imaging and microRNA data using supervised variational autoencoders". Under review at the *IEEE Journal of Biomedical and Health Informatics*. **(Chapter 6 of the dissertation)**.

PEER-REVIEWED CONFERENCE PROCEEDINGS

- **Kmetzsch, V.**, Becker, E., Saracino, D., Anquetil, V., Rinaldi, D., Camuzat, A., Gareau, T., Le Ber, I., Colliot, O., PREV-DEMALS study group, "A multimodal variational autoencoder for estimating progression scores from imaging and microRNA data in rare neurodegenerative diseases", *Proc. SPIE 12032, Medical Imaging 2022: Image Processing, 120321E (4 April 2022)* – Oral presentation. https://doi.org/10.1117/12.2607250 – hal-03576117

- Muñoz-Ramírez, V., **Kmetzsch, V.**, Forbes, F., Dojat, M., "Deep Learning Models to Study the Early Stages of Parkinson's Disease", *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1534-1537. doi: 10.1109/ISBI45749.2020.9098529

CONFERENCE ABSTRACTS

- **Kmetzsch, V.**, Becker, E., Saracino, D., Anquetil, V., Rinaldi, D., Camuzat, A., Gareau, T., Le Ber, I., Colliot, O., PREV-DEMALS study group, "Computing disease progression scores using multimodal variational autoencoders trained with neuroimaging and microRNA data", *Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM) 2022* – Oral presentation.

TALKS AND POSTERS

- Oral presentation – SPIE Medical Imaging 2022: Image Processing conference, San Diego, United States, March 2022. Video available on SPIE Digital Library.

- Oral presentation – Journées Ouvertes en Biologie, Informatique et Mathématiques (JOBIM), Rennes, France, July 2022.

- Poster – DKM Days, IRISA Rennes, France, February 2021.

Contents

List of Figures

List of Tables

List of Abbreviations

AA	Alzheimer's age
ABETA	Amyloid beta
AD	Alzheimer's disease
ADAS	Alzheimer's disease assessment scale
ADHD	Attention-deficit/hyperactivity disorder
ADNI	Alzheimer's disease neuroimaging initiative
AE	Autoencoder
AES	Apathy evaluation scale
AHBA	Allan Human Brain Atlas
ALS	Amyotrophic lateral sclerosis
ALS-FRS	ALS functional rating scale
ASL	Arterial spin labelling
AUC	Area under the curve
bvFTD	Behavioural variant FTD
C9orf72	Chromosome 9 open reading frame 72 gene
CBD	Corticobasal degeneration
CC BY	Creative Commons license, attribution
CC BY-NC-ND	Creative Commons license, attribution, non commercial, no derivatives
cDNA	Complementary DNA
CDRSB	Clinical dementia rating sum of boxes score
CI	Confidence interval
CSF	Cerebrospinal fluid
DNA	Deoxyribonucleic acid
ds-cDNA	Double-stranded complementary DNA
DTI	Diffusion tensor imaging
EDTA	Ethylenediamine tetraacetic acid
FAB	Frontal assessment battery
FBI	Frontal behavioral inventory
FDG-PET	Fluorodeoxyglucose PET
FTD	Frontotemporal dementia
FUS	RNA-binding protein FUS
FSL	FMRIB software library
GRN	Granulin precursor gene
HIPPO	Sum of the two lateral hippocampal volumes
IQR	Interquartile range

KEGG	Kyoto encyclopedia of genes and genomes
LSTM	Long short-term memory
MAPT	Microtubule associated protein tau gene
MCMC	Markov Chain Monte Carlo
miRNA	Micro-ribonucleic acid
MND	Motor neuron disease
MDRS	Mattis dementia rating scale
MMSE	Mini-mental state examination
MNI	Montreal Neurological Institute
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
nfvPPA	Nonfluent variant primary progressive aphasia
NGS	Next-generation sequencing
NODDI	Neurite orientation dispersion and density imaging
PCR	Polymerase chain reaction
PET	Positron emission tomography
PREV-DEMALS	Predict to prevent frontotemporal lobar degeneration and amyotrophic lateral sclerosis
PSP	Progressive supranuclear palsy
qRT-PCR	Quantitative real-time PCR
RAN	Repeat-associated non-AUG
RAVLT30	Rey auditory verbal learning test, 30 minute recall
RF	Radiofrequency
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROC	Receiver operating characteristic
ROI	Region of interest
RS-fMRI	Resting-state functional MRI
SPM	Statistical parametric mapping
svPPA	Semantic variant primary progressive aphasia
T1w [MRI]	T1-weighted [magnetic resonance imaging]
TADPOLE	The Alzheimer's disease prediction of longitudinal evolution
TBK1	TANK-binding kinase 1 gene
TCGA	The cancer genome atlas
TDP-43	Transactive response DNA-binding protein 43
TE	Echo time
TIV	Total intracranial volume
TR	Repetition time
TSV	Tab-separated values
UMI	Unique molecular identifier
VAE	Variational autoencoder

Chapter 1

Introduction

1.1 Context

Frontotemporal dementia (FTD) is a heterogeneous neurodegenerative disease, characterized by significant alterations in personality and social behavior, and associated with brain atrophy in the frontal and temporal lobes (Neary et al., 2005). Although a rare condition, FTD is the second most frequent early-onset dementia, after Alzheimer's disease, with an average age at onset being around 50 to 60 years (Seelaar et al., 2011). Currently, there are no effective treatments for this fatal disorder, which has terrible personal, familial, and social consequences.

Genetic factors have an important contribution to FTD, since a positive family history is observed in up to 50% of patients (DeJesus Hernandez et al., 2011). The most common causes of familial FTD are repeat expansions in the chromosome 9 open reading frame 72 (*C9orf72*) gene and mutations in the progranulin gene (*GRN*) (DeJesus Hernandez et al., 2011; Renton et al., 2011).

FTD shares disease mechanisms with amyotrophic lateral sclerosis (ALS), a devastating disorder caused by motor neuron degeneration, leading to progressive muscle atrophy and eventually complete paralysis (Pasinelli and Brown, 2006). The hexanucleotide repeat expansion in the *C9orf72* gene is the most frequent cause of both familial FTD and familial ALS identified to date. Both diseases may occur within the same family, or even the same patient (DeJesus Hernandez et al., 2011).

Most validated biomarkers in these diseases are used to distinguish patients from neurologically healthy controls. However, future therapeutic trials also need to monitor treatment responses, so research on potential progression biomarkers is becoming increasingly important (Meeter et al., 2017). Particularly, the study of genetic forms in the presymptomatic phase may provide valuable information about the early disease stages, before any irreversible neuronal loss has occurred (Rohrer et al., 2015).

In this context, the group led by Isabelle Le Ber at the Institut du Cerveau – Paris Brain Institute – ICM and the Institute for Memory and Alzheimer's Disease at the Pitié-Salpêtrière Hospital has put together some of the world's largest cohorts on familial variants of FTD/ALS. In each cohort, participants were divided into three groups: symptomatic mutation carriers (patients), neurologically healthy mutation

carriers (in the presymptomatic phase), and non-carriers (control group). Each subject underwent various examinations, including but not limited to structural magnetic resonance imaging (MRI) brain scans and blood samples.

Before the beginning of this PhD, the structural MRI scans were used to compute the volumes of specific brain regions of interest, an extensively applied technique for the study of neurodegenerative diseases. Statistical analyses of neuroimaging data have led to the identification of potential biomarkers of preclinical disease progression (Bertrand et al., 2018). However, neuroimaging alone cannot totally explain such complex disorders and reliably measure treatment response (Carreiro et al., 2015).

Blood samples were extracted to allow the quantification of circulating microRNA (miRNA) expression levels. MicroRNAs are a class of small ribonucleic acids (RNA) that regulate gene expression by degrading certain messenger RNAs or inhibiting their translation into proteins (Huntzinger and Izaurralde, 2011). Circulating miRNAs have been recently investigated as potential biomarkers for neurodegenerative diseases, with promising but so far conflicting results (Grasso et al., 2014).

From a clinical standpoint, the main goal of these cohort studies is to use multimodal data to identify robust biomarkers of FTD/ALS, in order to evaluate disease progression in future therapeutic trials.

1.2 Objectives

This thesis has two main objectives.

First, we aim to analyze miRNA expression levels in blood samples from presymptomatic mutation carriers, FTD/ALS patients and controls, to identify potential non-invasive biomarkers of disease progression. In particular, clinical trials would greatly benefit from robust preclinical biomarkers, that could assess the effect of treatments prior to the onset of symptoms and permanent brain damage.

Second, we intend to design a new approach to fuse multimodal data from miRNA expression and neuroimaging and evaluate disease progression using a cross-sectional study design. Indeed, research studies on rare neurodegenerative conditions, such as FTD and ALS, lack sufficient longitudinal data to apply existing disease progression models.

Specifically, our work seeks to elucidate the following research questions regarding genetic forms of frontotemporal dementia and amyotrophic lateral sclerosis:

- Are circulating miRNAs promising disease progression biomarkers?
- Could we estimate disease progression using only cross-sectional miRNA and neuroimaging data?

1.3 Contributions

The contributions of this thesis are three-fold.

First, we investigate the expression levels of plasma miRNAs in a cohort of *C9orf72* mutation carriers, and we identify four miRNAs likely to play a role in neurodegeneration and *C9orf72*-associated pathogenesis. Our study suggests that the expression levels of certain miRNAs are dysregulated during disease progression, and can be detected even before neurological symptoms appear.

Second, we analyze two homogeneous cohorts of *C9orf72* and *GRN* mutation carriers, to assess all circulating miRNA signatures identified in previous conflicting studies about FTD or ALS. Our findings indicate that miRNA expression is most likely mutation specific, since miRNAs previously revealed in sporadic/mixed cohorts can potentially serve as biomarkers in *C9orf72*-associated FTD or ALS, but not in *GRN*-associated FTD.

Finally, we propose a novel approach for assessing disease progression scores (DPS) using cross-sectional neuroimaging and miRNA data, which may be used in small samples like those observed in rare disorders. The technique is developed and tested using data from *C9orf72*-associated FTD and ALS, but it might be applied to other disorders as well. The method's capacity to reliably predict the DPS is demonstrated on synthetic data, while tests on a real-world dataset, in the absence of ground truth scores, reveal the classification of distinct diagnostic groups. The results of this work support that a single disease progression score, computed from cross-sectional data, might be used to represent a neurodegenerative disease progression.

1.4 Outline of this manuscript

Following this Introduction (Chapter 1), the manuscript consists of six more chapters:

- Chapter 2 describes the background related to this thesis, covering (1) genetic FTD and ALS, (2) miRNA data, (3) neuroimaging data, (4) integrative data analysis in neurodegenerative diseases, and (5) disease progression modeling and disease progression scores.
- Chapter 3 shows the miRNA and neuroimaging datasets used in this work.
- Chapter 4 presents a study analysing the expression of circulating miRNAs in a cohort of *C9orf72* carriers and their first-degree relatives, aiming at identifying potential plasma biomarkers in this genetic form of FTD and ALS.
- Chapter 5 extends the previous study by selecting all published articles that identify miRNA signatures associated with FTD and/or ALS, and testing them in an independent cohort focused on the *C9orf72* mutation and another cohort focused on the *GRN* mutation.

- Chapter 6 details the implementation of a novel framework, based on a supervised multimodal variational autoencoder, that infers disease progression scores from cross-sectional miRNA and neuroimaging data, evaluated on synthetic datasets and on a real-world dataset.
- Finally, Chapter 7 summarizes our contributions and outlines perspectives for future research.

In addition, the supplementary materials associated with the studies described in Chapters 4, 5 and 6 are presented in, respectively, Appendices A, B and C.

Chapter 2

Background

This chapter provides the background knowledge regarding the main concepts involved in this thesis. We initially present frontotemporal dementia and amyotrophic lateral sclerosis, focusing on their familial forms (Section 2.1). Next, we describe the usual preprocessing and feature extraction steps for microRNA (miRNA) expression data (Section 2.2) and structural neuroimaging data (Section 2.3). Then, we give an overview about integrative data analysis in neurodegenerative diseases, in particular the integration of transcriptomic and neuroimaging data (Section 2.4). Last, we briefly review the literature on disease progression modeling and disease progression scores (Section 2.5).

2.1 Frontotemporal dementia & amyotrophic lateral sclerosis

In this section, we present the two neurodegenerative diseases discussed in this thesis. We begin with a brief definition of frontotemporal dementia (section 2.1.1), followed by an explanation about its different familial forms (section 2.1.2), its overlap with amyotrophic lateral sclerosis (section 2.1.3), and finally an overview about the neuroimaging and fluid biomarkers associated with these diseases (section 2.1.4).

2.1.1 Frontotemporal dementia definition

Frontotemporal dementia (FTD) designates a group of clinically, pathologically and genetically heterogeneous disorders, that have in common the degeneration of the frontal and temporal lobes of the brain (Lashley et al., 2015). The annual incidence of these devastating conditions is estimated to be approximately 4 new cases per 100,000 people (Ratnavalli et al., 2002). Although rare, these disorders represent the second most common form of dementia in adults aged under 65 years, after Alzheimer's disease (AD) (Meeter et al., 2017).

FTD has two main clinical manifestations. First, behavioural variant FTD (bvFTD) is characterized by a progressive decline in cognition and social relationships, leading to impulsive actions, loss of empathy, compulsive behaviours, and deficits in executive functions (Rascovsky et al., 2011). Second, primary progression aphasia (PPA) is marked by an increasing language deficit that impairs daily activities, and is divided into semantic variant PPA (svPPA), nonfluent variant PPA (nfvPPA) and

lopogenic variant PPA (lvPPA) (Gorno-Tempini et al., 2011). Depending on the variant, PPA may lead to fluent speech with impaired comprehension of single words (svPPA), nonfluent speech with inability to use fundamental grammar and syntax (nfvPPA), or nonfluent speech characterized by compromised word retrieval (lvPPA) (Gorno-Tempini et al., 2011). Other clinical phenotypes include concomitant motor neuron disease (MND) in the disease course, that may manifest as amyotrophic lateral sclerois (ALS), corticobasal syndrome (CBD), or progressive supranuclear palsy (PSP) (Lashley et al., 2015).

Pathological heterogeneity is revealed by analyzing postmortem brain tissues collected from FTD patients. Different protein inclusions are observed, either of transactive response DNA-binding protein 43 (TDP-43), RNA-binding protein FUS, or microtubule associated protein tau (Lashley et al., 2015). The underlying pathology cannot be accurately predicted from the clinical phenotype (Meeter et al., 2017).

2.1.2 Familial forms

FTD may occur in sporadic or familial forms. However, a clear distinction between familial and sporadic disease is sometimes challenging, due to the difficulty to establish a reliable family history of cognitive disorders (Turner et al., 2017). Patients frequently have a family history of a comparable condition, usually in a pattern suggesting dominant inheritance (Lashley et al., 2015). Most familial forms of FTD are caused by mutations in one of three genes: *C9orf72* (encoding protein *C9orf72*), *GRN* (encoding progranulin), or *MAPT* (encoding microtubule associated protein tau) (Meeter et al., 2017). Other genes rarely associated with FTD include *VCP*, *TARDP*, and *TBK1* (Pottier et al., 2016).

Figure 2.1 shows the clinical, pathological and genetic heterogeneity of FTD. We observe that genetic forms have a predictable pathology: *GRN* mutations and *C9orf72* repeat expansions result in TDP-43 pathology, while the *MAPT* mutation causes tau pathology. However, a particular genetic mutation may result in distinct phenotypes.

2.1.3 Overlap with amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is a fatal, progressive, degenerative disorder characterized by the deterioration of motor neurons in the brain and spinal chord, leading to the complete paralysis of voluntary muscles, and eventually death induced by respiratory failure (Pasinelli and Brown, 2006). ALS is estimated to have an incidence of 2.1 new cases per 100,000 people per year (Chiò et al., 2013).

It is recognized that FTD and ALS form a disease spectrum, with remarkable pathological and genetic overlap, both conditions occurring sometimes within the same family, or even the same individual (Abramzon et al., 2020). According to cross-sectional studies, up to 50% of ALS patients also show cognitive impairment

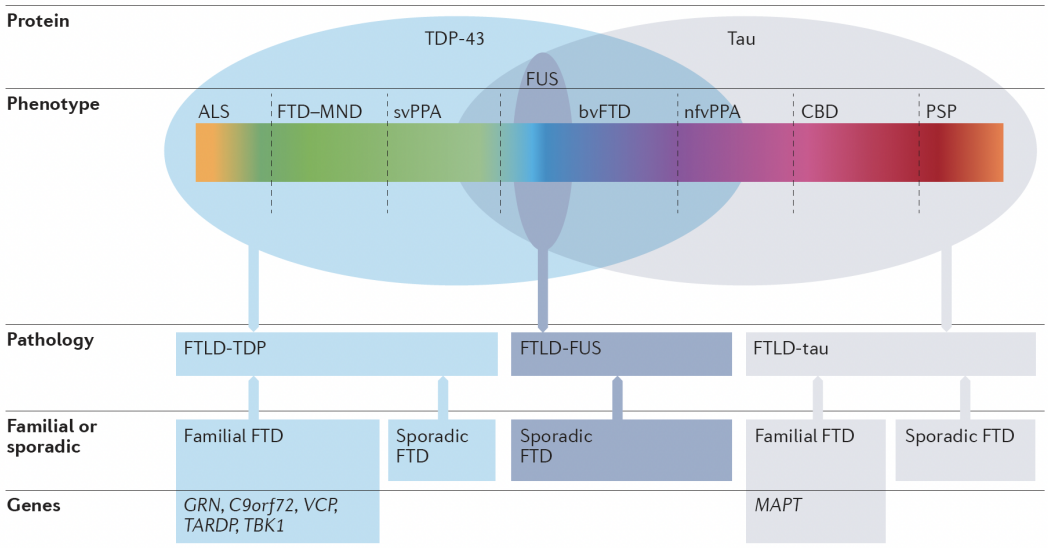


FIGURE 2.1: Clinical, pathological and genetic spectrum of FTD (reproduced from (Meeter et al., 2017), with permission from the publisher). bvFTD: behavioural variant FTD; CBD: corticobasal degeneration; FUS: RNA-binding protein FUS; MND: motor neuron disease; nfvPPA: nonfluent variant primary progressive aphasia; PSP: progressive supranuclear palsy; svPPA: semantic variant primary progressive aphasia; TDP-43: transactive response DNA-binding protein 43.

as a consequence of FTD, and motor dysfunction affects up to 30% of FTD patients (Burrell et al., 2011).

Pathological overlap between these disorders is supported by the fact that *TDP-43* represents the most frequent pathological protein aggregation in both FTD and ALS patients (DeJesus Hernandez et al., 2011). Genetic overlap is observed mainly because the *C9orf72* repeat expansion is the most common genetic abnormality observed in familial forms of both FTD and ALS (DeJesus Hernandez et al., 2011), with some other rarer genetic mutations, such as in *VCP*, *TARDP*, and *TBK1* also playing a role in this overlap (Abramzon et al., 2020). On the other hand, mutations in *GRN* and *MAPT* are linked only to FTD.

2.1.4 Biomarkers

The Biomarkers Definitions Working Group defines a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" (Biomarkers Definitions Working Group., 2001). The authors highlight the main applications of biomarkers, including their use as a diagnostic tool, indicator of disease staging, or measure of clinical response to a treatment (Biomarkers Definitions Working Group., 2001).

Since frontotemporal dementia and amyotrophic lateral sclerosis are heterogeneous disorders with no disease-modifying treatments to date, robust biomarkers

are crucial for selecting individuals to clinical trials and evaluating treatment results (Bertrand et al., 2018; Rohrer et al., 2015). Over the last two decades, a lot of effort went into identifying these biomarkers, with a particular focus on fluid biomaterial and neuroimaging features (Meeter et al., 2017). Ideally, biomarkers should be non-invasive, reliable, and inexpensive, and ought to enable diagnosis, monitoring of disease progression, and treatment response (Meeter et al., 2017).

Evidence suggests that alterations in certain biomarkers occur several years before symptoms appear, implying that the best time to treat FTD and ALS is before clinical presentation, when the least amount of irreversible neuronal loss has happened and cognitive function is still preserved (Bertrand et al., 2018; Rohrer et al., 2015). Studies focusing on familial forms of FTD and ALS are particularly important, since mutation carriers in the presymptomatic stage may give insight into the diseases' initial phases and improve therapy options (Rohrer et al., 2015). Below, we explore the main biomarkers associated with FTD and ALS in the literature, as well as their limitations.

Neuroimaging biomarkers

Most imaging studies in FTD have used volumetric T1-weighted MRI to investigate changes in gray matter structure, either to determine the volumes of specific brain regions of interest (ROI) or to evaluate the rate of atrophy in longitudinal studies (Meeter et al., 2017). Findings from investigations of FTD presymptomatic mutation carriers have shown gray matter atrophy at least ten years before expected disease onset (Bertrand et al., 2018; Rohrer et al., 2015).

To some degree, genetic forms can be distinguished by different patterns of gray matter atrophy at a group level, as shown by Figure 2.2, displaying the characteristic patterns of gray matter atrophy in FTD patients with different underlying mutations (Meeter et al., 2017).

Asymmetrical frontotemporoparietal atrophy is commonly seen in patients with *GRN* mutations, while most patients with a *C9orf72* repeat expansion exhibit a generalized symmetrical atrophy, and *MAPT* mutations cause symmetrical temporal atrophy (Meeter et al., 2017). Atrophy rates also vary between different genetic forms, since patients carrying the *GRN* mutation present significantly faster brain atrophy patterns than *C9orf72* expansion carriers (Whitwell et al., 2015). Even though these distinct patterns are visible in group-level studies, it is not possible to distinguish between patients with different underlying pathologies by analyzing gray matter atrophy alone (Whitwell and Josephs, 2012).

Recently, FTD imaging studies have also focused on white matter integrity using diffusion tensor imaging (DTI), which are more sensitive to detect the earliest changes in the brain, in comparison to gray matter atrophy (Bertrand et al., 2018). Neurite orientation dispersion and density imaging (NODDI) has also been explored, and seems to provide even higher sensitivity to changes in the presymptomatic phase than DTI (Wen et al., 2019). Other neuroimaging modalities used in

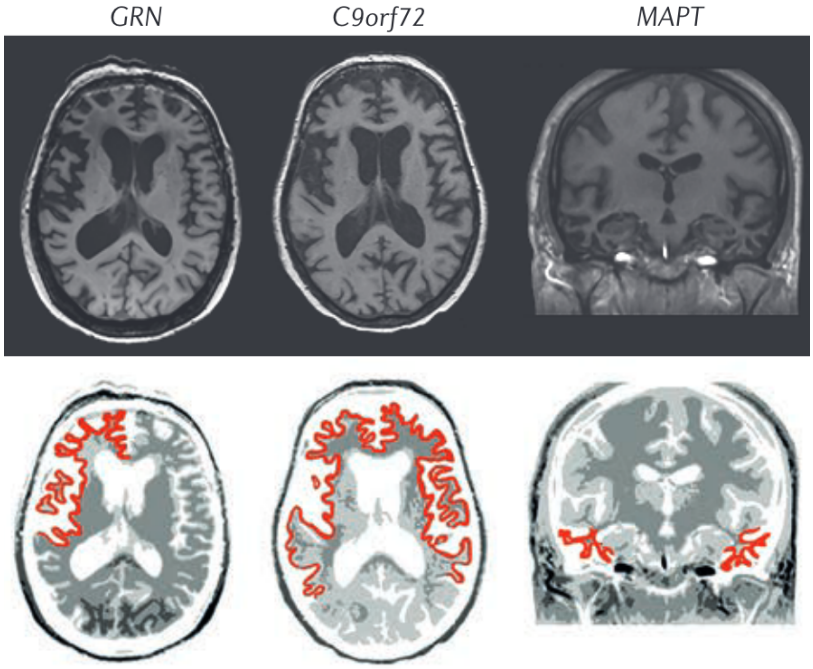


FIGURE 2.2: Gray matter atrophy in different genetic subtypes of frontotemporal dementia (reproduced from (Meeter et al., 2017), with permission from the publisher). The characteristic patterns of gray matter atrophy in patients carrying each mutation are highlighted in red.

FTD studies include positron emission tomography with 18F-fluorodeoxyglucose as the tracer (FDG-PET) to visualize alterations in brain metabolism, functional connectivity between brain regions measured with resting-state functional MRI (RS-fMRI), and arterial spin labelling (ASL) to create a tracer of cerebral blood flow (Meeter et al., 2017).

Even though there is converging evidence that neuroimaging modalities are useful to assess disease progression, their use in clinical practice is until now limited to classification between FTD, AD and controls (Meeter et al., 2017). Larger studies are necessary to validate robust biomarkers, before these modalities can be used at an individual level to evaluate the outcomes of clinical trials (Meeter et al., 2017).

Fluid biomarkers

The main idea behind fluid biomarkers is that changes in specific protein concentrations in different human fluid compartments, such as cerebrospinal fluid (CSF) or blood, can indicate pathophysiological changes in disease processes (Meeter et al., 2017).

CSF amyloid-β and tau have been extensively validated to distinguish between Alzheimer's disease and frontotemporal dementia (Rivero-Santana et al., 2017). These biomarkers have also demonstrated potential to differentiate individuals with different genetic or pathological FTD subtypes, but additional studies are necessary to validate these applications (Meeter et al., 2017).

Moreover, increased levels of neurofilaments reflect axonal damage, and it has been established that blood and CSF levels of neurofilaments are increased in FTD patients when compared to controls, and could potentially be used as biomarkers of disease progression (Rohrer et al., 2016; Saracino et al., 2021). However, these proteins are also overexpressed in other neurodegenerative diseases, so they must be combined with other disease-specific biomarkers to increase their clinical utility (Meeter et al., 2017).

Additionally, some gene-specific fluid biomarkers, such as progranulin levels in the CSF of *GRN* mutation carriers (Feneberg et al., 2016) and dipeptide-repeat proteins in the CSF of *C9orf72* expansion carriers (Su et al., 2014), have been identified as potential progression biomarkers of FTD and FTD/ALS respectively, but need to be combined with other biomarkers to have clinical value (Meeter et al., 2017).

Finally, recent studies have been focusing on miRNAs as potential biomarkers for several neurodegenerative diseases (Grasso et al., 2014), including FTD and ALS (Eitan and Hornstein, 2016; Gascon and Gao, 2014). MicroRNAs act as regulators of key biological functions in the central nervous system, such as synaptic plasticity and neurogenesis, and are preserved in biofluids such as CSF and blood (Grasso et al., 2014). Since miRNAs have an essential regulatory function in the brain, changes in their concentration levels could indicate physiological changes that precede neuronal cell loss (Watson et al., 2019).

The reasoning behind the link between miRNAs and FTD/ALS is that TDP-43 and FUS, two of the proteins frequently accumulated in the brains of FTD/ALS patients, are RNA-binding proteins that participate in DNA replication, mRNA translation, and miRNA synthesis (Buratti and Baralle, 2010). Therefore, the dysregulation of these proteins associated with FTD and ALS pathogenesis could lead to alterations in miRNAs expression levels (Gascon and Gao, 2014).

It is highly unlikely that a single miRNA could be a useful biomarker of a neurodegenerative disease, but a combination of miRNA biomarkers (a miRNA signature) could enable diagnosis and disease monitoring (Watson et al., 2019). Numerous studies have recently identified specific miRNAs as potential biomarkers of FTD (Denk et al., 2018; Grasso et al., 2019; Piscopo et al., 2018; Sheinerman et al., 2017) and ALS (De Felice et al., 2014; Dobrowolny et al., 2021; Freischmidt et al., 2015, 2014; Magen et al., 2021; Raheja et al., 2018; Sheinerman et al., 2017; Soliman et al., 2021; Takahashi et al., 2015; Tasca et al., 2016; Waller et al., 2017). Although promising, results among these different works are conflicting, for three main reasons: heterogeneous cohorts (genetic, sporadic or mixed), lack of independent validation cohorts in most cases, and study designs with different assumptions about which miRNAs to investigate. We tackle these issues by conducting a discovery study without *a priori* assumptions in an homogeneous cohort of *C9orf72* expansion carriers (Chapter 4), and a comprehensive validation study using two independent homogeneous cohorts of *C9orf72* and *GRN* mutation carriers (Chapter 5).

2.2 MicroRNA data

In this section, we begin by defining what is a miRNA (section 2.2.1). Then, we describe the main steps of a typical RNA sequencing (RNA-seq) experiment (section 2.2.2), and finally we explain how to apply RNA-seq to assess miRNA expression levels (section 2.2.3).

2.2.1 MicroRNA definition

The central dogma of molecular biology explains the flow of information contained in genes: DNA is transcribed into messenger RNA (mRNA), which is translated into proteins (Crick, 1970). As a consequence of encoding proteins via the genetic code, mRNA molecules were historically the most investigated RNA species (Kukurba and Montgomery, 2015).

Recently, novel classes of noncoding RNA have been identified, including miRNAs, which are small RNA molecules (average length of 22 nucleotides) that have a regulatory role: they repress the expression of certain mRNA targets by inducing translational inhibition, mRNA degradation, or both (Huntzinger and Izaurralde, 2011). One miRNA can target up to hundreds of mRNA transcripts, thus regulating the levels of several genes (Gascon and Gao, 2014). Likewise, a particular mRNA may be targeted by many different miRNAs (Watson et al., 2019).

Next-generation sequencing technologies, that allow the identification and quantification of RNA molecules in biological samples, have enabled increased rates of miRNA discovery in recent years (Kozomara and Griffiths-Jones, 2011). The most used of these technologies, called RNA sequencing, is described below.

2.2.2 RNA sequencing

The first studies measuring RNA expression levels relied on low-throughput technologies such as polymerase chain reaction (PCR), which are limited to quantifying individual transcripts (Becker-André and Hahlbrock, 1989). In the mid-1990s, methods evolved to high-throughput techniques such as microarrays, that quantify the abundance of large sets of transcripts, and allowed the first genome-wide analyses of gene expression patterns (Brown and Botstein, 1999). Nevertheless, these technologies have important limitations: they rely upon existing knowledge about the sequences to be analyzed, and they have a narrow dynamic range of detection of expression levels.

To overcome these limitations, a high-throughput next-generation sequencing (NGS) technology called RNA sequencing (RNA-Seq) was developed, and began to grow in popularity after 2008 (Lowe et al., 2017; Wang et al., 2009). RNA-Seq has clear advantages compared to other approaches: it does not require prior knowledge about the target sequences, it has a high accuracy even for very lowly expressed and very highly expressed transcripts, and it has high levels of reproducibility (Wang

et al., 2009). Since 2015, RNA-Seq surpassed microarrays as the most used technique for measuring RNA expression (Lowe et al., 2017).

Specific RNA-Seq protocols may differ in several aspects, such as transcript enrichment, fragmentation, and amplification (Lowe et al., 2017). Almost 100 different methods have been created based on the standard RNA-Seq protocol (Stark et al., 2019). Usually, the main steps of a typical RNA-Seq experiment are (1) extraction and isolation of RNA from the organism, (2) sequencing library preparation, (3) sequencing on a NGS platform, and (4) alignment of sequencing reads and quantification of transcripts (Kukurba and Montgomery, 2015; Lowe et al., 2017). These typical steps are displayed in Figure 2.3 and explained below, while more specific procedures concerning miRNA sequencing are detailed in section 2.2.3.

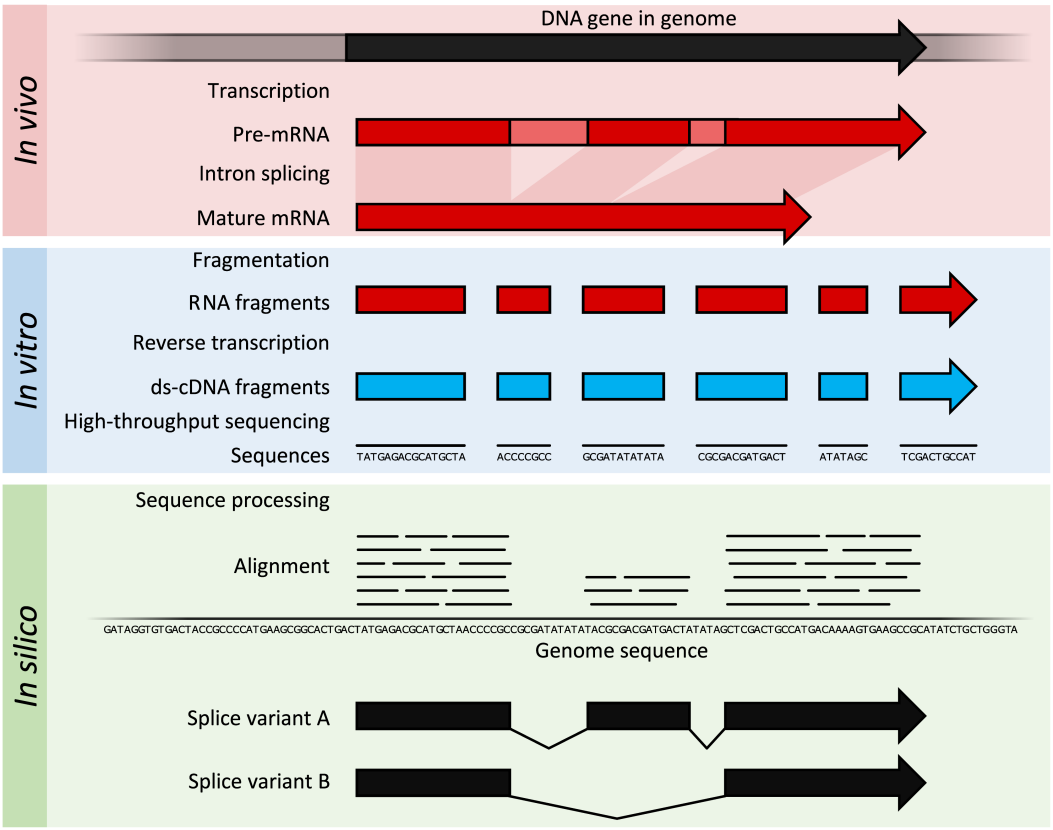


FIGURE 2.3: Main steps of the RNA Sequencing methodology (reproduced from (Lowe et al., 2017), CC BY 4.0). The three colored segments represent respectively the steps happening in the living organism, carried out in the laboratory, and performed by computational tools. Pre-mRNA: precursor messenger RNA; ds-cDNA: double-stranded complementary DNA.

RNA extraction and isolation

The first step for a successful RNA-Seq experiment is to isolate high-quality RNA from the analyzed biological samples (e.g. tissue, plasma) using a specific protocol, which is a critical factor to avoid incorrect conclusions in downstream analyses (Kukurba and Montgomery, 2015). Within organisms, the transcription process copies segments of DNA into precursor mRNA molecules, then the splicing process removes introns (noncoding sequences) and joins exons (coding sequences), creating mature mRNA molecules to be extracted (Lowe et al., 2017).

Sequencing library preparation

The next step is to construct RNA-Seq libraries. An appropriate library preparation protocol must be chosen, each with its own biases and limitations, in order to enrich or deplete the RNA sample to target specific RNA species (Kukurba and Montgomery, 2015). Usually, RNA molecules have to be fragmented, because messenger RNA molecules are longer than the typical read-lengths supported by next-generation sequencing methods (Lowe et al., 2017). Moreover, since most sequencing technologies require DNA libraries, it is necessary to convert RNA into stable double-stranded complementary DNA (ds-cDNA) (Kukurba and Montgomery, 2015). Next, DNA adaptors are attached to the ends of the complementary DNA fragments (Kukurba and Montgomery, 2015). Finally, before sequencing, PCR amplification may be applied to allow sequencing of low-input amounts of RNA (Lowe et al., 2017). In that case, to control for PCR amplification biases, unique molecular identifiers (UMIs) may be used to tag cDNA fragments before amplification, allowing the identification and removal of PCR duplicates (Stark et al., 2019).

Sequencing on a NGS platform

In recent years, although several next-generation sequencing platforms have been developed and are now commercially available, the Illumina platform is the most frequently applied in RNA-Seq experiments and has clearly dominated the industry (Kukurba and Montgomery, 2015). Released in 2006, Illumina technology provides a highly accurate tool for measuring RNA expression (single read accuracy of 99.9%), with typical read lengths from 50 to 300 base pairs (Lowe et al., 2017), and usual read depths of 10 to 30 million reads per sample (Stark et al., 2019). A recent survey indicated that more than 95% of the published RNA-Seq data has been generated using the Illumina sequencing technology (Stark et al., 2019). In order to output high-quality and biologically meaningful data, sequencing parameters such as read depth and read length must be carefully chosen, depending on the study characteristics (Stark et al., 2019). The standard RNA-Seq pipeline outputs the raw sequence data in FASTQ-format files. FASTQ is a standard text-based file format for storing sequencing read data, that combines both the nucleotide sequences and the

corresponding quality scores (Cock et al., 2010). The following steps are all computational.

Alignment and quantification of transcripts

Once the raw reads are available, the next step is to align them to an annotated reference genome, a task for which several approaches exist (Kukurba and Montgomery, 2015). Each alignment tool has advantages concerning performance and resource utilisation, therefore selecting the best method depends on the peculiarities of each RNA-Seq study (Kukurba and Montgomery, 2015). Regardless of the chosen alignment tool, a quality control step is essential to remove any abnormalities from the data, since sequence reads are not perfect (Lowe et al., 2017). After reads have been mapped to genomic locations, a quantification step is performed to obtain an expression matrix, where the rows represent features (genes or transcripts), the columns represent samples, and the values are read counts (Stark et al., 2019). More detailed quality assessment, alignment and quantification steps, specific to miRNA sequencing, are described in the next subsection.

2.2.3 MicroRNA sequencing

In addition to quantifying messenger RNA transcripts, RNA-Seq can be applied to analyze different populations of RNA, including miRNAs (Kukurba and Montgomery, 2015). Since miRNAs are shorter and less abundant than messenger RNA, specific protocols focused on miRNA sequencing have been developed and are commercially available as extraction kits (Kukurba and Montgomery, 2015).

From a computational point of view, multiple tools are available for miRNA sequencing analysis, which may cause uncertainty in how to define the most appropriate computation pipeline (Potla et al., 2021). The vast majority of RNA-Seq studies performed to date are focused on messenger RNAs, and the few studies that perform miRNA sequencing do not provide enough details about particular choices of their bioinformatics analysis (Potla et al., 2021).

Recently, a generic miRNA bioinformatics pipeline was proposed, using exclusively open source software, with the goal to improve clarity and reproducibility among miRNA sequencing studies (Potla et al., 2021). We thus describe below this computation pipeline, which has five major steps: (1) quality assessment of raw reads, (2) UMI extraction and adapter trimming, (3) alignment of reads to the mature miRNA sequences, (4) removal of PCR duplicates, and (5) creation of miRNA count tables.

Quality assessment of raw reads

It is critical to analyze if the raw reads are of sufficient quality for downstream analyses. Tools like FastQC (Andrews S. 2010[1]) may be applied to FASTQ format files,

[1] https://www.bioinformatics.babraham.ac.uk/projects/fastqc

generating summary graphs and tables that allow assessing the overall quality of each sample by checking for unusual patterns. For instance, one may observe the number of total reads, the quality scores of individual base sequences, the distribution of sequence lengths, the amount of duplicate sequences, and other criteria (Potla et al., 2021).

UMI extraction and adapter trimming

If in the previous step the raw reads are considered to be of sufficient quality, the next step is to extract the UMIs from the reads, and save them for later use. For this task, UMI-tools (Smith et al., 2017) removes UMIs from FASTQ files and appends them to each read name. Then, Cutadapt (Martin, 2011) finds and eliminates all adapter sequences, as well as discards reads that are either shorter (less than 18 nucleotides) or longer (more than 30 nucleotides) than the expected read lengths with some tolerance. After UMI extraction, adapter trimming, and read filtering, the remaining reads are of good quality and ready for alignment.

Alignment of reads to the mature miRNA sequences

Before describing the alignment of reads, it is useful to mention the importance of miRBase[2]. Next-generation sequencing technologies have contributed to an increasing rate of novel miRNA discovery, and the community maintained miRBase database is the most important online repository for all published miRNA sequences, curating a consistent nomenclature scheme and providing miRNA target predictions and validations (Kozomara and Griffiths-Jones, 2011). It is recommended to use the latest version of miRBase as the reference database. Alignment must be performed with strict criteria, to avoid matches of the same read across multiple mature miRNA sequences (Potla et al., 2021), so it is recommended to align reads using Bowtie (Langmead et al., 2009), allowing no mismatches. After the alignment of reads to the reference database, a second major quality-check must performed, since samples with too few aligned reads must be discarded to avoid biasing the results (Potla et al., 2021). To give an idea about the order of magnitude of a successful alignment, on average, a sequencing depth of 10 million reads per sample will yield at least 3 million aligned reads per sample (Potla et al., 2021).

Removal of PCR duplicates

Once the alignment step is complete for all samples, UMI-tools (Smith et al., 2017) is used again, this time to collapse read counts with the same UMIs, in order to count the number of reads corresponding to the original biological sample, before PCR amplification.

[2] https://www.mirbase.org

Creation of miRNA count tables

Finally, Samtools idxstats (Li et al., 2009) counts each miRNA in every sample, and records the counts in a tab-separated values (TSV) file, where each row corresponds to a mature miRNA, and each column represents a sample. These raw counts are then ready to be used in downstream differential expression analyses.

2.3 Neuroimaging data

In the last decades, analyses of neuroimaging data have been extensively performed to reveal neuroanatomical and functional differences between patients with neurodegenerative diseases and healthy controls (Orrù et al., 2012). One of the most noticeable changes in the brains of patients with dementia is gray matter atrophy resulting from the loss of neuronal cells, which is visualized by structural imaging techniques such as structural MRI (Cedazo-Minguez and Winblad, 2010). Among the existing structural MRI sequences, T1-weighted (T1w) MRI is one of the most frequently used to assess brain abnormalities caused by neurodegenerative diseases (Popuri et al., 2018; Tartaglia et al., 2011).

Indeed, in the clinical research cohorts that we had access, T1w MRI is the most commonly available sequence, and with the most reliable harmonization across centers. Furthermore, it is recommended to be systematically included in the clinical follow-up of patients with suspected dementia. This explains our focus on this imaging modality, even though other MRI sequences (including diffusion MRI and functional MRI) or imaging modalities (including PET) have the potential to provide more sensitive biomarkers.

In this section, we briefly describe the two main steps usually carried out prior to T1-weighted MRI data analysis: image acquisition (section 2.3.1) and image preprocessing (section 2.3.2) for feature extraction. For more details on T1w MRI, or on other MRI modalities unrelated to this thesis, one can refer to Bernstein et al., 2004; Haacke et al., 1999; McRobbie et al., 2006.

2.3.1 T1-weighted MRI acquisition

MRI scanners apply a uniform magnetic field to align the hydrogen nuclei within the tissues being examined, then generate radiofrequency (RF) pulses to perturb this alignment (Bitar et al., 2006). After a perturbation, the scanner measures the RF signals emitted from the realignment of the nuclei from different tissues, each with their own characteristic relaxation times, allowing the detection of locations with distinct concentrations of fat and water (Bitar et al., 2006).

To create tissue contrast on MRI, there are two key parameters: repetition time (TR) is the time between the application of two RF pulses, and echo time (TE) is the time between the application of a RF pulse and the echo detected (Bitar et al., 2006).

These parameters may be tuned to emphasize a particular type of contrast: in T1-weighted MRI, short TR and TE are used (Brown and Semelka, 1999). When exposed to a T1-weighted sequence, fluids generate a very low signal intensity, corresponding to darker regions in the resulting images (Brown and Semelka, 1999). Therefore, the cerebrospinal fluid appears in dark, gray matter in dark gray, and white matter in light gray (due to the high lipid fat content of the myelin)[3].

T1-weighted MR images clearly depict the anatomy of the brain, and for that reason have been extensively used to reveal patterns of brain atrophy associated with several neurodegenerative diseases (Tartaglia et al., 2011). Figure 2.4 displays axial slices from an FTD patient's brain, obtained with T1-weighted MRI, where focal brain atrophy is clearly observed.

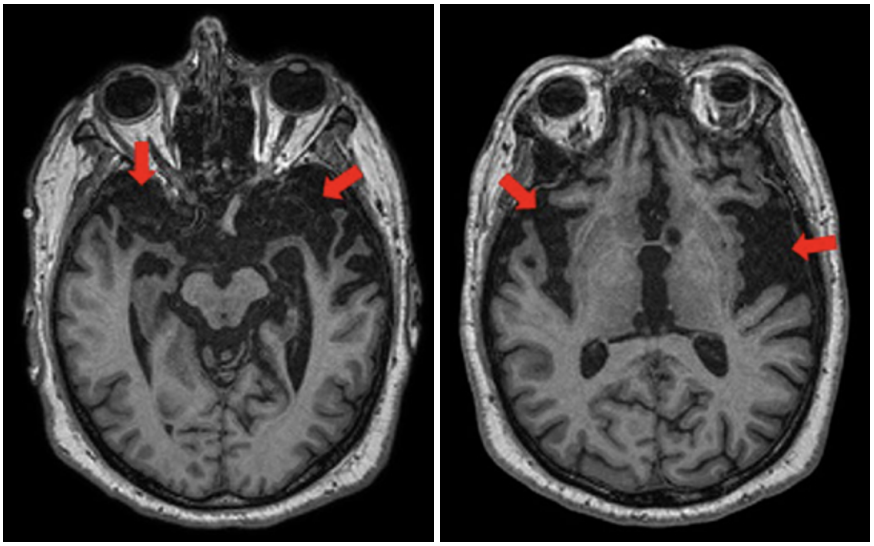


FIGURE 2.4: Axial views of T1-weighted MRI from an FTD patient (reproduced from (Ramirez et al., 2020), CC BY 4.0). The red arrows highlight the regions with significant brain atrophy.

2.3.2 T1-weighted MRI preprocessing

The quality of MR images is mainly affected by spatial resolution, contrast, and signal to noise ratio, the latter being mainly determined by scan time and substantially influenced by the patients' ability to remain still (Symms et al., 2004). Any head motion during the MRI exam, including the respiratory cycle and eye movements, can result in motion artifacts that degrade image quality (Symms et al., 2004). Artifacts may also be caused by the MR scanner itself, from the complex interactions between the main magnet, RF transmitter and receiver, and computer systems for image reconstruction (Zhuo and Gullapalli, 2006).

Therefore, preprocessing is an essential step to improve image quality for successful downstream analyses. The essential steps for T1w MRI preprocessing in the

[3] http://fmri.ucsd.edu/Howto/3T/structure.html

context of brain atrophy associated with neurodegenerative diseases are: (1) intensity non-uniformity correction, (2) intensity rescaling and standardization, (3) skull stripping, (4) image registration, and (5) feature extraction (Bertrand et al., 2018). Each of these steps is described below, and can be performed, for example, using the software packages FreeSurfer[4] (Fischl, 2012), the FMRIB Software Library (FSL)[5] (Jenkinson et al., 2012), or Statistical Parametric Mapping (SPM)[6] (Ashburner, 2012), three of the most popular software packages for processing human brain MRI scans.

Intensity non-uniformity correction

MR images can suffer from a phenomenon usually referred to as intensity non-uniformity, or bias field, characterized by a smooth intensity variation across the image (Vovk et al., 2007). These artifacts arise from imperfections in the image acquisition process, causing the intensity of the same tissue to vary depending on its location within the image, and therefore degrading the performance of automatic segmentation algorithms (Vovk et al., 2007). Several approaches exist to reduce or eliminate this intensity non-uniformity, among which the popular nonparametric nonuniform intensity normalization (N3) method (Sled et al., 1998), available in the FreeSurfer software package, or its variant N4ITK (Tustison et al., 2010), available in the Insight Toolkit (ITK)[7] (McCormick et al., 2014).

Intensity rescaling and standardization

There is no canonical scale for MR images intensities (except for quantitative MRI which are outside of our scope). Even considering images of the same body region from a particular patient, obtained with the same scanner within the same MRI protocol, absolute intensity values have no fixed numeric meaning, and only differences between intensities are relevant (Madabhushi and Udupa, 2005). This lack of standard for interpreting image intensities may compromise the quality of post-processing applications, therefore methods like histogram matching (Nyul et al., 2000) should be applied to transform images so that similar intensities have similar tissue meaning.

Skull stripping

MR brain images also contain extra-cranial or non-brain tissues from head scans, which pose a major obstacle for automatic segmentation algorithms (Kalavathi and Prasath, 2016). A preprocessing step, usually referred to as skull stripping, consists in isolating the brain from these non-brain tissues, a task for which several methods have been proposed (Kalavathi and Prasath, 2016). The approach used in the

[4] https://surfer.nmr.mgh.harvard.edu/
[5] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki
[6] https://www.fil.ion.ucl.ac.uk/spm/
[7] https://itk.org/

FreeSurfer preprocessing pipeline uses three groups of operations: intensity thresholding to obtain a preliminary mask, removal of false connections between brain and non-brain structures using graph cuts, and post-processing to smooth the final result (Sadananthan et al., 2010).

Image registration

MR brain image registration is a process that aligns multiple images to guarantee the spatial correspondence of brain regions across these images (Zhang et al., 2019). The adequate transformation (linear or non-linear) must be identified to ensure that corresponding image features are spatially aligned (Despotović et al., 2015). Linear registration may be rigid (6-parameter transformation corresponding to rotation and translation in three dimensions) or affine (12-parameter transformation corresponding to rotation, translation, scaling and skewing in three dimensions), and is commonly applied for registering images from the same subject (Despotović et al., 2015). To register images from different subjects, a non-linear algorithm is usually required (Despotović et al., 2015). Several methods have been developed for MRI registration (Oliveira and Tavares, 2014), among which the FSL tools FLIRT[8] (Greve and Fischl, 2009) and FNIRT[9] (Andersson et al., 2010), dedicated respectively to linear and non-linear registration. Frequently, all MR images from a given study are registered into a standard space such as the Montreal Neurological Institute (MNI space) (Evans et al., 1993), which enables the comparison of results across different studies (Despotović et al., 2015).

Feature extraction

After the preprocessing steps described above, different feature extraction pipelines may be implemented, according to the study design. For instance, volumetric segmentation pipelines may be applied to T1-weighted MR images to extract voxel-based anatomical features such as maps of tissue density (gray matter, white matter, and cerebrospinal fluid), and to compute the volumes of cortical and subcortical gray matter regions of interest (ROI) (Bertrand et al., 2018; Routier et al., 2021). Alternatively, cortical reconstruction pipelines may be used to obtain surface-based cortical thickness maps for each image (Popuri et al., 2018; Routier et al., 2021).

2.4 Integrative data analysis in neurodegenerative diseases

Neurodegenerative disorders such as frontotemporal dementia and amyotrophic lateral sclerosis are extremely complex processes. Data from single sources are unable to explain the intricate pathways involved in these diseases, therefore it is crucial to integrate data from different modalities, that provide complementary views,

[8] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT
[9] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FNIRT

to obtain more powerful models (Carreiro et al., 2015).

In this section, we begin by describing the strategies of data and model integration most commonly applied in multimodal data from neurodegenerative diseases (section 2.4.1). Next, we give an overview about the integration of neuroimaging and different types of omics data (section 2.4.2). Then, we focus on the emerging field of imaging transcriptomics, that integrate brain imaging and gene expression data (section 2.4.3). Finally, we narrow our attention to the few published studies that jointly analyzed miRNA and neuroimaging data (section 2.4.4).

2.4.1 Categories of data and model integration

Among the categories of data and model integration defined in (Azuaje, 2010), three are frequently used in multimodal studies dealing with neurodegenerative diseases: integration at the input level, heterogeneous data and model integration, and serial integration. These categories are described below, along with examples.

Integration at the input level

Data integration at the input level happens when features from different modalities are concatenated together into a single input vector, before being processed by the model (Azuaje, 2010). Synonyms for this method are early integration or feature-based integration (Behrad and Saniee Abadeh, 2022). This is the most frequently used data integration strategy, notably to build models using already established biomarkers (Carreiro et al., 2015). For instance, one study on Alzheimer's disease investigated multimodal features including cortical volumes from several brain regions extracted from structural MRI scans, concentrations of certain proteins measured in the cerebrospinal fluid, and results from cognitive and neuropsychological assessments (Cui et al., 2011). The authors performed feature selection to identify the optimal subset of features from each modality that independently achieved a high performance in classifying individuals with mild cognitive impairment from those with Alzheimer's disease. Then, they compared the results of support vector machine classifiers trained with single modalities (MRI, CSF or tests scores) and a model trained with input vectors concatenating all modalities. Their conclusion was that the simple concatenation of multimodal features at the input level was able to leverage complementary information and outperform all individual modalities.

Heterogeneous data and model integration

In this category, each modality is used to train a specific model, then the outputs of all models are fused together to give the final prediction (Azuaje, 2010). An ensemble of models built with data from different sources fits into this category (Carreiro et al., 2015). This method is also known as decision-level, model-based, or late integration (Behrad and Saniee Abadeh, 2022). One example of model within this category is the bi-level multi-source feature learning approach presented in (Xiang et al., 2014),

applied to features extracted from imaging scans (MRI, PET), CSF and plasma of Alzheimer's disease patients, mild cognitive impairment patients, and controls. In this study, the authors proposed to learn individual models for each data modality, then learn the appropriate weights for their combination, while extending their framework to handle missing data without the need for imputation. They explored the fact that, in medical applications, missing data is frequently block-wise, meaning that for each patient, a given modality is either available or non existent. They partitioned the dataset into non-disjoint subsets, according to the available data sources, in order to have complete data within each subset. This approach allowed the use of the whole dataset, even of samples with just one available data source, and the classification performance of the final model was greatly improved by including incomplete samples in the training set.

Serial integration

In a serial integration approach, the output of one model is fed as input to the next model, and so forth (Azuaje, 2010). This strategy may be used when subgroups of features or subjects are initially identified and then used by the subsequent models (Carreiro et al., 2015). For instance, one study on sporadic ALS consisted of a serial integration approach using first animal models, then human tissue, with the goal of identifying disease-relevant differentially expressed genes (Kudo et al., 2010). The authors performed a combination of microarray methods, initially identifying candidate genes in spinal cord motor neurons of transgenic mice, then evaluated the relevance of these genes as clinical biomarkers by testing their expression in the blood of mouse models, and finally provided a follow-up confirmation in human postmortem spinal cord tissues. Their serial integration approach to study a highly complex disease allowed the identification of 13 genes, encouraging their use as clinically useful biomarkers.

2.4.2 Integration of neuroimaging and omics data

The definition of *omics imaging*, according to (Antonelli et al., 2019), is "an emerging interdisciplinary field concerned with the integration of data collected from biomedical images and omics analyses". In this context, *omics* refers mainly to three popular fields, as illustrated by (Antonelli et al., 2019): *genomics* addresses the study of DNA sequences, often in comparison to a reference genome, focusing on the association between DNA mutations and observed phenotypes; *transcriptomics* studies the patterns of gene expression, either in the form of messenger RNAs or non-coding RNAs; and *proteomics* refers to the study of expressed proteins in different tissues.

Regarding neurological diseases, the vast majority of studies integrating imaging and omics data leverage data from Single Nucleotide Polymorphisms (SNPs) on the human genome, and thus fall into the category of *radiogenomics* (Kuo and Jamshidi,

2014) or *imaging genomics* (Thompson et al., 2010). In particular, the Alzheimer's Disease Neuroimaging Initiative (ADNI)[10] dataset contains multimodal data from MRI, PET, CSF biomarkers and SNPs from hundreds of participants, which has motivated several *imaging genomics* research initiatives. A comprehensive review of these studies is provided in (Antonelli et al., 2019).

However, nomenclature describing the landscape of imaging and omics data integration is not homogeneous. For example, one of the first studies combining neuroimaging and large-scale transcriptomic data was also framed in the field of *radiogenomics* by the authors, even though they did not examine DNA sequences and mutations (Zinn et al., 2011). Their research focused on correlating MRI features and gene expression patterns in patients with glioblastoma, using data obtained from The Cancer Genome Atlas (TCGA)[11], a publicly available database containing imaging and omics data from patients with several types of cancer.

Eventually, a more appropriate nomenclature for the combination of imaging and transcriptomic data to better understand disease mechanisms was proposed in (Katrib et al., 2016): *radiotranscriptomics*. In this review, the authors illustrate the potential applications of *radiotranscriptomics* in precision medicine, notably for increasing the knowledge about pathological processes in complex, multifactorial disorders such as cancer and neurodegeneration (Katrib et al., 2016).

More recently, the narrower field of *imaging transcriptomics* has emerged to study the correlation between neuroimaging features and regional variability in gene expression in brain tissues (Arnatkeviciute et al., 2019; Martins et al., 2021). This discipline is described in the next section.

2.4.3 Imaging transcriptomics

Recent advances in high-throughput sequencing technologies allowed to multiply the number of large-scale transcriptomic projects, leading to the creation of comprehensive datasets of gene expression across brain regions, many of them publicly available (Keil et al., 2018). For instance, one of the largest brain-wide gene expression atlases is the Allan Human Brain Atlas (AHBA), which contains expression levels from more than 20,000 genes observed in 3702 tissue samples covering nearly the entire brain (Arnatkeviciute et al., 2019; Sunkin et al., 2013).

Public databases such as the AHBA have motivated several studies integrating transcriptomic and neuroimaging data, aiming at understanding the spatial correlations of gene expression levels and neuroimaging-derived phenotypes (Martins et al., 2021). This emerging field of research, known as *imaging transcriptomics*, has already started to reveal how variations of gene expression in different parts of the brain relate to structural and functional characteristics, and vary during brain disease (Arnatkeviciute et al., 2019; Martins et al., 2021). Several examples of imaging

[10] https://adni.loni.usc.edu/
[11] https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

transcriptomics studies are described in (Arnatkeviciute et al., 2019), where the authors propose a standardized data processing pipeline to enable the comparison of results from different investigations.

It is important to note that the aforementioned studies focus on identifying correlations between transcriptomic and neuroimaging features, not to combine these features to improve prediction performance.

2.4.4 Integration of microRNA and neuroimaging data

In this section, we present the few published studies jointly analyzing miRNA and neuroimaging data.

One of the first works that simultaneously investigated miRNA expression and brain MRI focused on the association between these modalities, to better understand multiple sclerosis (MS) (Regev et al., 2017). The authors computed the Spearman correlation coefficients between the expression of serum miRNAs and measures of MS severity extracted from quantitative MRI (lesions and atrophy). Although several miRNAs were identified as associated with MRI measures, the associations were not significant after correcting for multiple comparisons using the false discovery rate (Regev et al., 2017). In a follow up work from the same research team, the authors defined four groups of patients with MS, based on MRI phenotypes characterizing lesion volume and brain atrophy, then identified a signature of 16 serum miRNAs differentially expressed among these groups. (Hemond et al., 2019).

Another study investigated the Pearson correlation between gray matter volumes (extracted from T1w MRI) and miRNA expression (measured in blood) in attention-deficit/hyperactivity disorder (ADHD) patients (Wang et al., 2020). The authors identified that gray matter volumes in some regions of the cingulate cortex were negatively correlated with the expressions of three miRNAs found in blood, and thus concluded that miRNA dysregulation likely participates in the pathophysiology of ADHD.

Similarly, the expression of certain miRNAs in serum was found to be significantly correlated (Pearson's correlation test) with brainstem volume in patients with Wolfram syndrome (Zmyslowska et al., 2020). The findings of this study demonstrated that a miRNA signature could be used as a non-invasive and easily accessible indicator of neurodegeneration in subjects suffering from this rare syndrome.

Another example of association study aiming at replacing MRI scans with blood tests was focused on cerebral malaria, an acute and often fatal neurological complication caused by severe malaria (Gupta et al., 2021). Since MRI scanners are rarely accessible in some malaria-endemic countries, the authors investigated if plasma miRNAs could be used, instead of expensive neuroimaging, to assess the progression of cerebral malaria. They used Spearman's correlation test to assess the correlation between MRI features (brain volume and apparent diffusion coefficient) and miRNA expression in plasma samples from patients with cerebral malaria. This

study identified miR-3158-3p as a promising biomarker for measuring cerebral malaria progression.

Recently, PET-derived features and miRNA expression levels in blood were jointly analyzed in cognitively normal individuals with subjective memory complaints, who have an increased risk to develop Alzheimer's disease (Vergallo et al., 2021). The authors initially selected brain-enriched miRNAs in a pilot study, then investigated their longitudinal association with brain metabolic uptake computed in 12 cortical regions of interest using FDG-PET. To this aim, they fitted linear mixed models, then tested the interaction effect between miRNA expression and time on metabolic uptake. The results of this study, which reported significant associations between miRNA expression levels in plasma and neuroimaging biomarkers, supported the role of miRNAs as candidate biomarkers of AD.

Remarkably, none of the previously described works leveraged the joint analysis of miRNA and neuroimaging data to enhance prediction performance: all of them were based on association studies of these modalities. To the best of our knowledge, the only published study that integrated miRNA data and MR imaging features to improve prediction is (Gallivanone et al., 2019), which did not focus on brain imaging, but on breast cancer differential diagnosis. The authors trained support vector machine classifiers integrating miRNA expression and imaging features at the input level, with the goal of distinguishing different breast cancer subtypes. They showed that the combination of miRNA and MRI data resulted in better classification performance than using any of these modalities alone.

2.5 Disease progression modeling and disease progression scores

To introduce the concepts of disease progression modeling and disease progression scores (DPS), we will turn to one of the most studied neurodegenerative disorders: Alzheimer's disease.

In a seminal work, Jack and colleagues proposed a hypothetical model of the temporal evolution of AD biomarkers, and their relation to disease staging (Jack et al., 2010). The authors explained that certain AD biomarkers begin to vary decades before any visible clinical symptoms, and that dementia is the end result of the accumulation of several pathological changes. Moreover, they showed evidence suggesting that biomarkers become abnormal in an ordered manner, and that understanding this time-dependent order is crucial for estimating disease stage.

Figure 2.5 depicts the hypothetical model presented in (Jack et al., 2010), which followed two main assumptions: (1) as the disease progresses, more biomarkers become abnormal, and (2) rates of change vary in a non-linear fashion, hypothesised to be in a sigmoid shape. Notably, no single biomarker can cover the full spectrum of AD progression, so a combination of multiple biomarkers is essential to estimate disease stage for a particular individual (Jack et al., 2010).

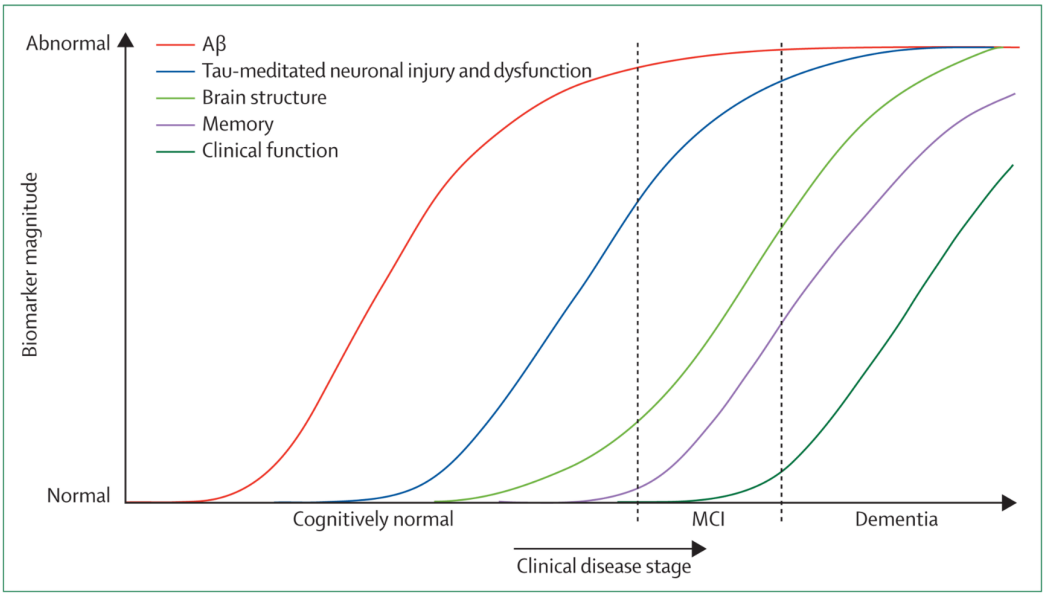


FIGURE 2.5: Model of the pathological cascade of AD biomarkers, for a single prototypical individual (reproduced from (Jack et al., 2010), with permission from the publisher). The horizontal axis represents progressive cognitive impairment, while the vertical axis indicates the level of abnormality of each biomarker. Aβ: amyloid beta; MCI: mild cognitive impairment.

A few years later, an update on this model supported its main assumptions, while changing the definition of the horizontal axis (Jack et al., 2013). Indeed, in empirical testing, the authors concluded that it was problematic to index subjects on the horizontal axis by a continuous measure of cognitive impairment. The main reason was that measures of cognitive decline are imprecise during the preclinical phase, which usually corresponds to more than half of the total disease duration. Jack and colleagues then proposed that the horizontal axis should represent the "distance traveled along the AD pathophysiological pathway". They explained that, ideally, this axis should be constructed using a large longitudinal dataset containing all relevant biomarkers, measured for decades at multiple time points, in several subjects. Since it is very difficult to build a dataset of this magnitude, the authors argued that the model should instead rely on piece-wise data from several subjects in different stages of the disease.

In order of increasing power, disease progression models could be built using purely cross-sectional data, cross-sectional imaging data with longitudinal follow-up of non-imaging data, or short-term longitudinal data contemplating all modalities (Jack et al., 2013). In particular, Jack and colleagues mentioned that multiple biomarkers could be non-linearly combined to produce a single horizontal-axis metric: a latent trait to represent the entire disease spectrum.

Although hypothetical, (Jack et al., 2013, 2010) have served as a foundation for future data-driven models for neurodegenerative diseases progression. Since then, several disease progression modeling approaches have been proposed, most of them

depending on large amounts of longitudinal data. Some examples include fitting logistic functions to biomarker trajectories (Jedynak et al., 2012; Mehdipour Ghazi et al., 2021), non-linear mixed-effects models (Koval et al., 2021; Schiratti et al., 2015, 2017), recurrent neural networks (Mehdipour Ghazi et al., 2019), and Gaussian processes (Lorenzi et al., 2019b). To the best of our knowledge, the only published disease progression models that work with cross-sectional data are event-based models (Fonteijn et al., 2012; Venkatraghavan et al., 2019). In the next sections, we briefly describe all these approaches.

2.5.1 Logistic functions and biomarker trajectories

One of the first studies that proposed a method for estimating a disease progression score (DPS) combining multiple biomarkers was (Jedynak et al., 2012). The motivation behind this work was to estimate an accurate measure of Alzheimer's disease progression in the early stages of the disease, when symptoms are not yet present and there is a higher chance of a successful therapeutic intervention. The authors developed a generic computational approach, and evaluated their models using the largest freely available longitudinal dataset of biomarkers of a neurodegenerative disease: the ADNI dataset.

In a nutshell, the main contribution of (Jedynak et al., 2012) is a method to compute a common temporal scale representing the disease's underlying biological and cognitive changes, and to assign to each subject, at each time point, a DPS within this scale. Their method relies on three main assumptions: (1) individuals follow a common disease progression, possibly with different rates of progression; (2) the evolution of all biomarkers is sigmoid-shaped; and (3) the rate of disease progression for a given individual is constant. Within this framework, the DPS of a given subject i is obtained by linearly transforming age:

$$DPS_i = \alpha_i \times age + \beta_i \tag{2.1}$$

where α_i and β_i are coefficients to be fitted for each subject. Regarding the biomarkers, each one of them is considered to be a sigmoidal function of the DPS, with the same dynamic across the population. The parameters defining each biomarker's sigmoid function (four parameters per function) and the parameters defining the linear transformations for each subject's DPS are fitted using alternating least squares. Details of the fitting algorithm are available in (Jedynak et al., 2012).

To evaluate their method, Jedynak and colleagues computed the DPS for 687 subjects from the ADNI database, using data collected from seven biomarkers measured over 3658 visits in total (average of 5.3 visits per subject). Figure 2.6 displays the normalized sigmoid functions for each of the seven biomarkers, plotted as a function of the DPS, as well as the DPS probability densities conditioned on the clinical status of each individual. Notably, this figure demonstrates that the DPS correlates with the

clinical status, since Alzheimer's patients tend to have a higher DPS than individuals with mild cognitive impairment, who tend to have a higher DPS than healthy controls. Moreover, these results support the hypothetical plot from (Jack et al., 2010) depicted in Figure 2.5.

The authors concluded that their data-driven approach, which creates a composite biomarker, could be used to stage patients and presymptomatic individuals of other neurodegenerative diseases, as long as enough longitudinal data is available (Jedynak et al., 2012).

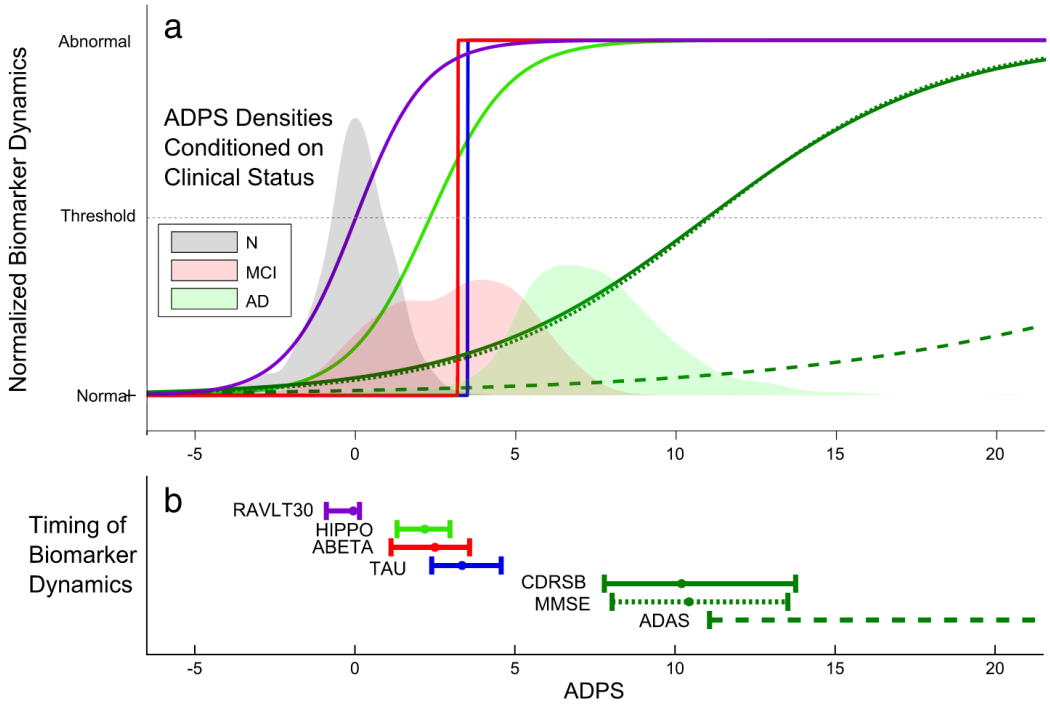


FIGURE 2.6: (a) Estimated biomarker dynamics of the ADNI population as a function of DPS, and conditional probability densities of the DPS given the clinical status of each individual. (b) Inflection point for each curve, along with 90% confidence intervals (reproduced from (Jedynak et al., 2012), with permission from the publisher). ADPS: Alzheimer's disease progression score; N: normal subjects; MCI: mild cognitive impairment; AD: Alzheimer's disease; RAVLT30: Rey auditory verbal learning test, 30 minute recall; HIPPO: sum of the two lateral hippocampal volumes; ABETA: Amyloid beta; TAU: tau protein; CDRSB: Clinical dementia rating sum of boxes score; MMSE: Mini-mental state examination score; ADAS: Alzheimer's disease assessment scale-cognitive subscale.

More recently, an extension to the aforementioned approach was proposed, with the main goal to address outliers and obtain more robust estimations (Mehdipour Ghazi et al., 2021). Mehdipour and colleagues argued that the use of a new logistic function (modified Stannard), along with M-estimation to fit the parameters of the logistic functions to the dynamics of each biomarker, would better fit the biomarker trajectories and reduce the influence of outliers. Additionally, the authors presented an end-to-end approach that estimates the trajectory of each biomarker,

orders events based on inflection points, and classifies the clinical status of each subject using the computed DPS.

Tests with 16 ADNI biomarkers were conducted, using data from 1518 subjects collected in a total of 9098 visits (on average, 6 visits per subject) (Mehdipour Ghazi et al., 2021). The diagnostic performance of the models was assessed by computing the multiclass area under the receiver operating characteristic curve (ROC-AUC) when classifying the clinical status of individuals from the test set with a Bayesian classifier, based exclusively on the DPS. Results showed state-of-the-art performance (average ROC-AUC of 0.937 on the Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE)[12] challenge). The authors also showed the generalizability of their approach by training with ADNI data and testing with an independent test set, for which a ROC-AUC of 0.929 was obtained.

2.5.2 Non-linear mixed-effects models

A generic Bayesian mixed-effects model to estimate progression of biological phenomena from longitudinal observations was presented in (Schiratti et al., 2015) and further extended in (Schiratti et al., 2017). Schiratti and colleagues argued that the temporal evolution of biomarkers could be modeled as *spatiotemporal* trajectories, since trajectories vary across individuals for two reasons: (1) intrinsic phenotypic differences (*spatial* variability), and (2) different ages of onset and paces of progression (*temporal* variability). Here, the term *spatial* is to be taken in a broad sense, since it does not refer to physical space within the brain, but to a more general abstract space of biomarkers. Thus, this framework is generic and has been applied to various types of data, from cognitive scores to brain images.

The authors proposed a data-driven approach to estimate the typical trajectory of biomarker changes within a population, as well as the *spatiotemporal* variability across individuals, using longitudinal datasets (Schiratti et al., 2017). They extended nonlinear mixed-effects models to deal with longitudinal observations lying on *Riemannian manifolds*, in order to decompose spatial and temporal variability. The fixed effects captured the typical group-average trajectory, while the random effects accounted for subject-specific variability of the trajectories (Schiratti et al., 2017).

This generic framework was later applied to build a spatiotemporal atlas of Alzheimer's disease progression, called AD Course Map (Koval et al., 2021). Koval and colleagues leveraged longitudinal data from ADNI to build typical progression curves for each studied biomarker, where the x-axis representing disease stage was named Alzheimer's Age (AA). The model's core idea was to compute three parameters per subject, in order to consider individual variability: (1) a time-shift translating the curves along the x-axis to account for different ages and disease onsets; (2)

[12] (https://tadpole.grand-challenge.org)

an acceleration factor scaling the x-axis to reflect different paces of disease progression; and (3) space-shifts for each biomarker, to change their ordering and account for different phenotypes (Koval et al., 2021).

AD Course Map was evaluated using data from the TADPOLE challenge, and was able to predict cognitive decline with a better accuracy than other 56 competing approaches, while achieving the same prediction error as the best performing method in the prediction of ventricular volume (Koval et al., 2021).

2.5.3 Recurrent neural networks

A nonparametric approach to disease progression modeling, based on recurrent neural networks (RNN), was presented in (Mehdipour Ghazi et al., 2019). Mehdipour Ghazi and colleagues argued that standard RNNs cannot be directly used in most longitudinal cohorts, because of frequent missing values, so they proposed an alternative formulation of backpropagation to handle incomplete data. In short, to minimize the influence of missing values, the authors modified the batch gradient descent algorithm used to train long short-term memory (LSTM) networks, introducing a weighted update rule depending on the number of available time points per biomarker.

Contrary to (Jedynak et al., 2012; Mehdipour Ghazi et al., 2021), the goal of this approach was not to compute disease progression scores, but instead to focus on predicting future biomarker values (Mehdipour Ghazi et al., 2019). The authors evaluated their end-to-end method by modeling the longitudinal dynamics of six imaging biomarkers from the ADNI dataset, collected from 742 individuals that had at least three visits. They computed the mean absolute values between estimated and actual biomarker values using their proposed approach, two standard LSTM networks with different data imputation techniques, and the parametric regression-based model from (Jedynak et al., 2012). Moreover, tests were conducted for different amounts of missing values. Results demonstrated the superiority of the proposed robust RNN to predict biomarker values, in comparison to all other tested models, up until 74% of the data were missing (Mehdipour Ghazi et al., 2019).

Finally, the authors concluded that using RNNs to model disease progression imposes fewer assumptions on the data and thus brings more flexibility, for instance supporting biomarkers with a non-monotonic behaviour (Mehdipour Ghazi et al., 2019).

2.5.4 Gaussian processes

A different modeling paradigm was proposed in (Lorenzi et al., 2019b), with the main goal of quantifying uncertainty in the estimation of biomarker trajectories and diagnostic predictions. The authors reformulated disease progression modeling within a probabilistic setting, based on Gaussian process regression, and assessed uncertainty with respect to missing observations, biomarkers, and follow-up visits.

To evaluate the proposed framework, the authors used longitudinal training data from 15 biomarkers measured in 200 subjects from ADNI, while the test set contained information from 582 subjects. The estimated AD biomarker progression was compatible with the hypothetical model from (Jack et al., 2010), and classification between controls and AD patients achieved an area under the ROC curve of 0.99.

The authors also demonstrated their framework's ability to quantify the uncertainty of single biomarkers in characterizing disease progression, thus allowing to compare the usefulness of different biomarkers (Lorenzi et al., 2019a). Finally, their findings also confirmed the intuition that the greater the amount of longitudinal information used to train the model, the lower the uncertainty in predicting the diagnosis of a given test subject.

2.5.5 Event-based models

Even though analyzing longitudinal data is in principle the best way to assess disease progression, building disease progression models from longitudinal data can be challenging. With the exception of common disorders such as AD and Parkinson's disease, large longitudinal datasets of neurodegenerative disorders are rare (Venkatraghavan et al., 2019). To overcome this limitation, Fonteijn and colleagues proposed event-based models (EBM), which describe disease progression as a sequence of discrete events that can be estimated from cross-sectional data (Fonteijn et al., 2012). In this context, an event may be defined as the appearance of a symptom (Fonteijn et al., 2012), or more generally as a biomarker transition from a normal to an abnormal value (Venkatraghavan et al., 2019).

Figure 2.7 illustrates hypothetical biomarker trajectories throughout a neurodegenerative disease progression, with each biomarker sequentially transitioning from a normal to an abnormal state. The inspiration from Jack's hypothetical model (Jack et al., 2010) is noteworthy. An event-based model aims at computing the sequence of biomarker transitions that maximizes the likelihood of the data given that sequence (Fonteijn et al., 2012).

The two main assumptions of an EBM are (1) biomarker values change monotonically throughout disease progression, and (2) all subjects follow the same ordering of events (Fonteijn et al., 2012). Without these assumptions, it would not be possible to infer the sequence of events in a particular cohort only from cross-sectional data (Fonteijn et al., 2012).

To put it formally, EBM consist of a set of N events $(E_1, \ldots, E_N)$ and a sequence $S = (s(1), \ldots, s(N))$, which determines the event ordering $(E_{s(1)}, \ldots, E_{s(N)})$, estimated from the cross-sectional dataset X containing biomarker values X_j for each subject $j = (1, \ldots, J)$ (Fonteijn et al., 2012). The model is fitted by identifying the sequence that maximizes $p(S|X)$, which can be written using Bayes' theorem:

$$p(S|X) = \frac{p(S)p(X|S)}{p(X)} \tag{2.2}$$

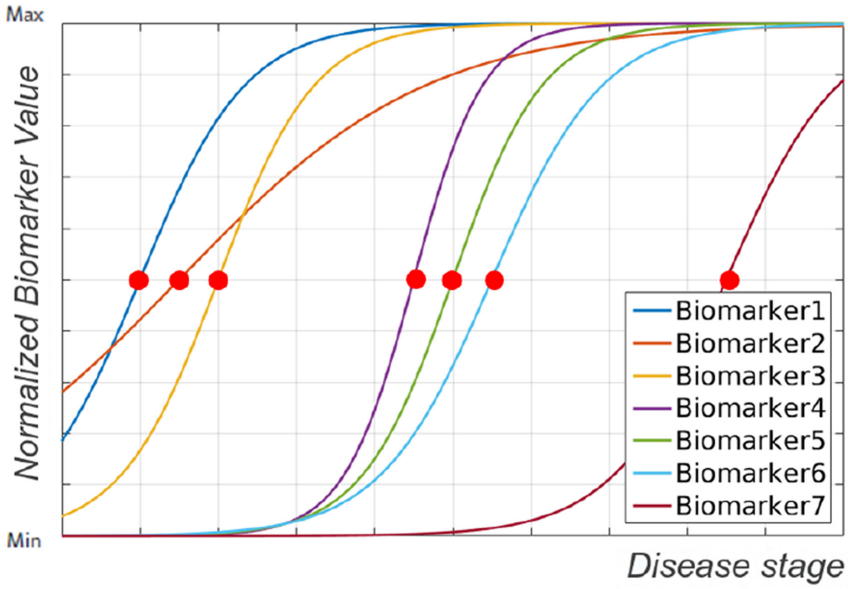


FIGURE 2.7: Hypothetical biomarker trajectories from a normal to an abnormal state (reproduced from (Venkatraghavan et al., 2019), CC BY-NC-ND). The red dots represent events, as defined in the EBM context.

Since the authors assume that all sequences S are equally likely, the problem of inferring S is equivalent to the maximum likelihood problem of maximizing $p(X|S)$. However, Fonteijn and colleagues instead use a Markov Chain Monte Carlo (MCMC) algorithm to obtain 1,000,000 samples from the posterior $p(S|X)$, and then compute the average position of events $\bar{S}$ in the samples (Fonteijn et al., 2012). Uncertainty in the ordering may be visualized in a positional variance diagram, as illustrated in Figure 2.8.

Once the event-based model is fitted, the estimated ordering of events can be used to infer a disease stage for a new subject, by finding the stage that maximizes the subject's data likelihood, given the estimated sequence of events (Fonteijn et al., 2012). Concretely, the estimated disease stage k for subject j is the one that maximizes $p(X_j|\bar{S}, k)$ (Fonteijn et al., 2012). This approach results in a discrete set of stages, with $N + 1$ possible values (corresponding to 0 to N abnormal biomarkers).

Recently, a new EBM framework called discriminative event-based model (DEBM) was proposed in order to scale the algorithm to a larger number of biomarkers, as well as to relax the assumption that all subjects in a dataset follow the same sequence of events (Venkatraghavan et al., 2019). The authors also improved disease staging by proposing the concept of relative distance between events.

In summary, the DEBM framework has three main steps: (1) Bayesian classifiers are independently trained for each biomarker to differentiate between normal and abnormal values; (2) orderings of biomarker abnormalities are estimated initially for each subject, then a central ordering is computed as the mean of subject-specific estimates by minimizing the probabilistic Kendall's Tau distances between the central ordering and the individual orderings; and (3) a patient staging algorithm uses

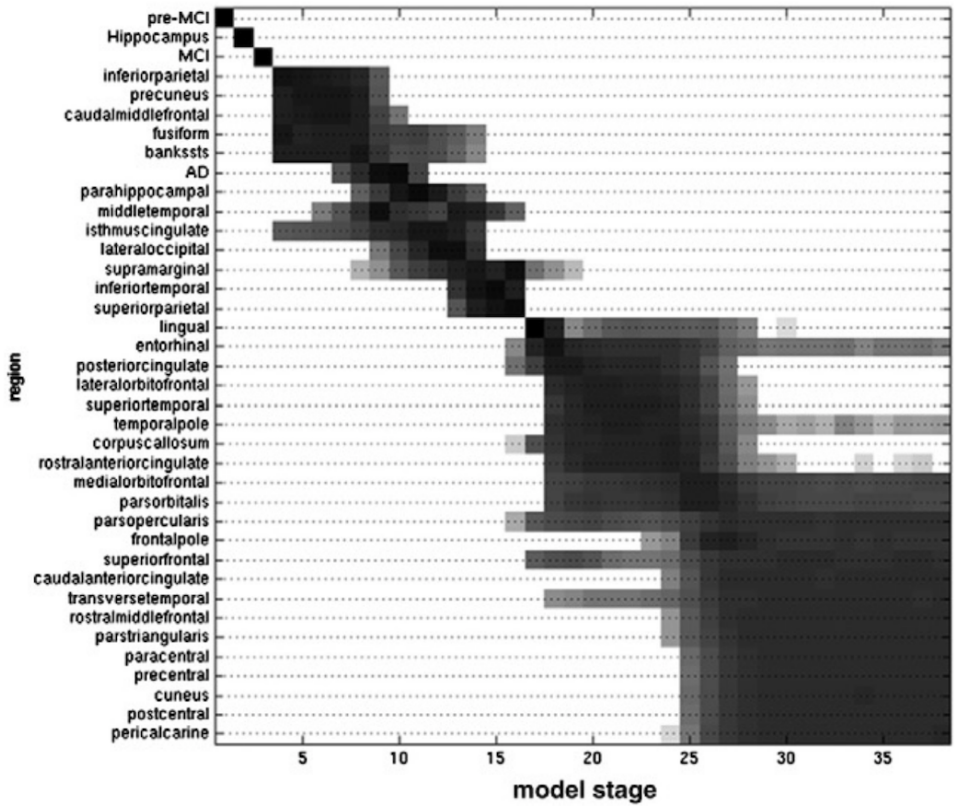


FIGURE 2.8: Event sequence computed by an EBM using regional atrophy values from T1-weighted MRI scans from a familial AD cohort (reproduced from (Fonteijn et al., 2012), with permission from the publisher). The regions are ordered according to the average position of events $\bar{S}$. Positional variance is depicted according to the number of times a region occupies each position over the samples obtained with MCMC. MCI: mild cognitive impairment; AD: Alzheimer's disease.

the central ordering of events and the relative distances between event-centers to compute continuous patient stages (Venkatraghavan et al., 2019).

To evaluate their framework, Venkatraghavan and colleagues used 88 features measured in 1737 subjects from the ADNI dataset at baseline, comprising gray matter volumes of 83 regions of interest, values from three CSF biomarkers, and two cognitive scores (Venkatraghavan et al., 2019). Since there was no ground-truth data for the event ordering, the authors relied on the classification performance of controls versus AD patients, using only the computed disease stages, as an indirect evaluation of the estimated event ordering.

Results demonstrated that DEBM distinguishes controls from patients consistently better than previously published EBM models (Venkatraghavan et al., 2019). Additionally, the authors were able to better classify MCI converters from non-converters, illustrating their framework's ability to identify individuals with a high risk of developing AD. Finally, experiments with simulated data supported that the continuous patient staging mechanism was more accurate than the traditional discrete staging approach from (Fonteijn et al., 2012).

Variations of event-based models have been successfully applied to model the progression of several neurodegenerative diseases, such as AD (Archetti et al., 2019; Firth et al., 2020; Fonteijn et al., 2012; Oxtoby et al., 2018; Venkatraghavan et al., 2019; Young et al., 2014), multiple sclerosis (Dekker et al., 2020; Eshaghi et al., 2018), Parkinson's disease (Oxtoby et al., 2021), Huntington's disease (Wijeratne et al., 2021), FTD (Ende et al., 2021; Panman et al., 2021), and ALS (Gabel et al., 2020). These are currently the state-of-the-art approaches to model disease progression from cross-sectional data.

Chapter 3

Datasets

In this chapter, we describe the datasets used in each study conducted during this PhD project. We present the three microRNA datasets in Section 3.1 and the neuroimaging dataset in Section 3.2.

3.1 MicroRNA datasets

The first miRNA dataset presented below (section 3.1.1) was used in the miRNA discovery study detailed in Chapter 4, and to assess the estimation of disease progression scores from multimodal neuroimaging and miRNA expression data in Chapter 6. The other two miRNA datasets (described in section 3.1.2 and section 3.1.3) were used in the comprehensive validation study in Chapter 5.

3.1.1 *C9orf72* carriers from PREV-DEMALS

The Predict to Prevent Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis (PREV-DEMALS)[1] project is a French multicentric, observational study focused on *C9orf72* expansion carriers and their first-degree relatives, who have a 50% chance to carry the mutation. The main objective of the PREV-DEMALS study is to identify biomarkers that allow an early diagnosis of *C9orf72*-associated disease, and to follow disease progression.

Subjects from this cohort were recruited and evaluated between 2015 and 2017. Participants underwent T1-weighted MRI, diffusion MRI, functional MRI, and FDG-PET scans, as well as blood sampling and neuropsychological tests. Not all subjects underwent all assessments, and in this thesis we used only features extracted from T1-weighted MRI and blood plasma miRNA expression.

Blood samples were stored in the Paris Brain Institute DNA and cell bank. MicroRNAs were extracted and sequenced in December 2019. The resulting dataset consisted of normalized log2 expression levels of **2576** miRNAs (corresponding to all miRNAs listed in mirBase[2] version 22) from **110** individuals divided into three groups:

- Control group: **43** non-carriers of the *C9orf72* expansion.

[1] https://clinicaltrials.gov/ct2/show/NCT02590276
[2] https://www.mirbase.org

- Presymptomatic group: **45** *C9orf72* expansion carriers without neurological symptoms.
- Patient group: **22** *C9orf72* expansion carriers with a diagnosis of FTD, ALS, or both.

3.1.2 *C9orf72* carriers from clinical practice

In addition to the subjects studied in the PREV-DEMALS cohort, *C9orf72* expansion carriers and controls were recruited through the French research network on FTD/ALS (Inserm RBM02-59) between 2011 and 2021.

These individuals underwent blood sampling, from which plasma miRNAs were extracted and sequenced in January 2022. The resulting dataset contained normalized log2 expression levels of **2656** miRNAs (corresponding to all miRNAs listed in mirBase version 22.1) from **77** subjects divided into three groups:

- Control group: **31** non-carriers (this group is shared with the *GRN* cohort from Predict-PGRN described in section 3.1.3).
- Presymptomatic group: **17** *C9orf72* expansion carriers without neurological symptoms.
- Patient group: **29** *C9orf72* expansion carriers with a diagnosis of FTD, ALS, or both.

3.1.3 *GRN* carriers from Predict-PGRN

The Natural History Characterization in Symptomatic and Asymptomatic Progranuline Gene Mutation Carriers (Predict-PGRN)[3] study is focused on the *GRN* mutation. The main goal of the Predict-PGRN study is to investigate whether biomarkers in presymptomatic *GRN* mutation carriers may be used for early diagnosis of *GRN*-associated FTD and to assess disease progression.

Individuals from this cohort were recruited and evaluated between 2011 and 2021, and underwent the same assessments as individuals from the PREV-DEMALS cohort.

From the Predict-PGRN cohort, we used in this thesis only miRNA expression data, extracted from blood plasma. Samples were stored in the Paris Brain Institute DNA and cell bank, and miRNA sequencing was performed in January 2022. The resulting dataset consisted of normalized log2 expression levels of **2656** miRNAs (corresponding to all miRNAs listed in mirBase version 22.1) from **89** subjects divided into three groups:

- Control group: **31** non-carriers (this group is shared with the *C9orf72* cohort from clinical practice described in section 3.1.2).

[3] https://clinicaltrials.gov/ct2/show/NCT04014673

- Presymptomatic group: **30** *GRN* mutation carriers without neurological symptoms.
- Patient group: **28** *GRN* mutation carriers with a diagnosis of FTD.

3.2 Neuroimaging dataset

The neuroimaging dataset described below (section 3.2.1) was used to evaluate the model presented in Chapter 6, which estimates disease progression scores from multimodal neuroimaging and miRNA expression data.

3.2.1 *C9orf72* carriers from PREV-DEMALS

Neuroimaging features consisted of gray matter volumes extracted from T1-weighted MR images, including cortical structures (68 regions of interest), subcortical structures (18 regions of interest) and the estimated total intracranial volume (TIV), thus resulting in **87** neuroimaging features. Of the 110 individuals recruited in the PREV-DEMALS cohort, only **91** underwent T1-weighted MRI scans. The composition of each group is as follows:

- Control group: **37** non-carriers of the *C9orf72* expansion.
- Presymptomatic group: **40** *C9orf72* expansion carriers without neurological symptoms.
- Patient group: **14** *C9orf72* expansion carriers with a diagnosis of FTD, ALS, or both.

Neuroimaging features were extracted as described in (Bertrand et al., 2018), notably using the Desikan atlas (Desikan et al., 2006) for the cortical ROIs and the Aseg nomenclature (Fischl et al., 2002) for the subcortical ROIs.

Chapter 4

Plasma microRNA signature in presymptomatic and symptomatic subjects with *C9orf72*-associated frontotemporal dementia and amyotrophic lateral sclerosis

This chapter has been published as an original research article in the *Journal of Neurology, Neurosurgery & Psychiatry* (Kmetzsch et al., 2021):

- **Kmetzsch, V.**, Anquetil, V., Saracino, D., Rinaldi, D., Camuzat, A., Gareau, T., Jornea, L., Forlani, S., Couratier, P., Wallon, D., Pasquier, F., Robil, N., de la Grange, P., Moszer, I., Le Ber, I., Colliot, O., Becker, E., PREV-DEMALS study group, "Plasma microRNA signature in presymptomatic and symptomatic subjects with C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis", *Journal of Neurology, Neurosurgery & Psychiatry*, 92(5):485-493 (2021). doi: 10.1136/jnnp-2020-324647 – hal-03046771

4.1 Abstract

Objective. To identify potential biomarkers of preclinical and clinical progression in *C9orf72*-associated disease by assessing the expression levels of plasma microRNAs (miRNAs) in *C9orf72* patients and presymptomatic carriers.

Methods. The PREV-DEMALS study is a prospective study including 22 *C9orf72* patients, 45 presymptomatic *C9orf72* mutation carriers and 43 controls. We assessed the expression levels of 2576 miRNAs, among which 589 were above noise level, in plasma samples of all participants using RNA sequencing (RNA-seq). The expression levels of the differentially expressed miRNAs between patients, presymptomatic carriers and controls were further used to build logistic regression classifiers.

Results. Four miRNAs were differentially expressed between patients and controls: miR-34a-5p and miR-345-5p were overexpressed, while miR-200c-3p and miR-10a-3p were underexpressed in patients. MiR-34a-5p was also overexpressed in presymptomatic carriers compared with healthy controls, suggesting that miR-34a-5p expression is deregulated in cases with *C9orf72* mutation. Moreover, miR-345-5p was also overexpressed in patients compared with presymptomatic carriers, which supports the correlation of miR-345-5p expression with the progression of *C9orf72*-associated disease. Together, miR-200c-3p and miR-10a-3p underexpression might be associated with full-blown disease. Four presymptomatic subjects in transitional / prodromal stage, close to the disease conversion, exhibited a stronger similarity with the expression levels of patients.

Conclusions. We identified a signature of four miRNAs differentially expressed in plasma between clinical conditions that have potential to represent progression biomarkers for *C9orf72*-associated FTD and ALS. This study suggests that dysregulation of miRNAs is dynamically altered throughout neurodegenerative diseases progression, and can be detectable even long before clinical onset.

4.2 Introduction

Frontotemporal dementia (FTD) designates neurodegenerative dementias characterized by progressive behavioral, executive and language impairments (Rascovsky et al., 2011). Amyotrophic lateral sclerosis (ALS) is a degenerative disease of motor neurons that leads to progressive muscle atrophy and motor deficit. FTD and ALS form a clinical continuum, as these two diseases may be associated in the same patients (FTD-ALS) or within families. They also share common pathophysiological mechanisms and genetic causes (Mackenzie et al., 2010). The most frequent genetic cause of familial FTD and ALS is a hexanucleotide (GGGGCC) repeat expansion in the chromosome 9 open reading frame 72 (*C9orf72*) gene (DeJesus Hernandez et al., 2011; Renton et al., 2011). This autosomal dominant mutation may cause neurodegeneration through *C9orf72* loss of function, aggregation of mutant RNA in nuclear foci and of dipeptide repeats generated by repeat-associated non-AUG (RAN) translation, ultimately leading to pathological inclusions of TAR-DNA binding protein 43 (TDP-43) (Mackenzie et al., 2014).

There are no effective treatments available in *C9orf72* disease to date, but several promising trials including antisense therapies are being developed. Presymptomatic *C9orf72* carriers represent an optimal target population for the development of new therapeutic interventions for FTD and ALS (Bertrand et al., 2018; Eisen et al., 2014). Therefore, it is of paramount importance to identify biomarkers of preclinical progression for FTD and ALS, which could be used to initiate and monitor potential disease-modifying treatments before any irreversible brain damage has occurred.

There is increasing evidence that microRNA (miRNA) expression in body fluids, such as plasma/serum (Grasso et al., 2014) or CSF (Schneider et al., 2018), correlates

with the diagnosis and progression of many neurodegenerative diseases, including FTD (Denk et al., 2018) and ALS (Freischmidt et al., 2014). MicroRNAs are a class of small noncoding RNAs that negatively regulate gene expression by promoting translational repression and messenger RNA degradation (Huntzinger and Izaurralde, 2011). Since TDP-43 promotes miRNA biogenesis (Buratti and Baralle, 2010), the dysregulation of TDP-43 activity associated with FTD and ALS pathogenesis could impact miRNA expression levels (Gascon and Gao, 2014). Notably, miRNAs originating from neurons and glial cells are released through extracellular vesicles, especially exosomes, and can be measured in different body fluids, including CSF and plasma (Li and Wang, 2019). Aberrant expression of miRNAs can be thus non-invasively detected in easily accessible body compartments, such as blood plasma, and potentially serve as biomarkers (Sohel, 2016).

Previous studies have explored selected plasma miRNAs as biomarkers for FTD and ALS (Sheinerman et al., 2017) or FTD (Grasso et al., 2019; Piscopo et al., 2018) using quantitative real-time PCR. Two of them have analyzed the expression of a limited number of candidate miRNAs: nine miRNAs linked with apoptosis (Piscopo et al., 2018) or 37 brain-enriched miRNAs (Sheinerman et al., 2017). A wider miRNA profiling study (Grasso et al., 2019) analyzed 752 miRNAs, as a first attempt to perform an unbiased assessment of circulating miRNAs in patients with FTD. In addition, a more recent study (Magen et al., 2020) assessed the expression levels of 2313 miRNAs in a merged cohort of FTD patients with different genetic forms (*C9orf72*, MAPT, GRN, TBK1) or with sporadic forms, by next generation RNA sequencing (RNA-seq). However, results among different studies have been conflicting so far, probably due to the heterogeneity of cohorts with respect to the underlying pathology (genetic or sporadic). Besides, these studies only compared healthy controls and symptomatic patients, focusing on evaluating potential diagnostic biomarkers. To date, no studies have evaluated plasma miRNAs as progression biomarkers for FTD or ALS in presymptomatic individuals.

The present work aims at investigating expression levels of plasma miRNAs in a large homogeneous genetic cohort of *C9orf72* mutation carriers, both in the presymptomatic and in the clinical phases, to identify potential non-invasive biomarkers of preclinical and clinical progression in *C9orf72*-associated FTD and ALS. We hypothesize that performing large scale RNA-seq analyses in plasma samples, without *a priori* assumptions, will reveal significant differences in miRNA expression levels between healthy controls, presymptomatic and symptomatic mutation carriers.

4.3 Material and Methods

4.3.1 Participants

PREV-DEMALS (https://clinicaltrials.gov/ Identifier: NCT02590276) is a national multicentric study focused on *C9orf72* mutation carriers. Between 2015 and 2017,

111 individuals were investigated with the same protocol in four French university hospitals (Paris, Limoges, Lille, Rouen), as previously described (Bertrand et al., 2018; Montembeault et al., 2020). Written informed consents were obtained from all participants, after approval obtained from the Comité de Protection des Personnes CPP Ile-De-France VI (CPP 68-15 and ID RCB 2015-A00856-43).

This cohort included 22 patients (15 FTD, 4 FTD/ALS, 3 ALS) carrying a *C9orf72* expansion and 89 asymptomatic first-degree relatives of *C9orf72* patients (who have 50%-risk to carry the mutation), out of 64 families. A pathogenic expansion was detected in 46 of them, denoted as the *presymptomatic group*. The control group was formed by the 43 asymptomatic individuals that did not carry an expansion.

At inclusion, each participant's cognitive and behavioral clinical status was assessed based on standardized interview with relatives, comprehensive neurological examination, an extensive neuropsychological battery assessing all cognitive domains (including, notably, MMSE, FAB, MDRS, Ekman faces tests) and behavioral scales (including Frontal Behavioral Inventory and Apathy Evaluation scale) (Table 4.1). The cognitive and behavioral evaluations and their scores have been described in more detail elsewhere (Bertrand et al., 2018; Montembeault et al., 2020) and in Appendix A Method A1. Neuromuscular function was thoroughly evaluated by means of quantitative motor testing according to Medical Research Council (MRC) muscle scale, assessment of upper and lower motor neuron signs, and administration of ALS-FRS (ALS functional rating scale), evaluating the degree of functional impairment. All participants underwent a systematic standardized interview to investigate the presence of cramps, fatigue, muscle pain, muscle weakness, muscle stiffness or fasciculations. Electromyography was proposed to the participants with even subtle motor signs or complaints.

One participant was excluded because mild cerebellar syndrome was detected at a neurological examination, after inclusion. Thus, the present study comprises 110 individuals (22 patients, 45 presymptomatic carriers, 43 healthy controls), all of which underwent plasma sampling at their inclusion. The demographic and clinical characteristics of the studied population are shown in Table 4.1.

The participants have then been clinically followed after their inclusion during a 3-year period, from 2017 to 2020. Four out of the 45 *C9orf72* presymptomatic carriers have developed subtle frontal cognitive and/or behavioral changes and/or motor signs/symptoms during this period, without fitting diagnostic criteria for FTD or ALS, suggesting they were in the transitional *prodromal* phase at the moment of or just after their inclusion visit. These cases are described in Appendix A Method A2. All analyses in the presymptomatic group were performed with (n=45) and without (n=41) the four prodromal subjects. We also analyzed these cases separately in an additional complementary approach.

TABLE 4.1: Demographic and clinical characteristics of the studied population. Values are expressed as mean ± standard deviation, or as number (%). Demographic characteristics were compared between groups using the χ^2 test for gender and Kruskal-Wallis with Dunn's test for numerical variables. Statistically significant p-values are in bold. ALS-FRS: ALS Functional Rating Scale; MMSE: Mini-Mental State Examination; MDRS: Mattis Dementia Rating Scale; FAB: Frontal Assessment Battery; FBI: Frontal Behavioral Inventory; AES: Apathy Evaluation Scale.

	Control (n=43)	**Presymptomatic (n=45)**	**Patient (n=22)**			
				χ^2 p-value		
Female gender	23 (53.5%)	28 (62.2%)	10 (45.4%)	.408		
				Kruskal-Wallis p-value	**Comparison**	**Dunn's test p-value**
Age at inclusion (years)	46.4 ± 13.5	41.8 ± 11.8	62.7 ± 10.5	**<.001**	Control vs. Presymptomatic Control vs. Patient Presymptomatic vs. Patient	.118 **<.001** **<.001**
ALS-FRS	39.5 ± 1.3	39.5 ± 1.9	33.4 ± 7.7	**<.001**	Control vs. Presymptomatic Control vs. Patient Presymptomatic vs. Patient	.827 **<.001** **<.001**
MMSE	29 ± 1.2	28.5 ± 1.4	17.8 ± 8.4	**<.001**	Control vs. Presymptomatic Control vs. Patient Presymptomatic vs. Patient	.183 **<.001** **<.001**
MDRS	142.1 ± 1.8	141.2 ± 3.0	97.3 ± 36.7	**<.001**	Control vs. Presymptomatic Control vs. Patient Presymptomatic vs. Patient	.431 **<.001** **<.001**
FAB	17 ± 1.2	17.2 ± 0.9	9.7 ± 5.3	**<.001**	Control vs. Presymptomatic Control vs. Patient Presymptomatic vs. Patient	.583 **<.001** **<.001**
Ekman faces test	30.1 ± 2.6	30.1 ± 2.3	18 ± 9.1	**.001**	Control vs. Presymptomatic Control vs. Patient Presymptomatic vs. Patient	.694 **<.001** **.001**
FBI	0.9 ± 1.8	1.5 ± 2.7	28.5 ± 15.2	**<.001**	Control vs. Presymptomatic Control vs. Patient Presymptomatic vs. Patient	.387 **<.001** **<.001**
AES	4.8 ± 3.9	6.5 ± 3.6	23.5 ± 13.1	**<.001**	Control vs. Presymptomatic Control vs. Patient Presymptomatic vs. Patient	.095 **<.001** **.004**

4.3.2 Plasma collection and preparation

Blood samples were collected on EDTA using the same standardized collection and handling procedures for all participants across the centers. The mean disease duration at sampling was 6.2 ± 4.0 years in the patients' group. All were in fasted state. All samples were centralized at the ICM DNA and cell bank, and processed using the same protocol. Plasma was extracted at room temperature after centrifugation of blood samples at 2500 rpm for 10 minutes. Aliquots of 1 mL were stored in polypropylene tubes at -80°C.

4.3.3 MiRNA extraction and sequencing

MiRNA extraction was performed with a miRNeasy Serum/Plasma Kit (Qiagen) following the manufacturer's instructions. We used 200 μL of plasma quickly melted

and directly added to 1 mL of Qiazol solution. MiRNAs were eluted in 14 μL of water; 5 μL were used for miRNA sequencing library preparation with QIAseq miRNA Library Kit (Qiagen) according to the manufacturer's protocol.

MiRNA sequencing was performed on Illumina NovaSeq 6000 in three independent batches, targeting a minimum of 10 million mapped reads per sample. Since batch effects may have a critical impact in high-throughput experiments, we randomly assigned each individual to one batch, equally distributing clinical status (control, presymptomatic, patient) and centers (Paris, Limoges, Lille, Rouen), to allow adjusting for batch effects during data analysis. Appendix A Table A1 describes the distribution of subjects across batches.

4.3.4 Raw reads to miRNA counts computation pipeline

Quality control of raw reads was performed with FastQC (Andrews S. 2010)[1]. UMI-tools (Smith et al., 2017) and Cutadapt (Martin, 2011) were used respectively to extract UMIs and suppress adapting sequences as well as polyA tails. The resulting sequences were aligned with Bowtie (Langmead et al., 2009) and sorted by genomic location with Samtools sort (Li et al., 2009). PCR bias was corrected with UMI-tools, its efficacy was assessed per chromosome with Samtools idxstats. After controlling for the overlap/ambiguity between miRNAs enrichment and Gencode annotation with FeatureCounts (Liao et al., 2014), miRNAs were counted with miRDeep2 (Friedländer et al., 2012).

4.3.5 Statistical analysis

Statistical analyses were performed using R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria). The differential expression of miRNAs between clinical groups was assessed with the R package EdgeR (Robinson et al., 2010). The analysis began with a count matrix with 2576 rows (one per miRNA i) and 110 columns (one per individual j). Only miRNAs considered above noise level (minimum count of 50 reads for at least one sample and a minimum total count of 1000) were retained for statistical analyses, reducing the count matrix to 589 rows. We assumed that miRNA counts followed a negative binomial distribution with mean μ_{ij} and dispersion ϕ_i and used generalized linear models to fit a log-linear model

$$\log_2 \mu_{ij} = x_j^T \beta_i \tag{4.1}$$

for each miRNA, where x_j is the vector of covariates that describes sample j and β_i is the vector of coefficients to be fitted for miRNA i. To control for possible batch, center, age and gender effects, we added these variables as covariates in the model, in addition of clinical status. Raw counts were normalized using a trimmed mean

[1] http:// www.bioinformatics.babraham.ac.uk/projects/fastqc

of M-values (Robinson and Oshlack, 2010). Once the models were fitted, quasi-likelihood (QL) F-test was employed to determine the subset of miRNAs differentially expressed between clinical conditions (miRNA signature). Statistical significance was set at level $\alpha = 0.05$ and p-values were adjusted for multiple testing using the Benjamini-Hochberg method.

4.3.6 Machine learning for binary classification

After the differentially expressed miRNAs were identified, we implemented logistic regression classifiers with L2 regularization in Python 3.8.0 using scikit-learn (Pedregosa et al., 2011) version 0.22.1. We used the expression levels of the miRNA signature as features to train binary classification models for each pairwise comparison between clinical status: controls versus presymptomatic individuals, controls versus patients and presymptomatic individuals versus patients. A stratified nested cross-validation strategy (Appendix A Figure A1) was chosen to find the optimal hyperparameter (L2 regularization coefficient) and to assess model performance using the area under the receiver operating characteristic curve (ROC AUC). We computed 90% confidence intervals (CIs) for the ROC AUC scores from 2000 bootstrap samples, by taking the 5^{th} and 95^{th} percentiles of the bootstrap distribution. Stratification with respect to clinical status was performed to preserve the proportion of healthy controls, presymptomatic subjects and patients in each fold.

4.3.7 Generalization analysis

Since the differentially expressed miRNAs were computed with the entire dataset, the test folds of the cross-validation were also used in the feature selection for our classification models, which could inflate prediction performance. To estimate this possible bias, we then incorporated feature selection in the nested 5-fold cross validation process: differentially expressed miRNAs were computed using only the outer cross-validation loop training data (four out of five folds) at each iteration. The nested cross-validation was repeated 100 times with different fold splits to assess the generalization performance of our classifiers.

4.3.8 Analysis of the transitional stage to clinical FTD/ALS disease

Since we hypothesized that the expression levels of differentially expressed miRNAs might provide information relevant to *C9orf72* disease progression, we designed an experiment to evaluate prediction performance of clinical conversion to FTD/ALS in presymptomatic carriers. A logistic regression classifier was fitted with the expression levels of differentially expressed miRNAs from controls and patients. We used a regular 5-fold cross-validation to determine the optimal hyperparameter (L2 regularization coefficient). Subsequently, this model was tested with the expression levels from the four known presymptomatic carriers who were in their transitional

stage to the clinical disease. Scores from 0 to 1 were provided for each subject, indicating proximity with the expression levels of controls (scores near 0) or patients (scores closer to 1).

4.3.9 Target prediction and pathway analysis

A target-gene based miRNA enrichment analysis was performed, to discover potential biological functions regulated by the differentially expressed miRNAs. We used the publicly available tool DIANA-miRPath v.3 (Vlachos et al., 2015), which implements an *in silico* miRNA target prediction algorithm (DIANA-microT-CDS) as well as an experimentally validated miRNA:gene interaction dataset (DIANA-TarBase v7.0). Both approaches were carried out to identify target genes and the associated Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, using the set of differentially expressed miRNAs as input. The enrichment analysis method consisted of Fisher's exact test (hypergeometric distribution) with Benjamini-Hochberg adjusted p-value threshold of 0.05, giving as output a union set of associated KEGG pathways.

4.4 Results

4.4.1 Differentially expressed miRNAs computed with the entire dataset

Table 4.2 displays all miRNAs identified as differentially expressed, for each pairwise comparison between clinical status, after correction for multiple comparisons. Four miRNAs were computed as differentially expressed between healthy controls and patients: miR-34a-5p and miR-345-5p were overexpressed, while miR-200c-3p and miR-10a-3p were underexpressed in symptomatic mutation carriers. Interestingly, miR-34a-5p was identified as significantly overexpressed also in presymptomatic mutation carriers compared to healthy controls, suggesting that miR-34a-5p expression is associated with *C9orf72* mutation status. Additionally, miR-345-5p was also significantly overexpressed in patients when compared to presymptomatic carriers. When removing the four prodromal subjects from the presymptomatic group, the same miRNAs were identified as differentially expressed, indicating that the differences between the presymptomatic and other groups were not mainly driven by the four prodromal subjects.

We considered these four miRNAs (miR-34a-5p, miR-345-5p, miR-200c-3p, miR-10a-3p) as our miRNA signature for further analyses. The complete output from EdgeR is available in Appendix A Table A2.

Figure 4.1 displays boxplots with the expression levels, for each clinical group, of the four miRNAs identified as differentially expressed. There is a clear difference in miR-34a-5p expression levels between controls and *C9orf72* expansion carriers (presymptomatic and symptomatic). Moreover, the other three identified miRNAs differentiate the mutation carriers at different stages of the pathology: miR-345-5p

TABLE 4.2: Differentially expressed miRNAs identified by EdgeR, after correction for multiple comparisons, for each pairwise comparison between clinical status: Control (n=43), Presymptomatic (n=45), Patient (n=22).

miRNA	log-fold change	*p*-value	adjusted *p*-value
Control vs. Presymptomatic			
miR-34a-5p	-1.433	5.251e-16	3.093e-13
Control vs. Patient			
miR-34a-5p	-1.239	1.650e-8	9.720e-6
miR-345-5p	-0.540	1.131e-5	3.330e-3
miR-200c-3p	0.333	3.109e-5	6.104e-3
miR-10a-3p	0.697	7.141e-5	1.051e-2
Presymptomatic vs. Patient			
miR-345-5p	-0.528	3.610e-5	2.126e-2

showed increased expression in patients, while miR-200c-3p and miR-10a-3p exhibited decreased expression. An expression heatmap of the miRNA signature is displayed in Appendix A Figure A2.

4.4.2 MiRNA signature to classify between clinical groups

To assess whether the identified miRNA signature could distinguish between clinical groups, we implemented logistic regression models using as features the expression levels of the four differentially expressed miRNAs (miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p). The area under the ROC curve (ROC AUC) for the classification of healthy controls and presymptomatic mutation carriers was 0.90 (90% CI 0.83 to 0.95), for controls and patients was 0.90 (90% CI 0.82 to 0.97) and to distinguish presymptomatic carriers and patients was 0.80 (90% CI 0.67 to 0.90) (Figure 4.2). The distributions of the bootstrapped ROC AUC scores are displayed in Appendix A Figure A3.

4.4.3 Generalization analysis

Since we used the entire dataset to identify the miRNA signature, including test data, classification performance could be inflated. In order to assess the generality of our classification scores, we then incorporated feature selection in the nested cross-validation scheme (Appendix A Figure A1), by using only the training data from the outer cross-validation loop to compute differentially expressed miRNAs. Figure 4.3 shows the distribution of miRNAs identified as differentially expressed after performing nested 5-fold cross-validation with 100 different fold splits. Notably, the most frequent miRNAs (highlighted in blue) correspond to the ones computed using the entire dataset: miR-34a-5p (500 occurrences) when comparing healthy controls and presymptomatic mutation carriers; miR-34a-5p, miR-345-5p, miR-200c-3p

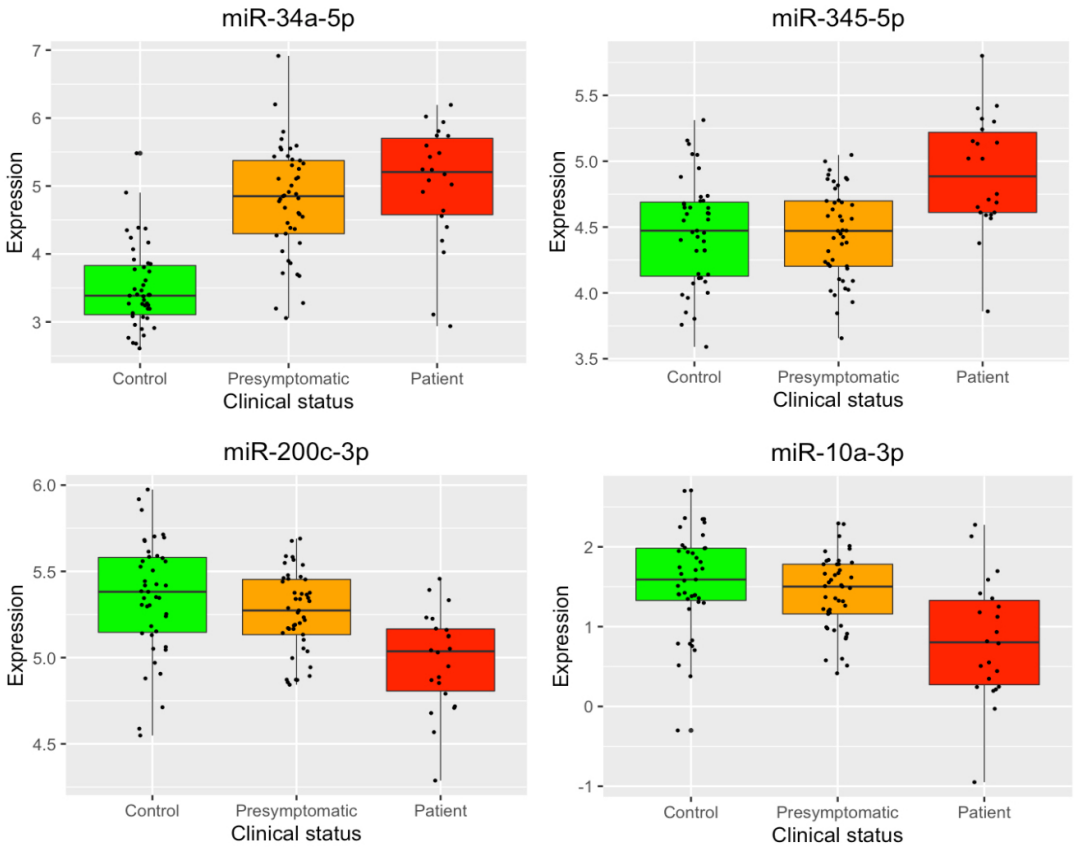


FIGURE 4.1: Boxplots depicting the normalised log2 expression levels of the four microRNAs identified as differentially expressed. Box boundaries represent the first and third quartiles and the median is indicated by the line dividing the interquartile range (IQR). The upper whiskers extend to the values that are within 1.5 × IQR over the third quartiles. The lower whiskers extend to the values that are within 1.5 × IQR under the first quartiles.

and miR-10a-3p (respectively 497, 335, 259 and 196 occurrences) for controls and patients; miR-345-5p (157 occurrences) when analyzing presymptomatic subjects and patients.

Regarding prediction performance, the average ROC AUC when classifying controls versus presymptomatic subjects was 0.88 (90% CI 0.83 to 0.91), for controls versus patients was 0.89 (90% CI 0.83 to 0.94) and for presymptomatic individuals versus patients was 0.67 (90% CI 0.52 to 0.77). The distributions of the ROC AUC scores computed with 100 different fold splits are displayed in Appendix A Figure A4.

4.4.4 Analysis of the transitional stage to clinical FTD/ALS disease

We evaluated the performance to predict the transitional stage to FTD/ALS disease by training a logistic regression classifier with the expression levels from patients and controls and testing with the expression levels of presymptomatic individuals. The probability scores computed for the four subjects in their transitional stage were

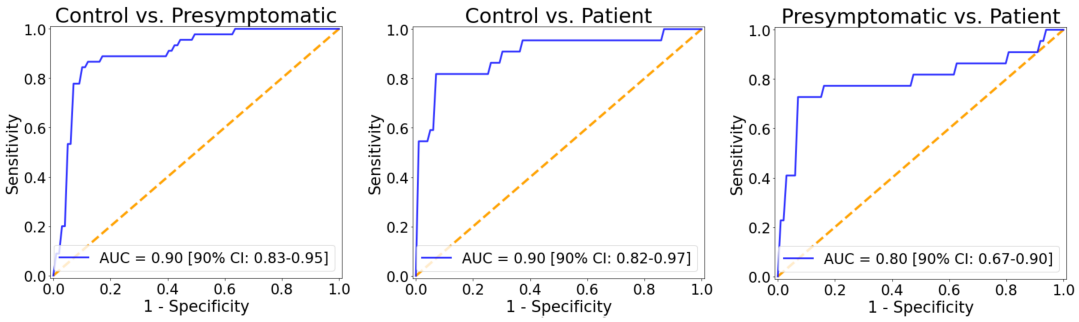


FIGURE 4.2: ROC (receiver operating characteristic) curves for each pairwise classification (control vs presymptomatic, control vs patient and presymptomatic vs patient) obtained with logistic regression using as features the expression levels of the microRNAs signature (miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p). Bootstrapped 90% CIs are reported in brackets. AUC: area under the ROC curve.

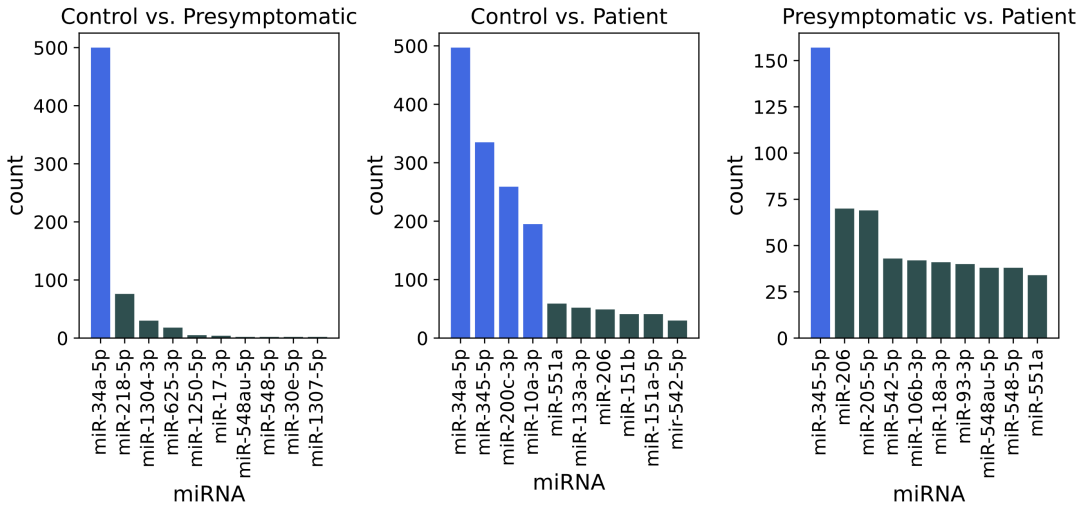


FIGURE 4.3: Number of times each miRNA was found differentially expressed, when performing a repeated 5-fold nested cross-validation for 100 times with different fold splits. In each step of the outer cross-validation loop, four of the five folds were used to identify differentially expressed miRNAs. Since one outer loop consists of five steps, and we performed 100 repetitions, 500 sets of miRNAs were computed for each pairwise comparison between groups, respectively: control vs presymptomatic, control vs patient and presymptomatic vs patient. MiRNAs from the signature computed with the entire data set are highlighted. miRNA: microRNA.

all above 0.50, indicating a stronger similarity with patients: 0.54, 0.75, 0.80 and 0.82. The distribution of probability scores for all presymptomatic subjects is displayed in Appendix A Figure A5.

4.4.5 Target prediction and pathway analysis

Using the four differentially expressed miRNAs (miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p) as input, we performed target prediction and pathway analysis with two methods available in DIANA-miRPath v.3. The *in silico* miRNA target prediction algorithm (microT-CDS) identified 31 influenced pathways (14 significant after Benjamini-Hochberg correction), while the experimentally supported approach (TarBase) resulted in 54 associated pathways (38 significant after Benjamini-Hochberg correction). Complete outputs concerning the list of the putative target genes and their related pathways are given in Appendix A Table A3 and A4. Table 4.3 reports the 13 pathways that were identified by both methods and have significant adjusted *p*-values in at least one of them.

TABLE 4.3: Results from pathway analysis using the four differentially expressed miRNAs as input. Only significant pathways for at least one approach are shown. Statistically significant *p*-values are in bold.

Category	KEGG pathway	*p*-value microT-CDS	*p*-value TarBase
Cancer	Proteoglycans in cancer	**7.941e-4**	**4.259e-8**
	MicroRNAs in cancer	**1.386e-3**	**3.356e-8**
	Glioma	6.554e-2	**1.423e-2**
	Renal cell carcinoma	**1.098e-2**	9.254e-2
	Small cell lung cancer	3.220e-1	**3.341e-2**
Cell signaling/apoptosis	Hippo signaling pathway	**4.556e-2**	**5.622e-4**
	TGF-beta signaling pathway	5.008e-2	**9.288e-4**
	Thyroid hormone signaling pathway	**2.132e-3**	**1.502e-2**
	FoxO signaling pathway	2.368e-1	**1.449e-2**
	Neurotrophin signaling pathway	**9.801e-3**	3.113e-1
Intermediary metabolism	Lysine degradation	**1.606e-2**	**7.882e-4**
	Glycosphingolipid biosynthesis - lacto and neolacto series	**3.885e-10**	**4.423e-2**
Meiosis	Oocyte meiosis	2.487e-1	**2.446e-3**

Appendix A Figure A6 shows miRNA versus KEGG pathways heatmaps, which depict the level of enrichment in significant KEGG pathways for the four differentially expressed miRNAs as computed by the two approaches.

4.5 Discussion

The present study aimed to identify fluid biomarkers by analyzing expression levels of plasma miRNAs without *a priori* knowledge in a large cohort of healthy controls, presymptomatic and symptomatic *C9orf72* carriers. We identified four miRNAs differentially expressed between clinical conditions: miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p. Significantly higher expression of miR-34a-5p was found in mutation carriers when compared with healthy controls, which suggests

that miR-34a-5p expression is deregulated in cases with *C9orf72* mutation. Additionally, we observed miR-345-5p expression to be significantly increased in patients when compared with presymptomatic carriers, which supports the correlation of miR-345-5p expression with the progression of *C9orf72*-associated disease. Finally, our results also suggest that miR-200c-3p and miR-10a-3p underexpression might be associated with full-blown disease as decreased expression levels were significant only between patients and healthy controls.

We used the expression levels of the miRNA signature to train logistic regression classifiers, which were able to differentiate individuals from different clinical groups with good predictive performance (Figure 4.2). Notably, presymptomatic and symptomatic *C9orf72* carriers were distinguished with ROC AUC of 0.80 (90% CI 0.67 to 0.90), which suggests the suitability of plasma miRNAs for following preclinical progression and determining disease onset. We believe that this score was lower in our generalization analysis (0.67, 90% CI 0.52 to 0.77) because the limited number of patients (22) led to a higher variability in the differentially expressed miRNAs in each step of the cross-validation loop (Figure 4.3). Furthermore, we have obtained promising results regarding prediction performance of conversion from the presymptomatic to the clinical stage of FTD/ALS. The four presymptomatic subjects in transitional stage exhibited scores above 0.50, denoting a stronger similarity with the expression levels of patients. Although preliminary, these results suggest that the expression levels of our miRNA signature might be used as early predictors of the *C9orf72* disease conversion.

Previous studies have shown the potential of miRNAs in serum, plasma or CSF as diagnostic biomarkers for FTD and ALS (Grasso et al., 2019; Magen et al., 2020; Piscopo et al., 2018; Schneider et al., 2018; Sheinerman et al., 2017), focusing on comparing healthy controls and patients. However, our findings differ from preceding results: only two miRNAs from our signature (miR-345-5p, miR-200c-3p) were identified as differentially expressed in one of these studies (Magen et al., 2020), none in the others (Grasso et al., 2019; Piscopo et al., 2018; Sheinerman et al., 2017). Results are conflicting probably due to restricted choices for the analyzed miRNAs (Piscopo et al., 2018; Sheinerman et al., 2017) and heterogeneous cohorts, either with sporadic forms (Grasso et al., 2019; Piscopo et al., 2018) or a mixture of sporadic and familial forms with different mutations (Magen et al., 2020). To the best of our knowledge, the present work is the first to compare the expression levels of plasma miRNAs between presymptomatic and symptomatic carriers focusing on *C9orf72* mutation, in addition to providing a plasma miRNA signature that may contribute to the assessment of preclinical progression for *C9orf72*-associated FTD and ALS. Table 4.4 displays a comparison among studies evaluating miRNAs from blood samples (serum or plasma) of FTD and/or ALS patients.

TABLE 4.4: Comparison of studies investigating miRNAs from blood samples (serum or plasma) of FTD and/or ALS patients. FTD: Frontotemporal Dementia; ALS: Amyotrophic Lateral Sclerosis; qRT-PCR: Quantitative Real-Time Polymerase Chain Reaction; RNA-seq: RNA Sequencing; * in serum; † in plasma; ‡ *SOD1, FUS, C9orf72, PFN1*; § *C9orf72, MAPT, GRN, TBK1*

	Freischmidt et al., 2014*	**Sheinerman et al., 2017†**	**Piscopo et al., 2018†**	**Grasso et al., 2019†**	**Magen et al., 2020†**	**This study†**
Disease	ALS	FTD, ALS	FTD	FTD	FTD, ALS	FTD, ALS
Cohort	Separate sporadic/genetic‡	Not mentioned	Sporadic	Sporadic	Mixed sporadic/genetic§	C9orf72
Patients, n= Discovery/replication	9/13 genetic 14 sporadic	50 FTD 50 ALS	54	10/48	52/117 FTD 115 ALS	22
Presymptomatic carriers, n=	18	-	-	-	-	45
Methods of analysis	Microarrays	37 selected miRNAs (qRT-PCR)	9 selected miRNAs (qRT-PCR)	752 selected miRNAs (qRT-PCR)	Large scale sequencing (RNA-seq)	Large scale sequencing (RNA-seq)
Major deregulated miRNAs	miR-4745-5p miR-3665 miR-1915-3p miR-4530 (validated from panel of 30 miRNAs)	miR-9/let-7e miR-7/miR-451 miR335-5p/let-5e (FTD) miR-206/miR-338-3p miR-9/miR-129-3p miR-335-5p/miR-338-3p (ALS)	miR-127-3p	miR-663a miR-502-3p miR-206	Panels of 20, 147, 121 miRNAs for each cohort	miR-34a-5p miR-345-5p miR-200c-3p miR-10a-3p

Overall, our work suggests that miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p are likely involved in neuronal degeneration and *C9orf72*-associated pathogenesis. Among the KEGG pathways identified in this study, some involved in neurodevelopment (Hippo signaling, FoxO signaling), inflammation (TGF-beta signaling), intracellular transduction (neurotrophin signaling), and apoptosis (TGF-beta, FoxO signaling) were relevant as previously shown to be involved in *C9orf72*-disease (Atkinson et al., 2015; Burberry et al., 2020; Farg et al., 2017). Accordingly, these four miRNAs have been previously linked with a range of neurodevelopmental processes, neuropsychiatric and neurodegenerative conditions (Berg et al., 2020; Chua and Tang, 2019; Cosín-Tomás et al., 2017; Fu et al., 2019). For instance, miR-200c and miR-34a family members are implicated in synaptic function, neuronal maturation, differentiation, and survival (Jauhari et al., 2018; Jin et al., 2012). Aberrant expression of miR-34a and miR-345 are also associated with neuronal apoptosis (Modi et al., 2016), whereas members of miR-10a family were found to be differentially expressed in the muscle tissue of ALS patients (Kovanda et al., 2018).

How these four miRNAs are implicated in *C9orf72*-associated pathogenesis, and their relevance in brain pathology are important questions to go further. So far, only few studies addressing miRNA dysregulation in brain tissues of FTD/ALS patients have been performed, and are summarized in Appendix A Table A5. They specifically addressed GRN-associated (Chen-Plotkin et al., 2012; Kocerha et al., 2011), sporadic FTD (Gascon et al., 2014; Hébert et al., 2013), sporadic (Jawaid et al., 2019) or mixed genetic-sporadic ALS patients (Helferich et al., 2018). Notably, there was

no miRNA dysregulation in common between the aforementioned studies, nor between any of those studies on the brain and ours on plasma. Those discrepancies may stem from the heterogeneity of the previous autoptic cohorts and the differences in the methods of miRNA expression analysis. Noteworthy, and differently from our investigation, none of the patient cohorts mentioned in Appendix A Table A5 were exclusively made up of *C9orf72* carriers. Additionally, the observed differences between brain tissue and plasma miRNA profiles may be due to the tissue-specific expression of miRNA on the one hand, and to the time-dependent variations of detectable miRNAs all along the disease course on the other. Due to the disease process itself and other potential confounding factors, significant changes in miRNA expression are likely to occur between a relatively early phase of the disease, in which plasma miRNAs may be used as biomarkers, and the ultimate disease stage, at the moment of brain sampling. At this point, further miRNA profiling studies on *C9orf72* brain tissue are needed to better understand whether tissue miRNAs correlate with plasma expression profiles and their contribution to the disease pathogenesis.

Regardless, it is noteworthy that some studies pointed towards a direct relationship between these miRNAs and *C9orf72* pathogenesis. *C9orf72* stands as a putative target of miR-34a-5p, likely acting as a negative regulator of *C9orf72* mRNA expression (Lal et al., 2011). Additionally, miR-200c-3p and miR-345-5p are down- and up-regulated, respectively, in the extracellular vesicles secreted by induced astrocytes obtained from *C9orf72* patients (Varcianna et al., 2019). Even if not completely explained so far, these important results parallel our study showing a comparable upregulation of miR-34a-5p and miR-345-5p and downregulation of miR-200c-3p in carriers, and provide converging evidence for a link between our set of miRNAs and *C9orf72*-pathogenesis, which will need further investigations.

Previous studies have provided the proof-of-concept that specific sets of miRNAs have the potential to serve as biomarkers of the preclinical/premanifest stages of other neurodegenerative diseases, such as ALS (Freischmidt et al., 2014), Huntington (Jin et al., 2012), and Prion diseases (Boese et al., 2016). Our study supports the usefulness of our four miRNAs as biomarkers of disease progression from the presymptomatic to the symptomatic phase of *C9orf72* disease. Nevertheless, some of them may be dysregulated in a broader range of neurodegenerative conditions. For instance, miR-345 and miR-200c-3p were also dysregulated during the presymptomatic stage of Prion (Boese et al., 2016) and Huntington's diseases (Jin et al., 2012), respectively. This would not prevent, however, their use in longitudinal monitoring of specific genetic neurodegenerative disorders, possibly in combination with other biomarkers. Together, all these studies and ours suggest that dysregulation of such miRNAs is dynamically altered throughout neurodegenerative diseases progression, and can be detectable even long before clinical onset.

The current study has limitations. Firstly, the significant age difference between patients and the other clinical groups may have introduced a confounding factor,

which we considered by including age as a covariate. Secondly, the absence of validation in other tissues or of a replication cohort means that further studies in independent cohorts are required to confirm our results, even though our generalization analysis confirmed the identified miRNA signature. Finally, the limited number of patients does not allow any conclusions about the correlation of plasma miRNAs and different disease phenotypes. Future work will explore longitudinal analyses of plasma miRNAs to assess their use as biomarkers of FTD and ALS progression.

In summary, the current work revealed significant differences in miRNA expression levels in plasma when comparing healthy controls, presymptomatic and symptomatic *C9orf72* mutation carriers. Specifically, we highlighted the potential of miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p expression levels in plasma as biomarkers of preclinical progression for *C9orf72*-associated FTD and ALS. Our results encourage the use of plasma miRNAs, possibly in combination with other markers, to improve the design of clinical trials for these neurodegenerative disorders.

Chapter 5

Circulating microRNA signatures as potential biomarkers of genetic frontotemporal dementia and amyotrophic lateral sclerosis

This chapter has been submitted as an original research article to *Molecular Psychiatry*:

- **Kmetzsch, V.**, Latouche, M., Saracino, D., Rinaldi, D., Camuzat, A., Gareau, T., the French research network on FTD/ALS, Le Ber, I., Colliot, O., Becker, E, "Circulating microRNA signatures as potential biomarkers in genetic frontotemporal dementia and amyotrophic lateral sclerosis". Submitted to *Molecular Psychiatry*.

5.1 Abstract

MicroRNAs are promising biomarkers of frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS), but discrepant results between different studies have so far hampered their use in clinical trials. We aim to assess all previously identified circulating microRNA signatures as potential biomarkers of genetic FTD and/or ALS, using homogeneous, independent validation cohorts of *C9orf72* and *GRN* mutation carriers. Between 2011 and 2021, 104 individuals carrying a *C9orf72* or a *GRN* mutation, along with 31 controls, were recruited through the French research network on FTD/ALS. All subjects underwent blood sampling, from which circulating microRNAs were extracted. We measured differences in the expression levels of 65 microRNAs, selected from 15 published studies about FTD or ALS, between controls, *C9orf72* presymptomatic subjects, and *C9orf72* patients. We also assessed differences in the expression levels of 30 microRNAs, selected from five studies about FTD, between controls, *GRN* presymptomatic subjects, and *GRN* patients. More than half (35/65) of the selected microRNAs were differentially expressed in the

C9orf72 cohort, while only a small proportion (5/30) of microRNAs were differentially expressed in the *GRN* cohort. In multivariate analyses, only individuals in the *C9orf72* cohort could be adequately classified (ROC AUC up to 0.98 for controls vs. presymptomatic subjects, 0.94 for controls vs. patients, and 0.77 for presymptomatic subjects vs. patients) with some of the signatures. Our results suggest that previously identified microRNAs using sporadic or mixed cohorts of FTD and ALS patients could potentially serve as biomarkers of *C9orf72*-associated disease, but not *GRN*-associated disease.

5.2 Introduction

Frontotemporal dementia (FTD) is a neurodegenerative disease characterized by brain atrophy in the frontal and temporal lobes, causing severe changes in personality and social behavior (Neary et al., 2005). The most prevalent genetic causes of FTD are GGGGCC repeat expansions in the *C9orf72* gene and mutations in the *GRN* gene (DeJesus Hernandez et al., 2011; Renton et al., 2011). FTD shares disease pathways with amyotrophic lateral sclerosis (ALS), a debilitating motor neuron disease that causes progressive motor deficit and muscle wasting (Pasinelli and Brown, 2006). The *C9orf72* hexanucleotide repeat expansion has been identified as the most common genetic cause of both familial FTD and ALS, as well as of their sporadic counterparts (DeJesus Hernandez et al., 2011).

There are currently no disease-modifying treatments that can stop the course of FTD or ALS. New therapeutic trials depend on robust progression biomarkers to assess treatment outcomes. The study of FTD/ALS genetic forms is particularly important, since asymptomatic mutation carriers may provide insights about the early disease stages, before any irreversible neuronal damage (Rohrer et al., 2015).

Among the potential non-invasive biomarkers of neurodegenerative diseases, circulating microRNAs (miRNAs) constitute a promising approach (Grasso et al., 2015). MicroRNAs are short noncoding RNAs that negatively regulate gene expression (Huntzinger and Izaurralde, 2011). There is increasing evidence of a link between miRNA expression levels and the diagnosis of FTD (Denk et al., 2018; Grasso et al., 2019; Kmetzsch et al., 2021; Piscopo et al., 2018; Sheinerman et al., 2017) and ALS (De Felice et al., 2014; Dobrowolny et al., 2021; Freischmidt et al., 2015, 2014; Kmetzsch et al., 2021; Magen et al., 2021; Raheja et al., 2018; Sheinerman et al., 2017; Soliman et al., 2021; Takahashi et al., 2015; Tasca et al., 2016; Waller et al., 2017). However, there are strong inconsistencies between the identified miRNA signatures in different studies. The examined cohorts are highly heterogeneous, most of them being sporadic or mixed cohorts of sporadic and genetic forms. Importantly, it is unclear which miRNAs are specific to a particular genetic mutation or might serve as biomarkers for several genetic forms. It is also uncertain whether miRNAs found in sporadic forms are differentially expressed in genetic forms. Furthermore, several of the published articles lacked an independent validation cohort, which also

might have caused disparity between results. This absence of convergence amongst different studies so far hinders the use of miRNAs in clinical trials.

The present work aims at testing circulating miRNA signatures identified in the literature, using two independent homogeneous cohorts of patients and presymptomatic carriers: one focused on *C9orf72* expansion carriers and another comprising *GRN* mutation carriers. For that purpose, we selected all published studies that identified specific miRNAs isolated from plasma or serum as potential biomarkers of FTD and/or ALS. With a preregistered study design, we investigated whether (1) miRNAs revealed in cohorts of sporadic patients (or in mixed cohorts with sporadic and genetic forms) may be biomarkers in *C9orf72* and/or *GRN* genetic forms, (2) miRNAs identified in a *C9orf72* cohort are validated in an independent *C9orf72* cohort, and (3) miRNAs discovered in a *C9orf72* cohort may be relevant in a *GRN* cohort.

We hypothesize that if a miRNA is a potential progression biomarker in a particular genetic form, it will be differentially expressed (adjusted p-value below 0.05) between controls and presymptomatic subjects, controls and patients, or presymptomatic subjects and patients in an independent cohort of subjects carrying that mutation. Moreover, we consider that a miRNA signature will constitute a promising biomarker if a logistic regression model (using these miRNAs as features) classifies subjects between clinical groups with an acceptable area under the ROC curve (above 0.70).

5.3 Material and Methods

This research was conducted according to the preregistration available in https://osf.io/4pw8f.

5.3.1 Participants of the validation cohorts

Between 2011 and 2021, 135 individuals were recruited through the French research network on FTD/ALS (Inserm RBM02-59) and investigated with the same protocol, as previously described in detail (Le Ber et al., 2006). All participants signed written informed consents. This study was approved by the Comité de Protection des Personnes CPP Ile-De-France VI (CPP 36-09 / ID RCB 2008-A01376-49 and CPP 68-15 / ID RCB 2015-A00856-43).

Two cohorts were studied. One cohort was focused on *C9orf72* mutation carriers, including 29 patients (20 FTD, 6 FTD/ALS and 3 ALS) and 17 carriers in the presymptomatic phase. Another cohort was focused on *GRN* mutation carriers, comprising 28 FTD patients and 30 presymptomatic carriers. The control group, shared between the two cohorts, was made up of 31 neurologically healthy individuals that did not carry any of these mutations. Table 5.1 displays the demographic characteristics of the studied cohorts.

TABLE 5.1: Demographic characteristics of the studied cohorts.

	C9orf72 patients	*C9orf72* presymptomatic carriers	*GRN* patients	*GRN* presymptomatic carriers	Controls
No.	29	17	28	30	31
Female, No. (%)	14 (48.3)	10 (58.8)	10 (35.7)	17 (56.7)	18 (58.1)
Age at inclusion, mean (SD)	66.2 (8.8)	51.7 (12.1)	62.9 (11.2)	42.5 (11)	47.1 (14.6)

Standardized interviews with family members, full neurological examinations, quantitative motor testing, and extensive neuropsychological tests measuring all cognitive domains were used to assess each participant's cognitive and clinical conditions. All subjects underwent blood tests, and collected samples were stored in the Paris Brain Institute (ICM) DNA and cell bank.

5.3.2 Plasma preparation, miRNA sequencing and computation pipeline

Blood samples from all participants were collected on EDTA following standardized collection and handling procedures. The mean disease duration at sampling was 4.9 (SD 3.8) years in the *C9orf72* patients' group and 3.2 (SD 1.4) years in the *GRN* patients' group. MiRNA extraction and sequencing were performed as previously reported (Kmetzsch et al., 2021), with the only differences being that plasma was slowly melted at 4°C before adding it to the QIAzol solution, and sequencing was performed in two batches.

Finally, the quantification of miRNAs was carried out according to recommendations by (Potla et al., 2021).

5.3.3 Selected studies

We aimed to find all papers that identified specific miRNAs extracted from human plasma or serum as potential biomarkers of FTD and/or ALS, excluding reviews and meta-analyses. We thus conducted the following search in PubMed[1], on March 10, 2022:

(microRNA[Title] OR microRNAs[Title] OR miR[Title] OR miRNA[Title]) AND (serum[Title] OR circulating[Title] OR plasma[Title]) AND (ALS[Title] OR FTD[Title] OR amyotrophic[Title] OR frontotemporal[Title] OR (neurodegenerative[Title] AND (frontotemporal[Title/Abstract] OR amyotrophic[Title/Abstract]))) NOT mice [Title/Abstract] NOT mouse[Title/Abstract] NOT extracellular vesicles[Title] NOT review[PT] NOT meta-analysis[PT] NOT (comment[PT])

This search yielded 19 results. Two papers (Brennan et al., 2019; Grasso et al., 2015) were excluded because they were review studies, one (Freischmidt et al., 2021)

[1] https:// pubmed.ncbi.nlm.nih.gov/

was discarded because it was focused on protein levels, and one (Xu et al., 2018) was excluded because it was focused on one microRNA from serum exosomes.

Our final selection therefore contained 15 articles. Three of these studies identified miRNA signatures exclusively associated with FTD (Denk et al., 2018; Grasso et al., 2019; Piscopo et al., 2018), one revealed a single miRNA signature for both FTD and ALS (Kmetzsch et al., 2021), one detected separate signatures for FTD and ALS (Sheinerman et al., 2017), and ten of them were entirely focused on ALS (De Felice et al., 2014; Dobrowolny et al., 2021; Freischmidt et al., 2015, 2014; Magen et al., 2021; Raheja et al., 2018; Soliman et al., 2021; Takahashi et al., 2015; Tasca et al., 2016; Waller et al., 2017). The selected papers, their associated microRNAs, diseases, cohort types, numbers of patients, and methods of analyses are displayed in Table 5.2. Since *C9orf72* expansions can cause both FTD and ALS, and *GRN* mutations only cause FTD, there are 16 miRNA signatures (FTD and/or ALS) to be tested with our *C9orf72* cohort and five miRNA signatures (FTD) to be tested with our *GRN* cohort.

Some miRNAs were identified by multiple studies, for instance miR-206 (Grasso et al., 2019; Sheinerman et al., 2017; Soliman et al., 2021; Tasca et al., 2016; Waller et al., 2017), but most miRNAs were found by a single study. Considering all the selected articles, the set of miRNAs associated with either FTD or ALS is composed of 65 miRNAs, and the set of miRNAs associated only with FTD is composed of 30 miRNAs.

Most of the selected articles were based on previous versions of the miRBase. Since we used miRBase version 22.1 in our computation pipeline, the following conversions were performed:

- miR-320a: miR-320a-5p plus miR-320a-3p
- miR-9: miR-9-5p plus miR-9-3p
- let-7e: let-7e-5p plus let-7e-3p
- miR-1: miR-1-5p plus miR-1-3p
- miR-133-a: miR-133-a-5p plus miR-133a-3p
- miR-27a: miR-27a-5p plus miR-27a-3p
- miR-7: miR-7-5p
- miR-451: miR-451a
- miR-129-3p: miR-129-1-3p

5.3.4 Differential expression

Differential expression analyses were performed using the R package EdgeR (Robinson et al., 2010). After microRNA extraction and sequencing, our dataset contained the expression levels of 2656 miRNAs (denoted by i) for each of the 135 subjects

TABLE 5.2: Selected studies investigating circulating microRNA expression (from serum or plasma) of patients with FTD or ALS.

Article	Disease	Cohort	Patients, No. (discovery/replication)	Presymtomatic carriers, No.	Method of analysis	Dysregulated miRNAs
(Grasso et al., 2019)	FTD	Sporadic	10/48 split of same cohort	-	qRT-PCR of 752 miRNAs	miR-663a, miR-502-3p, miR-206
(Piscopo et al., 2018)	FTD	Sporadic	54	-	qRT-PCR of 9 miRNAs linked with apoptosis	miR-127-3p
(Denk et al., 2018)	FTD	Sporadic	48	-	qRT-PCR of 96 miRNAs identified in preliminary study	let-7b-5p, let-7g-5p, miR-106a-5p, miR-106b-5p, miR-18b-5p, miR-223-3p, miR-26a-5p, miR-26b-5p, miR-301a-3p, miR-30b-5p, miR-146a-5p, miR-15a-5p, miR-22-3p, miR-320a, miR-320b, miR-92a-3p, miR-1246
(Kmetzsch et al., 2021)	FTD, ALS	Genetic (*C9orf72*)	22	45	RNA-sequencing of 2576 miRNAs	miR-34a-5p, miR-345-5p, miR-200c-3p, miR-10a-3p
(Sheinerman et al., 2017)	FTD, ALS	Unspecified	For each disease, 25/25 split of same cohort	-	qRT-PCR of 37 brain-enriched miRNAs	miR-9/let-7e, miR-7/miR-451, miR-335-5p/let-7e (FTD) and miR-206/miR-338-3p, miR-9/miR-129-3p, miR-335-5p/miR-338-3p (ALS)
(Magen et al., 2021)	ALS	Mixed sporadic and genetic (*C9orf72*)	126/122 split of same cohort	-	RNA-sequencing of 125 miRNAs identified in longitudinal study	miR-181a-5p, miR-181b-5p
(Soliman et al., 2021)	ALS	Mixed sporadic and genetic (unspecified mutation)	30	-	qRT-PCR of 7 miRNAs involved in ALS	miR-206, miR-143-3p, miR-142-3p
(Dobrowolny et al., 2021)	ALS	Mixed sporadic and genetic (unspecified mutation)	13/23	-	RNA-sequencing followed by qRT-PCR	miR-151a-5p, miR-199a-5p, miR-423–3p
(Raheja et al., 2018)	ALS	Mixed sporadic and genetic (*C9orf72*, SOD1)	23	-	qRT-PCR of 191 miRNAs identified on prior study	miR-29b-3p, miR-320c, miR-34a-5p, miR-29c-3p, miR-320a, miR-22-3p, miR-1, miR-133a-3p, miR-191-5p, miR-144-5p, miR-320b, miR-423-3p, miR-192-5p, miR-133b, miR-194-5p, miR-7-1-3p, miR-19a-3p, miR-425-5p, miR-145-5p, miR-144-3p
(Waller et al., 2017)	ALS	Sporadic	27/23	-	qRT-PCR of 750 miRNAs	miR-206, miR-143-3p, miR-374b-5p
(Tasca et al., 2016)	ALS	Sporadic	14	-	qRT-PCR of 9 muscle-specific, inflammatory, or angiogenic miRNAs	miR-206, miR-133a, miR-133b, miR-27a
(Takahashi et al., 2015)	ALS	Sporadic	16/48 split of same cohort	-	Microarrays, followed by qRT-PCR of 9 miRNAs	miR-4649-5p, miR-4299
(Freischmidt et al., 2015)	ALS	Sporadic	18/20	-	Microarrays of 1733 miRNAs, followed by qRT-PCR of 2 miRNAs	miR-1234-3p, miR-1825
(Freischmidt et al., 2014)	ALS	Separate sporadic and genetic (SOD1, FUS, *C9orf72*)	9/13 (genetic), 14 (sporadic)	18	Microarrays of 1733 miRNAs and qRT-PCR of 4 miRNAs	miR-4745-5p, miR-3665, miR-1915-3p, miR-4530
(De Felice et al., 2014)	ALS	Sporadic	10	-	qRT-PCR of 1 miRNA	miR-338-3p

(represented by j). First, we created two count matrices: one containing the miRNA counts from the *C9orf72* patients, presymptomatic subjects and controls, and another containing the miRNA counts from the *GRN* patients, presymptomatic individuals, and controls. Second, for each count matrix, we fit a log-linear model to each miRNA, following a negative binomial distribution with mean μ_{ij} and dispersion ϕ_i:

$$\log_2 \mu_{ij} = x_j^T \beta_i \tag{5.1}$$

where x_j denotes the covariates describing sample j and β_i denotes the coefficients to be fitted for miRNA i. In addition to the clinical group (control, presymptomatic or patient), we included batch, center, age and gender as covariates in the model. A trimmed mean of M-values (Robinson and Oshlack, 2010) was used to normalize raw counts. Finally, after the log-linear models were fitted, quasi-likelihood (QL) F-tests were used to identify the differentially expressed miRNAs.

Concretely, we tested each of the 65 miRNAs associated with either FTD or ALS in the literature, to identify which were differentially expressed between (a) controls vs. *C9orf72* presymptomatic subjects, (b) controls vs. *C9orf72* patients, and (c) *C9orf72* presymptomatic subjects vs. *C9orf72* patients. Additionally, we tested the 30 miRNAs associated with only FTD in the literature, to highlight which were differentially expressed between (d) controls vs. *GRN* presymptomatic subjects, (e) controls vs. *GRN* patients, and (f) *GRN* presymptomatic subjects vs. *GRN* patients.

All p-values were 2-tailed, and the level of statistical significance was set at 0.05, while p-values between 0.05 and 0.1 were considered as suggestive. The Benjamini-Hochberg (Benjamini and Hochberg, 1995) procedure was used to adjust p-values for multiple testing.

5.3.5 Binary classification

To test if the miRNA signatures described in the literature could discriminate between clinical groups, we trained L2-regularized logistic regression classifiers, using Python 3.8.5 with scikit-learn (Pedregosa et al., 2011) 0.23.2. We first organized the miRNA expression data into six datasets, one for each relevant pairwise comparison: (a) controls vs. *C9orf72* presymptomatic subjects, (b) controls vs. *C9orf72* patients, (c) *C9orf72* presymptomatic subjects vs. *C9orf72* patients, (d) controls vs. *GRN* presymptomatic subjects, (e) controls vs. *GRN* patients, and (f) *GRN* presymptomatic subjects vs. *GRN* patients. A total of 18 classifiers were trained for each of the comparisons (a), (b) and (c): 16 classifiers used as features each of the miRNA signatures identified in the literature, and two were trained with meta-signatures containing the differentially expressed miRNAs identified in the univariate analyses (a), (b) and (c), respectively with adjusted p-values lower than 0.05 and 0.1. In addition, seven classifiers were built for each of the comparisons (d), (e) and (f): five of them used as features each of the miRNA signatures associated with FTD in the literature, and two were trained with meta-signatures containing the differentially

expressed miRNAs identified in the univariate analyses (d), (e) and (f), respectively with adjusted *p*-values lower than 0.05 and 0.1.

Each logistic regression model was trained with a nested 5-fold cross-validation strategy, as previously detailed (Kmetzsch et al., 2021). We computed the area under the ROC curve for each model, as well as 90% empirical confidence intervals from 2000 bootstrap samples. A miRNA signature was considered a promising biomarker for a given comparison if the corresponding ROC AUC was above 0.70.

5.4 Results

5.4.1 Differential expression in the *C9orf72* cohort

The first analysis consisted of testing which miRNAs identified in our literature search were differentially expressed in the *C9orf72* cohort. Of the 65 selected miRNAs, **35** were significantly differentially expressed (adjusted *p*-values smaller than 0.05) in at least one comparison, while **nine** miRNAs had a *p*-value between 0.05 and 0.1. All differentially expressed miRNAs, including log-fold changes indicating the intensity of underexpression or overexpression, as well as computed *p*-values, are displayed in Table 5.3. The complete output from the differential expression analyses in the *C9orf72* cohort are displayed in Appendix B Supplementary Table 1.

The meta-signature of the *C9orf72* cohort containing all differentially expressed miRNAs with adjusted *p*-values lower than 0.05 in at least one comparison comprises the following 35 miRNAs: miR-34a-5p, miR-338-3p, miR-142-3p, miR-320a, miR-145-5p, miR-92a-3p, let-7g-5p, miR-199a-5p, miR-206, miR-30b-5p, miR-191-5p, miR-27a, miR-320b, miR-143-3p, miR-1246, miR-223-3p, miR-144-3p, miR-451, miR-194-5p, miR-144-5p, miR-29b-3p, miR-29c-3p, miR-192-5p, miR-19a-3p, miR-502-3p, miR-15a-5p, miR-374b-5p, miR-7-1-3p, miR-320c, miR-106b-5p, miR-146a-5p, miR-133b, let-7b-5p, miR-345-5p, and miR-22-3p. The meta-signature consisting of the miRNAs with adjusted *p*-values lower than 0.1 has nine more miRNAs: miR-151a-5p, miR-1234-3p, miR-26a-5p, miR-301a-3p, let-7e, miR-18b-5p, miR-106a-5p, miR-1915-3p, and miR-9.

5.4.2 Differential expression in the *GRN* cohort

The second analysis focused on identifying which of the 30 miRNAs linked with FTD in the literature were differentially expressed in the *GRN* cohort. In at least one comparison, five miRNAs were significantly differentially expressed (adjusted *p*-values lower than 0.05), whereas four miRNAs had a *p*-value between 0.05 and 0.1. Table 5.4 lists all differentially expressed miRNAs, including log-fold changes reflecting the degree of underexpression or overexpression and the calculated *p*-values, while Appendix B Supplementary Table 2 summarizes the complete results of the differential expression experiments in the *GRN* cohort.

TABLE 5.3: Differentially expressed miRNAs in the *C9orf72* cohort, for each pairwise comparison between the clinical groups. A positive log-fold change means that the miRNA is overexpressed in the first group. Controls (n=31), *C9orf72* presymptomatic subjects (n=17), and *C9orf72* patients (n=29). All miRNAs with adjusted *p*-values lower than 0.1 are displayed, and adjusted *p*-values lower than 0.05 are shown in bold.

miRNA	log-fold change	*p*-value	adjusted *p*-value
Controls vs. *C9orf72* presymptomatic subjects			
miR-34a-5p	-1.58	3.72E-10	**2.42E-08**
miR-338-3p	-0.79	3.48E-04	**9.53E-03**
miR-142-3p	-0.82	4.90E-04	**9.53E-03**
miR-320a	0.74	5.87E-04	**9.53E-03**
miR-145-5p	-0.94	2.29E-03	**2.53E-02**
miR-92a-3p	0.75	2.63E-03	**2.53E-02**
let-7g-5p	-0.46	2.73E-03	**2.53E-02**
miR-199a-5p	-1.13	3.62E-03	**2.62E-02**
miR-206	2.04	3.62E-03	**2.62E-02**
miR-30b-5p	-1.17	4.45E-03	**2.89E-02**
miR-191-5p	-0.44	5.34E-03	**3.00E-02**
miR-27a	-0.89	5.53E-03	**3.00E-02**
miR-320b	0.76	7.88E-03	**3.94E-02**
miR-143-3p	-0.67	9.46E-03	**4.22E-02**
miR-1246	1.10	9.73E-03	**4.22E-02**
miR-223-3p	-0.70	1.08E-02	**4.38E-02**
miR-144-3p	0.87	1.17E-02	**4.46E-02**
let-7b-5p	0.39	1.52E-02	5.50E-02
miR-151a-5p	-0.50	1.83E-02	6.25E-02
miR-1234-3p	1.40	2.01E-02	6.52E-02
miR-26a-5p	-0.49	2.11E-02	6.52E-02
miR-374b-5p	-0.80	2.45E-02	7.01E-02
miR-146a-5p	-0.68	2.48E-02	7.01E-02
miR-320c	0.58	3.60E-02	9.37E-02
miR-301a-3p	-0.46	3.60E-02	9.37E-02
Controls vs. *C9orf72* patients			
miR-34a-5p	-1.49	7.78E-08	**5.06E-06**
miR-451	2.20	2.87E-05	**6.80E-04**
miR-194-5p	1.55	3.21E-05	**6.80E-04**
miR-144-5p	2.68	4.19E-05	**6.80E-04**
miR-29b-3p	1.01	2.86E-04	**3.72E-03**
miR-29c-3p	0.80	6.61E-04	**7.16E-03**
miR-192-5p	0.99	9.39E-04	**8.72E-03**
miR-19a-3p	1.25	1.32E-03	**1.08E-02**
miR-502-3p	0.80	4.84E-03	**3.50E-02**
miR-15a-5p	0.62	6.70E-03	**4.36E-02**
miR-206	-1.81	7.92E-03	**4.68E-02**
let-7e	-0.74	9.91E-03	5.37E-02
miR-133b	1.11	1.45E-02	7.26E-02
miR-18b-5p	0.62	1.59E-02	7.37E-02
miR-106a-5p	0.69	1.88E-02	8.17E-02
miR-1915-3p	1.44	2.04E-02	8.29E-02
***C9orf72* presymptomatic subjects vs. *C9orf72* patients**			
miR-206	-3.85	2.35E-06	**9.04E-05**
miR-29b-3p	1.42	2.78E-06	**9.04E-05**
miR-30b-5p	1.90	3.15E-05	**6.83E-04**
miR-199a-5p	1.70	6.58E-05	**8.93E-04**
miR-27a	1.41	6.87E-05	**8.93E-04**
miR-29c-3p	0.92	1.86E-04	**2.01E-03**
miR-320a	-0.85	2.17E-04	**2.02E-03**
miR-374b-5p	1.42	2.87E-04	**2.33E-03**
miR-7-1-3p	1.08	4.35E-04	**3.14E-03**
miR-19a-3p	1.37	6.33E-04	**4.11E-03**
miR-338-3p	0.76	1.26E-03	**7.43E-03**
miR-145-5p	1.04	1.41E-03	**7.61E-03**
miR-142-3p	0.79	1.57E-03	**7.83E-03**
miR-320b	-0.94	1.86E-03	**8.61E-03**
miR-320c	-0.92	2.09E-03	**9.06E-03**
miR-106b-5p	0.72	2.97E-03	**1.21E-02**
miR-146a-5p	0.96	3.36E-03	**1.28E-02**
miR-133b	1.31	6.20E-03	**2.23E-02**
miR-223-3p	0.82	6.53E-03	**2.23E-02**
let-7b-5p	-0.46	7.16E-03	**2.33E-02**
miR-345-5p	0.61	7.69E-03	**2.38E-02**
miR-194-5p	1.01	8.56E-03	**2.53E-02**
miR-143-3p	0.68	1.13E-02	**3.19E-02**
miR-22-3p	0.36	1.77E-02	**4.80E-02**
miR-301a-3p	0.54	2.24E-02	5.83E-02
miR-144-5p	1.50	2.38E-02	5.91E-02
miR-9	0.46	2.45E-02	5.91E-02
miR-451	1.18	2.76E-02	6.40E-02

TABLE 5.4: Differentially expressed miRNAs in the *GRN* cohort, for each pairwise comparison between the clinical groups. A positive log-fold change means that the miRNA is overexpressed in the first group. Controls (n=31), *GRN* presymptomatic subjects (n=30), *GRN* patients (n=28). All miRNAs with adjusted *p*-values lower than 0.1 are displayed, and adjusted *p*-values lower than 0.05 are shown in bold.

miRNA	log-fold change	*p*-value	adjusted *p*-value
Controls vs. *GRN* presymptomatic subjects			
(No miRNA was differentially expressed between controls and *GRN* presymptomatic subjects)			
Controls vs. *GRN* patients			
miR-451	2.23	2.65E-06	**7.96E-05**
miR-15a-5p	0.77	3.03E-04	**4.54E-03**
miR-502-3p	0.82	1.73E-03	**1.73E-02**
miR-7	0.56	4.56E-03	**3.42E-02**
miR-18b-5p	0.64	7.39E-03	**4.44E-02**
miR-106a-5p	0.68	1.14E-02	5.72E-02
miR-92a-3p	0.51	1.57E-02	6.72E-02
miR-106b-5p	0.47	2.67E-02	9.44E-02
let-7b-5p	0.33	2.83E-02	9.44E-02
***GRN* presymptomatic subjects vs. *GRN* patients**			
miR-451	1.45	3.35E-03	6.92E-02
miR-7	0.61	4.61E-03	6.92E-02

The meta-signature of the *GRN* cohort containing all differentially expressed miRNAs with adjusted *p*-values lower than 0.05 is comprised of the following five miRNAs: miR-451, miR-15a-5p, miR-502-3p, miR-7, and miR-18b-5p. The meta-signature consisting of the miRNAs with adjusted *p*-values smaller than 0.1 has four more miRNAs: miR-106a-5p, miR-92a-3p, miR-106b-5p, and let-7b-5p.

Finally, Table 5.5 allows comparing the differentially expressed miRNAs between the *C9orf72* and the *GRN* cohorts.

5.4.3 Binary classification in the *C9orf72* cohort

The first set of logistic regression classifiers focused on the *C9orf72* cohort. We trained 18 classifiers for each pairwise comparison, using the 16 miRNA signatures identified in the literature and two meta-signatures obtained from the differential expression analyses. The ROC AUC results and the 90% confidence intervals obtained with 2000 bootstrap samples are displayed in Figure 5.1. The miRNA signatures with the largest ROC AUC were from (Kmetzsch et al., 2021) (0.98 for controls vs. presymptomatic subjects), (Raheja et al., 2018) (0.94 for controls vs. patients), and the meta-signature with p-value < 0.1 (0.77 for presymptomatic subjects vs. patients).

TABLE 5.5: Differentially expressed miRNAs for at least one pairwise comparison between clinical groups, considering both cohorts. The * indicates in which comparisons each miRNA was significantly differentially expressed (adjusted p-values below 0.05), while the (+) denotes adjusted p-values between 0.05 and 0.1.

miRNA	Controls vs. *C9orf72* presymptomatic subjects	Controls vs. *C9orf72* patients	*C9orf72* presymptomatic subjects vs. *C9orf72* patients	Controls vs. *GRN* presymptomatic subjects	Controls vs. *GRN* patients	*GRN* presymptomatic subjects vs. *GRN* patients
miR-34a-5p	*	*				
miR-338-3p	*		*			
miR-142-3p	*		*			
miR-320a	*		*			
miR-145-5p	*		*			
miR-92a-3p	*				(+)	
let-7g-5p	*					
miR-199a-5p	*		*			
miR-206	*	*	*			
miR-30b-5p	*		*			
miR-191-5p	*					
miR-27a	*		*			
miR-320b	*		*			
miR-143-3p	*		*			
miR-1246	*					
miR-223-3p	*		*			
miR-144-3p	*					
miR-451		*	(+)		*	(+)
miR-194-5p		*	*			
miR-144-5p		*	(+)			
miR-29b-3p		*	*			
miR-29c-3p		*	*			
miR-192-5p		*				
miR-19a-3p		*	*			
miR-502-3p		*			*	
miR-15a-5p		*			*	
miR-374b-5p	(+)		*			
miR-7-1-3p			*			
miR-320c	(+)		*			
miR-106b-5p			*		(+)	
miR-146a-5p	(+)		*			
miR-133b		(+)	*			
let-7b-5p	(+)		*		(+)	
miR-345-5p			*			
miR-22-3p			*			
miR-7					*	(+)
miR-18b-5p		(+)			*	
miR-151a-5p	(+)					
miR-1234-3p	(+)					
miR-26a-5p	(+)					
miR-301a-3p	(+)		(+)			
let-7e		(+)				
miR-106a-5p		(+)			(+)	
miR-1915-3p		(+)				
miR-9			(+)			

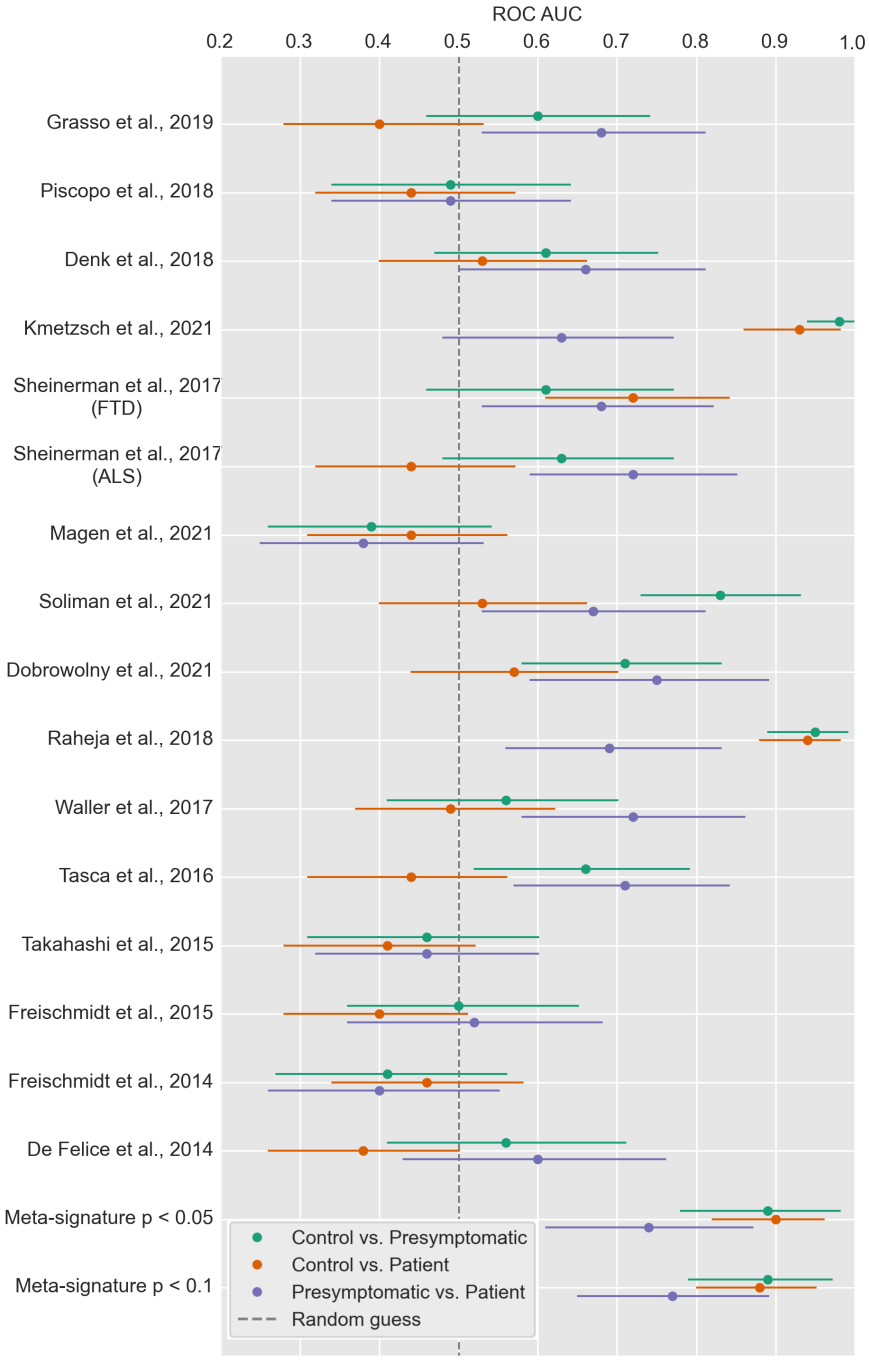


FIGURE 5.1: Area under the ROC curve results when classifying groups from the *C9orf72* cohort. The solid circles indicate the areas under the ROC curves obtained for each pairwise comparison using 18 different miRNA signatures. The whiskers denote empirical 90% confidence intervals obtained with 2000 bootstrap samples

5.4.4 Binary classification in the *GRN* cohort

The second set of classification experiments consisted of training seven logistic regression models for each pairwise comparison in the *GRN* cohort, using five miRNA signatures linked with FTD in the literature and two meta-signatures obtained in the

differential expression analyses. Figure 5.2 shows the ROC AUC results and the corresponding 90% confidence intervals computed with 2000 bootstrap samples. The miRNA signatures with the largest ROC AUC were (Grasso et al., 2019) (0.53 for controls vs. presymptomatic subjects), and the meta-signature with p-value < 0.1 (0.63 for controls vs. patients, and 0.72 for presymptomatic subjects vs. patients).

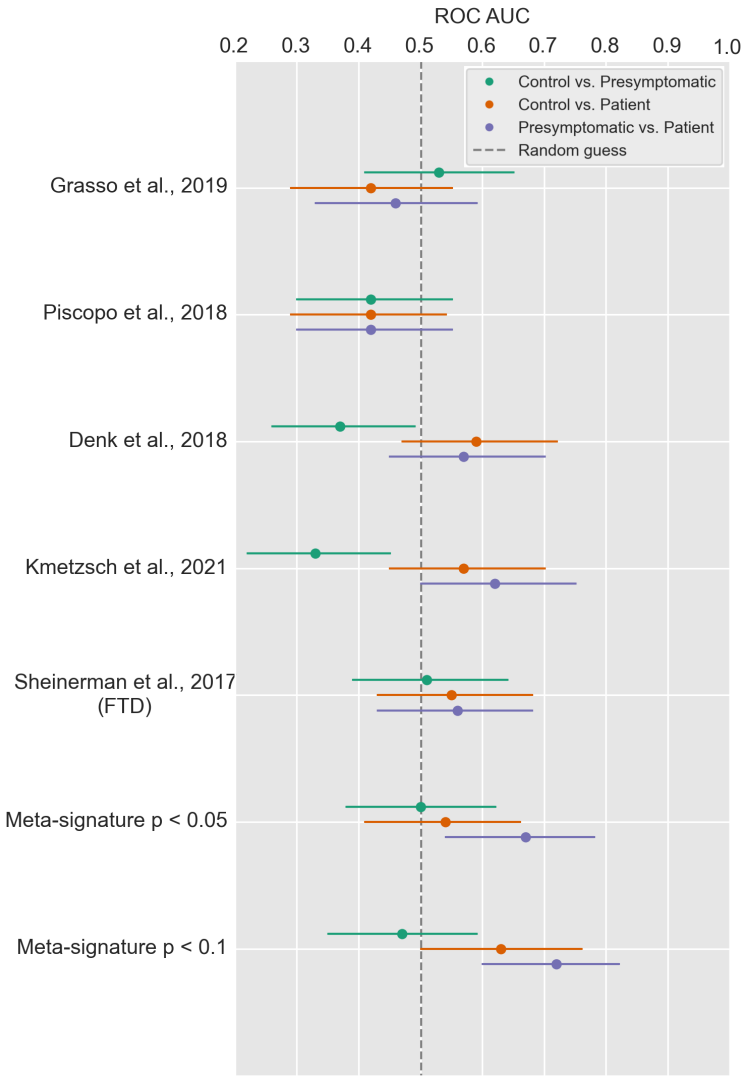


FIGURE 5.2: Area under the ROC curve results when classifying groups from the *GRN* cohort. The solid circles indicate the areas under the ROC curves obtained for each pairwise comparison using seven different miRNA signatures. The whiskers denote empirical 90% confidence intervals obtained with 2000 bootstrap samples

5.5 Discussion

The goal of this study was to assess all circulating miRNA signatures previously published in the literature as possible biomarkers of FTD and/or ALS, by testing them in two separate homogeneous cohorts of *C9orf72* and *GRN* mutation carriers,

comprising patients and presymptomatic subjects. The results of this work demonstrate that (1) several miRNAs identified in sporadic or mixed FTD/ALS cohorts could potentially be used as biomarkers of *C9orf72*-associated disease; (2) some miRNAs revealed in a *C9orf72* cohort are validated in an independent *C9orf72* cohort; and (3) most miRNAs associated with FTD in sporadic or mixed cohorts, or in a cohort of *C9orf72* mutation carriers, are not relevant biomarkers of *GRN*-associated disease.

First, differential expression results (Table 5.3) showed that more than half (35/65) of the miRNAs linked with FTD and/or ALS in the literature were significantly differentially expressed in the *C9orf72* cohort. Remarkably, only four of the 15 selected studies included *C9orf72* mutation carriers (Freischmidt et al., 2014; Kmetzsch et al., 2021; Magen et al., 2021; Raheja et al., 2018), three of which focused exclusively on ALS (Freischmidt et al., 2014; Magen et al., 2021; Raheja et al., 2018). Therefore, these outcomes reveal strong miRNA expression similarities between individuals with sporadic forms of FTD/ALS and *C9orf72*-associated disease. Classification results with the *C9orf72* cohort (Figure 5.1) also corroborate these findings, since half of the examined miRNA signatures (8/16) yielded at least one pairwise comparison with acceptable performance (ROC AUC above 0.70), and all comparisons employing the meta-signatures had satisfactory performance.

Next, we observed that a miRNA signature previously identified in a homogeneous *C9orf72* cohort (Kmetzsch et al., 2021) and another one revealed in a mixed cohort of sporadic and familial ALS (Raheja et al., 2018) displayed an outstanding result (ROC AUC above 0.90) when classifying controls vs. *C9orf72* presymptomatic subjects and controls vs. *C9orf72* patients (Figure 5.1). These two signatures have in common the presence of miR-34a-5p, which has the smallest adjusted p-value in the differential expression analyses regarding these comparisons (Table 5.3, respectively p-value = 2.42E-08 and p-value = 5.06E-06). In contrast, the performance of both of these signatures classifying *C9orf72* presymptomatic individuals from patients was unsatisfactory. Indeed, neither of them contained miR-206, which is the most differentially expressed miRNA in this comparison (Table 5.3, p-value = 9.04E-05). The overexpression of miR-206 in ALS patients had already been evidenced (Toivonen et al., 2014), and the results of the present work extend this association also to *C9orf72*-associated disease. Nevertheless, even using the expression levels of miR-206, the classification of *C9orf72* presymptomatic subjects vs. patients led to lower performances than comparisons involving the control group: the highest ROC AUC was 0.77, using the meta-signature with p-value < 0.1.

Finally, our results with the *GRN* cohort suggest that previously identified miRNAs have a weaker correlation with disease diagnosis and progression in this genetic form. Only a small proportion (5/30) of the miRNAs associated with FTD in previous studies was significantly differentially expressed in the *GRN* cohort (Table 5.4), and not a single miRNA was differentially expressed between controls and presymptomatic *GRN* carriers. Regarding the classification experiments, none of the

studied miRNA signatures in the *GRN* cohort exhibited an acceptable performance (Figure 5.2), and the only ROC AUC slightly above 0.70 was obtained when classifying *GRN* presymptomatic carriers and patients using the largest meta-signature (miRNAs with p-value < 0.1). One should note that none of the previous studies included *GRN* participants. Thus, our results demonstrate that miRNAs associated with sporadic FTD or genetic FTD due to *C9orf72* are not relevant for *GRN*-associated disease. Further studies are needed to determine if other miRNAs, not analyzed in the present paper, are useful in *GRN*-associated disease.

Validation studies using independent datasets, such as this one, are crucial to assess the utility of biomarker candidates, fostering research rigor and reproducibility. Notably, we carefully defined our research questions and analysis plan before data analysis, and preregistered our study. Preregistration has the strong benefit of leaving no flexibility for changes in analytical decisions after observing the data, which has been highlighted as a major source of false discoveries and replication failure (Nosek et al., 2018).

The main limitation of this work is the size of the studied cohorts, particularly the small group of *C9orf72* presymptomatic carriers (17) in comparison with the other groups, due to the rarity of genetic FTD. Additionally, due to the low number of *C9orf72* patients with different phenotypes (20 FTD, 6 FTD/ALS and 3 ALS), no conclusions can be drawn concerning the relationship between miRNAs and different disease manifestations. Future work will explore the combination of circulating microRNAs with other biomarkers, such as gray matter volume (Rohrer et al., 2015), white matter integrity (Bertrand et al., 2018), and neurofilament light chain level (Saracino et al., 2021). Multimodality will be crucial to accurately assess progression in *GRN*-associated FTD, and will likely improve the understanding of *C9orf72*-associated disease.

In summary, the present work revealed that most miRNAs previously identified in sporadic or mixed FTD/ALS cohorts are potential biomarkers of *C9orf72*-associated FTD/ALS, but not of *GRN*-associated FTD. Longitudinal studies are needed to confirm our findings, before circulating microRNAs can be used to evaluate *C9orf72*-associated disease progression in clinical trials.

Chapter 6

Disease progression score estimation from multimodal imaging and microRNA data using supervised variational autoencoders

This chapter is under review as an original research article at the *IEEE Journal of Biomedical and Health Informatics*:

- **Kmetzsch, V.**, Becker, E., Saracino, D., Rinaldi, D., Camuzat, A., Le Ber, I., Colliot, O., "Disease progression score estimation from multimodal imaging and microRNA data using supervised variational autoencoders". Under review at the *IEEE Journal of Biomedical and Health Informatics*.

6.1 Abstract

Frontotemporal dementia and amyotrophic lateral sclerosis are rare neurodegenerative diseases with no effective treatment. The development of biomarkers allowing an accurate assessment of disease progression is crucial for evaluating new therapies. Concretely, neuroimaging and transcriptomic (microRNA) data have been shown useful in tracking their progression. However, no single biomarker can accurately measure progression in these complex diseases. Additionally, large samples are not available for such rare disorders. It is thus essential to develop methods that can model disease progression by combining multiple biomarkers from small samples. In this paper, we propose a new framework for computing a disease progression score (DPS) from cross-sectional multimodal data. Specifically, we introduce a supervised multimodal variational autoencoder that can infer a meaningful latent space, where latent representations are placed along a disease trajectory. A score is computed by orthogonal projections onto this path. We evaluate our framework with multiple synthetic datasets and with a real dataset containing 14 patients, 40

presymptomatic genetic mutation carriers and 37 controls from the PREV-DEMALS study. There is no ground truth for the DPS in real-world scenarios, therefore we use the area under the ROC curve (AUC) as a proxy metric. Results with the synthetic datasets support this choice, since the higher the AUC, the more accurate the predicted simulated DPS. Experiments with the real dataset demonstrate better performance in comparison with a state-of-the-art approach. The proposed framework thus leverages cross-sectional multimodal datasets with small sample sizes to objectively measure disease progression, with potential application in clinical trials.

6.2 Introduction

Frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS) are rare neurodegenerative disorders that have devastating personal and social consequences. Progressive cognitive and behavioural changes, emotional instability, and language impairment are the main symptoms of FTD (Rascovsky et al., 2011). ALS is a motor neuron disease characterized by gradual muscle wasting, ultimately leading to disability (Pasinelli and Brown, 2006). FTD and ALS may be sporadic (no previous family history) or genetically inherited. The most common genetic cause of FTD and ALS is a hexanucleotide repeat expansion in the *C9orf72* gene (DeJesus Hernandez et al., 2011; Renton et al., 2011). These fatal conditions can sometimes coexist in *C9orf72*-mutated individuals, and have no cure or standard treatment to date.

Carriers of the *C9orf72* mutation that do not present clinical symptoms are considered presymptomatic, since they have a very high probability of manifesting FTD and/or ALS later in life. Clinical trials for potential therapies are likely to be most effective at this presymptomatic stage, before any irreversible brain damage has occurred. However, the evaluation of new treatments depends on an accurate measure of disease progression, which is not evident without observable symptoms. Therefore, it is crucial to identify biomarkers to assess disease progression in presymptomatic subjects. Indeed, previous work has shown the relevance of neuroimaging (Bertrand et al., 2018; Rohrer et al., 2015) and transcriptomic (microRNA) (Kmetzsch et al., 2021) biomarkers for a better understanding of *C9orf72*-associated disease in presymptomatic carriers. Nevertheless, when these modalities are analyzed separately, they provide only an incomplete picture of these complex neurodegenerative diseases. It is thus essential to develop methods that leverage the complementary information available from different modalities to accurately measure disease progression. As different biomarkers characterise distinct disease stages, various biomarkers can be combined to represent the entire disease course with a single measure, commonly referred in the literature as the *disease progression score* (DPS).

The idea of computing disease progression scores falls within the larger topic of modeling disease progression. In the past years, many approaches have been developed for data-driven modeling of disease progression, such as event-based models (EBM) (Fonteijn et al., 2012; Venkatraghavan et al., 2019), different algorithms fitting

logistic functions to biomarker trajectories (Jedynak et al., 2012; Mehdipour Ghazi et al., 2021), non-linear mixed-effects models (Koval et al., 2021; Schiratti et al., 2017), a vertex-wise model of brain diseases fitted with expectation-maximisation (Marinescu et al., 2019), Gaussian processes (Lorenzi et al., 2019a), topological profiles reflecting brain connectivity (Garbarino et al., 2019), Bayesian multi-task learning (Aksman et al., 2019), and recurrent neural networks (Mehdipour Ghazi et al., 2019).

Most of these approaches require longitudinal data. For instance, the authors of (Jedynak et al., 2012) assume that the longitudinal dynamic of each biomarker can be represented as a sigmoidal function of the DPS. They propose a joint optimization algorithm to compute the DPS, fit one sigmoid function per biomarker using alternating least squares, and apply their work to hundreds of patients with Alzheimer's disease (AD). Similarly, a more recent method (Mehdipour Ghazi et al., 2021), also applied to AD, uses M-estimation to map each subject's age to a DPS, jointly fitting generalized logistic functions to the longitudinal dynamics of biomarkers as functions of the DPS. Schiratti and colleagues (Schiratti et al., 2017) propose a general non-linear mixed-effects model for longitudinal data based on concepts from Riemannian geometry. The application of this framework to AD, called AD Course Map (Koval et al., 2021), allowed to map each subject to their corresponding disease stage. The authors of (Lorenzi et al., 2019a) propose a probabilistic approach based on Gaussian process regression from time-series of biomarker measurements. Yet another framework, named Data-driven Inference of Vertexwise Evolution (DIVE) (Marinescu et al., 2019) consists in identifying clusters of vertex-wise biomarker measurements in the brain, and estimating representative trajectories for these clusters. Finally, (Mehdipour Ghazi et al., 2019) uses recurrent neural networks to predict biomarker values without parametric assumptions about trajectories, with application to AD. To the best of our knowledge, the only disease modeling approaches that infer a DPS from cross-sectional data are EBM (Fonteijn et al., 2012; Venkatraghavan et al., 2019). These models explore the temporal sequence in which biomarkers become abnormal in the course of a disease. They have been successfully applied to a variety of diseases including AD (Archetti et al., 2019; Firth et al., 2020; Fonteijn et al., 2012; Oxtoby et al., 2018; Venkatraghavan et al., 2019; Young et al., 2014), multiple sclerosis (Dekker et al., 2020; Eshaghi et al., 2018), Parkinson's disease (Oxtoby et al., 2021), Huntington's disease (Wijeratne et al., 2021) as well as FTD (Ende et al., 2021; Panman et al., 2021) and ALS (Gabel et al., 2020). However, in these works, EBMs were applied to a relatively small number of features (typically 10-50) and it is unknown if they would perform well in higher dimensions.

Despite the recognized importance of estimating neurodegenerative diseases progression, research has tended to focus mostly on higher prevalence conditions. Existing solutions are thus inadequate to model rare diseases with high-dimensional cross-sectional data, for three main reasons. First, we observe that longitudinal data is needed for the vast majority of approaches. However, *C9orf72*-associated FTD and ALS are slowly progressive conditions in the presymptomatic phase, which hinders

the collection of meaningful longitudinal data. Second, most published methods benefit from large samples, which are not available for very low prevalence disorders such as genetic FTD and ALS. Finally, it is unclear if event-based models, the only methods suitable for cross-sectional data, can be robustly applied to high-dimensional microRNA expression data, which comprise hundreds of biomarkers.

In this paper, we thus present a novel framework to estimate disease progression scores for rare neurodegenerative disorders using only cross-sectional data. To that purpose, we introduce a new supervised multimodal variational autoencoder (VAE) trained with neuroimaging and microRNA data. Our working hypothesis is that disease progression scores may be modelled as underlying latent traits. Concretely, we aim to learn a meaningful latent space, where the relative positions of latent representations indicate the distance travelled along the disease pathophysiological pathway.

VAEs are powerful generative models that project data into a low-dimensional regularized latent space (Kingma and Welling, 2014). These models have been previously used with multimodal data (Antelmi et al., 2019; Cheng et al., 2022; Xu et al., 2021), but not for the purpose of inferring a DPS. Usually VAEs are trained in an unsupervised manner. However, extensions have been proposed for semi-supervised (Berkhahn et al., 2019; Kingma et al., 2014; N et al., 2017) or supervised (Ji et al., 2021) tasks. These studies demonstrate that providing supervision to the model imposes specific semantics on the latent space, resulting in more meaningful and robust representations. In our context, explicit labels (control, presymptomatic, patient) are already available for all subjects. We thus add supervision during training, leveraging this information to improve the separation of the groups in the latent space. Additionally, we propose to split high-dimensional (neuroimaging and microRNA data) and low-dimensional (demographic information) modalities. Our model thus couples two neural networks with different inputs: (1) an encoder/decoder that learns a latent space from the high-dimensional features, and (2) a classifier having as input the latent variables concatenated with the low dimensional features, useful for the classification task. As no ground truth is available for the DPS in real-world scenarios, we evaluate our models with a proxy metric: the area under the ROC curve (AUC) for each pairwise classification between clinical groups, computed using only the inferred DPS.

A preliminary version of this work has been published at the SPIE Medical Imaging 2022 conference (Kmetzsch et al., 2022). Compared to the conference version, the present paper introduces the following novelties: (1) a supervised instead of a standard unsupervised VAE approach, (2) data split between low-dimensional and high-dimensional modalities, (3) disease trajectory computation in the latent space using principal curves instead of straight lines, (4) additional experiments with multiple synthetic datasets, (5) a comparison with event-based models, and (6) an ablation study.

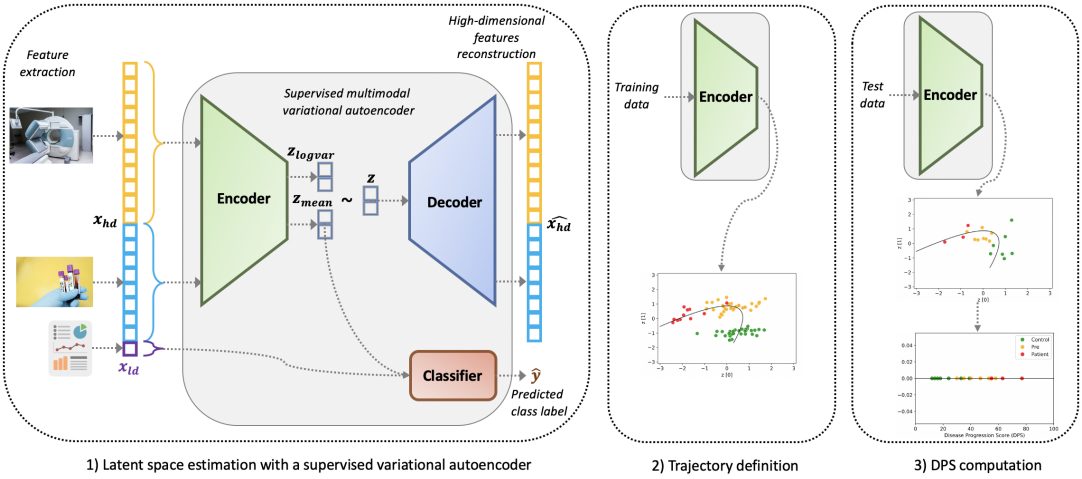


FIGURE 6.1: Illustration of the proposed framework for disease progression scores (DPS) computation. 1) High-dimensional (neuroimaging and microRNAs expression data) and low-dimensional (demographic information) features are extracted; the former are fed to the encoder, the latter are concatenated with latent codes and fed to the classifier. 2) Once the model is trained, all training examples are encoded in the latent space and a principal curve is calculated to define the disease trajectory. 3) Test examples are encoded in the latent space and the latent representations are orthogonally projected onto the previously computed curve; the DPS correspond to their coordinates along the curve.

The manuscript is organized as follows. Section 6.3 explains our proposed framework, section 6.4 describes the analyzed datasets, section 6.5 details our experiments and corresponding outcomes, and finally section 6.6 examines the meaning of our results and highlights the broader implications of our study.

6.3 Methodology

We consider a dataset $(\mathcal{X}, \mathcal{Y}) = \{(x_1, y_1), ..., (x_n, y_n)\}$. The i-th subject is characterized by a feature vector $x_i \in \mathbb{R}^m$ and a label $y_i \in \{0, 1, 2\}$ denoting the clinical group (control, presymptomatic, patient). Our aim is to estimate a DPS, denoted as $v_i \in [1, 100]$ (the interval for the scores is arbitrary), where a greater score corresponds to a higher disease severity. To that purpose, we assume that the observations have corresponding latent variables $z_i \in \mathcal{R}^\ell$. We will thus aim to estimate a latent representation and the DPS will be computed from a trajectory in the latent space.

Our framework is composed of three main steps, as illustrated in Fig. 6.1. First, we propose a supervised multimodal variational autoencoder to estimate the latent space. We leverage the fact that participants belong to different groups to introduce some supervision in order to improve the VAE training. The model aims at simultaneously reconstructing the data and classifying the participants. We propose to

split low-dimensional sociodemographic data (denoted $\mathcal{X}_{ld}$, used only for the classification) from high-dimensional multimodal neuroimaging and transcriptomic data (denoted $\mathcal{X}_{hd}$, used both for reconstruction and classification). Second, we build a curve representing disease trajectory in the latent space. Finally, data from new subjects, not included in the training set, are encoded in the latent space and projected onto this trajectory, in order to obtain their DPS.

In this section, we first explain the three main steps of our framework, then we describe implementation details.

6.3.1 Supervised multimodal VAE

A variational autoencoder (VAE) (Kingma and Welling, 2014) is a generative model that learns the training data distribution $p(x)$ using a latent representation model:

$$p(x) = \int p(x|z)p(z)dz,$$

where z is a continuous latent variable living in a lower dimensional space and $p(z)$ is its prior distribution, commonly a Gaussian with zero mean and identity covariance matrix. The solution of the inference problem to describe the latent space is given by deriving the posterior $p(z|x)$. However, there is no closed-form solution for complex real-world datasets. Therefore, VAEs introduce the idea of learning a variational approximation $q_\phi(z|x)$ of the true posterior, in the form of a neural network referred to as the *encoder*. The encoder maps data x to a mean vector z_{mean} and a log-variance vector z_{logvar}, that parametrize a Gaussian distribution from which we obtain the latent representation z. VAEs are also equipped with a generative function $p_\theta(x|z)$, parametrized by a neural network referred to as the *decoder*. The decoder transforms the latent representation z back to the original input space.

During training, the vanilla VAE aims at maximizing the variational lower bound of the marginal log-likelihood, known as the evidence lower bound (ELBO). This is equivalent to minimizing a loss function with two terms:

$$\mathcal{L} = \mathcal{L}_r(x, \hat{x}) + \mathcal{L}_{KL}(q_\phi(z|x), p(z)).$$

The first term is the reconstruction error between the input data x and the reconstructed data $\hat{x}$, typically a mean squared error (MSE). The second term is the Kullback-Leibler divergence between the approximated posterior $q_\phi(z|x)$ and the prior distribution $p(z)$, acting as a regularization term.

We propose to insert a supervised branch in the vanilla VAE architecture in order to exploit the fact that our samples have different diagnostic labels, even though their DPS is unknown. Denoting y as the true class label and $\hat{y}$ as the predicted class label, we define our training objective as:

$$\mathcal{L} = \alpha_1 \cdot \mathcal{L}_r(x, \hat{x}) + \alpha_2 \cdot \mathcal{L}_{KL}(q_\phi(z|x), p(z)) + \alpha_3 \cdot \mathcal{L}_c(y, \hat{y}),$$

where $\mathcal{L}_r$ and $\mathcal{L}_{KL}$ correspond to the ELBO in vanilla VAEs and $\mathcal{L}_c$ is a cross-entropy term that penalizes the classification error. The hyperparameters α_k control the relative weights between the different loss terms ($\sum_{k=1}^{3} \alpha_k = 1$).

Before training, we split the high-dimensional modalities (miRNA expression and neuroimaging) from the low-dimensional (demographic information). As it will be mentioned later in the datasets description, we consider one low-dimensional feature and $m-1$ high-dimensional features, although the same concepts can be applied to more low-dimensional features. So we use $m-1$ features to feed the encoder and one feature concatenated to the latent code to feed the classifier. Features are rescaled from 0 to 1. Our encoder consists of fully-connected layers of sizes $(m-1) \rightarrow 50 \rightarrow 2$, meaning our latent space is 2-dimensional. The decoder is implemented with fully-connected layers of sizes $2 \rightarrow 50 \rightarrow (m-1)$. The nonlinear activation function is the leaky rectified linear unit (ReLU) in all layers except the decoder's last layer which uses a sigmoid function to constrain the output between 0 and 1. The classifier network has one fully connected layer of $3 \rightarrow 3$ units, with a softmax function to normalize the output to probabilities over the predicted classes. We use the mean squared error as the reconstruction loss $\mathcal{L}_r$ and the cross-entropy as the classification loss $\mathcal{L}_c$.

6.3.2 Trajectory definition

Once the model is trained, the next step is to encode the training data in the latent space. We then compute the straight line passing through the centroids of the control and patient clusters. This straight line could be used in downstream analyses as a rudimentary disease trajectory in the latent space. Instead, we obtain an improved nonlinear trajectory by using this line as initialization for the principal curve algorithm (Hastie and Stuetzle, 1989). A principal curve is a smooth one-dimensional curve passing through the *middle* of given data points. The algorithm detailed in (Hastie and Stuetzle, 1989) finds a nonparametric curve by iteratively minimizing the orthogonal distances to the points until convergence.

6.3.3 DPS computation

Once the disease trajectory curve is computed in the latent space, we can encode the test data. The next step is to orthogonally project the latent codes onto the computed curve. The DPS $v_i \in [1, 100]$ for each subject is the coordinate of their projection along this curve, 1 corresponding to the beginning and 100 to the end of the curve. The pseudo-code from model training to DPS computation is shown in Algorithm 1.

6.3.4 Implementation details

The hyperparameters of the training objective were set as $\alpha_1 = 0.2$, $\alpha_2 = 0.2$, and $\alpha_3 = 0.6$. The loss function was optimized using Adam (Kingma and Ba, 2017), with a learning rate of 10^{-3}, batches of 32 observations and 250 epochs.

Algorithm 1 DPS computation from latent representation

Input: features $\mathcal{X} = \{x_i\}_{i=1}^{n} \in \mathbb{R}^m$, labels $\mathcal{Y} = \{y_i\}_{i=1}^{n} \in \{0,1,2\}$, training set indices I_{tr} and test set indices I_{te} for one data split into training and test set.
Output: DPS $\{v_i\}_{i=I_{te}}$ of the subjects in the test set.
/* first step: supervised VAE training */
for epoch in [1,250] **do**
  Sample batches $(\mathcal{X}_j, \mathcal{Y}_j)$ from $(\mathcal{X}_{I_{tr}}, \mathcal{Y}_{I_{tr}})$
  for each batch $(\mathcal{X}_j, \mathcal{Y}_j)$ **do**
    $\mathcal{X}_{hd}, \mathcal{X}_{ld} \leftarrow$ split_high_low_dimension$(\mathcal{X}_j)$
    $\mathcal{Z}_{mean}, \mathcal{Z}_{logvar} \leftarrow$ encoder$(\mathcal{X}_{hd})$
    Draw latent codes $\mathcal{Z} \sim \mathcal{N}(\mathcal{Z}_{mean}, e^{\mathcal{Z}_{logvar}})$
    $\hat{\mathcal{Y}}_y \leftarrow$ classifier(concatenate$(\mathcal{X}_{ld}, \mathcal{Z}_{mean})$)
    $\hat{\mathcal{X}}_{hd} \leftarrow$ decoder$(\mathcal{Z})$
    $\mathcal{L}_r \leftarrow$ mean_squared_error$(\mathcal{X}_{hd}, \hat{\mathcal{X}}_{hd})$
    $\mathcal{L}_{KL} \leftarrow$ kl_divergence$(\mathcal{N}(\mathcal{Z}_{mean}, e^{\mathcal{Z}_{logvar}}), \mathcal{N}(0, I))$
    $\mathcal{L}_c \leftarrow$ cross_entropy$(\mathcal{Y}_y, \hat{\mathcal{Y}}_y)$
    $\mathcal{L} \leftarrow \alpha_1 \cdot \mathcal{L}_r + \alpha_2 \cdot \mathcal{L}_{KL} + \alpha_3 \cdot \mathcal{L}_c$
    Compute gradients, update network to minimize $\mathcal{L}$
  end for
end for
/* second step: trajectory definition */
$\mathcal{Z}, \_\_ \leftarrow encoder(\mathcal{X}_{I_{te}})$
$c_{control} \leftarrow$ mean$(\{\mathcal{Z}_j : y_j == 0\})$
$c_{patient} \leftarrow$ mean$(\{\mathcal{Z}_j : y_j == 2\})$
$pc \leftarrow$ principal_curve$(c_{control}, c_{patient}, degree = 2)$
/* third step: DPS computation */
for i in I_{te} **do**
  $z_{pc} \leftarrow$ projection of z_i into pc
  $v_i \leftarrow$ coordinate of $z_{pc} \in [0, 100]$
end for
return $\{v_i\}_{i=I_{te}}$

We carried out the experiments on a computer equipped with a 2.4 GHz Intel Quad-Core Core i5 processor and 16 GB of RAM. Models were implemented in Python 3.8.5 using PyTorch 1.8.1 and Scikit-learn 0.23.2 (Pedregosa et al., 2011). For the principal curves computation, we used the implementation provided in the Python package pcurvepy 0.0.10[1], specifying 2 as the degree of the smoothing spline.

6.4 Datasets

6.4.1 Synthetic datasets

Since ground truth disease progression scores are not available in real-world scenarios, we created synthetic datasets to better evaluate the proposed framework. Multiple datasets were generated, with different noise levels and distinct proportions of features correlating with the DPS.

[1] https://pypi.org/project/pcurvepy/

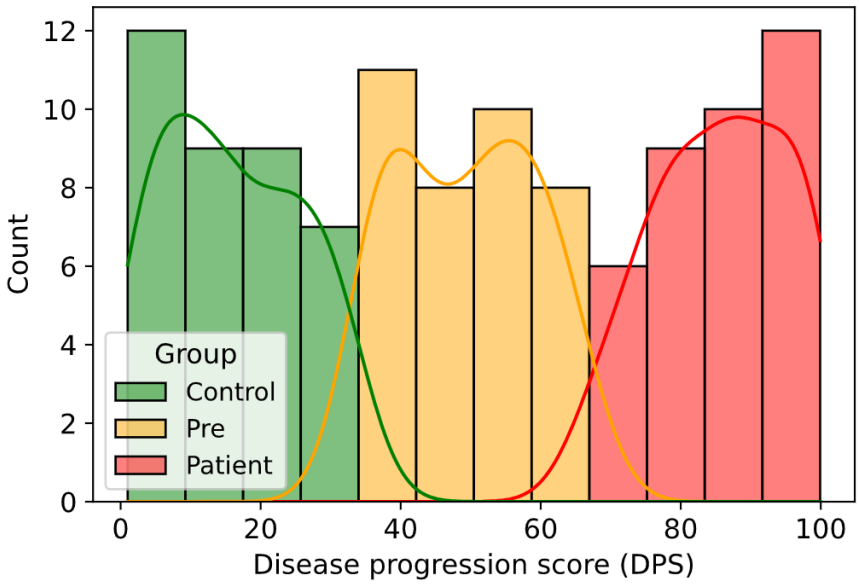


FIGURE 6.2: Synthetic ground truth disease progression scores $\{v_i\}_{i=1}^{n} \in [0, 100)$ for $n = 111$ subjects (37 subjects per group).

Let $Y \in \{0, 1, 2\}$ indicate the class labels (respectively control, presymptomatic and patient). We created $n = 111$ synthetic participants (a number close to that of our real dataset) with class labels denoted by y_i $(i = 1, ..., 111)$,

$$y_{i=1,...,37} = 0$$
$$y_{i=38,...,74} = 1$$
$$y_{i=75,...,111} = 2.$$

Next, we modeled the disease progression scores as continuous random variables following uniform distributions. Let $V \in [1, 100)$ represent the DPS values. We defined the conditional distribution of the DPS given the class labels as follows:

$$V|Y = 0 \sim \mathrm{U}[1, 34)$$
$$V|Y = 1 \sim \mathrm{U}[34, 67)$$
$$V|Y = 2 \sim \mathrm{U}[67, 100)$$

We then sampled the corresponding DPS v_i from the conditional distributions defined above. The obtained disease progression scores are displayed in Fig. 6.2.

Once the synthetic ground truth DPS were created, we generated multiple datasets $\mathcal{D} \in \mathbb{R}^{n \times m}$ containing $n = 111$ participants and $m = 160$ features. In order to simulate two modalities, features were initially sampled from two distributions: half from a negative binomial distribution (typical of miRNA expression data) and half from a normal distribution (representative of various real-world datasets). We denote the columns of $\mathcal{D}$ by $C_1, ..., C_m$. The format of the synthetic datasets is illustrated in Fig. 6.3.

Each created dataset had a distinct proportion of features correlating with the

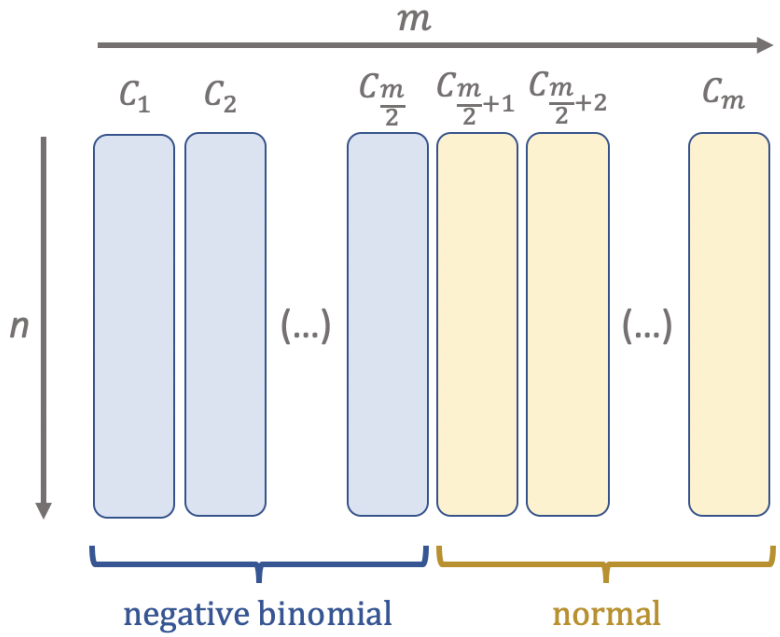


FIGURE 6.3: Format of the synthetic datasets $\mathcal{D} \in \mathbb{R}^{n \times m}$ containing m features from n individuals. Half of the features are initially sampled from a negative binomial distribution and half from a normal distribution.

DPS and different noise levels. The number of features from each modality to positively and negatively correlate with the DPS is denoted as f, and the standard deviation of the added zero-mean Gaussian noise as s. We used $f = \{0, 2, 5, 10, 15, 20, 25, 30, 35, 40\}$ and $s = \{0.001, 0.2, 0.5, 0.8, 1, 5\}$ and thus obtained a total of 60 synthetic datasets. The set of operations performed for their generation is shown in Appendix C Algorithm 1.

6.4.2 Real dataset

Participants were recruited through the PREV-DEMALS (https://clinicaltrials.gov, ID NCT02590276) study, a French multicentric prospective cohort focused on *C9orf72* expansion carriers. Written informed consents were obtained from all participants. The study was approved by the ethics committee (Comité de Protection des Personnes CPP Ile-De-France VI, CPP 68-15 and ID RCB 2015-A00856-43). A detailed description of this cohort and its demographic profile can be found in (Kmetzsch et al., 2021).

We included 110 individuals in our analyses, divided into three groups, according to their clinical status:

- Patient group: 22 symptomatic (15 FTD, 4 FTD/ALS and 3 ALS) carriers of a pathogenic *C9orf72* expansion;
- Presymptomatic group: 45 asymptomatic carriers;
- Control group: 43 asymptomatic non-carriers.

The dataset comprised multimodal data including microRNA (miRNA) sequencing data and neuroimaging data. These two modalities are described below.

MicroRNA data

MicroRNAs are a class of small noncoding RNAs that negatively regulate gene expression (Huntzinger and Izaurralde, 2011). MicroRNAs expression in blood plasma has been shown to correlate with the diagnosis and progression of many neurodegenerative diseases (Grasso et al., 2014), including FTD and ALS. All individuals included in this cohort underwent plasma sampling, from which miRNA sequencing was performed. Plasma collection and preparation, miRNA extraction and sequencing, quality control and the computational pipeline to obtain the miRNA counts are detailed in (Kmetzsch et al., 2021). The initial miRNA dataset contained expression levels for all miRNAs mapped in the human genome (2576 miRNAs). We retained the 589 miRNAs with expression profiles above noise level (minimum total count of 1000 reads and at least 50 reads for one sample). A trimmed mean of M-values (Robinson and Oshlack, 2010) implemented in the R package EdgeR (Robinson et al., 2010) was used to normalize the raw counts.

Neuroimaging data

Neuroimaging data consisted of gray matter volumes extracted from T1-weighted anatomical magnetic resonance imaging (MRI), including the estimated total intracranial volume (TIV), 68 cortical regions of interest (ROIs) using the Desikan atlas and 18 subcortical ROIs using the Aseg nomenclature, thus resulting in 87 neuroimaging features. The TIV was used to normalize the volume of each ROI,

$$NV_{ROI} = \frac{TIV_m \times V_{ROI}}{TIV},$$

where V_{ROI} is the original volume of the ROI, NV_{ROI} is the corresponding normalized volume and TIV_m is the average TIV computed across all subjects. The MRI acquisition parameters, quality check and processing pipeline are thoroughly described in (Bertrand et al., 2018).

Only 91 subjects (14 patients, 40 presymptomatic carriers and 37 controls) had MRI scans collected. Hence, we divided our dataset into two subsets: 19 subjects that only had miRNA data available, and 91 subjects with multimodal neuroimaging and miRNA data. The former subset was used as a discovery set for miRNA feature selection: we used these 19 individuals to perform differential expression analysis (as described in (Kmetzsch et al., 2021)). The 68 miRNAs with the lowest p-values were selected for all downstream analyses.

Lastly, we also included age as demographic information for all subjects. So the total dimension of each feature vector was $m = 87 + 68 + 1 = 156$.

6.5 Experiments and results

6.5.1 Synthetic datasets

We applied our framework to 60 synthetic datasets (described in Section 6.4.1) with different noise levels and distinct number of features correlating with the ground truth DPS. Each synthetic dataset was divided into a training set of 90 subjects (30 per clinical group) and a test set of 21 individuals (7 per group). We trained one model per dataset, using the same hyperparameters as the experiments with the real dataset. After training each model, we computed the DPS for the subjects from the test set. We then calculated the Spearman correlations between the simulated ground-truth scores and the predicted scores. Finally, we evaluated the ROC AUC for each pairwise comparison between the three simulated clinical groups.

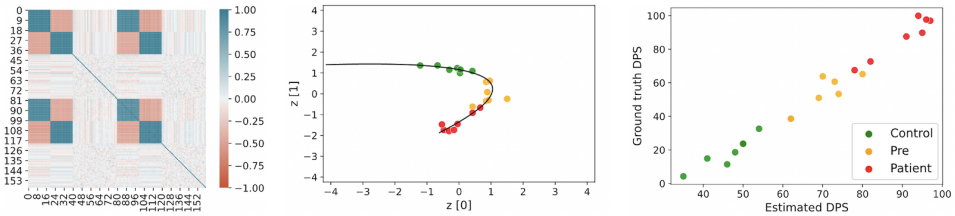
Fig. 6.4 presents the computed trajectories and the DPS obtained when 50% of the features are correlated (25% positively and 25% negatively correlated) with the disease progression, for different noise levels. The correlation matrices illustrate the strength of the relationships between the simulated features, for all investigated noise levels.

The results of the Spearman correlation between the estimated DPS and the ground truth data, as well as the average ROC AUC scores for the three pairwise comparison between groups, are showed in Fig. 6.5. As expected, we can observe that the DPS is very well estimated for lower noise levels and higher proportion of relevant features, while the performances decrease when the noise level becomes very high and when only few features are correlated with the DPS. Importantly, we observe that the Spearman correlation of the DPS and the ROC AUC have similar behaviors, indicating that the ROC AUC of pairwise comparisons is a reasonable proxy to evaluate the DPS, as will be done with the real dataset.

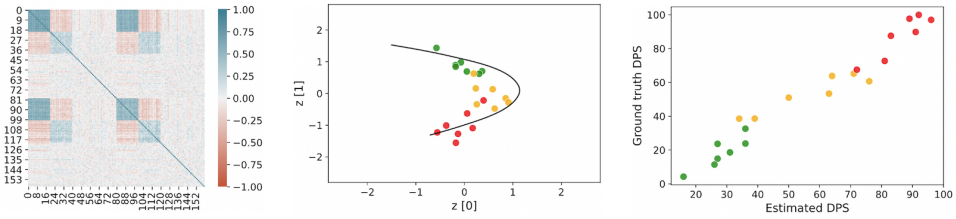
6.5.2 Real dataset

Experiments with the real dataset (described in Section 6.4.2) were carried out with a cross-validation of 100 stratified randomized folds, using data from the 91 subjects that underwent both plasma sampling and MRI scans. For each fold split, we trained a model using 73 training subjects, and then computed the DPS for the 18 individuals in the test set. Fig. 6.6 displays an example of the latent space trajectory computed with one representative training data split, the corresponding test set projected in the latent space, and the obtained disease progression scores.

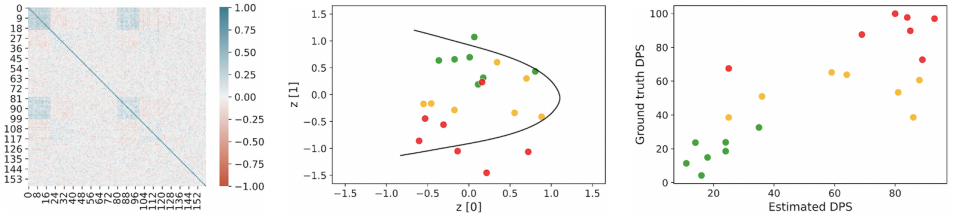
Unlike for the synthetic dataset, there is no ground truth for the DPS in the real dataset. We thus applied a proxy metric to assess model performance: using only the inferred DPS, we did pairwise comparisons between the clinical groups and computed the corresponding areas under the ROC curves. Specifically, we present the following experiments: (1) evaluation of the proposed method, (2) comparison with



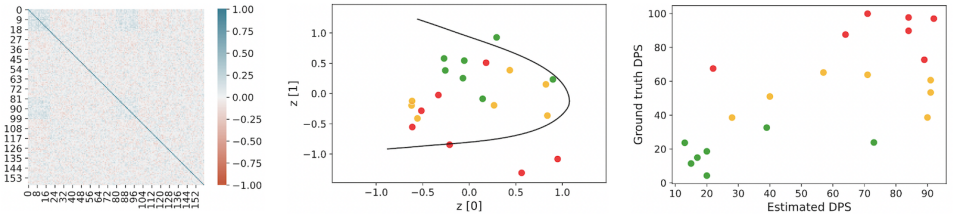
(A) Noise standard deviation = 0.001



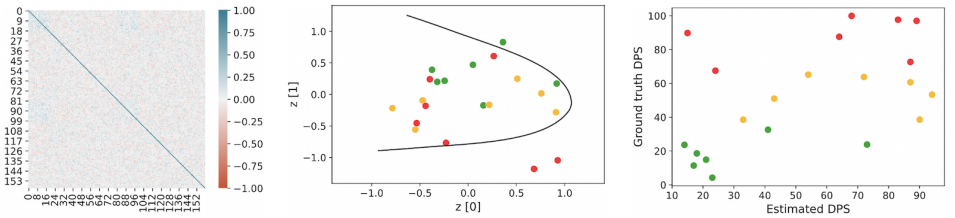
(B) Noise standard deviation = 0.2



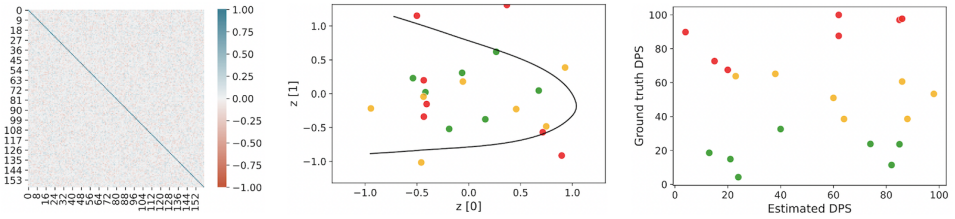
(C) Noise standard deviation = 0.5



(D) Noise standard deviation = 0.8



(E) Noise standard deviation = 1



(F) Noise standard deviation = 5

FIGURE 6.4: Results on synthetic data when 50% of the features are correlated with the disease progression score. The rows indicate different noise levels (zero-mean Gaussian noise with different standard deviations). Each column displays, respectively: (1) correlation matrices showing the strength of the relationships between the simulated features, (2) inferred trajectories and test sets projected in the latent space, and (3) estimated DPS vs. ground truth DPS.

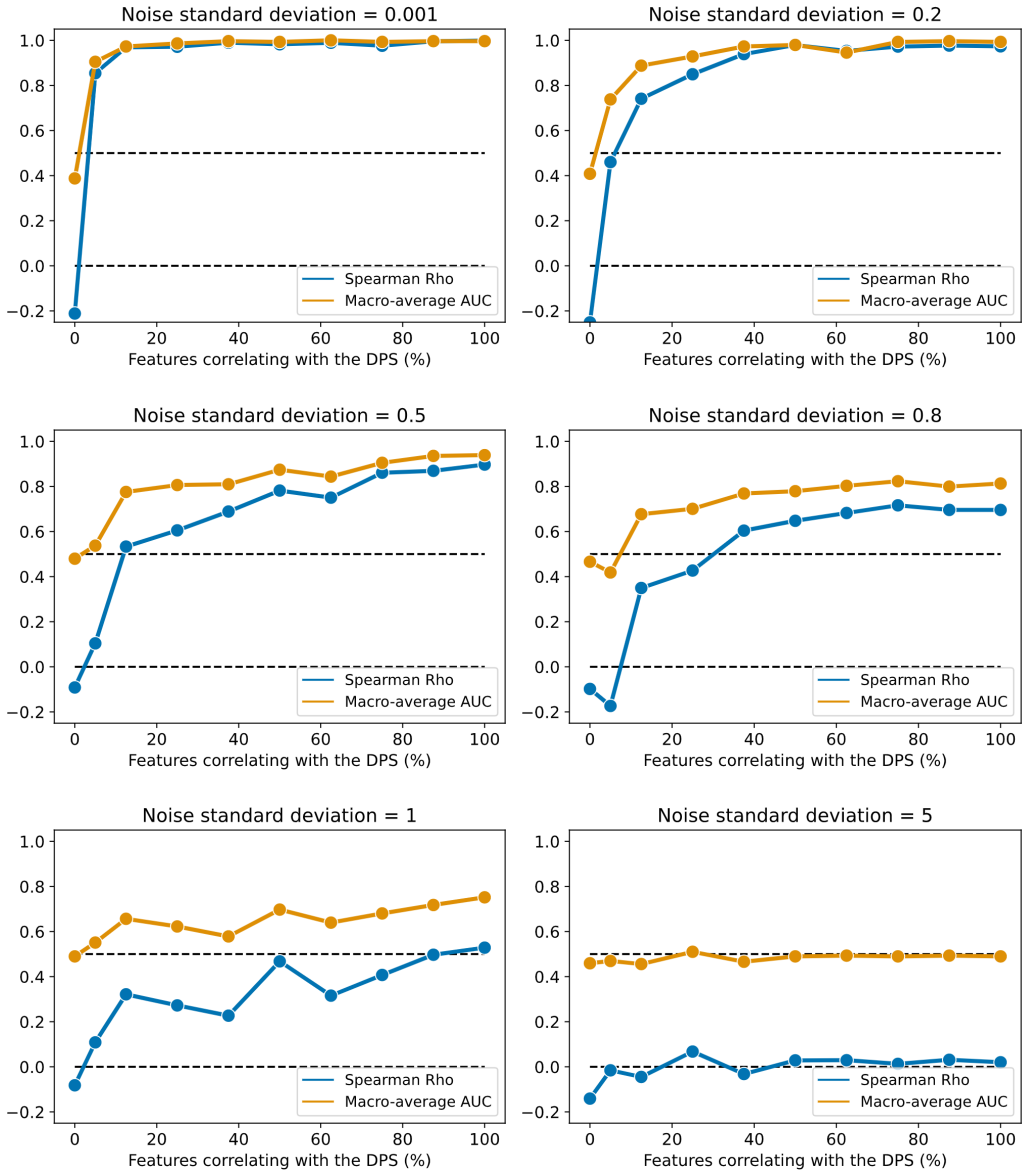
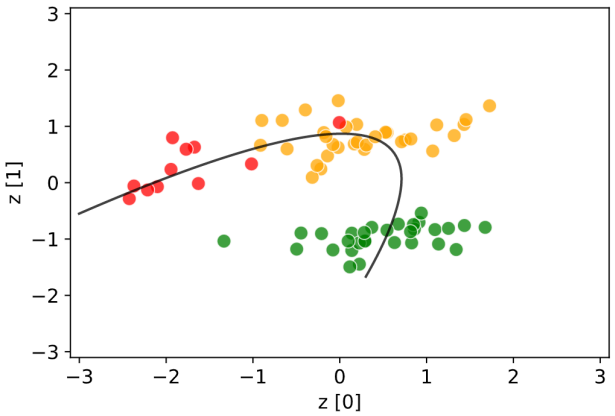
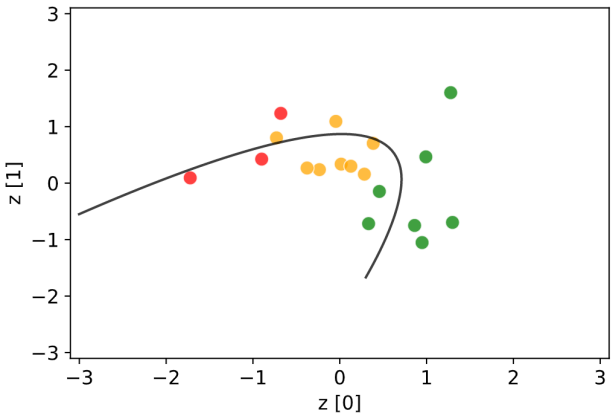


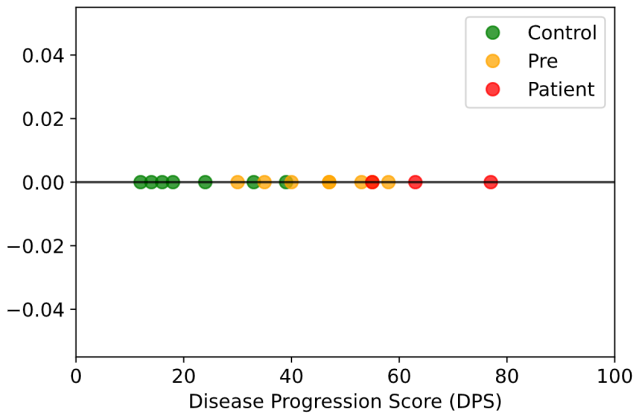
FIGURE 6.5: Results on synthetic data. Macro-average ROC AUC and Spearman correlation between ground truth and estimated DPS, for different noise levels (zero-mean Gaussian with 0.001, 0.2, 0.5, 0.8, 1, and 5 as standard deviation) and several proportions (0% to 100%) of features correlating with the disease progression score. Random chance is denoted by the dashed lines (ROC AUC = 0.5 and Spearman Rho = 0).



(A) Training data projected in the latent space



(B) Test data projected in the latent space



(C) Computed scores for the test data

FIGURE 6.6: Results on real data. (a) Training data projected in the latent space and the corresponding computed trajectory for one of the 100 fold splits. (b) Test data projected in the latent space, along with the previously computed trajectory. (c) Scores computed after the projection of the latent representation of the test data onto the trajectory.

a state-of-the-art method for modeling disease progression, the discriminative event-based model, (3) ablation study, and (4) variation of hyperparameters.

Evaluation of the proposed method

First, we used the DPS computed in each fold to build ROC curves for the three pairwise comparisons between clinical groups. The average ROC curves are shown in Fig. 6.7. The ROC AUC for the classification of controls and presymptomatic subjects was 0.74 ± 0.13, for controls and patients was 0.98 ± 0.05 and to distinguish presymptomatic carriers and patients was 0.96 ± 0.07. These results reveal that it is harder to differentiate controls from presymptomatic individuals than it is to distinguish between patients and the other two groups. The histogram displayed in Fig. 6.8 illustrates the disease progression scores computed over all 100 test folds (18 subjects per test fold, corresponding to 1800 DPS). The distribution shapes highlight a clear separation between the patient group and the other groups. The distribution of the DPS for the presymptomatic group is more spread, which was expected as this group is the most heterogeneous. Some presymptomatic subjects are very far from onset and the neurodegenerative process has barely begun, they are thus closer to controls. Other presymptomatic subjects are closer to disease onset and thus their DPS is closer to that of patients.

Comparison with DEBM

Next, we compared our results to a discriminative event-based model (DEBM) (Venkatraghavan et al., 2019), a method that also infers a DPS from cross-sectional data. For that experiment, the same cross-validation strategy of 100 stratified folds was applied. We built the DEBM models and computed the DPS using the Python package pyebm[2] 2.0.3. Table 6.1 displays the corresponding ROC AUC results for each pairwise comparison. We can observe that our model achieves a substantially better classification performance for all pairwise comparisons. Additionally, our approach used less computing time: our framework took 2 seconds per fold for training and DPS computation, while the DEBM algorithm took on average 180 seconds per fold.

TABLE 6.1: Results on real data: comparison between our approach and a discriminative event-based model (DEBM) (Venkatraghavan et al., 2019). ROC AUC (mean $\pm$ standard deviation) over 100 stratified splits.

Comparison	Our model	DEBM
Control vs. Pre	$\mathbf{0.74 \pm 0.13}$	0.67 ± 0.14
Control vs. Patient	$\mathbf{0.98 \pm 0.05}$	0.76 ± 0.17
Pre vs. Patient	$\mathbf{0.96 \pm 0.07}$	0.65 ± 0.17

[2] https://pypi.org/project/pyebm/

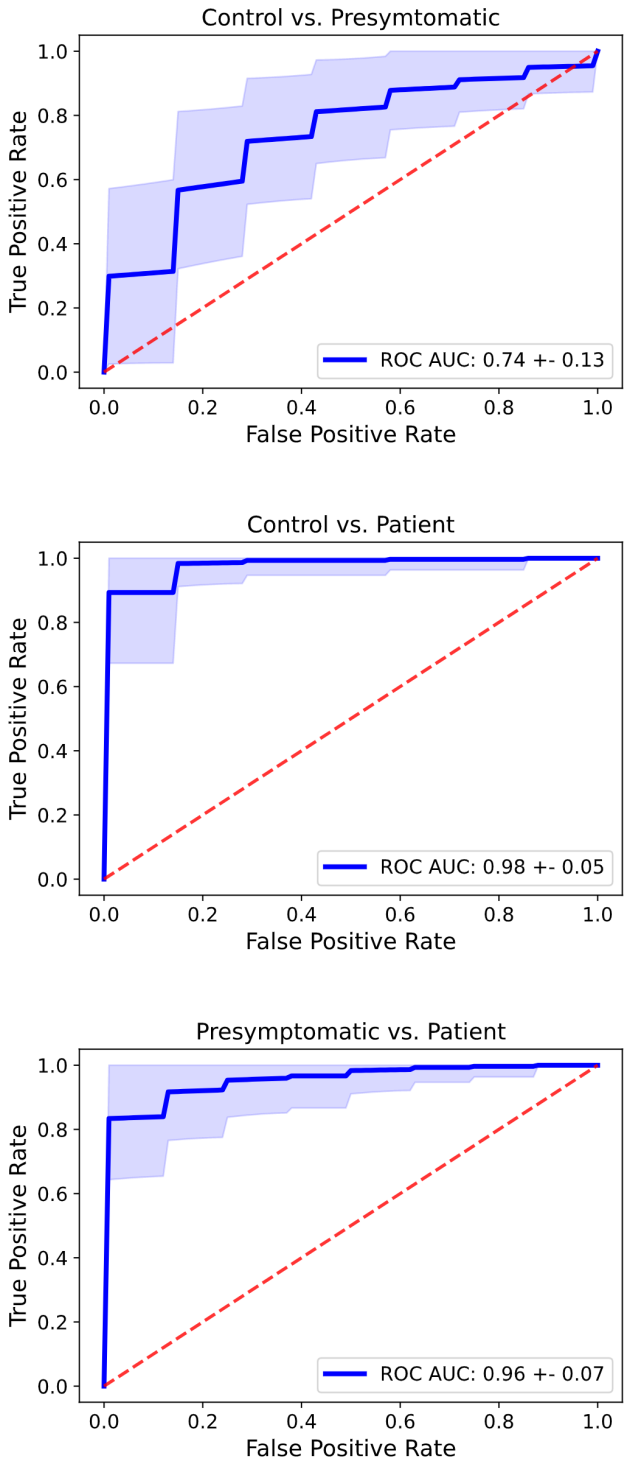


FIGURE 6.7: Results on real data. Average ROC (receiver operating characteristic) curves for each pairwise comparison between clinical groups, over 100 stratified splits. The shaded areas correspond to one standard deviation. The areas under the ROC curves (ROC AUC) are shown as mean $\pm$ standard deviation. Random chance is indicated by the dashed line.

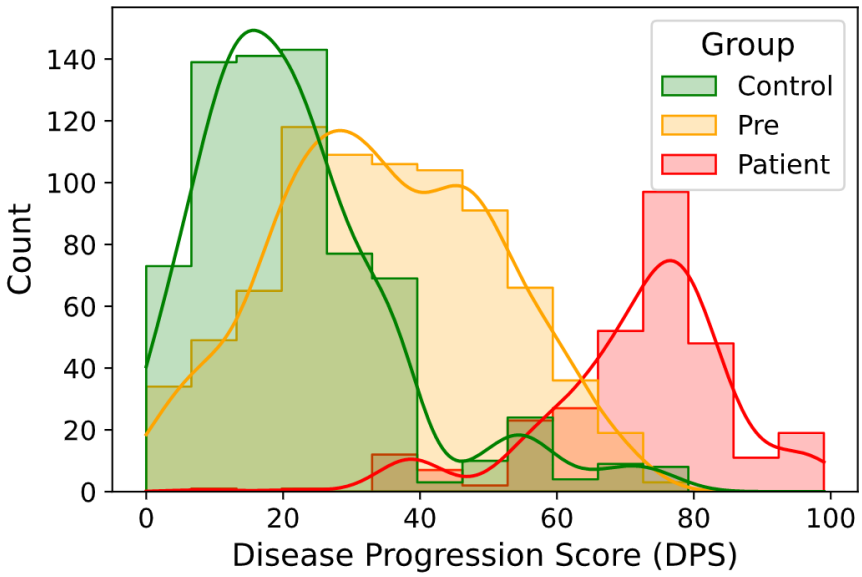


FIGURE 6.8: Results on real data. Histogram of the disease progression scores (DPS) inferred for 18 test subjects over 100 stratified splits. The distribution shapes are approximated with kernel density estimates.

Ablation study

Afterwards, to investigate the impact of certain components of our framework, we conducted an ablation study. We changed some elements of the proposed approach to obtain three alternative models:

- Linear instead of curved trajectory: rather than computing the trajectory in the latent space using principal curves, we simply used a straight line.
- No supervised branch: we removed the classification component of the loss function, thus performing unsupervised training.
- Joint low-dimensional modality: we concatenated the low-dimensional modality (demographic information) with the high-dimensional modalities (neuroimaging and miRNA expression) in the encoder input, and used only the latent codes as input for the classifier.

For each alternative model, we conducted the same cross-validation strategy of 100 stratified folds, computing the DPS for the test sets and the corresponding areas under the ROC curves. The results, displayed in Table 6.2, show that the proposed model has a better and more stable performance in all comparisons, with the highest average ROC AUC and lowest standard deviation among the splits.

Variation of hyperparameters

Finally, we checked whether our results were robust to reasonable changes in the hyperparameters. Notably, we tested different numbers of hidden units in the fully-connected layers, and different combinations of the relative weights between the

TABLE 6.2: Results on real data: ablation study. ROC AUC results (mean ± standard deviation) for the proposed model and three alternative models from the ablation study, respectively using a linear instead of a curved trajectory, removing the classification branch, and concatenating the low-dimensional modality with the high-dimensional ones.

Comparison	Proposed model	Linear trajectory	No supervision	Joint low-dim.
Control vs. Pre	**0.74 ± 0.13**	0.62 ± 0.15	0.67 ± 0.15	0.72 ± 0.15
Control vs. Patient	**0.98 ± 0.05**	0.93 ± 0.12	0.96 ± 0.06	0.95 ± 0.17
Pre vs. Patient	**0.96 ± 0.07**	0.93 ± 0.11	0.94 ± 0.10	0.91 ± 0.18

loss terms. These results are summarized in Table 6.3 and Table 6.4. The slightly different but overall similar results demonstrate that our hyperparameter choice is not overfitting the data.

TABLE 6.3: Results on real data. ROC AUC results (mean ± standard deviation) over 100 stratified splits when changing the number of units of the hidden layers. Original results, with 50 units, are shown in bold.

Hidden units	**50**	100	80	25
Control vs. Pre	**0.74 ± 0.13**	0.73 ± 0.13	0.71 ± 0.12	0.71 ± 0.13
Control vs. Patient	**0.98 ± 0.05**	0.98 ± 0.04	0.97 ± 0.05	0.98 ± 0.05
Pre vs. Patient	**0.96 ± 0.07**	0.96 ± 0.06	0.96 ± 0.06	0.96 ± 0.06

TABLE 6.4: Results on real data. ROC AUC results (mean ± standard deviation) over 100 stratified splits when changing the weights of the loss function terms. Original results, with α_1=0.2, α_2=0.2, α_3=0.6, are shown in bold.

Weights α_k	**0.2, 0.2, 0.6**	0.1, 0.1, 0.8	0.1, 0.2, 0.7	0.3, 0.2, 0.5
Control vs. Pre	**0.74 ± 0.13**	0.72 ± 0.12	0.73 ± 0.12	0.72 ± 0.14
Control vs. Patient	**0.98 ± 0.05**	0.97 ± 0.08	0.97 ± 0.06	0.98 ± 0.05
Pre vs. Patient	**0.96 ± 0.07**	0.94 ± 0.10	0.95 ± 0.09	0.96 ± 0.07

6.6 Discussion

In this paper, we proposed a new approach for estimating disease progression scores from cross-sectional neuroimaging and transcriptomic data that is applicable in small samples, which are typically found in rare diseases. The approach was designed and evaluated on data from *C9orf72*-associated FTD and ALS, but is potentially applicable to other diseases. Results on synthetic data demonstrated the ability of the method to accurately estimate the DPS, and experiments on real data, in the absence of ground truth DPS, showed the separation of different diagnostic classes. The findings of this study supported the usefulness of supervised variational autoencoders to infer disease trajectories from cross-sectional multimodal data, indicating that a

single disease progression score may be used to represent progression of neurodegenerative diseases. Remarkably, our results revealed that the DPS may be inferred using only cross-sectional data from a small sample of subjects.

Experiments with a cohort of *C9orf72*-mutation carriers demonstrate that subjects from the same clinical groups (patients, presymptomatic individuals and controls) are clustered together in the latent space (Fig. 6.6), allowing the inference of a disease trajectory. After training the model, data from new individuals is encoded in the latent space and orthogonally projected onto this trajectory to compute the DPS. Notably, using only the computed DPS, we are able to classify presymptomatic subjects and patients with an average ROC AUC of 0.96 over 100 stratified fold splits (Fig. 6.7). Of the three possible pairwise comparisons between clinical groups, this is the most relevant. It illustrates how much the DPS reflects the degree of disease progression in mutation carriers. Unsurprisingly, it is harder to differentiate between controls and presymptomatic individuals, as indicated by the average ROC AUC of 0.74 and displayed in Fig. 6.8. This stems from the fact that, during earlier disease stages, most biomarker levels are closer to normal ranges, so the presymptomatic class is more heterogeneous.

To the best our of knowledge, event-based models are the only published methods to compute disease progression scores from cross-sectional data, other approaches requiring longitudinal data. The comparison summarized in Table 6.1 reveals that our approach resulted in considerably higher ROC AUC than DEBM for all pairwise classifications. This suggests that the proposed approach is more suitable than event-based models for DPS computation with high-dimensional features, such as microRNA data. Indeed, published studies using event-based models explored a substantially lower number of features. For instance, in Alzheimer's disease, EBM experiments were carried our with 13 to 50 biomarkers (Archetti et al., 2019; Firth et al., 2020; Fonteijn et al., 2012; Oxtoby et al., 2018; Venkatraghavan et al., 2019; Young et al., 2014). Studies focusing on FTD analyzed 21 (Panman et al., 2021) or 7 (Ende et al., 2021) biomarkers, while multiple sclerosis was investigated with 25 (Dekker et al., 2020) or 24 (Eshaghi et al., 2018) biomarkers. Other conditions such as Parkinson's disease (Oxtoby et al., 2021), ALS (Gabel et al., 2020) and Huntington's disease (Wijeratne et al., 2021) were modeled with respectively 42, 19 and 8 biomarkers. Nevertheless, the EBM model presents useful additional features, beyond the computation of DPS. In particular, it can provide a temporal ordering of when the different biomarkers become abnormal, which is useful for understanding disease progression. Moreover, a balance has to be found between the number of features and the number of subjects in each dataset. Indeed, we also had to perform feature selection to decrease the number of microRNAs in our study. It should be noted that this feature selection was unbiased, since it was performed using a completely separate set of participants that was not used in the rest of the study. The proposed framework was able to achieve a good performance with 156 features and less than a hundred subjects, thus demonstrating its potential for dealing with

higher dimensional datasets.

An ablation study evaluated the impact of different components of our approach (Table 6.2). We observed that each component positively impacted the framework's performance. First, it can be seen that a curved trajectory better fits the disease pathway in the latent space when compared to a straight line. The use of principal curves has been inspired from their application in a similar task: pseudotime inference for single-cell transcriptomics, as shown in (Street et al., 2018). In that context, pseudotime represents an underlying temporal variable driving a smooth transition between cellular states, and principal curves are used to infer a trajectory in a low-dimensional space. Second, it is clear that the addition of supervision with a classifier branch improves the separation between clinical groups in the latent space. Rather than discrete clusters, our experiments demonstrate that latent representations are placed along a continuous path. Specifically, supervision adds meaning to the relative positions between points in the latent space. Finally, results show the contribution of splitting high and low-dimensional features. When using the low-dimensional features concatenated with the latent codes as inputs to the classifier, the model's performance is enhanced. The same pattern is observed in (Ji et al., 2021), although in a totally different context (failure detection in robotics). Concretely, a low-dimensional feature can directly contribute to the classifier, without the need for encoding.

Regarding the experiments with simulated datasets, it is crucial to highlight the relationship of the average ROC AUC with the Spearman correlation between ground truth and estimated DPS (Fig. 6.5). The simulation supports that the higher the ROC AUC, the more accurate the predicted DPS. Therefore, for real-world scenarios without ground truth DPS, our choice of the ROC AUC as proxy metric is corroborated. Furthermore, evidence was found that the models do not overfit the data, since it is clear that larger noise levels lead to poorer results, eventually equivalent to random chance. The effect of noise is further illustrated in Fig. 6.4. We observe that lower noise levels induce more evident clusters and more meaningful trajectories in the latent space. Consequently, the estimated DPS are closer to the ground truth. These simulations also confirm one intuition behind our model: the more features correlate with disease progression, the closer the estimated DPS are to the ground truth.

Our study has the following limitations. First, there is no ground truth for the progression scores in real datasets. Although the experiments with synthetic data showed that the ROC AUC is an adequate proxy metric, long-term follow-up of patients will be necessary to assess the accuracy of the computed DPS. For instance, we need follow-up data to confirm the hypothesis that a higher DPS implies an earlier disease onset for a presymptomatic subject. Another limitation was the lack of a replication cohort. This will be necessary to further support the clinical relevance of our findings. Future work will concentrate on the integration of more data sources, such as positron emission tomography (PET) scans and neurofilament light chain

(NfL) levels in blood.

In conclusion, we proposed a new approach to measure disease progression from multimodal imaging and microRNA data in rare neurodegenerative disorders using only cross-sectional data. Even though we focused on *C9orf72*-associated FTD and ALS, our framework is generic. It has the potential to be useful for a variety of other diseases, enabling the evaluation of novel treatments even when only cross-sectional data from small cohorts are available.

Chapter 7

Conclusion and perspectives

The goals of this interdisciplinary thesis, combining biomedical and machine learning research, were (1) to assess circulating microRNAs as progression biomarkers of genetic frontotemporal dementia and amyotrophic lateral sclerosis, and (2) to propose a method to estimate disease progression using cross-sectional multimodal data from small samples. Specifically, clinical trials testing novel treatments for FTD and ALS need non-invasive and robust biomarkers to measure disease progression in subjects without or with mild clinical symptoms. We thus conducted three studies. First, we analyzed the expression levels of plasma miRNAs in a cohort of *C9orf72* expansion carriers, to investigate whether circulating miRNAs are promising progression biomarkers of *C9orf72*-associated FTD and ALS. Second, we used independent homogeneous cohorts of *C9orf72* and *GRN* mutation carriers, to perform a comprehensive evaluation of all previously identified miRNA signatures associated with FTD and/or ALS and to determine their usefulness in assessing the progression of genetic FTD and ALS. Third, we designed, implemented, and tested a new generic model to estimate a disease progression score (DPS) from cross-sectional multimodal datasets with small sample sizes, and demonstrated its usefulness with miRNA and neuroimaging datasets obtained from a cohort of *C9orf72* expansion carriers.

The contributions of this thesis, as well as future directions for our work, are summarized below.

MicroRNAs are potential preclinical progression biomarkers of *C9orf72*-associated FTD and ALS

Before this thesis, several studies had observed circulating miRNAs as potential biomarkers of FTD and ALS. However, the lack of concordance between results undermined the reliability in miRNAs for assessing disease progression in clinical trials. Previous investigations were performed with heterogeneous cohorts (patients with sporadic disease forms, or mixed cohorts with sporadic and genetic disease forms) and prior assumptions about the subset of miRNAs to be analyzed. Our study presented in Chapter 4 was the first to evaluate the expression levels of plasma miRNAs in a cohort focused on *C9orf72* expansion carriers, without *a priori* assumptions. After performing large scale RNA-sequencing analyses in plasma samples from 110 individuals (22 patients, 45 presymptomatic subjects, and 43 controls),

we identified four miRNAs differentially expressed between groups: miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p. In the absence of an independent validation cohort, a generalization analysis demonstrated a satisfactory prediction performance when classifying subjects between clinical groups. Additionally, target prediction and pathway analyses identified relevant pathways, previously mentioned in the literature as involved in *C9orf72*-associated disease. This first study therefore highlighted the potential of circulating miRNAs as progression biomarkers of *C9orf72*-associated FTD and ALS.

Remarkably, our comprehensive validation study, described in Chapter 5, reinforced the usefulness of circulating miRNAs as biomarkers of *C9orf72*-associated disease. In this work, we evaluated all previously identified circulating miRNA signatures associated with FTD or ALS (from 15 papers, including our study from Chapter 4), using a homogeneous, independent validation cohort of *C9orf72* expansion carriers (29 patients, 17 presymptomatic carriers, and 31 controls). Results indicated an outstanding prediction performance (average ROC AUC greater than 0.90) when using the miRNA signature identified in Chapter 4 to classify controls vs. *C9orf72* patients and controls vs. *C9orf72* presymptomatic subjects. The most challenging problem was to classify *C9orf72* presymptomatic individuals vs. *C9orf72* patients, task for which the levels of miR-206 were found to be crucial. Interestingly, although miR-206 was not significantly differentially expressed in our discovery study from Chapter 4, it was the second most frequent miRNA identified in this comparison in the generalization analysis (Figure 4.3), with an adjusted p-value close to significance level (p-value = 0.06, Appendix A Table A2). Indeed, miR-206 illustrates the idea from (Bzdok et al., 2020), that features relevant for prediction are often not identified as statistically significant in biomedical datasets.

As a result of both our discovery and our validation studies, we encourage the use of plasma miRNAs as non-invasive progression biomarkers of *C9orf72*-associated FTD and ALS. In particular, the broader meta-signature proposed in Chapter 5 might be useful for *C9orf72*-disease progression assessment in clinical trials.

MicroRNA signatures identified in sporadic or mixed cohorts of FTD and ALS patients are useful as biomarkers of *C9orf72* disease, but not *GRN* disease

The extensive validation study presented in Chapter 5 demonstrated that more than half of the investigated miRNAs (35/65) were differentially expressed in at least one pairwise comparison in the *C9orf72* cohort. Moreover, when analyzing the prediction performance of the previously published miRNA signatures, half of them (8/16) resulted in an average ROC AUC above 0.70 for at least one classification task. Strikingly, the vast majority of the studied miRNA signatures had been identified in sporadic or mixed cohorts. Therefore, these findings indicated miRNA expression similarities between subjects with sporadic FTD or sporadic ALS and individuals carrying the *C9orf72* expansion.

Nevertheless, these miRNA expression similarities were not observed between previous studies on FTD and our validation cohort of *GRN* mutation carriers. Indeed, only a small minority (5/30) of miRNAs previously linked to FTD were found to be differently expressed in the *GRN* cohort. In addition, none of the five miRNA signatures investigated in this cohort exhibited an adequate classification performance. Therefore, our data suggest that the studied miRNAs are not useful biomarkers of *GRN*-associated FTD. Furthermore, our findings corroborate the heterogeneity of complex disorders such as FTD and ALS, indicating that miRNA expression profiles are most likely mutation specific.

This work illustrates the importance of validation studies in independent cohorts, which could be more encouraged and valued by the research community, particularly by scientific journals. In addition, it is noteworthy that the preregistration of our validation study has shown to be highly beneficial for several reasons, as thoroughly explained in (Nosek et al., 2018). First, preregistration imposes a detailed study design beforehand, thus avoiding false discoveries due to a possibly too flexible statistical analysis. Second, describing the details of the study before conducting the analyses clearly separates confirmatory and exploratory aspects, improving the credibility of results. Last, the few additional steps required to preregister a study on a registry such as The Open Science Framework[1] add very little burden, and actually save time afterwards during data analysis.

Disease progression scores can be estimated from cross-sectional neuroimaging and microRNA data from small samples

Several approaches have been proposed in the literature to model disease progression in a data-driven fashion, most of them depending on longitudinal data. Rare neurodegenerative disorders, such as FTD and ALS, could not be modeled with such approaches, for lack of longitudinal datasets with sufficient sample sizes. We thus presented (in Chapter 6) a novel framework to estimate disease progression using cross-sectional datasets from small samples. The main hypothesis of this work was that disease progression may be modeled as a latent trait: the disease pathophysiological pathway was represented by a trajectory in the latent space, and relative positions in this trajectory indicated disease progression scores. The main challenge in developing this framework was the absence of ground truth data for the DPS. Hence, we proposed a proxy metric to evaluate our models: classification performance (ROC AUC) was computed using the DPS as the only feature. Experiments with synthetic data corroborated the choice for the proxy metric, since the ROC AUC was positively correlated with the accuracy in estimating the simulated DPS. Moreover, analyses with a real multimodal dataset of miRNA and neuroimaging data demonstrated a good classification performance using only the inferred DPS as a feature: average ROC AUC of 0.74 for controls vs. presymptomatic subjects, 0.98 for

[1] https:// osf.io

controls vs. patients, and 0.96 for presymptomatic subjects vs. patients. The only disease progression models supporting cross-sectional data found in the literature were event-based models. When compared to our results, a discriminative event-based model yielded substantially lower performance (Table 6.1).

Although we used data from a cohort focused on *C9orf72* mutation carriers, the framework is generic and could be applied to other cross-sectional datasets. We believe that the proposed approach might be useful to infer a disease progression score for other rare disorders, for which longitudinal datasets with appropriate sample sizes are not available, and previously published methods cannot be applied.

Perspectives

There are several future directions for our work.

First, a straightforward path for future research is to analyze larger cohorts of *C9orf72* expansion carriers. It has been demonstrated that for predictive applications, such as biomarkers discovery and validation, larger samples lead to better generalization performance estimation (Varoquaux, 2018). For instance, with a larger miRNA expression dataset from *C9orf72* carriers, we could more precisely assess the generalization capabilities of the meta-signatures proposed in Chapter 5, or even identify other miRNA signatures with better prediction power. Moreover, training the model presented in Chapter 6 with larger multimodal datasets would improve DPS estimation on unseen data, an essential step towards deploying the model in a clinical setting. Furthermore, the investigation of a larger number of patients carrying the *C9orf72* expansion could shed a light on the differences in miRNA expression observed between different phenotypes (FTD, FTD/ALS, and ALS), and possibly improve the stratification of participants in clinical trials.

Since FTD and ALS are rare conditions, it is unlikely that a single entity (university, research institution, or hospital) can bring together a cohort with hundreds of patients and presymptomatic individuals. It is therefore crucial to investigate alternative strategies to increase sample size. One solution could be to explore federated learning, a decentralized machine learning paradigm that allows different institutions to share their medical datasets while preserving data protection and patient privacy (Rieke et al., 2020). Federated learning has recently received much attention, due to its potential benefits in machine learning for healthcare applications (Li et al., 2020; Prayitno et al., 2021). Further research could then be undertaken to deploy our proposed disease progression score model from Chapter 6 in a federated learning setting, with a considerably larger and more representative dataset, which would very likely lead to more robust results.

Additionally, as discussed in Chapter 5, the findings of our validation study showed that most miRNAs identified in the literature as potential FTD biomarkers are not relevant for measuring the progression of *GRN*-associated disease. However, since *GRN* mutation carriers were not included in any of the prior investigations,

further work is needed to establish whether other miRNAs may be useful in *GRN*-associated FTD. An independent cohort of subjects carrying the *GRN* mutation is required, so that one cohort is used to identify miRNAs as biomarker candidates (such as was performed in Chapter 4 for the *C9orf72* mutation), and another cohort is employed to validate these miRNAs.

Furthermore, adding new modalities to the DPS computation would most likely improve model performance, since different modalities bring complementary information about the disease process (Carreiro et al., 2015). Futures studies could include blood levels of neurofilaments (Saracino et al., 2021), features extracted from positron emission tomography (PET) scans (Meeter et al., 2017), diffusion tensor imaging (DTI) (Bertrand et al., 2018), or neurite orientation dispersion and density imaging (NODDI) (Wen et al., 2019). A relevant question would be to investigate which modalities offer the best prediction performance. Multiple DPS models could be trained with different combinations of modalities, and a proxy metric (ROC AUC) comparison could indicate which modalities are most strongly correlated with disease progression.

Lastly, the validation of our proposed framework for DPS computation (Chapter 6) will require longitudinal follow-up over a long period of time, to determine whether presymptomatic individuals with higher scores will have earlier disease onsets. Even though this thesis focused on two rare neurodegenerative disorders, further experiments with datasets from more prevalent conditions, such as Alzheimer's disease, could be performed to evaluate the accuracy of the estimated DPS. For instance, we could take cross-sectional baseline data from the Alzheimer's Disease Neuroimaging Initiative (ADNI)[2] dataset to estimate the DPS of individuals with mild cognitive impairment at baseline. Then, we could analyze longitudinal data to verify whether a higher DPS at baseline implies an earlier Alzheimer's disease onset. Additionally, training our model with baseline data from ADNI would allow a comparison with disease progression models that leverage longitudinal data (Section 2.5). Clearly, it is expected that an approach estimating disease progression from cross-sectional data will be less accurate than a method using longitudinal data (Jack et al., 2013). Nevertheless, this comparison would be important to determine *how much* less accurate is our DPS model, for different training sample sizes. The results of this research could further support the clinical relevance of disease progression models trained with cross-sectional data from small cohorts.

Concluding remarks

In summary, this interdisciplinary thesis has demonstrated that circulating miRNAs are useful as non-invasive biomarkers of genetic frontotemporal dementia and amyotrophic lateral sclerosis associated with the expansion in the *C9orf72* gene. Moreover, we have conceived and implemented a methodology for estimating disease

[2] https://adni.loni.usc.edu/

progression scores from cross-sectional multimodal data, and tested it with miRNA and neuroimaging features extracted from a cohort of *C9orf72* expansion carriers. Taken together, the findings of this thesis represent an encouraging step in improving the design of clinical trials for these rare but devastating neurodegenerative disorders.

Appendix A

Plasma microRNA signature in presymptomatic and symptomatic subjects with *C9orf72*-associated frontotemporal dementia and amyotrophic lateral sclerosis

This appendix is the supplementary material of the Chapter 4, published as a journal article in the *Journal of Neurology, Neurosurgery & Psychiatry* (Kmetzsch et al., 2021):

- **Kmetzsch, V.**, Anquetil, V., Saracino, D., Rinaldi, D., Camuzat, A., Gareau, T., Jornea, L., Forlani, S., Couratier, P., Wallon, D., Pasquier, F., Robil, N., de la Grange, P., Moszer, I., Le Ber, I., Colliot, O., Becker, E., PREV-DEMALS study group, "Plasma microRNA signature in presymptomatic and symptomatic subjects with C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis", *Journal of Neurology, Neurosurgery & Psychiatry*, 92(5):485-493 (2021). doi: 10.1136/jnnp-2020-324647 – hal-03046771.

SUPPLEMENTARY MATERIAL

Method A1. Neuropsychological protocol

The PREV-DEMALS cognitive evaluation included standardized neuropsychological tests to investigate all cognitive domains, and in particular frontal lobe functions. The scores were provided previously (Bertrand et al., 2018).

Briefly, global cognitive efficiency was evaluated by means of Mini-Mental State Examination (MMSE) and Mattis Dementia Rating Scale (MDRS). Frontal executive functions were assessed with Frontal Assessment Battery (FAB), forward and backward digit spans, Trail Making Test part A and B (TMT-A and TMT-B), Wisconsin Card Sorting Test (WCST), and Symbol-Digit Modalities test. Hayling Sentence Completion Test was used to assess cognitive inhibition. Ekman faces test and Faux-pas test evaluated emotional assessment and social cognition.

Episodic verbal memory was assessed with the Free and Cued Selective Reminding Test (FCSRT), whereas visual memory with the Benson figure recall and identification test. Boston Naming Test (BNT), phonological and semantic fluencies, and Camel & Cactus test were used to evaluate language skills. Visuospatial abilities were assessed by means of Benson figure copy and the cube analysis task from the Visual Object and Space Perception (VOSP) battery. Limb-kinetic and ideo-motor apraxias were evaluated using a French validated scale.

Method A2. Description of the four cases at the transitional stage

Four *C9orf72* presymptomatic carriers have developed frontal cognitive and/or behavioral changes and/or subtle motor signs/symptoms during a 3-year follow-up period, without fitting diagnostic criteria for FTD or ALS, suggesting they were in the transitional 'prodromal' phase at the moment of or just after their baseline visit. They are described below.

The case 1 was a right-handed 42-year-old female with 13 years of schooling. Neurological examination, behavioral scores (FBI: 0; AES 9/42), and cognitive scores were normal (MDRS 138/144; FAB 18/18; WCST 20/20; forward/backward digit spans: 7/4), except for isolated moderate decrease of the faux-pas test score (21/30). Ekman test score was normal (31/35). At follow-up three years later (45 years), cognitive scores exhibited attentional deficit (forward/backward digit spans: 5/3), perseverations and social cognition deficit (faux-pas test 21/30). The Ekman score was 28/35. Neurological examination revealed upper and lower limbs brisk reflexes with propagation of reflexes.

The case 2 was a left-handed 47-year-old male with 15 years of education. When included in the study, neurological examination and behavior evaluation were normal. Cognitive scores (MMSE: 30; MDRS 134/144; WCST: 20/20) and behavioral scores (FBI: 4; AES score: 2) were normal except for mild impairment of social cognition (faux-pas test: 19/30). The frontal adapted version of CDR®+NACC-FTLD scored 0.5 at baseline. At clinical follow-up evaluation, three years later (50 years), he exhibited inappropriate familiarity and joviality, without any other behavioral changes. The AES score mildly increased (9/42) and the CDR®+NACC-FTLD reached a score of 1.0. Cognitive tests revealed attentional deficit (forward/backward digit spans: 5/3), slow processing of information and decline in several cognitive scores, among which MDRS (130/144) and faux-pas test (18/30). Motor evaluation revealed cramps and rare fasciculations.

The case 3 was a right-handed 69-year-old male with 11 years of schooling. At inclusion, neurological and behavioral evaluations were normal (FBI: 6/42; AES: 13). MDRS

(135/142) and WCST (18/20) were normal but executive dysfunction (FAB: 12/18) and social cognition deficit (faux-pas test: 18/30, Ekman 29/35) were present. Two years later, he presented loss of interests and apathy (AES 19, cut-off >13), irritability, familiarity and disinhibition (FBI: 12), and imitation behavior. Cognitive tests exhibited attentional (forward/backward digit spans: 3/3) and planning difficulties. The WCST was scored 9/20. Scores of MDRS, FAB, and faux-pas test were stable. Ekman score was 26/35. Motor examination was normal.

The case 4 was a right-handed female of high education level (17 years). She was included in the study at age 64. Neurological examination was normal at inclusion, as well as behavior. The FBI scored 0 and the AES scored 8. Cognitive scores were within normal ranges or just above the lower limits according to her age. MDRS scored 138/144, MMSE 26/30 (2 errors in attention subtest) FAB 17/18, WCST 18/20, and TMT was completely normal (0 errors). Reevaluation at age 67, three years after her inclusion in the study, evidenced decline in several cognitive scores, with the occurrence of executive dysfunction, deficit in mental flexibility and perseverations. WCST scored 9/20, MMSE 24/30 and FAB 15/18. She made 10 errors at the TMT. She presented global slowness, scoring 12 at the AES apathy scale. Her relative did not mention any other behavioral disorders (FBI score: 0). At motor evaluation, she exhibited decreased Achilles tendon reflexes, cramps and rare fasciculations never noticed before.

Table A1. Clinical status and center proportion in each batch

Table A1. Number of subjects analyzed in each RNA-seq batch, by clinical status (above) and by center (below).

Batch	Clinical status			Total
	Control	Presymptomatic	Patient	
Batch 1	25 (39.7%)	27 (42.9%)	11 (17.4%)	63 (100%)
Batch 2	9 (37.5%)	10 (41.7%)	5 (20.8%)	24 (100%)
Batch 3	9 (39.1%)	8 (34.8%)	6 (26.1%)	23 (100%)

Batch	Center				Total
	Paris	Rouen	Limoges	Lille	
Batch 1	51 (81.0%)	5 (7.9%)	5 (7.9%)	2 (3.2%)	63 (100%)
Batch 2	18 (75%)	2 (8.3%)	3 (12.5%)	1 (4.2%)	24 (100%)
Batch 3	15 (65.2%)	3 (13.0%)	4 (17.4%)	1 (4.4%)	23 (100%)

Table A2. Complete output from EdgeR

Table A2. Complete output from EdgeR, for each pairwise comparison between clinical groups (control vs. presymptomatic, control vs. patient, presymptomatic vs. patient). The columns show the 589 analyzed miRNAs (above noise level), the log-fold change when comparing the clinical groups, the average log-counts per million, the unadjusted p-values and finally the adjusted p-values after Benjamini-Hochberg.

miRNA	log-fold change	log-counts per million	*p*-value	adjusted *p*-value
		Control vs. Presymptomatic		
miR-34a-5p	-1.433	4.676	5.251e-16	3.093e-13
miR-218-5p	1.475	1.231	0.000	0.081
miR-1250-5p	-0.310	0.691	0.001	0.209
miR-625-3p	0.495	8.394	0.002	0.228
miR-548au-5p	-0.275	14.473	0.005	0.458
miR-548am-c-o-5p	-0.283	14.295	0.005	0.458
miR-30e-5p	-0.153	10.655	0.006	0.458
miR-1307-5p	-0.241	5.708	0.006	0.458
miR-200c-3p	0.150	5.273	0.007	0.458
miR-5010-3p	0.206	1.408	0.008	0.458
miR-17-3p	-0.238	3.215	0.010	0.511
miR-361-5p	0.124	8.171	0.014	0.605
miR-199a-5p	-0.260	8.817	0.016	0.605
miR-125b-5p	0.245	8.080	0.017	0.605
miR-23b-5p	0.327	1.691	0.018	0.605
miR-548h-5p	-0.669	6.935	0.019	0.605
miR-29b-3p	-0.423	9.739	0.022	0.605
miR-10a-5p	0.179	7.108	0.026	0.605
miR-23b-3p	0.140	9.599	0.028	0.605
miR-32-5p	-0.280	5.289	0.029	0.605
miR-10a-3p	0.269	1.500	0.030	0.605
let-7g-3p	-0.359	0.220	0.031	0.605
miR-509-3p	0.470	2.288	0.032	0.605
miR-942-5p	0.140	4.976	0.033	0.605
miR-197-3p	0.207	6.745	0.033	0.605
miR-142-3p	-0.218	13.185	0.033	0.605
miR-21-3p	-0.316	3.394	0.034	0.605
miR-301a-3p	-0.186	5.377	0.034	0.605
miR-4742-3p	0.235	1.378	0.035	0.605
miR-5584-5p	0.286	0.672	0.037	0.605
miR-625-5p	0.278	6.770	0.038	0.605
miR-375	0.567	2.730	0.038	0.605
miR-330-5p	-0.111	3.616	0.041	0.605
miR-100-5p	0.466	2.346	0.041	0.605
miR-19a-3p	-0.286	5.810	0.043	0.605
miR-101-3p	-0.227	11.459	0.047	0.605
miR-582-5p	-0.498	0.885	0.047	0.605
miR-27a-3p	-0.306	9.739	0.048	0.605
miR-548y	-1.266	-0.318	0.055	0.605
miR-548aq-3p	-0.265	9.450	0.056	0.605
miR-153-3p	-0.360	0.310	0.056	0.605
miR-377-5p	-0.350	-0.095	0.057	0.605
miR-191-5p	0.116	12.565	0.057	0.605
miR-548t-5p	-0.407	1.936	0.057	0.605
miR-29a-3p	-0.401	9.472	0.059	0.605
let-7c-5p	0.178	9.528	0.060	0.605
miR-548ay-5p	-0.214	13.403	0.060	0.605
miR-215-5p	0.276	2.194	0.060	0.605
miR-874-3p	-0.160	1.960	0.061	0.605
miR-548d-5p	-0.210	12.885	0.061	0.605
miR-29c-3p	-0.338	8.568	0.061	0.605
miR-136-5p	-0.395	2.770	0.061	0.605
miR-19b-3p	-0.323	10.256	0.063	0.605
miR-487a-3p	-0.442	0.530	0.063	0.605
miR-548ag	-0.300	6.333	0.063	0.605
miR-548p	-0.785	1.823	0.064	0.605
miR-326	-0.197	6.590	0.064	0.605
miR-369-3p	-0.445	4.919	0.065	0.605
miR-551b-3p	-0.258	1.815	0.065	0.605
miR-30d-3p	-0.213	1.671	0.065	0.605
miR-135a-5p	-0.439	1.965	0.066	0.605
miR-33a-5p	-0.261	1.653	0.068	0.605
miR-181d-5p	0.171	3.947	0.068	0.605
miR-154-5p	-0.410	3.881	0.070	0.605
let-7d-3p	0.116	6.419	0.073	0.605
miR-376b-3p	-0.410	3.330	0.074	0.605
miR-548u	-0.576	1.210	0.074	0.605
miR-142-5p	-0.235	10.854	0.074	0.605
miR-6873-3p	0.297	5.014	0.075	0.605
miR-18a-5p	-0.137	7.212	0.076	0.605
miR-133b	-0.493	-0.138	0.079	0.605
miR-1260b	0.191	3.783	0.080	0.605
let-7i-3p	-0.231	2.370	0.081	0.605
miR-548j-3p	-0.355	0.393	0.081	0.605
miR-138-5p	-0.316	0.000	0.082	0.605

miR-148b-5p	-0.113	2.775	0.083	0.605
miR-31-5p	-0.473	2.641	0.085	0.605
miR-424-3p	0.218	0.968	0.086	0.605
miR-6511a-3p	0.198	4.865	0.086	0.605
miR-3194-3p	-0.580	0.002	0.088	0.605
miR-339-5p	-0.139	11.621	0.092	0.605
miR-106b-5p	-0.271	4.658	0.092	0.605
miR-181b-3p	-0.397	0.083	0.093	0.605
miR-125a-3p	0.248	0.692	0.095	0.605
miR-1306-3p	0.210	2.215	0.096	0.605
miR-598-3p	-0.256	4.565	0.097	0.605
miR-130b-5p	0.167	5.414	0.097	0.605
miR-574-3p	0.151	5.403	0.098	0.605
miR-548ah-3p	-0.782	2.757	0.098	0.605
miR-1249	0.346	0.351	0.099	0.605
miR-873-5p	-0.418	1.352	0.099	0.605
miR-496	-0.337	1.970	0.100	0.605
miR-590-5p	-0.316	0.651	0.100	0.605
miR-106b-3p	0.155	6.995	0.101	0.605
miR-18b-5p	-0.205	2.410	0.101	0.605
miR-7110-3p	0.366	-0.067	0.102	0.605
let-7e-5p	0.212	9.484	0.102	0.605
miR-223-5p	0.119	7.682	0.103	0.605
miR-337-5p	-0.311	3.963	0.104	0.605
miR-889-3p	-0.322	3.340	0.104	0.605
miR-1268a	-1.055	3.542	0.104	0.605
miR-377-3p	-0.306	2.438	0.105	0.606
miR-22-3p	-0.084	8.495	0.106	0.607
miR-205-5p	0.390	2.270	0.111	0.629
miR-324-5p	-0.103	5.454	0.115	0.638
miR-3688-3p	-0.209	1.689	0.116	0.638
miR-548j-5p	0.139	9.735	0.118	0.638
miR-6513-3p	0.097	1.835	0.119	0.638
miR-5583-3p	0.225	0.528	0.119	0.638
miR-660-5p	-0.149	5.628	0.121	0.638
miR-219a-5p	-0.278	1.164	0.123	0.638
miR-222-3p	0.115	8.603	0.123	0.638
miR-2277-5p	0.130	0.413	0.127	0.638
miR-340-5p	-0.197	6.743	0.128	0.638
miR-616-3p	0.219	0.543	0.129	0.638
miR-339-3p	-0.109	5.037	0.129	0.638
miR-95-3p	-0.247	2.108	0.130	0.638
miR-1304-3p	0.717	3.254	0.133	0.638
miR-136-3p	-0.399	3.534	0.133	0.638
miR-144-3p	-0.283	5.293	0.135	0.638
miR-548al	-0.842	-0.392	0.140	0.638
miR-4286	-0.132	10.665	0.141	0.638
miR-21-5p	-0.171	12.294	0.141	0.638
miR-369-5p	-0.229	5.217	0.142	0.638
miR-378i	0.126	4.394	0.145	0.638
miR-651-5p	-0.210	1.314	0.145	0.638
miR-1285-5p	-0.472	8.438	0.147	0.638
miR-338-3p	-0.175	5.028	0.150	0.638
miR-548q	0.214	4.336	0.150	0.638
miR-3198	0.313	0.544	0.151	0.638
miR-130a-3p	-0.142	7.658	0.154	0.638
miR-624-5p	-0.271	0.116	0.156	0.638
miR-7641	0.282	5.298	0.157	0.638
miR-30b-5p	-0.312	7.651	0.159	0.638
miR-664a-3p	0.115	4.261	0.160	0.638
miR-362-5p	0.134	2.457	0.161	0.638
miR-30a-5p	-0.126	6.368	0.161	0.638
miR-99a-5p	0.168	4.162	0.162	0.638
miR-376a-5p	-0.331	0.104	0.162	0.638
miR-301a-5p	-0.207	1.602	0.162	0.638
miR-655-3p	-0.392	-0.112	0.163	0.638
miR-582-3p	-0.297	1.716	0.164	0.638
miR-301b	-0.159	1.088	0.164	0.638
miR-320c	0.260	8.553	0.164	0.638
miR-381-3p	-0.233	5.537	0.165	0.638
miR-548f-3p	-0.289	5.197	0.165	0.638
miR-320d	0.219	7.345	0.166	0.638
miR-3913-5p	-0.139	2.448	0.169	0.638
miR-15a-5p	-0.146	9.211	0.169	0.638
miR-193a-5p	0.390	3.243	0.169	0.638
miR-497-5p	-0.174	1.523	0.171	0.638
miR-548ap-5p	0.118	9.773	0.172	0.638
miR-221-3p	0.096	11.513	0.174	0.638
miR-9-3p	-0.171	3.953	0.175	0.638
miR-664b-5p	0.211	1.474	0.176	0.638
miR-122-5p	0.452	11.047	0.177	0.638
miR-556-3p	-0.202	1.254	0.178	0.638
miR-93-3p	0.112	4.195	0.178	0.638
miR-299-3p	-0.322	0.096	0.178	0.638
miR-3138	0.224	2.724	0.179	0.638
miR-487a-5p	-0.238	0.384	0.181	0.638
let-7b-5p	0.170	12.514	0.182	0.638
miR-374b-3p	-0.174	1.235	0.183	0.638
miR-320b	0.252	8.897	0.184	0.638
miR-3605-3p	0.210	1.394	0.184	0.638
miR-3928-3p	0.166	0.931	0.186	0.638
miR-23c	0.113	0.292	0.187	0.638
miR-133a-3p	-0.278	7.345	0.187	0.638
miR-6852-5p	0.222	2.746	0.188	0.638
miR-18a-3p	0.146	2.363	0.189	0.638
miR-9-5p	-0.168	5.867	0.190	0.638
miR-548e-5p	-0.258	1.838	0.190	0.638

miR-27b-5p	0.133	1.148	0.191	0.638
miR-25-5p	0.181	2.573	0.192	0.638
miR-664b-3p	0.132	2.162	0.192	0.638
miR-199b-5p	-0.278	4.252	0.193	0.638
miR-154-3p	-0.327	0.419	0.193	0.638
miR-125a-5p	0.199	9.141	0.193	0.638
let-7f-1-3p	-0.225	0.491	0.194	0.638
miR-376c-5p	-0.379	1.518	0.196	0.640
miR-378g	0.251	0.027	0.198	0.640
let-7a-5p	0.131	17.835	0.198	0.640
miR-145-5p	0.172	6.318	0.199	0.640
miR-1273d	-1.451	4.498	0.201	0.643
miR-29b-1-5p	0.154	0.138	0.204	0.647
miR-146b-5p	-0.142	9.310	0.205	0.647
miR-103a-2-5p	-0.266	-0.340	0.209	0.647
miR-376b-5p	-0.370	1.487	0.210	0.647
miR-140-5p	-0.106	6.331	0.210	0.647
miR-411-5p	-0.225	5.491	0.211	0.647
miR-3620-5p	0.702	3.024	0.212	0.647
let-7f-2-3p	-0.256	0.250	0.213	0.647
miR-10b-5p	0.226	4.540	0.214	0.647
miR-941	0.156	9.475	0.214	0.647
miR-374a-3p	-0.192	2.987	0.217	0.647
miR-323b-3p	-0.201	4.932	0.219	0.647
miR-542-3p	-0.221	2.341	0.219	0.647
miR-3161	0.757	-0.951	0.221	0.647
miR-664a-5p	0.142	5.295	0.222	0.647
miR-885-5p	0.370	0.498	0.227	0.647
miR-92a-3p	0.174	13.677	0.227	0.647
miR-6859-5p	0.193	3.735	0.228	0.647
miR-1270	0.226	3.003	0.228	0.647
miR-378a-3p	0.098	9.082	0.229	0.647
miR-539-3p	-0.309	1.473	0.229	0.647
miR-376c-3p	-0.268	5.292	0.229	0.647
miR-1268b	-0.555	0.512	0.230	0.647
miR-766-5p	0.111	2.287	0.230	0.647
miR-3613-5p	0.223	9.506	0.230	0.647
miR-92b-5p	0.321	-0.009	0.232	0.647
miR-132-3p	0.091	3.504	0.233	0.647
miR-505-5p	0.165	3.558	0.234	0.647
miR-760	0.262	1.310	0.234	0.647
miR-4507	1.115	1.601	0.235	0.647
miR-3173-5p	0.234	1.011	0.245	0.665
miR-548aa	-0.271	10.786	0.245	0.665
miR-548t-3p	-0.271	10.786	0.245	0.665
miR-4772-3p	-0.289	1.699	0.247	0.665
miR-382-5p	-0.212	8.383	0.247	0.665
miR-5095	-1.849	8.523	0.251	0.670
miR-545-3p	-0.166	1.909	0.251	0.670
miR-361-3p	-0.075	6.970	0.253	0.672
miR-576-5p	0.107	6.414	0.255	0.675
miR-186-3p	-0.112	0.571	0.257	0.676
miR-155-5p	0.079	8.603	0.262	0.677
miR-590-3p	-0.212	6.770	0.262	0.677
miR-382-3p	-0.210	3.232	0.262	0.677
miR-744-5p	0.149	7.986	0.263	0.677
miR-5187-5p	0.170	2.108	0.263	0.677
miR-376a-3p	-0.239	6.560	0.265	0.677
miR-20a-5p	-0.097	9.087	0.266	0.677
miR-32-3p	-0.151	0.143	0.273	0.690
miR-545-5p	-0.199	1.482	0.274	0.690
miR-495-5p	-0.198	-0.270	0.279	0.690
miR-92b-3p	0.184	3.086	0.280	0.690
miR-409-5p	-0.166	3.705	0.280	0.690
miR-27a-5p	0.170	2.220	0.280	0.690
miR-671-3p	0.158	3.284	0.282	0.690
miR-2355-5p	-0.175	0.457	0.283	0.690
miR-619-5p	0.846	11.202	0.284	0.690
miR-641	-0.081	1.385	0.284	0.690
miR-98-3p	-0.192	0.779	0.285	0.690
miR-23a-5p	0.186	1.003	0.286	0.690
miR-221-5p	0.070	5.310	0.288	0.690
miR-589-5p	0.082	3.684	0.288	0.690
miR-6842-3p	0.103	1.371	0.288	0.690
miR-628-5p	-0.092	4.677	0.296	0.707
let-7d-5p	0.090	10.427	0.299	0.708
miR-30e-3p	-0.111	7.059	0.299	0.708
miR-454-5p	0.059	3.780	0.302	0.711
miR-548e-3p	-0.132	2.682	0.303	0.711
miR-99b-3p	0.151	2.773	0.308	0.719
miR-29a-5p	-0.115	0.831	0.312	0.719
miR-627-5p	-0.075	0.873	0.313	0.719
miR-544a	-0.367	0.669	0.315	0.719
miR-186-5p	-0.086	8.853	0.317	0.719
miR-365a-3p	0.189	4.151	0.318	0.719
miR-365b-3p	0.189	4.151	0.318	0.719
miR-548a-3p	-0.142	5.676	0.318	0.719
miR-618	0.314	-0.303	0.320	0.719
miR-1185-2-3p	-0.203	0.123	0.320	0.719
miR-548h-3p	-0.274	8.937	0.321	0.719
miR-548z	-0.274	8.937	0.321	0.719
let-7g-5p	-0.053	10.980	0.326	0.725
miR-6734-5p	0.166	0.555	0.326	0.725
miR-23a-3p	0.061	11.962	0.328	0.726
miR-103a-3p	0.051	15.095	0.330	0.726
miR-550a-3-5p	0.116	2.398	0.332	0.726
miR-548ak	-0.536	2.923	0.334	0.726

miR-758-3p	-0.186	1.734	0.335	0.726
miR-181a-5p	0.102	10.817	0.335	0.726
miR-4477b	-0.222	0.399	0.335	0.726
miR-5193	0.104	1.173	0.338	0.730
miR-134-5p	-0.187	7.091	0.340	0.732
miR-374a-5p	-0.135	7.594	0.342	0.732
miR-1277-3p	-0.146	1.948	0.345	0.734
miR-501-3p	0.147	2.746	0.345	0.734
miR-548an	-0.264	1.688	0.347	0.734
miR-28-3p	0.059	10.686	0.352	0.737
miR-1301-3p	0.130	5.720	0.353	0.737
miR-548d-3p	-0.277	1.169	0.354	0.737
miR-2115-3p	0.243	0.295	0.355	0.737
miR-1	-0.241	8.318	0.356	0.737
miR-423-3p	0.124	8.841	0.356	0.737
miR-450b-5p	-0.153	2.048	0.357	0.737
miR-1306-5p	0.240	2.765	0.361	0.743
miR-629-5p	0.097	6.404	0.363	0.743
miR-379-5p	-0.153	7.187	0.364	0.743
miR-7706	0.159	-0.037	0.366	0.743
miR-20a-3p	-0.083	1.922	0.366	0.743
miR-320a	0.145	11.779	0.371	0.743
miR-126-5p	-0.103	11.057	0.372	0.743
miR-148b-3p	-0.077	9.436	0.374	0.743
miR-548k	-0.082	1.443	0.377	0.743
miR-493-3p	-0.149	5.508	0.378	0.743
miR-548av-5p	-0.082	1.444	0.378	0.743
miR-4677-3p	-0.107	0.746	0.380	0.743
miR-99b-5p	0.118	7.865	0.380	0.743
miR-3615	0.175	5.538	0.381	0.743
miR-27b-3p	-0.088	10.051	0.381	0.743
miR-424-5p	-0.141	3.748	0.382	0.743
miR-199a-3p	-0.086	15.594	0.383	0.743
miR-199b-3p	-0.086	15.594	0.383	0.743
let-7a-3p	-0.155	4.497	0.384	0.743
miR-335-3p	0.107	5.333	0.385	0.743
miR-548a-5p	-0.212	1.849	0.390	0.746
miR-628-3p	0.072	5.661	0.391	0.746
miR-141-3p	-0.104	3.127	0.392	0.746
miR-429	-0.092	0.270	0.393	0.746
miR-30a-3p	-0.061	3.079	0.394	0.746
miR-192-5p	-0.088	7.507	0.394	0.746
miR-3143	-0.097	-0.186	0.395	0.746
miR-1273c	0.106	1.024	0.397	0.746
miR-296-5p	-0.139	1.358	0.398	0.746
miR-1277-5p	-0.124	12.138	0.402	0.749
miR-16-2-3p	-0.115	3.631	0.402	0.749
miR-1303	0.278	4.458	0.403	0.749
miR-4446-3p	0.120	3.069	0.408	0.756
miR-127-5p	-0.147	1.654	0.411	0.758
miR-210-3p	-0.070	2.393	0.414	0.758
miR-574-5p	0.407	14.568	0.415	0.758
miR-450a-5p	-0.176	0.853	0.415	0.758
miR-3200-3p	-0.171	0.193	0.417	0.758
miR-548f-5p	-0.382	0.762	0.419	0.758
miR-1307-3p	0.142	7.329	0.420	0.758
miR-548az-5p	-0.231	5.839	0.423	0.758
miR-3120-3p	-0.073	2.583	0.424	0.758
miR-500a-3p	0.065	4.766	0.425	0.758
miR-1294	0.127	1.946	0.426	0.758
miR-1226-3p	0.119	1.073	0.426	0.758
miR-1185-1-3p	-0.150	3.358	0.427	0.758
miR-6513-5p	0.074	0.012	0.428	0.758
miR-487b-3p	-0.132	5.402	0.428	0.758
miR-340-3p	-0.055	3.655	0.434	0.763
miR-7849-3p	0.090	0.247	0.434	0.763
miR-191-3p	0.057	3.805	0.439	0.763
miR-4659a-3p	0.100	0.124	0.442	0.763
miR-5585-3p	-0.861	6.719	0.443	0.763
miR-548ar-5p	-0.300	7.739	0.444	0.763
miR-106a-5p	-0.098	1.761	0.445	0.763
miR-548l	0.089	1.877	0.446	0.763
miR-3617-5p	0.124	0.442	0.448	0.763
miR-206	0.336	1.104	0.449	0.763
miR-26b-5p	-0.061	12.787	0.452	0.763
miR-4433b-3p	-0.174	2.351	0.452	0.763
miR-744-3p	-0.155	-0.457	0.455	0.763
miR-548ab	-0.320	4.211	0.455	0.763
miR-652-5p	-0.103	1.446	0.455	0.763
let-7b-3p	0.078	1.579	0.456	0.763
miR-548x-3p	-0.740	2.828	0.457	0.763
miR-539-5p	-0.189	0.162	0.458	0.763
miR-185-5p	-0.041	9.811	0.459	0.763
miR-183-5p	0.142	5.546	0.461	0.763
miR-548g-3p	-0.546	1.192	0.462	0.763
miR-409-3p	-0.123	8.698	0.462	0.763
miR-335-5p	-0.090	7.461	0.462	0.763
miR-1229-3p	0.115	0.765	0.466	0.763
miR-1255b-5p	0.121	7.593	0.466	0.763
miR-30b-3p	0.068	0.223	0.469	0.763
miR-1908-5p	0.183	2.300	0.470	0.763
miR-502-3p	0.065	3.800	0.470	0.763
miR-548aj-5p	-0.267	7.504	0.472	0.763
miR-548g-5p	-0.267	7.504	0.472	0.763
miR-548x-5p	-0.267	7.504	0.472	0.763
miR-1185-5p	-0.156	2.420	0.477	0.768
miR-6772-3p	0.116	0.690	0.478	0.768

miR-6500-3p	-0.300	-0.223	0.479	0.768
miR-15b-3p	-0.092	5.648	0.480	0.768
miR-328-3p	0.114	5.945	0.481	0.768
miR-1972	-0.999	5.510	0.486	0.770
miR-425-5p	0.050	10.913	0.486	0.770
miR-148a-5p	0.075	0.933	0.486	0.770
miR-331-3p	-0.064	4.398	0.490	0.771
miR-16-5p	-0.049	17.408	0.491	0.771
miR-181a-3p	-0.090	4.944	0.491	0.771
miR-654-3p	-0.114	7.125	0.495	0.775
miR-548n	-0.071	4.212	0.498	0.777
miR-548ap-3p	-0.159	10.374	0.501	0.777
miR-378a-5p	0.083	2.841	0.501	0.777
miR-379-3p	-0.124	1.468	0.503	0.777
miR-548av-3p	-0.207	3.903	0.504	0.777
miR-550a-5p	0.087	1.899	0.506	0.777
miR-378c	-0.064	5.108	0.508	0.777
miR-6803-3p	0.151	-0.112	0.509	0.777
miR-665	-0.139	0.677	0.510	0.777
miR-196a-5p	-0.149	2.899	0.511	0.777
let-7i-5p	0.041	13.259	0.511	0.777
miR-483-5p	0.252	1.834	0.512	0.777
miR-4659b-3p	-0.063	0.948	0.521	0.787
miR-378d	-0.241	0.023	0.524	0.787
miR-584-3p	-0.084	0.426	0.524	0.787
miR-145-3p	0.089	1.763	0.525	0.787
miR-182-5p	0.085	7.556	0.525	0.787
miR-671-5p	0.084	4.210	0.526	0.787
miR-548ax	-0.065	5.062	0.530	0.791
miR-1233-3p	-0.379	-0.544	0.533	0.792
miR-130b-3p	-0.041	5.063	0.534	0.792
miR-431-3p	-0.111	0.193	0.535	0.792
miR-25-3p	0.046	10.901	0.536	0.792
miR-548ae	-0.626	3.953	0.538	0.792
miR-6511b-3p	0.076	2.078	0.539	0.792
miR-584-5p	0.065	9.252	0.541	0.793
miR-20b-5p	-0.057	5.620	0.547	0.799
miR-548as-5p	-0.162	3.448	0.551	0.801
miR-224-5p	0.123	7.850	0.551	0.801
miR-1290	0.167	4.532	0.554	0.804
miR-642a-3p	0.118	0.484	0.556	0.804
miR-1179	0.056	1.832	0.558	0.804
miR-30d-5p	0.029	11.785	0.559	0.804
miR-28-5p	-0.035	6.656	0.560	0.804
miR-324-3p	-0.039	4.372	0.561	0.804
miR-139-3p	0.118	4.524	0.563	0.806
miR-150-3p	0.107	1.296	0.567	0.807
miR-181c-3p	-0.055	3.538	0.567	0.807
miR-423-5p	0.126	10.597	0.572	0.811
miR-452-5p	-0.106	2.671	0.574	0.811
miR-22-5p	-0.057	3.999	0.574	0.811
miR-4433b-5p	0.100	7.929	0.575	0.811
miR-485-3p	-0.133	5.831	0.585	0.821
miR-2110	0.100	3.338	0.589	0.821
miR-425-3p	-0.033	6.606	0.591	0.821
miR-126-3p	-0.032	14.192	0.591	0.821
miR-190a-5p	-0.068	5.571	0.592	0.821
miR-6741-3p	0.106	0.231	0.592	0.821
miR-431-5p	-0.086	7.576	0.593	0.821
miR-769-3p	0.075	0.674	0.594	0.821
miR-30c-1-3p	0.036	2.582	0.597	0.824
miR-548w	-0.140	5.896	0.601	0.827
miR-3158-3p	-0.118	1.499	0.609	0.836
miR-542-5p	0.072	1.306	0.612	0.836
miR-503-5p	0.056	3.907	0.614	0.836
miR-2355-3p	-0.038	3.959	0.615	0.836
miR-195-5p	-0.053	3.376	0.616	0.836
miR-495-3p	-0.093	3.399	0.616	0.836
miR-337-3p	-0.111	2.407	0.620	0.839
miR-29b-2-5p	0.031	1.605	0.621	0.839
miR-329-3p	-0.076	5.645	0.625	0.843
miR-17-5p	-0.053	8.427	0.629	0.844
miR-363-3p	-0.055	6.300	0.631	0.844
miR-4448	0.078	4.223	0.632	0.844
miR-5189-3p	0.140	0.435	0.632	0.844
miR-451a	-0.077	11.183	0.635	0.846
miR-342-5p	0.053	3.944	0.640	0.848
miR-139-5p	-0.070	6.068	0.640	0.848
miR-548at-5p	-0.088	4.401	0.641	0.848
miR-374b-5p	-0.064	6.920	0.647	0.854
miR-4301	-0.207	2.532	0.650	0.856
miR-30c-5p	-0.049	11.296	0.652	0.857
miR-4732-3p	0.134	3.928	0.654	0.858
miR-493-5p	-0.075	5.498	0.657	0.858
miR-548o-3p	0.153	2.775	0.658	0.858
miR-1273g-3p	-0.575	11.153	0.661	0.858
miR-190b	0.026	3.267	0.661	0.858
miR-421	0.021	5.199	0.663	0.858
miR-96-5p	-0.093	3.856	0.663	0.858
miR-194-5p	-0.055	7.337	0.666	0.860
miR-485-5p	-0.101	3.492	0.674	0.860
miR-1302	-0.295	2.800	0.674	0.860
miR-98-5p	0.029	8.594	0.675	0.860
miR-5009-5p	-0.076	0.291	0.676	0.860
miR-532-5p	-0.038	5.499	0.676	0.860
miR-26b-3p	-0.026	3.623	0.677	0.860
miR-3064-5p	0.036	0.924	0.678	0.860

miR-5010-5p	0.102	0.210	0.678	0.860
miR-203a	0.087	-0.173	0.682	0.860
miR-6511b-5p	0.183	1.158	0.683	0.860
miR-1304-5p	-0.282	0.167	0.683	0.860
miR-766-3p	-0.039	5.380	0.684	0.860
miR-494-3p	-0.073	5.857	0.686	0.861
miR-3180	-0.174	0.868	0.692	0.865
miR-3180-3p	-0.174	0.868	0.692	0.865
miR-4732-5p	0.115	3.999	0.701	0.875
miR-3187-3p	0.060	0.369	0.705	0.875
miR-4435	0.064	1.570	0.705	0.875
miR-1273h-5p	0.056	11.451	0.709	0.875
miR-483-3p	0.132	1.205	0.709	0.875
miR-1285-3p	-0.041	8.508	0.710	0.875
miR-144-5p	-0.078	7.018	0.710	0.875
miR-1255a	0.075	4.833	0.714	0.878
miR-4785	-0.038	0.526	0.716	0.878
miR-589-3p	0.028	1.137	0.718	0.878
miR-181b-5p	0.035	8.044	0.719	0.878
miR-196b-5p	0.023	5.658	0.720	0.878
miR-299-5p	-0.058	2.468	0.724	0.882
miR-223-3p	0.025	14.649	0.727	0.882
miR-486-3p	0.068	5.430	0.729	0.884
miR-204-5p	-0.062	1.457	0.732	0.884
miR-181a-2-3p	-0.025	5.645	0.733	0.884
miR-93-5p	0.017	11.309	0.735	0.885
miR-1273h-3p	0.044	7.277	0.740	0.889
miR-1299	-0.190	4.630	0.741	0.889
miR-548am-3p	-0.232	2.613	0.742	0.889
miR-143-5p	0.055	1.979	0.747	0.893
let-7f-5p	0.022	16.225	0.751	0.895
miR-514a-3p	-0.097	0.606	0.755	0.897
miR-491-5p	0.027	3.945	0.756	0.897
miR-504-5p	0.042	0.213	0.757	0.897
miR-1287-5p	0.025	3.050	0.758	0.897
miR-185-3p	0.034	4.607	0.762	0.898
miR-1292-5p	-0.057	0.211	0.764	0.898
miR-1468-5p	-0.027	0.552	0.767	0.898
miR-1180-3p	0.077	2.024	0.767	0.898
miR-152-3p	-0.025	7.210	0.768	0.898
miR-150-5p	-0.056	10.272	0.769	0.898
miR-652-3p	-0.020	7.109	0.773	0.898
miR-330-3p	-0.017	3.864	0.773	0.898
miR-432-5p	0.057	7.959	0.774	0.898
miR-4645-3p	0.028	0.811	0.775	0.898
miR-15b-5p	-0.030	9.487	0.781	0.904
miR-26a-5p	-0.017	15.391	0.785	0.905
miR-6770-3p	0.059	0.142	0.787	0.905
miR-486-5p	0.061	14.249	0.787	0.905
miR-6516-5p	0.030	1.586	0.788	0.905
miR-342-3p	0.030	10.283	0.790	0.905
miR-454-3p	0.017	8.563	0.793	0.906
miR-1233-5p	-0.164	-0.221	0.794	0.906
miR-146a-5p	-0.019	12.934	0.795	0.906
miR-484	0.039	7.854	0.800	0.909
miR-877-3p	-0.065	2.150	0.801	0.909
miR-3179	0.070	-0.272	0.809	0.915
miR-487b-5p	-0.043	0.849	0.809	0.915
miR-370-3p	-0.053	5.773	0.812	0.916
miR-1246	0.063	5.861	0.815	0.916
miR-3133	0.114	-0.564	0.815	0.916
miR-148a-3p	-0.022	10.002	0.817	0.916
miR-26a-2-3p	0.037	-0.153	0.818	0.916
miR-433-3p	-0.043	1.880	0.824	0.920
miR-3679-5p	0.030	2.384	0.827	0.922
miR-500b-5p	-0.036	2.709	0.829	0.922
miR-505-3p	0.011	4.359	0.833	0.922
miR-7-5p	-0.021	10.227	0.834	0.922
miR-24-3p	0.011	12.655	0.834	0.922
miR-3960	-0.107	3.467	0.834	0.922
miR-127-3p	0.032	7.007	0.838	0.925
miR-181c-5p	0.025	3.046	0.842	0.927
miR-128-3p	-0.012	10.346	0.846	0.927
miR-550a-3p	0.032	3.973	0.849	0.927
miR-151a-5p	0.017	11.573	0.849	0.927
miR-378f	0.021	1.820	0.850	0.927
miR-3065-5p	0.022	2.573	0.850	0.927
miR-146b-3p	-0.016	3.582	0.851	0.927
miR-151b	0.017	11.596	0.853	0.927
miR-200a-3p	0.016	2.514	0.855	0.927
miR-151a-3p	-0.009	12.424	0.863	0.927
miR-548i	-0.068	3.203	0.866	0.927
miR-4454	0.028	9.159	0.866	0.927
miR-19b-1-5p	-0.015	0.495	0.868	0.927
miR-4685-3p	0.034	0.669	0.868	0.927
miR-769-5p	-0.008	4.654	0.868	0.927
miR-6511a-5p	0.076	1.038	0.868	0.927
miR-548ai	0.104	0.527	0.870	0.927
miR-570-5p	0.104	0.527	0.870	0.927
miR-708-5p	-0.071	2.615	0.870	0.927
miR-329-5p	-0.031	2.442	0.874	0.929
miR-146a-3p	0.109	2.291	0.877	0.929
miR-548aq-5p	0.103	2.543	0.877	0.929
miR-4326	0.042	1.874	0.881	0.931
miR-345-5p	-0.012	4.607	0.884	0.931
miR-4662a-5p	-0.019	1.907	0.884	0.931
miR-532-3p	-0.018	2.783	0.887	0.931

miR-143-3p	0.024	9.208	0.887	0.931
miR-6859-3p	-0.050	0.719	0.890	0.932
miR-3065-3p	-0.018	0.404	0.892	0.932
miR-1271-5p	-0.013	1.919	0.893	0.932
miR-5096	-0.072	11.722	0.901	0.939
miR-412-5p	-0.039	2.206	0.910	0.947
miR-1271-3p	-0.063	-0.768	0.914	0.948
miR-548b-5p	-0.016	7.793	0.915	0.948
miR-132-5p	0.010	0.644	0.917	0.948
miR-543	-0.015	3.379	0.919	0.948
miR-3074-5p	0.018	1.854	0.920	0.948
miR-576-3p	0.008	5.501	0.922	0.948
miR-3140-3p	0.011	0.699	0.923	0.948
miR-3180-5p	0.037	-0.295	0.924	0.948
miR-323a-3p	-0.014	4.078	0.927	0.949
miR-1289	0.028	-0.148	0.932	0.953
miR-7976	-0.008	1.801	0.940	0.959
miR-877-5p	-0.005	2.378	0.941	0.959
miR-107	-0.004	10.028	0.947	0.963
miR-7851-3p	-0.013	3.495	0.953	0.968
miR-29c-5p	-0.004	4.579	0.955	0.968
miR-7-1-3p	-0.006	2.452	0.959	0.971
miR-140-3p	-0.002	9.249	0.971	0.981
miR-500a-5p	-0.005	3.026	0.975	0.981
miR-200b-3p	-0.003	2.188	0.976	0.981
miR-26a-1-3p	0.002	2.985	0.978	0.981
miR-1296-5p	-0.002	2.226	0.978	0.981
miR-551a	-0.003	-0.321	0.988	0.989
miR-548c-3p	0.000	2.905	1.000	1.000

miRNA	**log-fold change**	**log-counts per million**	***p*-value**	**adjusted *p*-value**
Control vs. Patient				
miR-34a-5p	-1.239	4.676	1.650E-08	9.720E-06
miR-345-5p	-0.540	4.607	1.131E-05	0.003
miR-200c-3p	0.333	5.273	3.109E-05	0.006
miR-10a-3p	0.697	1.500	7.141E-05	0.011
miR-151a-5p	0.411	11.573	0.001	0.096
miR-151b	0.406	11.596	0.001	0.096
miR-551a	-1.102	-0.321	0.001	0.096
miR-133a-3p	-0.949	7.345	0.002	0.119
miR-206	-1.883	1.104	0.002	0.119
miR-133b	-1.200	-0.138	0.002	0.123
miR-542-5p	-0.636	1.306	0.003	0.161
miR-296-5p	-0.637	1.358	0.006	0.262
miR-660-5p	-0.378	5.628	0.006	0.262
miR-18a-3p	-0.440	2.363	0.007	0.262
miR-223-3p	-0.276	14.649	0.007	0.262
miR-205-5p	-1.031	2.270	0.008	0.262
miR-361-3p	-0.251	6.970	0.008	0.262
let-7b-3p	-0.407	1.579	0.008	0.262
miR-30a-5p	-0.334	6.368	0.009	0.262
miR-93-3p	-0.315	4.195	0.009	0.262
let-7f-5p	0.246	16.225	0.011	0.296
miR-532-3p	-0.480	2.783	0.011	0.296
miR-144-3p	-0.677	5.293	0.012	0.296
miR-425-3p	-0.215	6.606	0.012	0.300
miR-550a-3p	-0.641	3.973	0.013	0.300
miR-10a-5p	0.282	7.108	0.013	0.300
miR-17-5p	0.363	8.427	0.014	0.314
miR-451a	-0.569	11.183	0.016	0.326
miR-7976	-0.396	1.801	0.016	0.326
miR-196b-5p	0.212	5.658	0.017	0.326
miR-135a-5p	-0.682	1.965	0.017	0.326
miR-106b-3p	-0.322	6.995	0.018	0.340
miR-148a-5p	-0.354	0.933	0.021	0.375
miR-25-3p	-0.248	10.901	0.022	0.377
miR-1289	-1.046	-0.148	0.023	0.385
miR-324-3p	-0.214	4.372	0.024	0.387
miR-5583-3p	0.460	0.528	0.024	0.387
miR-30b-3p	0.295	0.223	0.026	0.391
miR-23b-3p	0.194	9.599	0.028	0.391
miR-326	-0.326	6.590	0.028	0.391
miR-3158-3p	-0.743	1.499	0.028	0.391
miR-146a-5p	0.214	12.934	0.029	0.391
miR-532-5p	-0.283	5.499	0.029	0.391
miR-1303	-1.127	4.458	0.029	0.391
miR-378a-5p	-0.396	2.841	0.030	0.391
miR-16-5p	-0.218	17.408	0.032	0.403
miR-155-5p	0.215	8.603	0.032	0.403
miR-3615	-0.647	5.538	0.033	0.403
miR-26b-5p	0.235	12.787	0.034	0.403
miR-140-3p	-0.168	9.249	0.036	0.418
miR-331-3p	0.268	4.398	0.036	0.418
miR-484	-0.463	7.854	0.037	0.418
miR-4785	-0.304	0.526	0.038	0.418
miR-374a-3p	0.435	2.987	0.040	0.432
miR-224-5p	-0.584	7.850	0.040	0.432
miR-502-3p	-0.265	3.800	0.042	0.440
miR-26b-3p	-0.179	3.623	0.043	0.440
miR-769-5p	-0.130	4.654	0.044	0.443
miR-412-5p	1.005	2.206	0.045	0.443
miR-4286	-0.248	10.665	0.045	0.443
miR-6513-5p	0.252	0.012	0.047	0.443
miR-339-5p	-0.233	11.621	0.047	0.443
miR-30e-5p	-0.151	10.655	0.048	0.443
miR-28-5p	0.161	6.656	0.049	0.443

miR-183-5p	-0.553	5.546	0.049	0.443
miR-664a-3p	0.222	4.261	0.051	0.443
miR-339-3p	-0.196	5.037	0.051	0.443
miR-618	-0.905	-0.303	0.051	0.443
miR-363-3p	-0.320	6.300	0.052	0.443
miR-5584-5p	0.373	0.672	0.053	0.443
miR-130a-3p	0.257	7.658	0.054	0.444
miR-192-5p	-0.286	7.507	0.055	0.444
miR-6513-3p	0.167	1.835	0.055	0.444
miR-203a	-0.524	-0.173	0.056	0.445
miR-642a-3p	-0.565	0.484	0.057	0.448
miR-139-5p	0.384	6.068	0.058	0.452
miR-148b-3p	0.223	9.436	0.060	0.455
miR-340-3p	0.180	3.655	0.060	0.455
miR-96-5p	-0.563	3.856	0.062	0.455
miR-10b-5p	-0.515	4.540	0.062	0.455
miR-514a-3p	0.836	0.606	0.065	0.456
miR-374a-5p	0.353	7.594	0.066	0.456
miR-550a-3-5p	-0.320	2.398	0.068	0.456
miR-196a-5p	-0.585	2.899	0.068	0.456
miR-1233-5p	-1.587	-0.221	0.069	0.456
miR-26a-5p	0.158	15.391	0.069	0.456
miR-1180-3p	-0.699	2.024	0.070	0.456
miR-627-5p	-0.189	0.873	0.070	0.456
miR-486-5p	-0.615	14.249	0.071	0.456
miR-98-5p	0.169	8.594	0.071	0.456
miR-3679-5p	-0.350	2.384	0.071	0.456
miR-4685-3p	-0.526	0.669	0.072	0.456
miR-629-5p	-0.272	6.404	0.073	0.456
miR-365a-3p	-0.506	4.151	0.075	0.456
miR-365b-3p	-0.506	4.151	0.075	0.456
miR-370-3p	-0.548	5.773	0.075	0.456
miR-22-3p	-0.129	8.495	0.075	0.456
miR-154-3p	0.617	0.419	0.081	0.478
miR-4732-3p	-0.787	3.928	0.081	0.478
miR-3120-3p	0.215	2.583	0.081	0.478
miR-26a-1-3p	0.163	2.985	0.084	0.485
miR-664b-3p	0.243	2.162	0.084	0.485
miR-186-3p	-0.234	0.571	0.086	0.492
miR-146b-3p	-0.205	3.582	0.087	0.492
miR-500b-5p	-0.405	2.709	0.089	0.498
miR-941	-0.319	9.475	0.090	0.498
miR-500a-5p	-0.362	3.026	0.093	0.506
miR-29b-2-5p	0.145	1.605	0.093	0.506
miR-590-3p	0.419	6.770	0.096	0.506
miR-7706	-0.429	-0.037	0.096	0.506
miR-150-5p	-0.472	10.272	0.097	0.506
miR-454-3p	0.150	8.563	0.097	0.506
miR-1277-5p	0.339	12.138	0.098	0.506
miR-665	-0.484	0.677	0.098	0.506
miR-548u	0.763	1.210	0.101	0.508
miR-93-5p	-0.115	11.309	0.101	0.508
miR-628-5p	0.195	4.677	0.102	0.508
miR-550a-5p	-0.314	1.899	0.103	0.508
miR-548j-5p	0.202	9.735	0.103	0.508
miR-548ap-5p	0.197	9.773	0.103	0.508
miR-204-5p	0.418	1.457	0.105	0.513
miR-1296-5p	-0.158	2.226	0.107	0.516
miR-4677-3p	0.267	0.746	0.109	0.518
miR-652-5p	0.301	1.446	0.109	0.518
miR-199b-5p	-0.489	4.252	0.111	0.523
miR-103a-3p	0.115	15.095	0.113	0.524
miR-425-5p	-0.160	10.913	0.114	0.524
miR-486-3p	-0.452	5.430	0.115	0.524
miR-500a-3p	-0.183	4.766	0.117	0.524
miR-4645-3p	0.212	0.811	0.117	0.524
miR-539-5p	0.517	0.162	0.119	0.524
miR-589-5p	-0.171	3.684	0.119	0.524
miR-769-3p	-0.317	0.674	0.119	0.524
miR-495-3p	0.390	3.399	0.119	0.524
miR-1277-3p	0.327	1.948	0.121	0.524
miR-548am-c-o-5p	0.214	14.295	0.122	0.524
miR-92a-3p	-0.321	13.677	0.122	0.524
miR-194-5p	-0.285	7.337	0.124	0.530
miR-21-3p	-0.308	3.394	0.126	0.534
miR-548au-5p	0.205	14.473	0.127	0.535
miR-338-3p	-0.257	5.028	0.128	0.535
miR-766-5p	0.195	2.287	0.131	0.538
miR-335-5p	0.248	7.461	0.134	0.538
miR-152-3p	0.169	7.210	0.134	0.538
miR-199a-3p	0.198	15.594	0.136	0.538
miR-199b-3p	0.198	15.594	0.136	0.538
miR-616-3p	-0.313	0.543	0.136	0.538
miR-26a-2-3p	0.331	-0.153	0.136	0.538
miR-874-3p	-0.175	1.960	0.136	0.538
miR-539-3p	0.506	1.473	0.137	0.538
miR-150-3p	-0.405	1.296	0.141	0.546
miR-32-3p	0.273	0.143	0.141	0.546
miR-181b-5p	-0.199	8.044	0.142	0.548
miR-424-5p	-0.337	3.748	0.143	0.549
miR-491-5p	-0.180	3.945	0.147	0.554
miR-1292-5p	-0.399	0.211	0.147	0.554
miR-27a-5p	0.329	2.220	0.148	0.554
miR-590-5p	0.371	0.651	0.149	0.554
miR-92b-3p	-0.346	3.086	0.151	0.556
miR-3200-3p	-0.439	0.193	0.153	0.556
miR-501-3p	-0.323	2.746	0.154	0.556

miR-1273d	-1.478	4.498	0.155	0.556
miR-16-2-3p	-0.272	3.631	0.156	0.556
miR-340-5p	0.244	6.743	0.158	0.556
miR-382-3p	0.367	3.232	0.159	0.556
let-7f-1-3p	-0.336	0.491	0.159	0.556
miR-185-3p	-0.222	4.607	0.159	0.556
let-7i-3p	-0.254	2.370	0.160	0.556
miR-431-3p	-0.343	0.193	0.161	0.556
miR-375	-0.564	2.730	0.161	0.556
miR-3180-5p	-0.853	-0.295	0.161	0.556
miR-545-5p	0.338	1.482	0.165	0.564
miR-1273h-3p	-0.261	7.277	0.166	0.564
miR-548ab	0.855	4.211	0.168	0.566
miR-30c-1-3p	0.129	2.582	0.169	0.566
miR-92b-5p	-0.536	-0.009	0.169	0.566
miR-1255a	0.392	4.833	0.171	0.570
miR-423-5p	-0.444	10.597	0.175	0.574
miR-576-5p	-0.183	6.414	0.175	0.574
miR-126-3p	0.109	14.192	0.176	0.574
miR-1307-5p	-0.161	5.708	0.177	0.576
miR-3180	-0.899	0.868	0.180	0.578
miR-3180-3p	-0.899	0.868	0.180	0.578
miR-423-3p	-0.252	8.841	0.181	0.578
miR-766-3p	-0.179	5.380	0.183	0.582
miR-378d	-0.691	0.023	0.184	0.582
miR-7-5p	0.181	10.227	0.190	0.598
miR-138-5p	-0.323	0.000	0.191	0.600
miR-548am-3p	1.170	2.613	0.194	0.604
miR-106a-5p	0.226	1.761	0.196	0.606
miR-548ay-5p	-0.201	13.403	0.197	0.607
miR-378i	-0.158	4.394	0.200	0.614
miR-582-3p	-0.385	1.716	0.201	0.614
miR-30d-5p	-0.090	11.785	0.203	0.614
miR-25-5p	-0.254	2.573	0.205	0.614
miR-3928-3p	-0.227	0.931	0.206	0.614
miR-221-5p	0.114	5.310	0.209	0.614
let-7b-5p	-0.226	12.514	0.213	0.614
miR-5193	0.191	1.173	0.214	0.614
miR-3688-3p	-0.227	1.689	0.214	0.614
miR-223-5p	-0.127	7.682	0.214	0.614
miR-2115-3p	-0.478	0.295	0.215	0.614
miR-4732-5p	-0.560	3.999	0.216	0.614
miR-1908-5p	-0.442	2.300	0.217	0.614
miR-7641	-0.356	5.298	0.218	0.614
miR-582-5p	-0.430	0.885	0.219	0.614
miR-641	0.128	1.385	0.219	0.614
miR-5009-5p	-0.330	0.291	0.220	0.614
miR-190b	-0.101	3.267	0.221	0.614
miR-3617-5p	0.260	0.442	0.221	0.614
miR-9-3p	0.209	3.953	0.222	0.614
miR-6511a-3p	0.196	4.865	0.222	0.614
miR-4446-3p	0.245	3.069	0.222	0.614
miR-1233-3p	1.024	-0.544	0.223	0.614
miR-29a-3p	-0.360	9.472	0.227	0.619
miR-1250-5p	-0.155	0.691	0.228	0.619
let-7e-5p	0.217	9.484	0.229	0.619
miR-548d-5p	-0.185	12.885	0.229	0.619
miR-487a-3p	0.382	0.530	0.237	0.638
miR-193a-5p	-0.485	3.243	0.243	0.648
miR-548d-3p	0.454	1.169	0.243	0.648
miR-130b-5p	0.165	5.414	0.246	0.648
miR-497-5p	-0.196	1.523	0.248	0.648
miR-1972	2.620	5.510	0.249	0.648
miR-6873-3p	-0.289	5.014	0.249	0.648
miR-4448	-0.265	4.223	0.249	0.648
miR-624-5p	-0.302	0.116	0.250	0.648
miR-374b-3p	0.198	1.235	0.254	0.655
miR-20b-5p	-0.151	5.620	0.254	0.655
miR-495-5p	-0.277	-0.270	0.260	0.665
miR-301b	0.174	1.088	0.261	0.665
miR-24-3p	-0.078	12.655	0.264	0.665
miR-671-5p	-0.210	4.210	0.264	0.665
miR-548c-3p	-0.556	2.905	0.266	0.665
miR-429	-0.167	0.270	0.267	0.665
miR-2110	-0.296	3.338	0.267	0.665
miR-7849-3p	0.175	0.247	0.268	0.665
miR-19b-3p	-0.263	10.256	0.270	0.668
miR-330-3p	0.090	3.864	0.272	0.670
miR-376a-5p	-0.350	0.104	0.274	0.672
let-7g-3p	-0.248	0.220	0.278	0.676
miR-3613-5p	-0.275	9.506	0.280	0.676
miR-545-3p	0.217	1.909	0.281	0.676
let-7g-5p	0.080	10.980	0.283	0.676
miR-15a-5p	-0.157	9.211	0.283	0.676
miR-181d-5p	0.136	3.947	0.285	0.676
miR-1307-3p	-0.265	7.329	0.286	0.676
miR-126-5p	0.166	11.057	0.286	0.676
miR-574-3p	0.133	5.403	0.288	0.676
miR-221-3p	0.102	11.513	0.288	0.676
miR-362-5p	-0.142	2.457	0.288	0.676
miR-130b-3p	0.095	5.063	0.290	0.678
miR-1271-5p	-0.140	1.919	0.291	0.679
miR-548e-5p	-0.272	1.838	0.297	0.688
miR-3065-3p	-0.179	0.404	0.300	0.691
miR-29a-5p	0.160	0.831	0.300	0.691
miR-95-3p	-0.236	2.108	0.303	0.695
miR-30e-3p	0.149	7.059	0.305	0.695

miR-31-5p	-0.387	2.641	0.305	0.695
miR-483-5p	-0.617	1.834	0.315	0.712
miR-144-5p	-0.293	7.018	0.316	0.712
miR-5010-5p	-0.365	0.210	0.318	0.714
miR-320a	-0.232	11.779	0.320	0.717
miR-1285-3p	-0.156	8.508	0.322	0.718
miR-5096	-0.816	11.722	0.324	0.718
miR-378a-3p	-0.114	9.082	0.324	0.718
miR-1246	-0.379	5.861	0.326	0.718
miR-320b	-0.272	8.897	0.327	0.718
miR-5585-3p	-1.718	6.719	0.329	0.720
miR-548a-5p	0.308	1.849	0.334	0.728
miR-6516-5p	0.140	1.586	0.335	0.728
miR-1468-5p	0.124	0.552	0.340	0.733
miR-320c	-0.261	8.553	0.340	0.733
miR-3143	-0.152	-0.186	0.344	0.740
miR-378c	-0.129	5.108	0.346	0.740
miR-18b-5p	-0.162	2.410	0.348	0.741
miR-191-5p	0.079	12.565	0.351	0.741
miR-100-5p	-0.311	2.346	0.351	0.741
miR-186-5p	-0.110	8.853	0.351	0.741
miR-432-5p	-0.256	7.959	0.352	0.741
miR-125b-5p	0.136	8.080	0.356	0.742
miR-6852-5p	-0.223	2.746	0.357	0.742
miR-143-5p	0.225	1.979	0.357	0.742
let-7f-2-3p	-0.257	0.250	0.358	0.742
miR-504-5p	0.175	0.213	0.361	0.743
miR-1185-2-3p	0.249	0.123	0.361	0.743
miR-1271-3p	0.667	-0.768	0.367	0.749
miR-671-3p	-0.187	3.284	0.368	0.749
miR-1255b-5p	-0.217	7.593	0.368	0.749
miR-125a-3p	-0.189	0.692	0.369	0.749
miR-1294	-0.198	1.946	0.378	0.760
miR-376a-3p	0.259	6.560	0.378	0.760
miR-548f-5p	-0.617	0.762	0.378	0.760
miR-320d	-0.202	7.345	0.381	0.761
miR-548y	-0.913	-0.318	0.382	0.761
miR-548e-3p	0.154	2.682	0.383	0.761
miR-760	-0.272	1.310	0.384	0.761
miR-450b-5p	-0.205	2.048	0.385	0.762
let-7a-3p	-0.213	4.497	0.387	0.762
miR-376b-3p	0.260	3.330	0.392	0.770
miR-431-5p	-0.187	7.576	0.394	0.771
miR-136-5p	-0.235	2.770	0.397	0.771
miR-151a-3p	0.064	12.424	0.398	0.771
miR-29b-3p	-0.212	9.739	0.398	0.771
miR-182-5p	-0.158	7.556	0.401	0.774
miR-15b-3p	-0.148	5.648	0.407	0.780
miR-548ax	0.114	5.062	0.408	0.780
miR-20a-3p	0.103	1.922	0.408	0.780
miR-548as-5p	0.295	3.448	0.411	0.784
miR-1290	-0.334	4.532	0.418	0.795
miR-452-5p	-0.198	2.671	0.420	0.795
miR-1299	-0.665	4.630	0.421	0.795
miR-374b-5p	0.150	6.920	0.430	0.810
miR-4507	1.007	1.601	0.434	0.810
miR-30c-5p	0.117	11.296	0.436	0.810
miR-376c-3p	0.233	5.292	0.439	0.810
miR-454-5p	-0.061	3.780	0.440	0.810
miR-199a-5p	-0.111	8.817	0.442	0.810
miR-1304-5p	0.685	0.167	0.444	0.810
miR-6772-3p	0.171	0.690	0.445	0.810
miR-4772-3p	-0.276	1.699	0.445	0.810
miR-483-3p	-0.400	1.205	0.448	0.810
miR-20a-5p	0.089	9.087	0.448	0.810
miR-369-3p	0.240	4.919	0.449	0.810
miR-411-5p	0.183	5.491	0.449	0.810
miR-23a-5p	-0.189	1.003	0.449	0.810
miR-873-5p	-0.269	1.352	0.450	0.810
miR-4301	-0.455	2.532	0.451	0.810
let-7c-5p	0.099	9.528	0.455	0.815
miR-29c-3p	-0.183	8.568	0.461	0.816
miR-329-3p	-0.157	5.645	0.463	0.816
miR-27a-3p	-0.152	9.739	0.466	0.816
miR-1229-3p	0.163	0.765	0.466	0.816
let-7d-3p	-0.066	6.419	0.467	0.816
miR-3138	-0.170	2.724	0.471	0.816
miR-181a-5p	-0.106	10.817	0.473	0.816
miR-17-3p	-0.089	3.215	0.473	0.816
miR-548g-3p	0.711	1.192	0.473	0.816
miR-153-3p	0.182	0.310	0.474	0.816
miR-1306-5p	-0.256	2.765	0.479	0.816
miR-450a-5p	-0.216	0.853	0.480	0.816
miR-1273g-3p	-1.238	11.153	0.481	0.816
miR-548aa	-0.229	10.786	0.481	0.816
miR-548t-3p	-0.229	10.786	0.481	0.816
miR-548f-3p	0.207	5.197	0.481	0.816
miR-503-5p	-0.112	3.907	0.481	0.816
miR-548ap-3p	-0.231	10.374	0.482	0.816
miR-142-5p	-0.126	10.854	0.482	0.816
miR-1260b	0.105	3.783	0.485	0.816
miR-181a-2-3p	0.073	5.645	0.486	0.816
miR-1268a	-0.657	3.542	0.487	0.816
miR-4326	-0.270	1.874	0.489	0.816
miR-154-5p	-0.214	3.881	0.490	0.816
miR-143-3p	0.167	9.208	0.490	0.816
miR-9-5p	0.117	5.867	0.493	0.816

miR-342-3p	-0.109	10.283	0.493	0.816
miR-210-3p	-0.083	2.393	0.495	0.816
miR-548a-3p	0.130	5.676	0.496	0.816
miR-3194-3p	-0.309	0.002	0.498	0.817
miR-2355-5p	-0.150	0.457	0.501	0.819
miR-381-3p	-0.152	5.537	0.502	0.819
miR-548h-3p	-0.265	8.937	0.505	0.819
miR-548z	-0.265	8.937	0.505	0.819
miR-1268b	0.419	0.512	0.512	0.827
miR-487a-5p	0.159	0.384	0.513	0.827
miR-22-5p	0.089	3.999	0.514	0.827
miR-191-3p	-0.065	3.805	0.522	0.837
miR-218-5p	-0.374	1.231	0.523	0.837
miR-889-3p	0.169	3.340	0.525	0.837
miR-1301-3p	-0.122	5.720	0.528	0.837
miR-619-5p	0.709	11.202	0.529	0.837
miR-377-5p	-0.152	-0.095	0.531	0.837
miR-140-5p	-0.072	6.331	0.531	0.837
miR-3161	0.517	-0.951	0.532	0.837
miR-99b-3p	0.130	2.773	0.533	0.837
miR-548o-3p	-0.280	2.775	0.537	0.840
miR-548x-3p	0.760	2.828	0.537	0.840
miR-299-3p	0.190	0.096	0.539	0.840
miR-625-5p	0.114	6.770	0.541	0.840
miR-6500-3p	0.374	-0.223	0.543	0.840
miR-134-5p	0.162	7.091	0.543	0.840
miR-29c-5p	-0.053	4.579	0.549	0.845
miR-3187-3p	-0.138	0.369	0.550	0.845
miR-485-5p	-0.194	3.492	0.551	0.845
miR-6859-5p	-0.135	3.735	0.554	0.847
miR-215-5p	0.120	2.194	0.557	0.848
miR-18a-5p	0.060	7.212	0.559	0.848
miR-548aq-3p	0.109	9.450	0.560	0.848
miR-548q	0.122	4.336	0.561	0.848
miR-185-5p	0.044	9.811	0.562	0.848
miR-4477b	0.180	0.399	0.564	0.848
miR-1302	0.605	2.800	0.567	0.848
miR-99a-5p	-0.097	4.162	0.568	0.848
miR-3605-3p	-0.128	1.394	0.572	0.848
miR-5095	1.128	8.523	0.573	0.848
miR-328-3p	-0.123	5.945	0.575	0.848
miR-7851-3p	-0.174	3.495	0.577	0.848
miR-329-5p	-0.149	2.442	0.581	0.848
miR-6741-3p	0.153	0.231	0.582	0.848
miR-125a-5p	0.119	9.141	0.583	0.848
miR-548b-5p	0.111	7.793	0.584	0.848
miR-99b-5p	0.103	7.865	0.585	0.848
miR-324-5p	-0.048	5.454	0.588	0.848
miR-3198	-0.168	0.544	0.589	0.848
miR-424-3p	-0.097	0.968	0.589	0.848
miR-197-3p	-0.073	6.745	0.590	0.848
miR-299-5p	0.123	2.468	0.590	0.848
miR-598-3p	-0.113	4.565	0.590	0.848
miR-1285-5p	0.259	8.438	0.591	0.848
miR-548j-3p	0.149	0.393	0.592	0.848
miR-494-3p	0.130	5.857	0.595	0.848
miR-548k	0.068	1.443	0.595	0.848
miR-3179	0.206	-0.272	0.596	0.848
miR-1270	-0.143	3.003	0.597	0.848
miR-379-3p	0.129	1.468	0.597	0.848
miR-548t-5p	-0.150	1.936	0.602	0.848
miR-379-5p	0.116	7.187	0.602	0.848
miR-195-5p	0.073	3.376	0.603	0.848
miR-3620-5p	0.390	3.024	0.604	0.848
miR-33a-5p	0.097	1.653	0.605	0.848
miR-142-3p	0.070	13.185	0.608	0.849
miR-21-5p	0.080	12.294	0.608	0.849
miR-3064-5p	0.060	0.924	0.610	0.850
miR-146a-3p	0.565	2.291	0.613	0.850
miR-4435	-0.119	1.570	0.614	0.850
miR-548n	0.073	4.212	0.615	0.850
miR-548av-5p	0.064	1.444	0.617	0.851
miR-181b-3p	-0.159	0.083	0.620	0.853
miR-487b-5p	-0.121	0.849	0.621	0.853
miR-330-5p	-0.036	3.616	0.628	0.860
miR-32-5p	0.084	5.289	0.631	0.860
miR-2355-3p	-0.050	3.959	0.632	0.860
miR-107	-0.041	10.028	0.633	0.860
miR-7-1-3p	0.078	2.452	0.634	0.860
miR-744-3p	-0.132	-0.457	0.638	0.863
miR-190a-5p	0.077	5.571	0.642	0.864
miR-3173-5p	-0.137	1.011	0.644	0.864
miR-342-5p	0.074	3.944	0.647	0.864
miR-6803-3p	-0.148	-0.112	0.647	0.864
miR-1	-0.163	8.318	0.647	0.864
miR-3960	-0.367	3.467	0.648	0.864
miR-628-3p	0.053	5.661	0.649	0.864
miR-146b-5p	0.070	9.310	0.650	0.864
miR-106b-5p	0.097	4.658	0.652	0.865
miR-6842-3p	-0.061	1.371	0.658	0.867
miR-98-3p	-0.108	0.779	0.659	0.867
miR-625-3p	-0.098	8.394	0.660	0.867
miR-132-3p	-0.048	3.504	0.660	0.867
miR-103a-2-5p	0.126	-0.340	0.661	0.867
miR-4742-3p	0.068	1.378	0.662	0.867
miR-4433b-3p	-0.137	2.351	0.664	0.867
miR-885-5p	-0.203	0.498	0.670	0.874

miR-551b-3p	0.079	1.815	0.674	0.875
miR-548i	-0.250	3.203	0.675	0.875
miR-3133	-0.280	-0.564	0.688	0.880
miR-548aj-5p	-0.202	7.504	0.689	0.880
miR-548g-5p	-0.202	7.504	0.689	0.880
miR-548x-5p	-0.202	7.504	0.689	0.880
miR-543	0.081	3.379	0.689	0.880
miR-181c-3p	-0.053	3.538	0.689	0.880
miR-148b-5p	-0.035	2.775	0.690	0.880
miR-1179	-0.051	1.832	0.693	0.880
miR-4662a-5p	0.072	1.907	0.693	0.880
miR-181a-3p	0.070	4.944	0.693	0.880
let-7i-5p	0.033	13.259	0.701	0.886
miR-122-5p	-0.203	11.047	0.703	0.886
miR-548ar-5p	-0.201	7.739	0.706	0.886
miR-101-3p	-0.058	11.459	0.706	0.886
miR-145-5p	-0.073	6.318	0.709	0.886
miR-487b-3p	-0.084	5.402	0.710	0.886
miR-377-3p	0.093	2.438	0.712	0.886
miR-23b-5p	-0.071	1.691	0.713	0.886
miR-548ah-3p	0.227	2.757	0.715	0.886
miR-19b-1-5p	0.046	0.495	0.716	0.886
miR-548p	0.203	1.823	0.717	0.886
miR-548ai	0.297	0.527	0.718	0.886
miR-570-5p	0.297	0.527	0.718	0.886
miR-141-3p	-0.060	3.127	0.720	0.888
miR-222-3p	-0.036	8.603	0.726	0.893
miR-548aq-5p	0.291	2.543	0.731	0.897
let-7a-5p	0.049	17.835	0.733	0.898
miR-1185-1-3p	0.082	3.358	0.740	0.900
miR-323a-3p	-0.069	4.078	0.740	0.900
miR-301a-3p	0.039	5.377	0.740	0.900
miR-200a-3p	0.039	2.514	0.746	0.904
miR-409-3p	0.074	8.698	0.747	0.904
miR-27b-5p	0.044	1.148	0.748	0.904
miR-181c-5p	0.056	3.046	0.749	0.904
miR-496	0.086	1.970	0.756	0.910
miR-6511b-3p	0.053	2.078	0.757	0.910
miR-1185-5p	0.090	2.420	0.762	0.911
miR-421	0.021	5.199	0.764	0.911
let-7d-5p	0.036	10.427	0.764	0.911
miR-1304-3p	-0.211	3.254	0.767	0.911
miR-1287-5p	0.033	3.050	0.770	0.911
miR-337-5p	-0.074	3.963	0.772	0.911
miR-200b-3p	-0.036	2.188	0.772	0.911
miR-29b-1-5p	-0.050	0.138	0.773	0.911
miR-505-5p	0.056	3.558	0.774	0.911
miR-548ag	-0.063	6.333	0.775	0.911
miR-544a	-0.137	0.669	0.776	0.911
miR-382-5p	0.070	8.383	0.778	0.911
miR-548h-5p	0.110	6.935	0.780	0.911
miR-3140-3p	-0.044	0.699	0.780	0.911
miR-556-3p	0.056	1.254	0.781	0.911
miR-4454	0.065	9.159	0.783	0.912
miR-574-5p	-0.198	14.568	0.786	0.913
miR-4659a-3p	0.047	0.124	0.795	0.921
miR-4433b-5p	0.064	7.929	0.797	0.923
miR-942-5p	-0.022	4.976	0.807	0.932
miR-509-3p	-0.079	2.288	0.808	0.932
miR-378f	-0.037	1.820	0.811	0.933
miR-376b-5p	0.094	1.487	0.812	0.933
miR-376c-5p	0.091	1.518	0.816	0.933
miR-19a-3p	0.044	5.810	0.817	0.933
miR-128-3p	0.019	10.346	0.818	0.933
miR-651-5p	0.044	1.314	0.819	0.933
miR-576-3p	0.024	5.501	0.821	0.933
miR-139-3p	-0.064	4.524	0.822	0.933
miR-127-3p	-0.047	7.007	0.824	0.934
miR-589-3p	0.024	1.137	0.827	0.935
miR-3913-5p	0.029	2.448	0.834	0.940
miR-542-3p	-0.052	2.341	0.835	0.940
miR-7110-3p	0.063	-0.067	0.839	0.941
miR-1226-3p	0.043	1.073	0.840	0.941
miR-664a-5p	-0.033	5.295	0.840	0.941
miR-1306-3p	0.035	2.215	0.842	0.941
miR-323b-3p	-0.044	4.932	0.846	0.941
miR-877-5p	0.019	2.378	0.846	0.941
miR-548an	0.077	1.688	0.847	0.941
miR-335-3p	-0.030	5.333	0.858	0.951
miR-23c	0.021	0.292	0.859	0.951
miR-127-5p	0.040	1.654	0.867	0.958
miR-548ak	0.119	2.923	0.881	0.967
miR-708-5p	0.090	2.615	0.882	0.967
miR-132-5p	0.019	0.644	0.883	0.967
miR-136-3p	0.052	3.534	0.886	0.967
miR-758-3p	0.036	1.734	0.887	0.967
miR-664b-5p	-0.031	1.474	0.888	0.967
miR-485-3p	0.047	5.831	0.889	0.967
miR-584-5p	-0.020	9.252	0.889	0.967
miR-30b-5p	0.040	7.651	0.892	0.967
miR-505-3p	-0.010	4.359	0.893	0.967
miR-548av-3p	-0.055	3.903	0.893	0.967
miR-654-3p	0.029	7.125	0.900	0.968
miR-4659b-3p	0.017	0.948	0.901	0.968
miR-23a-3p	0.010	11.962	0.904	0.968
miR-145-3p	0.024	1.763	0.905	0.968
miR-548ae	0.165	3.953	0.905	0.968

miR-493-5p	0.027	5.498	0.905	0.968
miR-378g	0.033	0.027	0.905	0.968
miR-5187-5p	0.024	2.108	0.910	0.969
miR-5189-3p	0.046	0.435	0.911	0.969
miR-6859-3p	-0.055	0.719	0.912	0.969
miR-584-3p	-0.019	0.426	0.913	0.969
miR-877-3p	0.039	2.150	0.916	0.970
miR-1249	0.032	0.351	0.919	0.970
miR-493-3p	-0.023	5.508	0.919	0.970
miR-15b-5p	0.013	9.487	0.928	0.978
miR-369-5p	0.018	5.217	0.932	0.980
miR-548al	0.066	-0.392	0.936	0.983
miR-301a-5p	-0.015	1.602	0.940	0.983
miR-5010-3p	-0.008	1.408	0.942	0.983
miR-6734-5p	-0.016	0.555	0.946	0.983
miR-2277-5p	0.035	0.413	0.947	0.983
miR-655-3p	0.025	-0.112	0.948	0.983
miR-3074-5p	-0.017	1.854	0.949	0.983
miR-361-5p	-0.004	8.171	0.950	0.983
miR-548az-5p	0.025	5.839	0.951	0.983
miR-28-3p	0.005	10.686	0.954	0.983
miR-744-5p	-0.010	7.986	0.956	0.983
miR-219a-5p	0.013	1.164	0.957	0.983
miR-3065-5p	-0.008	2.573	0.958	0.983
miR-548l	-0.008	1.877	0.959	0.983
miR-548at-5p	0.013	4.401	0.959	0.983
miR-30d-3p	0.005	1.671	0.974	0.994
miR-337-3p	0.009	2.407	0.975	0.994
miR-30a-3p	-0.003	3.079	0.976	0.994
miR-6511a-5p	-0.016	1.038	0.980	0.994
miR-1273h-5p	-0.004	11.451	0.983	0.994
miR-433-3p	-0.005	1.880	0.984	0.994
miR-652-3p	-0.002	7.109	0.985	0.994
miR-1273c	0.003	1.024	0.986	0.994
miR-148a-3p	-0.002	10.002	0.986	0.994
miR-6511b-5p	0.010	1.158	0.987	0.994
miR-409-5p	0.002	3.705	0.990	0.996
miR-6770-3p	-0.003	0.142	0.992	0.996
miR-548w	0.003	5.896	0.994	0.996
miR-27b-3p	0.000	10.051	0.998	0.998

miRNA	**log-fold change**	**log-counts per million**	***p*-value**	**adjusted *p*-value**
Presymptomatic vs. Patient				
miR-345-5p	-0.528	4.607	3.610E-05	0.021
miR-205-5p	-1.421	2.270	0.000	0.065
miR-206	-2.218	1.104	0.000	0.065
miR-18a-3p	-0.586	2.363	0.001	0.083
miR-93-3p	-0.426	4.195	0.001	0.083
miR-106b-3p	-0.477	6.995	0.001	0.083
miR-548am-c-o-5p	0.497	14.295	0.001	0.083
miR-548au-5p	0.480	14.473	0.001	0.083
miR-551a	-1.098	-0.321	0.002	0.097
miR-218-5p	-1.849	1.231	0.002	0.097
miR-542-5p	-0.709	1.306	0.002	0.097
let-7b-3p	-0.484	1.579	0.003	0.137
miR-151a-5p	0.394	11.573	0.004	0.150
miR-151b	0.389	11.596	0.004	0.150
miR-223-3p	-0.300	14.649	0.005	0.190
miR-374a-3p	0.627	2.987	0.005	0.190
miR-130a-3p	0.399	7.658	0.005	0.190
miR-375	-1.131	2.730	0.006	0.195
miR-1303	-1.405	4.458	0.007	0.224
miR-148a-5p	-0.428	0.933	0.009	0.224
miR-548u	1.339	1.210	0.009	0.224
miR-10b-5p	-0.741	4.540	0.009	0.224
miR-17-5p	0.416	8.427	0.009	0.224
miR-618	-1.220	-0.303	0.010	0.224
miR-25-3p	-0.293	10.901	0.010	0.224
miR-3615	-0.822	5.538	0.010	0.224
miR-625-3p	-0.593	8.394	0.010	0.224
miR-154-3p	0.944	0.419	0.011	0.224
miR-26b-5p	0.296	12.787	0.012	0.226
miR-378a-5p	-0.478	2.841	0.012	0.226
miR-550a-3p	-0.674	3.973	0.012	0.226
miR-590-5p	0.686	0.651	0.013	0.232
miR-941	-0.475	9.475	0.015	0.232
miR-616-3p	-0.532	0.543	0.015	0.232
miR-331-3p	0.332	4.398	0.015	0.232
miR-10a-3p	0.428	1.500	0.016	0.232
miR-365a-3p	-0.696	4.151	0.016	0.232
miR-365b-3p	-0.696	4.151	0.016	0.232
miR-502-3p	-0.330	3.800	0.016	0.232
miR-340-5p	0.441	6.743	0.016	0.232
miR-487a-3p	0.824	0.530	0.017	0.232
miR-374a-5p	0.488	7.594	0.017	0.232
miR-224-5p	-0.706	7.850	0.017	0.232
miR-148b-3p	0.301	9.436	0.018	0.232
miR-550a-3-5p	-0.436	2.398	0.018	0.232
miR-532-3p	-0.462	2.783	0.019	0.237
miR-590-3p	0.630	6.770	0.019	0.237
miR-340-3p	0.235	3.655	0.021	0.240
miR-183-5p	-0.695	5.546	0.022	0.240
miR-629-5p	-0.370	6.404	0.023	0.240

miR-539-3p	0.815	1.473	0.023	0.240
miR-7976	-0.387	1.801	0.023	0.240
miR-1289	-1.074	-0.148	0.023	0.240
miR-642a-3p	-0.683	0.484	0.024	0.240
miR-628-5p	0.286	4.677	0.024	0.240
miR-100-5p	-0.778	2.346	0.025	0.240
miR-200c-3p	0.183	5.273	0.025	0.240
miR-6873-3p	-0.586	5.014	0.025	0.240
miR-92a-3p	-0.495	13.677	0.025	0.240
miR-223-5p	-0.246	7.682	0.025	0.240
miR-203a	-0.610	-0.173	0.026	0.240
miR-146a-5p	0.232	12.934	0.026	0.240
miR-28-5p	0.196	6.656	0.026	0.240
let-7f-5p	0.224	16.225	0.028	0.259
miR-7706	-0.588	-0.037	0.030	0.264
miR-3120-3p	0.288	2.583	0.030	0.264
miR-378i	-0.285	4.394	0.030	0.264
miR-589-5p	-0.253	3.684	0.031	0.264
miR-484	-0.502	7.854	0.031	0.264
miR-382-3p	0.578	3.232	0.032	0.269
miR-1277-3p	0.473	1.948	0.032	0.269
miR-32-3p	0.424	0.143	0.033	0.270
miR-1277-5p	0.463	12.138	0.033	0.270
miR-376b-3p	0.671	3.330	0.034	0.270
miR-133a-3p	-0.671	7.345	0.034	0.270
miR-4677-3p	0.374	0.746	0.036	0.271
miR-9-3p	0.380	3.953	0.037	0.271
miR-545-5p	0.537	1.482	0.037	0.271
miR-139-5p	0.454	6.068	0.038	0.271
miR-92b-5p	-0.857	-0.009	0.039	0.271
miR-369-3p	0.685	4.919	0.039	0.271
miR-92b-3p	-0.530	3.086	0.040	0.271
miR-3928-3p	-0.393	0.931	0.040	0.271
let-7b-5p	-0.395	12.514	0.040	0.271
miR-296-5p	-0.499	1.358	0.040	0.271
miR-514a-3p	0.933	0.606	0.040	0.271
miR-25-5p	-0.435	2.573	0.041	0.271
miR-576-5p	-0.290	6.414	0.042	0.271
miR-7641	-0.638	5.298	0.042	0.271
miR-301b	0.333	1.088	0.043	0.271
miR-196b-5p	0.190	5.658	0.043	0.271
miR-500a-3p	-0.248	4.766	0.043	0.271
miR-451a	-0.491	11.183	0.044	0.271
miR-153-3p	0.543	0.310	0.044	0.271
miR-652-5p	0.405	1.446	0.044	0.271
miR-425-3p	-0.182	6.606	0.045	0.271
miR-193a-5p	-0.874	3.243	0.045	0.271
miR-412-5p	1.044	2.206	0.045	0.271
miR-374b-3p	0.372	1.235	0.046	0.271
miR-539-5p	0.706	0.162	0.046	0.271
miR-199a-3p	0.283	15.594	0.047	0.271
miR-199b-3p	0.283	15.594	0.047	0.271
miR-550a-5p	-0.401	1.899	0.048	0.274
miR-142-3p	0.288	13.185	0.048	0.274
miR-140-3p	-0.166	9.249	0.050	0.278
miR-425-5p	-0.210	10.913	0.050	0.278
miR-32-5p	0.365	5.289	0.051	0.278
miR-501-3p	-0.470	2.746	0.051	0.278
miR-125a-3p	-0.437	0.692	0.052	0.279
miR-362-5p	-0.275	2.457	0.053	0.283
miR-4732-3p	-0.921	3.928	0.053	0.283
miR-197-3p	-0.280	6.745	0.054	0.284
miR-23b-5p	-0.398	1.691	0.054	0.284
miR-335-5p	0.338	7.461	0.055	0.284
miR-1180-3p	-0.775	2.024	0.056	0.287
miR-26a-5p	0.175	15.391	0.057	0.291
miR-641	0.209	1.385	0.058	0.291
miR-423-3p	-0.376	8.841	0.059	0.293
miR-548aq-3p	0.374	9.450	0.059	0.293
miR-5010-3p	-0.214	1.408	0.060	0.293
miR-486-5p	-0.676	14.249	0.060	0.293
let-7d-3p	-0.182	6.419	0.061	0.293
miR-495-3p	0.483	3.399	0.064	0.307
miR-3679-5p	-0.380	2.384	0.069	0.318
miR-769-3p	-0.392	0.674	0.070	0.318
miR-3613-5p	-0.498	9.506	0.070	0.318
miR-320c	-0.521	8.553	0.070	0.318
miR-301a-3p	0.225	5.377	0.070	0.318
miR-4685-3p	-0.560	0.669	0.072	0.318
miR-532-5p	-0.246	5.499	0.072	0.318
miR-2115-3p	-0.721	0.295	0.072	0.318
miR-320b	-0.524	8.897	0.073	0.318
miR-769-5p	-0.123	4.654	0.073	0.318
miR-150-3p	-0.511	1.296	0.073	0.318
miR-548h-5p	0.779	6.935	0.074	0.318
miR-545-3p	0.383	1.909	0.074	0.318
miR-18a-5p	0.197	7.212	0.075	0.318
miR-33a-5p	0.358	1.653	0.075	0.318
miR-548ab	1.175	4.211	0.076	0.318
miR-361-3p	-0.176	6.970	0.076	0.318
miR-324-3p	-0.176	4.372	0.077	0.319

miR-889-3p	0.491	3.340	0.077	0.319
miR-93-5p	-0.132	11.309	0.078	0.319
miR-133b	-0.707	-0.138	0.078	0.319
miR-3158-3p	-0.626	1.499	0.079	0.321
miR-204-5p	0.480	1.457	0.082	0.327
miR-6852-5p	-0.445	2.746	0.083	0.327
miR-106a-5p	0.324	1.761	0.084	0.327
miR-486-3p	-0.521	5.430	0.084	0.327
miR-4785	-0.266	0.526	0.084	0.327
miR-320d	-0.422	7.345	0.084	0.327
miR-548d-3p	0.731	1.169	0.084	0.327
miR-361-5p	-0.128	8.171	0.085	0.328
miR-378a-3p	-0.212	9.082	0.086	0.328
miR-548j-3p	0.504	0.393	0.087	0.332
miR-29a-5p	0.275	0.831	0.091	0.345
miR-551b-3p	0.337	1.815	0.092	0.345
let-7g-5p	0.133	10.980	0.093	0.347
miR-30e-3p	0.260	7.059	0.094	0.348
miR-942-5p	-0.162	4.976	0.097	0.356
miR-423-5p	-0.571	10.597	0.098	0.357
miR-126-3p	0.141	14.192	0.100	0.357
miR-26b-3p	-0.152	3.623	0.100	0.357
miR-370-3p	-0.494	5.773	0.101	0.357
miR-30b-3p	0.226	0.223	0.101	0.357
miR-181b-5p	-0.234	8.044	0.101	0.357
miR-1908-5p	-0.625	2.300	0.102	0.357
miR-411-5p	0.408	5.491	0.102	0.357
miR-376a-3p	0.498	6.560	0.104	0.357
miR-424-3p	-0.315	0.968	0.104	0.357
miR-660-5p	-0.229	5.628	0.104	0.357
miR-19a-3p	0.330	5.810	0.105	0.357
miR-126-5p	0.269	11.057	0.105	0.357
miR-509-3p	-0.548	2.288	0.106	0.357
miR-26a-1-3p	0.161	2.985	0.107	0.357
miR-760	-0.534	1.310	0.108	0.357
miR-152-3p	0.194	7.210	0.109	0.357
miR-1233-3p	1.403	-0.544	0.109	0.357
miR-1185-2-3p	0.452	0.123	0.109	0.357
miR-30a-5p	-0.208	6.368	0.110	0.357
miR-376c-3p	0.501	5.292	0.110	0.357
miR-30d-5p	-0.119	11.785	0.111	0.357
miR-106b-5p	0.368	4.658	0.112	0.357
miR-16-5p	-0.169	17.408	0.112	0.357
miR-491-5p	-0.207	3.945	0.112	0.357
miR-487a-5p	0.397	0.384	0.113	0.359
miR-548f-3p	0.497	5.197	0.115	0.361
miR-9-5p	0.285	5.867	0.115	0.361
miR-299-3p	0.512	0.096	0.116	0.361
miR-500a-5p	-0.358	3.026	0.117	0.361
miR-548p	0.989	1.823	0.118	0.361
miR-671-3p	-0.345	3.284	0.118	0.361
miR-3138	-0.395	2.724	0.119	0.363
miR-1307-3p	-0.407	7.329	0.120	0.366
miR-363-3p	-0.265	6.300	0.122	0.367
miR-377-3p	0.400	2.438	0.122	0.367
miR-185-3p	-0.256	4.607	0.123	0.368
miR-320a	-0.378	11.779	0.125	0.368
miR-146b-3p	-0.190	3.582	0.125	0.368
miR-1233-5p	-1.423	-0.221	0.125	0.368
miR-548e-3p	0.286	2.682	0.125	0.368
miR-1273h-3p	-0.305	7.277	0.127	0.370
miR-99a-5p	-0.266	4.162	0.132	0.380
miR-1296-5p	-0.156	2.226	0.132	0.380
miR-21-5p	0.252	12.294	0.132	0.380
miR-548a-5p	0.521	1.849	0.135	0.385
miR-671-5p	-0.294	4.210	0.139	0.395
miR-548am-3p	1.402	2.613	0.140	0.396
miR-20a-5p	0.186	9.087	0.141	0.397
miR-500b-5p	-0.369	2.709	0.142	0.398
miR-3180-5p	-0.890	-0.295	0.144	0.401
miR-190b	-0.127	3.267	0.145	0.401
miR-496	0.423	1.970	0.145	0.401
miR-96-5p	-0.469	3.856	0.146	0.401
miR-3198	-0.481	0.544	0.151	0.414
miR-548ah-3p	1.009	2.757	0.152	0.414
miR-454-5p	-0.120	3.780	0.154	0.415
miR-150-5p	-0.416	10.272	0.155	0.415
miR-23a-5p	-0.375	1.003	0.155	0.415
miR-98-5p	0.141	8.594	0.156	0.415
miR-130b-3p	0.136	5.063	0.156	0.415
miR-20a-3p	0.186	1.922	0.159	0.418
miR-4732-5p	-0.674	3.999	0.159	0.418
miR-3605-3p	-0.338	1.394	0.159	0.418
miR-2110	-0.396	3.338	0.161	0.420
miR-144-3p	-0.394	5.293	0.162	0.420
miR-4448	-0.343	4.223	0.162	0.420
miR-454-3p	0.132	8.563	0.168	0.432
miR-222-3p	-0.152	8.603	0.169	0.432
miR-6859-5p	-0.327	3.735	0.170	0.432
miR-1268b	0.974	0.512	0.170	0.432
miR-7-5p	0.203	10.227	0.170	0.432

miR-1294	-0.324	1.946	0.171	0.433
miR-1285-5p	0.731	8.438	0.174	0.438
miR-483-5p	-0.869	1.834	0.176	0.440
miR-181a-5p	-0.208	10.817	0.179	0.447
miR-6513-5p	0.178	0.012	0.181	0.449
miR-30d-3p	0.218	1.671	0.181	0.449
miR-1255b-5p	-0.338	7.593	0.187	0.459
miR-548a-3p	0.272	5.676	0.187	0.459
miR-5095	2.977	8.523	0.189	0.463
miR-146b-5p	0.212	9.310	0.194	0.471
miR-4645-3p	0.185	0.811	0.195	0.471
miR-1972	3.618	5.510	0.196	0.471
miR-155-5p	0.135	8.603	0.196	0.471
miR-1270	-0.368	3.003	0.198	0.472
miR-103a-2-5p	0.391	-0.340	0.198	0.472
miR-1306-5p	-0.496	2.765	0.199	0.472
miR-196a-5p	-0.436	2.899	0.199	0.472
miR-192-5p	-0.197	7.507	0.200	0.472
miR-134-5p	0.349	7.091	0.203	0.477
miR-26a-2-3p	0.293	-0.153	0.208	0.486
miR-29b-2-5p	0.114	1.605	0.211	0.492
miR-651-5p	0.255	1.314	0.214	0.497
miR-1301-3p	-0.252	5.720	0.220	0.507
miR-330-3p	0.107	3.864	0.220	0.507
miR-5010-5p	-0.466	0.210	0.224	0.510
miR-132-3p	-0.139	3.504	0.224	0.510
miR-4477b	0.403	0.399	0.224	0.510
miR-136-3p	0.450	3.534	0.225	0.510
miR-548ax	0.179	5.062	0.226	0.510
miR-194-5p	-0.230	7.337	0.229	0.510
miR-122-5p	-0.655	11.047	0.231	0.510
miR-145-5p	-0.245	6.318	0.231	0.510
miR-24-3p	-0.088	12.655	0.232	0.510
miR-556-3p	0.258	1.254	0.232	0.510
miR-182-5p	-0.243	7.556	0.232	0.510
miR-1292-5p	-0.342	0.211	0.233	0.510
miR-665	-0.345	0.677	0.233	0.510
miR-3173-5p	-0.371	1.011	0.238	0.519
miR-548as-5p	0.457	3.448	0.241	0.522
miR-379-5p	0.269	7.187	0.241	0.522
miR-1304-3p	-0.927	3.254	0.248	0.533
miR-3913-5p	0.168	2.448	0.248	0.533
miR-885-5p	-0.573	0.498	0.249	0.533
miR-369-5p	0.247	5.217	0.254	0.540
miR-376c-5p	0.470	1.518	0.256	0.540
miR-432-5p	-0.313	7.959	0.257	0.540
miR-6842-3p	-0.164	1.371	0.258	0.540
miR-1290	-0.501	4.532	0.258	0.540
miR-17-3p	0.149	3.215	0.258	0.540
miR-219a-5p	0.291	1.164	0.259	0.541
miR-1250-5p	0.155	0.691	0.261	0.542
miR-191-3p	-0.122	3.805	0.261	0.542
miR-382-5p	0.282	8.383	0.263	0.544
miR-29b-1-5p	-0.204	0.138	0.264	0.544
miR-376b-5p	0.464	1.487	0.266	0.545
miR-5583-3p	0.235	0.528	0.267	0.545
miR-30b-5p	0.352	7.651	0.268	0.545
miR-548k	0.150	1.443	0.268	0.545
miR-1468-5p	0.150	0.552	0.274	0.555
miR-548av-5p	0.146	1.444	0.282	0.568
miR-548g-3p	1.256	1.192	0.286	0.574
miR-1246	-0.442	5.861	0.287	0.574
miR-374b-5p	0.214	6.920	0.288	0.574
miR-1255a	0.317	4.833	0.288	0.574
miR-6500-3p	0.674	-0.223	0.290	0.574
miR-548x-3p	1.500	2.828	0.292	0.574
miR-185-5p	0.085	9.811	0.292	0.574
miR-548al	0.909	-0.392	0.295	0.577
miR-655-3p	0.417	-0.112	0.295	0.577
miR-30c-5p	0.166	11.296	0.298	0.581
miR-627-5p	-0.113	0.873	0.300	0.581
miR-101-3p	0.168	11.459	0.302	0.581
miR-664b-5p	-0.242	1.474	0.304	0.581
miR-3180	-0.726	0.868	0.304	0.581
miR-3180-3p	-0.726	0.868	0.304	0.581
miR-548c-3p	-0.556	2.905	0.304	0.581
miR-4742-3p	-0.167	1.378	0.309	0.589
miR-328-3p	-0.238	5.945	0.311	0.592
miR-664a-5p	-0.176	5.295	0.313	0.592
miR-379-3p	0.253	1.468	0.315	0.595
miR-548ag	0.236	6.333	0.320	0.601
miR-22-5p	0.146	3.999	0.320	0.601
miR-503-5p	-0.168	3.907	0.327	0.609
miR-766-3p	-0.140	5.380	0.327	0.609
miR-1249	-0.314	0.351	0.329	0.611
miR-199a-5p	0.149	8.817	0.330	0.611
miR-330-5p	0.075	3.616	0.337	0.622
miR-1304-5p	0.967	0.167	0.338	0.622
miR-548n	0.144	4.212	0.343	0.628
miR-1306-3p	-0.175	2.215	0.343	0.628
miR-431-3p	-0.232	0.193	0.346	0.631

miR-30c-1-3p	0.093	2.582	0.350	0.637
miR-483-3p	-0.532	1.205	0.352	0.637
miR-1271-3p	0.730	-0.768	0.353	0.637
miR-337-5p	0.237	3.963	0.359	0.644
miR-7110-3p	-0.303	-0.067	0.360	0.644
miR-301a-5p	0.192	1.602	0.361	0.644
miR-1271-5p	-0.127	1.919	0.362	0.644
miR-151a-3p	0.074	12.424	0.362	0.644
miR-1185-1-3p	0.232	3.358	0.364	0.646
miR-664a-3p	0.108	4.261	0.367	0.649
miR-5009-5p	-0.255	0.291	0.370	0.651
miR-4286	-0.116	10.665	0.370	0.651
miR-181a-2-3p	0.098	5.645	0.372	0.651
miR-34a-5p	0.194	4.676	0.373	0.651
miR-548o-3p	-0.433	2.775	0.374	0.651
miR-5096	-0.744	11.722	0.375	0.651
miR-10a-5p	0.103	7.108	0.376	0.651
miR-6803-3p	-0.300	-0.112	0.379	0.655
miR-3065-3p	-0.161	0.404	0.384	0.661
miR-195-5p	0.126	3.376	0.387	0.665
miR-186-3p	-0.123	0.571	0.395	0.674
miR-181a-3p	0.160	4.944	0.395	0.674
miR-409-3p	0.197	8.698	0.400	0.675
miR-3200-3p	-0.268	0.193	0.400	0.675
miR-625-5p	-0.164	6.770	0.402	0.675
miR-326	-0.129	6.590	0.403	0.675
miR-342-3p	-0.139	10.283	0.404	0.675
miR-548an	0.340	1.688	0.404	0.675
miR-758-3p	0.222	1.734	0.405	0.675
miR-548t-5p	0.257	1.936	0.406	0.675
miR-103a-3p	0.064	15.095	0.406	0.675
miR-148b-5p	0.077	2.775	0.407	0.675
miR-378d	-0.451	0.023	0.410	0.678
miR-1302	0.901	2.800	0.411	0.679
miR-339-3p	-0.086	5.037	0.414	0.679
miR-190a-5p	0.144	5.571	0.415	0.679
miR-135a-5p	-0.243	1.965	0.417	0.679
miR-29b-3p	0.211	9.739	0.417	0.679
miR-3187-3p	-0.198	0.369	0.418	0.679
miR-744-5p	-0.159	7.986	0.420	0.681
miR-1185-5p	0.246	2.420	0.425	0.684
miR-424-5p	-0.195	3.748	0.425	0.684
miR-494-3p	0.203	5.857	0.425	0.684
miR-377-5p	0.198	-0.095	0.427	0.685
miR-409-5p	0.168	3.705	0.431	0.691
miR-1179	-0.107	1.832	0.433	0.692
miR-16-2-3p	-0.157	3.631	0.436	0.694
miR-299-5p	0.181	2.468	0.437	0.694
miR-6513-3p	0.070	1.835	0.441	0.696
miR-335-3p	-0.137	5.333	0.443	0.696
miR-215-5p	-0.157	2.194	0.444	0.696
miR-127-5p	0.187	1.654	0.445	0.696
miR-2277-5p	-0.095	0.413	0.445	0.696
miR-339-5p	-0.093	11.621	0.446	0.696
miR-664b-3p	0.111	2.162	0.453	0.705
miR-574-5p	-0.605	14.568	0.454	0.705
miR-548ak	0.655	2.923	0.456	0.705
miR-378g	-0.219	0.027	0.456	0.705
miR-4326	-0.312	1.874	0.459	0.708
miR-23c	-0.091	0.292	0.469	0.722
miR-125b-5p	-0.109	8.080	0.472	0.723
miR-181b-3p	0.238	0.083	0.473	0.723
miR-4435	-0.183	1.570	0.475	0.725
miR-6734-5p	-0.182	0.555	0.479	0.729
miR-6516-5p	0.110	1.586	0.481	0.731
miR-144-5p	-0.214	7.018	0.489	0.740
miR-27a-3p	0.154	9.739	0.491	0.740
miR-323b-3p	0.158	4.932	0.491	0.740
miR-1285-3p	-0.115	8.508	0.494	0.743
miR-199b-5p	-0.212	4.252	0.500	0.748
miR-27a-5p	0.160	2.220	0.500	0.748
miR-20b-5p	-0.094	5.620	0.504	0.752
miR-143-5p	0.171	1.979	0.506	0.752
miR-504-5p	0.133	0.213	0.508	0.754
miR-5187-5p	-0.146	2.108	0.514	0.761
miR-598-3p	0.143	4.565	0.518	0.764
miR-1307-5p	0.080	5.708	0.523	0.770
miR-542-3p	0.169	2.341	0.531	0.780
miR-548ap-5p	0.079	9.773	0.536	0.785
miR-3617-5p	0.136	0.442	0.538	0.785
miR-766-5p	0.084	2.287	0.539	0.785
miR-154-5p	0.196	3.881	0.539	0.785
miR-27b-3p	0.088	10.051	0.543	0.788
miR-654-3p	0.143	7.125	0.547	0.790
miR-139-3p	-0.182	4.524	0.548	0.790
miR-27b-5p	-0.088	1.148	0.550	0.790
miR-29c-3p	0.155	8.568	0.550	0.790
miR-548b-5p	0.127	7.793	0.552	0.790
miR-548az-5p	0.256	5.839	0.553	0.790
miR-23b-3p	0.054	9.599	0.557	0.792
miR-324-5p	0.055	5.454	0.559	0.792

miR-22-3p	-0.045	8.495	0.559	0.792
miR-3194-3p	0.272	0.002	0.560	0.792
miR-4446-3p	0.125	3.069	0.563	0.795
miR-142-5p	0.109	10.854	0.567	0.797
miR-136-5p	0.160	2.770	0.568	0.797
let-7c-5p	-0.079	9.528	0.568	0.797
miR-143-3p	0.144	9.208	0.571	0.797
miR-28-3p	-0.054	10.686	0.571	0.797
miR-4659b-3p	0.080	0.948	0.573	0.798
miR-30a-3p	0.058	3.079	0.575	0.798
miR-548l	-0.098	1.877	0.575	0.798
miR-1299	-0.475	4.630	0.577	0.798
miR-584-5p	-0.085	9.252	0.579	0.798
miR-23a-3p	-0.051	11.962	0.581	0.798
miR-493-3p	0.126	5.508	0.581	0.798
let-7a-5p	-0.082	17.835	0.585	0.802
miR-1273c	-0.103	1.024	0.588	0.802
miR-1260b	-0.086	3.783	0.589	0.802
miR-5193	0.087	1.173	0.589	0.802
miR-29c-5p	-0.049	4.579	0.594	0.806
miR-3133	-0.394	-0.564	0.596	0.806
miR-505-5p	-0.109	3.558	0.597	0.806
miR-485-3p	0.180	5.831	0.601	0.810
miR-7849-3p	0.085	0.247	0.609	0.819
miR-5585-3p	-0.858	6.719	0.618	0.830
miR-7851-3p	-0.161	3.495	0.627	0.835
miR-548ae	0.791	3.953	0.627	0.835
miR-7-1-3p	0.084	2.452	0.627	0.835
miR-548j-5p	0.063	9.735	0.628	0.835
miR-4662a-5p	0.091	1.907	0.632	0.837
miR-429	-0.075	0.270	0.632	0.837
let-7g-3p	0.112	0.220	0.638	0.842
miR-338-3p	-0.082	5.028	0.641	0.844
miR-19b-1-5p	0.061	0.495	0.645	0.846
miR-221-5p	0.044	5.310	0.645	0.846
miR-544a	0.230	0.669	0.646	0.846
miR-543	0.096	3.379	0.648	0.846
miR-431-5p	-0.101	7.576	0.650	0.847
miR-378c	-0.065	5.108	0.655	0.852
miR-493-5p	0.103	5.498	0.660	0.855
miR-5584-5p	0.087	0.672	0.662	0.855
let-7f-1-3p	-0.111	0.491	0.663	0.855
miR-329-5p	-0.118	2.442	0.663	0.855
miR-146a-3p	0.456	2.291	0.669	0.860
let-7d-5p	-0.054	10.427	0.672	0.862
miR-548q	-0.092	4.336	0.677	0.867
miR-191-5p	-0.037	12.565	0.681	0.870
miR-107	-0.037	10.028	0.682	0.870
miR-4301	-0.249	2.532	0.685	0.872
miR-873-5p	0.149	1.352	0.692	0.878
miR-1268a	0.398	3.542	0.698	0.884
miR-337-3p	0.121	2.407	0.706	0.892
miR-329-3p	-0.081	5.645	0.711	0.896
miR-1273g-3p	-0.663	11.153	0.712	0.896
miR-548w	0.143	5.896	0.716	0.897
miR-3620-5p	-0.312	3.024	0.717	0.897
miR-127-3p	-0.079	7.007	0.717	0.897
miR-548at-5p	0.101	4.401	0.719	0.897
miR-452-5p	-0.092	2.671	0.721	0.897
miR-381-3p	0.081	5.537	0.725	0.898
miR-378f	-0.058	1.820	0.726	0.898
miR-128-3p	0.031	10.346	0.727	0.898
miR-125a-5p	-0.079	9.141	0.727	0.898
miR-1226-3p	-0.076	1.073	0.735	0.903
miR-584-3p	0.064	0.426	0.735	0.903
miR-548av-3p	0.152	3.903	0.736	0.903
miR-3179	0.136	-0.272	0.739	0.904
miR-548y	0.353	-0.318	0.742	0.904
miR-548f-5p	-0.235	0.762	0.742	0.904
miR-98-3p	0.084	0.779	0.744	0.904
miR-3143	-0.055	-0.186	0.745	0.904
miR-3140-3p	-0.055	0.699	0.746	0.904
miR-3960	-0.260	3.467	0.752	0.907
miR-487b-5p	-0.078	0.849	0.754	0.907
miR-495-5p	-0.079	-0.270	0.755	0.907
miR-145-3p	-0.065	1.763	0.755	0.907
miR-15b-3p	-0.056	5.648	0.765	0.918
miR-548i	-0.182	3.203	0.776	0.927
miR-485-5p	-0.093	3.492	0.776	0.927
miR-4659a-3p	-0.054	0.124	0.780	0.929
miR-140-5p	0.034	6.331	0.782	0.929
miR-15b-5p	0.043	9.487	0.783	0.929
miR-582-3p	-0.088	1.716	0.784	0.929
miR-3161	-0.240	-0.951	0.786	0.929
miR-1273h-5p	-0.060	11.451	0.788	0.929
miR-505-3p	-0.022	4.359	0.789	0.929
miR-181d-5p	-0.035	3.947	0.792	0.930
miR-877-3p	0.105	2.150	0.793	0.930
miR-323a-3p	-0.055	4.078	0.794	0.930
miR-6511b-5p	-0.173	1.158	0.798	0.932
miR-200b-3p	-0.033	2.188	0.799	0.932

miR-141-3p	0.044	3.127	0.802	0.933
miR-708-5p	0.161	2.615	0.807	0.937
miR-19b-3p	0.060	10.256	0.809	0.938
miR-18b-5p	0.043	2.410	0.812	0.940
miR-6772-3p	0.055	0.690	0.816	0.941
miR-877-5p	0.024	2.378	0.817	0.941
let-7a-3p	-0.058	4.497	0.823	0.946
miR-31-5p	0.085	2.641	0.826	0.947
miR-548aq-5p	0.187	2.543	0.827	0.947
miR-5189-3p	-0.093	0.435	0.831	0.947
miR-548ai	0.194	0.527	0.832	0.947
miR-570-5p	0.194	0.527	0.832	0.947
miR-450b-5p	-0.053	2.048	0.834	0.947
miR-487b-3p	0.048	5.402	0.835	0.947
miR-548ap-3p	-0.072	10.374	0.837	0.947
miR-1	0.078	8.318	0.838	0.947
miR-1229-3p	0.048	0.765	0.840	0.948
miR-6770-3p	-0.062	0.142	0.844	0.950
miR-3064-5p	0.024	0.924	0.846	0.951
miR-186-5p	-0.024	8.853	0.848	0.951
miR-582-5p	0.068	0.885	0.853	0.954
miR-3065-5p	-0.030	2.573	0.854	0.954
miR-200a-3p	0.023	2.514	0.855	0.954
miR-652-3p	0.018	7.109	0.857	0.954
miR-181c-5p	0.031	3.046	0.866	0.962
miR-548ar-5p	0.098	7.739	0.868	0.962
miR-6741-3p	0.046	0.231	0.875	0.962
miR-628-3p	-0.019	5.661	0.877	0.962
miR-548d-5p	0.025	12.885	0.878	0.962
miR-4454	0.037	9.159	0.881	0.962
miR-576-3p	0.017	5.501	0.882	0.962
miR-433-3p	0.038	1.880	0.885	0.962
miR-148a-3p	0.020	10.002	0.889	0.962
miR-574-3p	-0.018	5.403	0.890	0.962
miR-29a-3p	0.041	9.472	0.894	0.962
miR-4433b-5p	-0.036	7.929	0.895	0.962
miR-6511a-5p	-0.091	1.038	0.895	0.962
miR-3074-5p	-0.035	1.854	0.897	0.962
miR-6511b-3p	-0.023	2.078	0.898	0.962
miR-497-5p	-0.022	1.523	0.900	0.962
miR-548aa	0.043	10.786	0.900	0.962
miR-548t-3p	0.043	10.786	0.900	0.962
let-7i-3p	-0.024	2.370	0.901	0.962
miR-450a-5p	-0.040	0.853	0.901	0.962
miR-619-5p	-0.136	11.202	0.902	0.962
miR-874-3p	-0.015	1.960	0.903	0.962
miR-342-5p	0.021	3.944	0.903	0.962
miR-548aj-5p	0.066	7.504	0.907	0.962
miR-548g-5p	0.066	7.504	0.907	0.962
miR-548x-5p	0.066	7.504	0.907	0.962
miR-624-5p	-0.031	0.116	0.908	0.962
miR-2355-3p	-0.012	3.959	0.913	0.965
miR-4433b-3p	0.037	2.351	0.914	0.965
miR-2355-5p	0.025	0.457	0.916	0.965
miR-210-3p	-0.013	2.393	0.918	0.965
miR-99b-3p	-0.021	2.773	0.922	0.968
miR-3688-3p	-0.018	1.689	0.926	0.971
let-7i-5p	-0.008	13.259	0.930	0.973
miR-4507	-0.108	1.601	0.938	0.978
miR-744-3p	0.023	-0.457	0.938	0.978
miR-548ay-5p	0.012	13.403	0.942	0.979
miR-99b-5p	-0.014	7.865	0.943	0.979
miR-15a-5p	-0.011	9.211	0.944	0.979
miR-132-5p	0.009	0.644	0.945	0.979
miR-1287-5p	0.008	3.050	0.950	0.980
miR-221-3p	0.006	11.513	0.950	0.980
miR-376a-5p	-0.019	0.104	0.955	0.983
miR-548e-5p	-0.015	1.838	0.958	0.985
miR-95-3p	0.011	2.108	0.963	0.988
miR-21-3p	0.009	3.394	0.967	0.991
miR-589-3p	-0.004	1.137	0.971	0.992
miR-4772-3p	0.013	1.699	0.971	0.992
miR-30e-5p	0.003	10.655	0.974	0.993
miR-138-5p	-0.007	0.000	0.980	0.993
let-7e-5p	0.005	9.484	0.980	0.993
miR-548h-3p	0.009	8.937	0.983	0.993
miR-548z	0.009	8.937	0.983	0.993
miR-1273d	-0.027	4.498	0.983	0.993
miR-181c-3p	0.002	3.538	0.987	0.994
miR-130b-5p	-0.002	5.414	0.987	0.994
miR-421	-0.001	5.199	0.991	0.995
miR-6859-3p	-0.005	0.719	0.992	0.995
miR-6511a-3p	-0.001	4.865	0.993	0.995
let-7f-2-3p	-0.001	0.250	0.998	0.998

Table A3. List of the putative target genes

Table A3. List of the putative target genes for each miRNA from our signature (miR-34a-5p, miR-345-5p, miR-200c-3p, miR-10a-3p), as indicated by DIANA-miRPath v.3 using TarBase (experimentally supported approach).

miR-34a-5p								
ARHGAP1	TMEM33	HUWE1	HNRNPU	KIAA1897(hsa)	MMP9	SLC36A1	KIAA0899(hsa)	PCYOX1
BMI1	POLI	MAP4	TAB2	MOCS2	MRPS2	RDRC(hsa)	SLC30A3	SCML2
TMSB10	LRSAM1	ETFDH	ANO9	RAD2(hsa)	PTGER2	A2RP(hsa)	LRRC46	RAD17
KDRF(hsa)	ZNF76	FRS2	FXN	ZFPL1	LZTFL1	PRSS1	SGPP1	IL6
RGP1	SLC37A3	BIN1	TCF19	MED16	ABCD1	STAR	HDAC8	CMPK(hsa)
TGDS	REEP6	TRIM7	SAPL(hsa)	GDA	POLR1B	KIF11	FAM111A	PPFIA1
NA	MTERFD2	LMNTD2	APOO	NCEH1	AF131216.5	ELMOD1	ATP1A1	RNF141
CDC46(hsa)	VPS37D	KLHL17	PPP1R18	TPM4	BCL7B	CITED2	KIAA0004(hsa)	OK/SW-cl.29(hsa)
ALR(hsa)	H2AFM(hsa)	AP5Z1	COQ9	BRCA1	DIAPH1	NLGN2	H4/K(hsa)	TRIB1
GOLPH3L	D3S48E(hsa)	SPTB2(hsa)	C17orf99	AGBL5	CHST6	KDELR1	SNF8	CTSH
VPS4A	ETV6	CAPZA2	PNPLA8	THEM4	MLST8	NUDT22	ST20	hCG_2005629(hsa)
NOTUM	ADCY7	DPYSL4	H4FI(hsa)	GOLGA7	PTP4A2	NPIPB4	NUCKS(hsa)	H1F4(hsa)
ZDHHC18	BCAT2	GOLGA8B	GLI4	GPC3	FGD3	P1725(hsa)	CNOT6L	MAPK1
RTN4	MMGT1	KLF12	UBE2I	LAD1	COG5	SURF4	TACSTD2	ADTB2(hsa)
SEMA3G	LIMD2	UTS2	MAGED1	LYPD8	PPP1R14C	N4BP4(hsa)	AS3MT	CLTB
C17orf85	FGFR3	MRPS12	SLC16A4	FAF2	CHERP	PSMD9	SNX5	IMP3
ARPC5	H2AFC(hsa)	RAD54B	VEPH1	GGA3	TK1	MPI	EHD1	AMER1
ERLIN1	STX5	DMBT1	GXYLT1	LPCAT4	ZDHHC4	WIPF3	KNSL1(hsa)	FNDC3A
EPB41L4A	H2AFX	STXBP5	C6orf141	MCM2	RNASEK	ANX4(hsa)	FLOT2	SDCBP
NEK3	LTN1	KIF2A	WIZ	RP11-452K12.5-012(hsa)	ORMDL2	SKA1	SLC2A12	SLAIN2
COPS7A	FAM86C1	TPCN2	SNX15	EHBP1	MECP2	RAD1	CLSTN1	PLOD1
PPP1CA	ARF5	H2AFG(hsa)	EZH2	SMG6	MAPRE3	PPP6R2	SSBP3	PAX9
E2F7	C19orf66	TIAF1	CDK6	SYNC	ZFP36L1	WBSCR20(hsa)	USP3	ANO7
TRIB3	LPHN2	LOXL2	HRG	POU6F1	ARV1	KIAA1914(hsa)	CHKA	KANK3
DHRS4-AS1	PTPRN2	SPDEF	SORBS1	NUP214	H3FB(hsa)	PUS7	RSBN1	LRRC41
H3FI(hsa)	SLC25A15	GTF3C1	USP30	ANK3	CORIN	ST7	KCNK6	SUCLG1
TMSA(hsa)	REPS1	BTBD19	RHEB	CROT	PDGFRL	SHARPIN	PICH(hsa)	RNF168
SLC44A2	C16orf58	GDF15	TFAP2A	PSD3	ARID1A	HNGS1(hsa)	ITPR3	TLN1
SLC2A14	KCNE3	CLAPM1(hsa)	ACAA2	LARP1	UBA1	PYGO2	KIAA0731(hsa)	PRO1777(hsa)
FBXO5	SCAF4	PLEKHH3	DCAF10	C15orf65	PANK2	PLXNB1	MPP2	C1QL1
SYNGR4	IP6K1	TBC1D17	SPTAN1	FH	STRAP	SLC12A9	SERF1A	RAD51AP1
PFN1	CRKL	LTBP2	VPS37B	HRIHFB2091(hsa)	FAM49B	EEF1G	TRIM21	HIST1H2AC
SCLT1	HIST1H1C	H2BFQ(hsa)	SLC25A29	RSRC1	CCDC82	IGF2BP3	FAM198A	KNSTRN
KIAA0765(hsa)	PTPN23	DRAM1	ZDHHC19	PRDX4	PRDX1	G3(hsa)	VWA1	DYRK1B
KRT32	ATXN2L	CDCA8	SP1507(hsa)	MLLT1	TMEM86B	NFRKB	ZNF573	ERCC6L
RNF114	AK4	MCM4	FOXN2	STK17B	SLC7A8	ELAVL1	H4FG(hsa)	RRAGC
C12orf73	NAT8L	AC013268.5.1	FYN	HNRNPUL2	FAM126A	FAM178B	PXT1	UACA
RBM14	C6orf136	HIST2H3D	TMEM104	NCAM2	CCDC116	PHKA1	EMILIN2	TRIT1
NSL1	HP1BP3	RBM23	NQO1	ITFG2	F8	ZBTB21	ZER1	C22orf29
SLC25A44	HN1	SLC26A11	UREB1(hsa)	ZNF697	AP2A2	SNCG	ALDH9A1	TMPRSS4
NA	GPRC5C	KNS1(hsa)	KIAA1307(hsa)	LYPD6	BEND4	CCDC88A	EFTUD1	L3MBTL3
HSPA13	WHSC1L1	SETD3	HACE1	RNASEH1	SMARCA1	CDC20	CTTNBP2NL	NFATC1
FCHO1	ANKRD13D	NTN1	PDIA6	DMWD	DOK7	EMP2	H4FC(hsa)	RNASEH2B
KLHL28	ARHGAP33	DNAJC10	PUS10	TRIM59	MYPOP	AP2A1	APOL2	DKFZp686N2176(hsa)
NDUFC1	UVSSA	ORMDL3	IGF2BP1	GNPI(hsa)	KXD1	CCDC50	hCG_17321(hsa)	FGF7

OAZ3	CNOT4	AREG	MDH2	H4/N(hsa)	MYL9	RP5-824I19__A.1-002(hsa)	hCG_33495(hsa)	RFC3
FAM155B	STK4	C7orf50	BCAS4	PGM2L1	CENPO	ARSG	UBR7	PRO0992(hsa)
RNF11	ARTN	GOT1	RSRP1	CCAR1	CTBS	DKFZp686L08115(hsa)	TPP1	TMPPE
BRAF	VTA1	LACS2(hsa)	USP54	WDHD1	MTPN	PPP1R12C	ACSM3	SPT5(hsa)
FHL2	DNASE1L1	FAM69A	SLC35A4	CENPK	STX17	PM227(hsa)	ZHX2	FHIT
HGSNAT	RP11-835E18.2	STK11	SNAPC2	MAL2	SOCS5	CALD1	ZO2(hsa)	RPL27A
POU2F1	MAP2K2	PAPOLG	S100A11	DDX24	TFRC	ADNP	RLF	AIP
SFT2D1	c-met(hsa)	PPP3R1	WDR62	CSRP1	CABLES1	LCLAT1	FUT10	RANBP9
UBASH3B	SLC25A46	RDX	TGFB1	EBF3	PFDN1	NF1	SEC22C	DHCR7
FASN	RASA2	IFT122	SERINC1	GSN	VIM	PI2(hsa)	IGSF3	SYNGR1
CDH26	PPP2R3A	HIST1H4C	LIPT1	DMD	DSN1	JPH1	XPO4	DIAPH2
CDC6	ZNF775	HERC6	ZNF789	CRTC2	BMP2K	TMEM219	ZNRF2	HIST1H2AB
HIST1H1A	MAP7D3	TOB1	DNAJC9	CSE1L	ABHD4	NAA11	WBSCR20A(hsa)	SERPINE1
ACTB	RBM45	ZFP36	NDFIP1	CNIH1	RNF34	IGBP1	SLC29A4	HMMR
CTB-1144G6.4-006(hsa)	POLDIP2	INPP5F	MARK3	KANSL1	CNRIP1	ERCC4	ZNF311	ZNF33A
BLZF1	UFC1	ADAT2	RPIA	KIAA0010(hsa)	CXCL11	SERPINB1	ZNF585A	NKIRAS2
DCLRE1A	REXO1	SYNJ2BP	CASP2	RNF145	MAPKAPK3	FZD1	QDPR	VPS35
EAF1	PIGX	CRYA1(hsa)	GDIL(hsa)	POP7	TMED10	KDSR	KIAA0922	SCARB2
ACAS2L(hsa)	FEN1	XB130(hsa)	BLOC1S3	USP2	STIL	STK25	DHX9	FAM73B
TTC7B	SNP70(hsa)	KRT15	CREB3	KBTBD4	FOXA1	MYT1	CPLX1	GAS1
SNAP23	COA1	WAC	SHMT2	TSKU	CBLC	FAM45A	ANKRD28	KIF4(hsa)
DUSP4	MORC2	MIEN1	GBF1	CERS2	WDR73	CDM(hsa)	PCYOX1L	AGTRAP
PHLPP2	FYCO1	IQGAP1	COX16	ANP32A	DLL1	MED28	TCEA2	PLAU
MTO1	CDC8(hsa)	UBTD1	ZNF189	MAP3K14	HIST1H3C	UBAP2	DERL1	ASB9
PRLR	BID	NRBP2	ZC3H3	SYNGR2	SNCA	EMR2	LMNA	SMYD5
MPHOSPH8	CLIC4	GAN	CORO1C	SLC9A1	HRIHFB2157(hsa)	ZNF398	NDC1	CIRBP
FOXJ2	BTRC	PSME1	RDH5	SP4	KPNA1	TMEM106C	SLC7A2	A2D(hsa)
BSCL2	NPRL3	TRIP11	MAP3K11	CDH1	RCN(hsa)	PQLC3	CANT1	CDK16
C15orf13(hsa)	C1orf74	ARHGAP32	MCM7	PARP6	CORO1B	YAF2	AUH	MTFR2
DDAH2	PRR3	ASXL1	HIP1R	RAB11FIP2	SLC25A39	RNASEH2C	CCDC74A	LPCAT3
LIG1	CYR61	WNK1	FBL	LHPP	PIM1	DTNA	TULP4	CYBRD1
HLN(hsa)	CHD1	MNT	TCP1	CENPA	TPA1(hsa)	SLCO4A1	FN1	MGAT4A
ZSCAN5A	FAM195B	RUFY3	AP1M1	PITPNM3	IL17RB	ABCC3	ACSL1	CYB5R3
SEC16A	HIST1H2AE	PEX16	FKBP1A	TGFB1I1	SRPR	PTK6	ARHGEF9	SCN1B
MSAP(hsa)	SIRT6	NRTN	CHD4	C9orf72	ZNF692	HIBADH	AMACR	CCNG2
PPP1R37	LYRM5	GSE1	CXCL2	CHKB	KIAA1462	GTF2IRD2	IFFO1	HELLS
RAB3GAP1	ACER3	FTSJ1	DENND1A	RNF106(hsa)	CALML4	YY1	SEPT2	SNX17
H3FJ(hsa)	TMEM39A	GPATCH2	SMARCD1	ZC4H2	PRRT1	KIAA1228(hsa)	UBN2	ABTB2
ACKR3	HIST2H4(hsa)	PMF1	ARHGDIB	LIME1	ARNTL2	DAMA-236L13.16-003(hsa)	H4F2(hsa)	GTSE1
ZEB2	RCAN1	CDC5L	C11orf57	FECH	ZC3H12C	HSD11B2	IPO8	BTG1
MLK4	ATE1	VPS29	MBD1	PITPNM2	COL1A1	C10orf88	GATAD2B	C11orf80
UNRIP(hsa)	FAIM3	PBX1	SIRPA	SMC4	DNM1	CDC21(hsa)	SGTA	PIDD1
AFAP1L2	CCNL2	C16orf46	TYRO3	SLC30A8	PRKACA	KIAA1491(hsa)	ANXA11	SPIRE2
KIAA0164(hsa)	NTE(hsa)	EXD2	ZNF707	SFXN1	ZNF496	ARID5B	SPTA2(hsa)	ITSN1
STAG3L3	CCNA2	CRTC1	PPIA	MYH14	VCP	B4GALT3	SNAI3	USP22
NETO2	HIST3H3	NBR1	MTMR12	RAP1GAP	DNAJC15	TESK1	EPS15L1	A2LG(hsa)
TROAP	CLEC2D	RPL37	ZNF304	UBL7	ME1	DUSP7	RRP12	SIX2
TCF3	FBP11(hsa)	OPA3	FOXRED2	PIKFYVE	CACNB1	VSNL1	RAB22A	ZSCAN9
XXbac-BPG296P20.4-014(hsa)	SEL1L3	SEMA4F	ERP44	PPP1R12A	HCN3	H3FD(hsa)	RAD54L2	RNF26
H4FN(hsa)	RMI1	ARID1B	MSN	C5orf55	hCG_2023614(hsa)	PIGB	PKN2	hCG_1685949(hsa)
CASP8AP2	RSP1(hsa)	KIAA0432(hsa)	ASF1B	AFAR(hsa)	SUPT5H	KLHDC8B	AKTIP	UGT8
FAM195A	FAM111B	IRGQ	PACS2	DPPA4	LGALS3	PLD2	SEMA4D	VAT1
HIST1H3B	HNRNPD	SRCAP	SNIP1	TH1(hsa)	SYT1	JAG1	GRAP2	GCFC2
SNRPD3	SPC25	H4/A(hsa)	ITGB5	CMIP	NYMEL3(hsa)	DUSP3	GSTT2B	HSPC225(hsa)
DPP3	NT5C2	CANPL1(hsa)	PIP5K1B	NAV1	C10orf69(hsa)	KIAA1846(hsa)	TNC	MPP3
HIPK2	CHST11	SLC17A5	MORF4L2	FOXL2	PARD3B	C11orf54	C2orf27A	H3.3B(hsa)
TMEM63B	PITPNC1	TRPS1	MAPK13	RER(hsa)	IRF3	KPNA3	CRIPT	TLDC1
CNOT6	SH3BP5	RAF1	TRAM2	MB	DUT	JADE2	C21orf59	TSPYL2

TMED8	NXPH4	GPSM1	DNF15S2(hsa)	C3orf33	C6orf106	C3orf58	RNF182	FKBP1B
SMAP2	LZIC	ADO	ZNF107	RIF1	CDC42EP4	MDM4	AP3S2	HSPB11
HSD17B6	STXBP6	WNK2	ACTA2	SF3B3	FNBP3(hsa)	LARS	EIF4G3	NUCB1
LACS(hsa)	UCK2	FNBP1L	EPC1	CDKN1C	MOGS	SLC2A4RG	TMEM30A	CAPN5
DHRS7B	CDC47(hsa)	CDCA4	RHPN1	TYMS	LYSMD2	RTCA	ASTN2	PATL1
MPC1	APLNR	PRKRIR	PAX8	TMED1	XRCC3	KIAA0661(hsa)	BAG6	ZCCHC17
CRTC3	ERO1L	C19orf60	RAB17	ICA1	ZMYM2	SEMA4B	CDKN1A	MRNP41(hsa)
MEN1	DCTN5	FRK	RAP6(hsa)	ANAPC4	APLP2	OGFRL1	NAGPA	TMUB1
SCYL3	PLC1(hsa)	REM2	H2AFD(hsa)	KCTD21	C19orf48	MYO7B	SGPP2	LRRC40
RNF44	MSJ1(hsa)	MRPL2	CASKIN2	AGO1	DNAJA3	SSR3	BCLAF1	FACL4(hsa)
CTDSP1	SFMBT2	MST1R	MTMR10	HCST	FAM20B	DKFZp434G1035(hsa)	SWT1	FAM129A
MOV10	TCF4	FES	PP6R3(hsa)	NECAP1	CD86	CSNK1G1	MAP2K1	ENTPD4
DSCR4	NUP153	ATP7A	CDABP0131(hsa)	TRERF1	NDUFV3	ADI1	PCIF1	TRIM32
DHDH	MIER2	GYLTL1B	PDS5B	SLN	WASF2	GOT2	STMN1	DNMBP
GPR63	TLR1	GINS3	DBF4B	FBXL6	PIG4(hsa)	SWS1(hsa)	UHMK1	VPS72
SOX4	MAP3K3	EXOSC10	CAPG	HOTAIR	CARF(hsa)	FNIP1	DCAF15	MTCL1
HID1	ZNFX1	MMS19L(hsa)	UMK(hsa)	GGA2	GDE1	SLC39A6	RHOF	LMF1
H3FC(hsa)	SRL300(hsa)	KDELR3	FGFBP1	RP11-664D7.4	TBC1D31	SYNJ2	TYW5	ALG3
CDIP1	AP1G2	PLIC2(hsa)	RSPH3	NAV2	NFKBIA	BTF3L4	RALB	RTN4IP1
FBXL3	TSPAN18	TCIRG1	ALKBH5	CLAPB1(hsa)	FBXO10	FUT11	POFUT1	TAF3
AFAR3(hsa)	FUT3	ZNF764	ZNF337	DLG4	MYO5A	SMIM13	NAPEPLD	TMEM79
FBXO11	SOCS2	NBN	TMEM256-PLSCR3	PKP4	NENF	PROSC	WTAP	WDR83OS
PYCR1	XTP3TPA(hsa)	LIMCH1	EPAS1	CYB5M(hsa)	MYCBP	PRKCB	PIG30(hsa)	ZNRF3
PKIA	MICAL1	TIAM2	LACS1(hsa)	GBX1	PEREC1(hsa)	KLC1	IRAK2	CRISPLD2
PRO2286(hsa)	FACL1(hsa)	RANBP10	NAP1L5	BBIP1	BEGAIN	CDC73	SYNE2	TAX1BP3
PHKA2	GTF2H1	CIT	LBR	SMG9	AK3	MRPL24	SLC25A19	TRIM13
CHST12	INHBB	FBXO4	CKAP5	RFXAP	RNF19A	RBPMS	ATP8B2	TMUB2
ESYT1	NME4	B3GALT4	ADAM19	H3FA(hsa)	CENPB	PPA2	C3orf38	ZNF623
OAS3	GNAS	C7orf43	hCG_41078(hsa)	H2BFR(hsa)	FAHD1	UNQ2441/PRO5003/PRO9924(hsa)	ZNF354A	RP11-511I2.1-003(hsa)
DTL	PCLO	TRIM62	TBC1D30	HIST2H2AA4	H3F2(hsa)	SPRY1	C13orf10(hsa)	H3FT(hsa)
SDCCAG16(hsa)	IP6K2	ZNF664	NREP	CLIP1	NME7	TNRC18	CLSPN	KIAA1279
FARP1	C1QTNF6	ARFGAP3	KCNE2	NOTCH1	EIF4G1	HPRT(hsa)	JRK	MYBL2
DDAH(hsa)	KIAA0784(hsa)	GMNN	SLC35A2	GPR143	MIA3	CDC23	CHTF8	RUFY1
GMIP	CORO6	MAPK8IP1	HIST1H2BD	SELM	H3FK(hsa)	TRPC4AP	TNFRSF12A	TMEM143
TNK1	GOSR1	ATPAF1	SH3BGRL3	RP4-697E16.3-004(hsa)	RBBP5	MARK1	C20orf27	KIAA0224(hsa)
NFIX	CIB2	ATG4D	TFPI	RSU1	NIPA2	MTERFD3	DOCK9	ROR2
KEL	MKL1	FBRS	KIAA1612(hsa)	CAPN12	DAMA-236L13.16-004(hsa)	CXorf3(hsa)	AAK1	ARHGEF26
RNF183	WARS	PIGM	ERGIC2	FMNL2	CEMIP	SERF2	TNRC6A	AKNA
CDKN2C	SIX3	ECH1	IKZF1	GANAB	H4/E(hsa)	ECD	BCAN	SCAMP4
CRLF1	LAMA5	VWCE	C4orf46	PLCG1	MED31	PEAR1	TRAPPC1	GMEB1
SLC48A1	OBFC1	SLC25A51	DST	FAM89B	CITED1	ARHGAP29	TP53INP2	TNPO3
PHF19	BRPF3	BCL11A	HIST1H2BA	POLR2A	DENND6A	YBX1	FBXO18	PDK2
SCAMP2	ZNF10	NT5(hsa)	SSX3	TMEFF1	KIT	FYB	ULBP2	ANKRD10
ARG2	THAP9	ZNF501	ABHD2	BTG2	LY6E	STK16	SLC2A1	KCNQ1
MYBL1	LEKR1	CGI-82(hsa)	EFNA5	POLR2M	FCER1G	REPS2	HMCES	ZBTB25
CWF19L1	TMOD2	TSGA10	KIAA0101	DNAL1	HIRA	PSMC3IP	ARF6	DKFZp586K2222(hsa)
RHEBL1	RGS4	CIDEC	HES1	PDLIM1	RBL1	NMT1	NA	SLC4A11
TRIM11	PPME1	NT5E	PLAGL1	ZNF395	HIST1H2BG	ILF2	KHDC1	DADB-70P7.10-002(hsa)
TXNL1	SLC25A3	ERLIN2	BMP8B	C1RL	ACSF2	NSMCE4A	AKR7A2	DNAJC2
ATG7	TAF9B	ZSCAN22	VASN	ENO1	XYLT2	GGT1	C1orf109	HSPC075(hsa)
H3FM(hsa)	CRK	ZNF507	CENPL	NCOR2	SCNM1	MELK	SAMD5	C14orf159
A2LP(hsa)	SRP54	ANAPC8(hsa)	KBTBD6	SLC16A14	JMJD1C	LAMC1	IRX3	NEIL3
SIN3B	RED(hsa)	ZNF706	PPM1M	TSC2	VEGFC	FOXF1	MBNL1	ATF7
C18orf21	OTX2	LIN9	FAM204A	ARL15	PLIN3	RAI14	PRC1	PCBP4
TBCD	GUCY1A3	BCL2	ZNF746	DGKZ	HNRNPM	RAB40C	WDR4	H2BFH(hsa)
DTYMK	ALDH9(hsa)	CAT	OSGEPL1	B4GALT1	RBM26	SFR1	C14orf182	GPR161
ARL6IP5	SSH1	CPSF6	PDS5A	H3.3A(hsa)	ZBED3	RPS6KA3	AZIN2	HBP(hsa)
HMG20B	GATA3	DAGLB	ALDH6A1	HPS1	KCTD7	FAM124A	TMEM201	STK40

TUBGCP3	TRIM37	PDXK	TSH2B(hsa)	PABPC1L	KE04(hsa)	FBXW4	NANP	MBD5
FKBP3	CFL1	CD151	ZNF575	PAQR7	ABHD14A	NFE2L1	NFIA	KPNB1
PSD4	COL12A1	UBA52	ADIPOR2	AK6	PDE12	NEDD4	KLRD1	RBAF600(hsa)
SPARC	IQGAP3	LRP11	PAST(hsa)	JUN	TTK	NDUFB2	NP95(hsa)	MYRF
PHF7	FKSG27(hsa)	ARSB	FAM60A	MCM3AP	AAMP	CNPPD1	SRSF11	BRWD1
UGCG	ARF3	SLC1A1	CAPRIN1	KIAA1521(hsa)	ERCC1	FOXJ3	HIST2H2AC	CBX5
LPPR2	CC2D1A	NRSN2	MTRF1L	ZDHHC2	SAMD14	ZNF414	ABHD12	KLF4
SAR1A	POLR3F	HIST2H4B	SLC3A2	RPS12	DIS3L2	SFI1	COBLL1	STARD4
GUCD1	TS(hsa)	SPT5H(hsa)	IER5	FAM3A	SUOX	MKI67	ATP2B4	TTPAL
STAT3	EMILIN3	PTPRM	ARAF	GPR19	NCAPG	TSPAN14	OGFOD1	NLRP11
TAGLN	RNF123	M7V1(hsa)	LYPLAL1	CRTAP	UBE2Q1	POC5	NOL10	RAD9A
HYPJ(hsa)	NDUFAF6	MAP2K3	EMS1(hsa)	PPP1R15B	C1orf213	AMHR2	ANKHD1-EIF4EBP3	PODXL
TRIM68	ZNF114	HSD17B10	DDX10	RBMS2	TRIM33	TAF1A	MAMLD1	HMGCL
PDE4B	ARMC8	PHB	MTFR1L	PDLIM2	CREBZF	IGF1	TEAD2	DMKN
MAN2A2	ARRDC4	CD40	SPIN4	WDR33	TGOLN2	RRP1B	GNG4	MXRA7
E2EPF(hsa)	ELANH2(hsa)	SLC45A3	KIAA1109	HIST2H2AA3	IL1RN	LMAN1	PDZK1IP1	TBL1XR1
HPCAL1	SIRT1	MYH9	ZBTB20	IL9R	SLC39A9	KAT2B	BHLHB9	FANCB
NUFIP1	HOXB8	HBP1	CCDC64	CDK3	CMTM4	PAXBP1	TNFAIP1	HIST1H3J
PDGFRA	SLC35G1	FAM20A	MYSM1	AMZ2	CPT2	KLC2	DISP1	SMOX
PTPRK	MED14	TPBG	DCTN2	C19orf54	GTF3A	LRRC8E	ZWILCH	NAA60
TSPAN1	THBS1	RFX1	ABI2	TMEM173	WDR45B	APEX2	ATG9A	DYNC1I2
EFCAB14	STOM	CD3D	NFX1	DSCR3	MAPT	ACD	RPMS13(hsa)	ATP6V0A2
HS1BP3	TBC1D23	IMMT	SHISA4	BABAM1	MUC1	PRKAR1A	HSD17B12	ATRN
ARGLU1	RUNX1	BIRC5	DPM2	FAM209B	DAG1	PEF1	LINC00337	FABP3
DIP2A	LACS4(hsa)	TUBB	KRTAP2-3	NCAPD2	RNGTT	VMA21	EIF2S2	HIST2H3A
PRKCQ	IPO9	SLC31A1	IFNAR2	FAM208B	ASNA1	KIAA0109(hsa)	ALG13	DAN26(hsa)
SSX1	EIF4A3	KIAA0430	SP2	TOP2A	SF3A2	MYADM	PRKCH	H4FM(hsa)
FERMT2	ZNF521	BCAM	CREM	CCND1	HIST1H4H	PROSER1	HHLA3	ASXL2
ENPP3	EIF3H	SNX9	CRAT	IMPAD1	SMKR1	SAP30L	DADB-333F21.4-002(hsa)	GPS1
TOX4	HIST3H2A	MACF1	GPR183	MAN2B1	INTS9	PIG28(hsa)	MON1B	SECISBP2L
RNF144A	KLRC3	PTPN21	C12orf10	GPC1	ZNF263	CSNK1E	EIF4EBP2	LYPLA1
POMZP3	USP6NL	CCNF	KIF12	JAK2	ARL6IP6	SLC39A13	LSM12	ESAM
CSNK2A2	MTMR6	TAF9	PREB	GRPEL2	PAFAH1B3	RP11-452K12.5-010(hsa)	INA	KLHL18
NBL1	PTRF	H4/G(hsa)	NCL	ILF3	MSANTD2	DESI1	MLXIP	TGFBR2
CACNA1A	SPATA33	SNTB2	POC1A	ZNF133	SAP130	SA2(hsa)	HIGD2A	SLC1A5
ELL2	GTF3C4	TMTC3	NAPG	TNFAIP2	TTLL1	CYB561A3	R3HDM4	COLGALT1
MIER1	TMPK(hsa)	WIPF1	NAB1	HIST1H3I	PABPC1	EP300	ZNF524	PCBD1
EEF2K	ANXA3	EME1	RPLP1	PFN2	TBC1D25	POM121	IFI35	DADB-70P7.10-003(hsa)
BAF155(hsa)	HDAC7	NUP98	FAM222B	TSR2	MR1	MIDN	RRM2	SLC39A1
TOB2	ZMYND11	PHACTR4	NOGO(hsa)	WBSCR27	CDYL	SUN1	VGLL4	PTCD3
CDH2	HIST1H2AM	CDA03(hsa)	MT1F	SRK(hsa)	MCMBP	AMOTL2	CAD	FERMT1
C3orf80	LGR4	CEP83	CMTR2	SMAD4	CFL2	BRE	NOH61(hsa)	RAB14
G3BP1	C21orf58	ANKIB1	C7orf49	SLC16A5	ACADVL	SRM300(hsa)	SLC35G2	TFG
E2F1	PTCH1	MAGEB2	UNG	ST6GAL1	RBM47	HDLBP	SLC39A10	MAML1
HSPA1A	GS3786(hsa)	SH3GL1	RBM38	SLC35D1	RPS18	DDAH1	CKAP2	ZNF641
RAB3GAP2	ECK(hsa)	M11S1(hsa)	SPEG	FNDC3B	CDCA7	SLC10A7	CG1(hsa)	TMEM200A
RPS23	AGO3	NRGN	KLF10	SEPHS1	RBBP4	ANKMY1	C21orf2	PON2
PDCD4	DBN1	CHRNB4	SNPH	C2CD5	KLHDC3	GAB1	TGIF2	ZFAND3
ARHGEF5	MMP15	C22orf23	SLC25A1	ICMT	ID1	ZNF35	XXbac-BCX40G17.4-002(hsa)	MGAT4B
MASTL	CABLES2	DCBLD2	SFXN2	RLTPR	GPR3	PPARG	HIST1H2BB	MED22
PCDC5RP(hsa)	RAD51	APH1A	RASSF7	CTNNB1	APBA2	TMEM245	RABEPK	PVRL1
MMS19	PIP5K1C	PPP6C	MPP5	PLA2G12A	MRPL10	ZBTB3	LDLR	STRADA
ZNF318	RAB25	NR4A2	NT5CP(hsa)	DLGAP3	NOL11	ARL5B	SLX4	RPL15
TUFT1	Nbla10545(hsa)	UBL3	SLC35B3	RHNO1	MCPH1	PRIMA1	PDAP1	AMN1
COMMD1	H1F1(hsa)	CCDC14	CCDC85B	BAF170(hsa)	PANX1	KCNAB2	SLC29A3	ATP2C2
TBC1D13	ABCB9	ATP6V1D	TMEM181	KIAA0153(hsa)	TERT	PIGT	TMEM25	PLA2G15

ASB1	SAC3D1	AAGAB	G6PD	RP4-657E11.7-008(hsa)	NYAP1	CNTNAP3B	IL17RC	PRMT3
PRSS3	KIAA1522	GYG(hsa)	RP11-244N20.4-002(hsa)	SLC10A1	HIST2H2AA(hsa)	TFCP2	DONSON	PTGES3
FAM167A	DHCR24	FAM127B	NECAP2	THOC6	ZNF426	NANOS1	GDPD5	ZBTB47
NPDC1	CEP57L1	FZR1	USP12	VTI1B	KLHL7	NOTCH3	TXNL4B	MARCH3
PRR22	PRRG4	FAM86B1	INSIG1	NTMOD(hsa)	NEO1	SLC44A1	RPRD1A	TTC6
RTTN	NINJ1	KIFC2	H2AFP(hsa)	RAE1	ARMCX3	CCDC169	PER2	ATP5SL
MORF4L1	ZNF681	DNAJB6	FAM104A	CRY2	IPO12(hsa)	DNAJC21	TFE3	SNRNP70
CDON	RP11-49N14.8-004(hsa)	RAVER2	MARC2	MSH6	QRSL1	SAMD10	MYL6	RAPH1
LDHD	IMPDH1	AGO2	DFFA	C6orf1	AFAR1(hsa)	NCAPH	H2BFF(hsa)	UBE2C
CRYAA	TTLL12	MCM5	TBC1D10A	TUBB3	TEAD1	TAX1BP1	SH2D2A	PTPRH
ADAM10	CLCN3	LIN28B	CBR3	RNF25	PXK	GNAQ	DCAF4	HSPB4(hsa)
NAT6	LENG8	IL21R	PAQR4	ASNS	LRRC8C	BCCIP	CHRNA10	FADS3
CAPN2	PRSS8	IMP4	CBFB	SNRPA	hCG_2018597(hsa)	FAM73A	NUP155	GCNT2
CXCL16	ATMIN	PKLR	PRADC1	RRP8	BAZ2A	PPP2CB	MOB3C	XXbac-BPG296P20.4-010(hsa)
EMP1	CDKAL1	AFF4	GCH1	RHBDF1	CHIC1	RPUSD3	HK1	BAD
HIST1H3A	AMPD3	FBXO6	SIDT2	NEDD8	SMARCC1	H1F5(hsa)	PP781(hsa)	RPI(hsa)
VASH2	RPSA	ANKRD17	TP53	WDR76	STRN3	CDK17	ASCT2(hsa)	RPA2
FBXO32	ATAD2	STAG2	BMP1	RNF208	HIST1H2AA	ZNF780A	PARD3	PPP1R3E
HIST1H2AJ	PWP2	TIMM8A	RP1-71L16.2.1	MAPRE2	SKI	TRMT10B	HEATR5A	DEF8
ATP5D	ZNRF1	ATPIF1	RNF187	CCNE2	NADK	SNF2B(hsa)	SPC24	MAF1
UROS	EIF4E3	NTNG1	EPB41L1	DSCC1	ANKRD27	SELPLG	DKK1	USH2A
OSGIN1	TINF2	FAM203A	VGL(hsa)	AXIN2	NT5M	CFL(hsa)	LRP3	BRE1B(hsa)
ZNF222	C11orf30	STT3A	EXOC4	TMEM56	CSNK2A1	TWISTNB	ADARB1	MTHFSD
ST6GALNAC4	EDEM2	HNRNPA0	NLE1	DAAP-21F2.8-002(hsa)	MOB4	KNSL6(hsa)	ZBTB4	ELK1
EPHB2	C10orf54	C1orf86	RBPMS2	ATP1B3	CFLAR	DKFZp434N101(hsa)	TRIM2	COQ2
FAM53B	AKIRIN2	GRHL2	PLEKHA5	GFPT1	KIAA0839(hsa)	CHML	FIP1L1	CMPK1
FBXO46	TPR	PLA2G6	TMEM140	INPP4A	GLUT3(hsa)	PSMD7	IFI27L1	SLC30A6
CEP152	ANKRD52	DOLPP1	ACS4(hsa)	FASTK	IFIT5	CELSR3	GORASP2	ARHGAP22
NFKB1	MAP2K7	IRF2BP2	FGFRL1	SLC25A13	FKBP8	RP11-16N10.1-004(hsa)	HMGCS1	MTDH
SAFB2	NSUN5	KEO4(hsa)	PLA2G10	SMARCC2	GGCX	GLTSCR1	PTPN12	GPR137C
IRF4	PVRL2	NA	B3GALNT1	CCDC163P	CDKN2AIP	ABCA7	H4/D(hsa)	LPGAT1
EXT1	MOB3A	DGCR2	TARS2	TAPBP	DHX38	SEPT3	OPN1SW	PRR11
DHPR(hsa)	H4FK(hsa)	RASIP1	UBAC1	RFTN1	AKR7A4(hsa)	CLN6	EXOSC6	ENO2
PCOLCE2	H1F2(hsa)	FBXL2	C10orf11	E2F5	WDTC1	ADTAA(hsa)	BARD1	ZNF428
MANEA	TNXB	GCOM1	AGFG2	TYRP1	TFCP2L1	POLD(hsa)	SMPD1	TNRC6C
PTPRU	SPATA2	SLC38A7	MOCS3	SSFA2	JAK3	ALAD	GAREML	ST6GALNAC1
S100A16	ELK3	HIST2H3C	C11orf23(hsa)	ST3GAL2	SLC13A3	MRE11A	LIMA1	N4BP1
NTNG2	TMEM184B	MCFD2	LITAF	HECTD3	NAMPT	RHBDF2	LMF2	ZNF282
TOP3A	PPP2CA	FAM103A1	KNS(hsa)	PIGC	XAP4(hsa)	JDP2	SHANK2	FANCI
MET	SPATA31A7	MYO1C	PPP1R11	C17orf53	RPL5	ARHGAP42	MTAP	MCM10
ATP6V0E1	PITHD1	PRDM11	DDX21	SAP30BP	PRPF38B	TRPM4	WAVE2(hsa)	RNF216
SLC6A6	SLC22A18	TM9SF4	PHF6	STRBP	MTSS1	ATF7IP	PIGF	WDR96
UTP14A	H4/J(hsa)	C9orf16	FAM178A	LARP(hsa)	PLEKHG5	IGFBP2	SGSM2	SYVN1
FSTL3	XK	RP11-265M18.2	PRRC2C	SLC8B1	CPEB2	BRD3	ZFP41	RHBG
TROVE2	INPP5J	BHLHE41	CLEC4M	MRPS30	CNKSR3	ACRBP	RP11-197M22.1-003(hsa)	KIAA1715
SLC10A3	SMIM20	MEF2D	MCTS1	AK9	TMCC1	GPIAP1(hsa)	EIF5A2	ORAI3
SLC39A8	H3F3(hsa)	CDKN2B	CDC42EP1	TIGD6	CAND2	C18orf54	CCNI	MYLK
FGD1	SCPEP1	MTMR9	MTURN	C1orf85	CYSRT1	KIAA0030(hsa)	TMEM263	GTF2F1
SPRYD4	GABRG3	SRM	SEMA3F	TSEN15	COMTD1	TGM2	PRODH2	VPS54
CBFA2T3	FLOT1	INTS8	SEC61A2	GIGYF2	CDK18	WDR3	HMGN4	ACCS
FEZ1	SMARCA4	MOB1B	ANXA5	KIAA1207(hsa)	FGF9	S100A3	HNRNPL	D3F15S2(hsa)
hCG_39482(hsa)	GTF3C3	SLC25A20	C9orf69	PIGL	PTPN11	TMEM92	RUFY2	WRN
FANCD2	KIAA0312(hsa)	C20orf193(hsa)	DCTN4	MPHOSPH6	POM121C	GRK6	ERGIC1	SUV39H1
GTF2E2	PLCD3	STAG1	ASH1L	C1orf61	HGS	ANKLE1	PPIG	BAZ1B
H4/C(hsa)	SSX5	MFAP2	MBD3	QARS	TMEM246	GJC1	MAP1A	MCM3
ZNF273	GNPDA1	PTPRN	ADAMTS2	H3F3AP6	MAD2L2	NLRC5	NAV3	PID(hsa)

SSH2	BCL2L13	URI1	S100A2	MRPS25	GCF2(hsa)	RP11-108M9.3	CCNJL	PNISR
UBP1	FAM218A	SRGAP2C	SSX2	DDX11	CDC42SE2	SDHA	CCNK	TMBIM6
HFE2	FGFR4	GPATCH2L	MND1	ALAS2	C2orf49	SEZ6L2	HAUS5	KDM4C
ADCY1	CPEB4	PRP16(hsa)	PCSK7	MAOA	KLK1	MAGI1	ID3	CCNL1
SLC44A5	L3HYPDH	CDCA5	METTL23	C1orf210	FARP2	CNOT1	STAT1	RCH2(hsa)
RRAGD	CALB2	TIP47(hsa)	HLA-DMA	CDV3	RNF181	ZNF281	HIPK1	GINM1
MYB	SP140L	DHX37	SRSF1	PIG1(hsa)	XXyac-R12DG2.2	EFNB1	ENY2	H1FX
CALCOCO2	NOL3	DAP	MYNN	INPP5K	GPS2	PTP4A3	AHNAK	NNT
IGSF9	GAPDH	TRMT61A	PURB	KLHL32	NPNT	RP11-269F19.1-004(hsa)	SEPT6	ZNF703
TOM1L2	INTS3	CTAG2	CSMD2	HYAL3	PHC2	ADCK4	RIMS3	RHOG
HNF4A	HDAC1	CTSD	CCDC97	FAM92A1	COX4I1	UBE2D3	HIST2H2BE	ECE1
NDST1	UBIAD1	ARID4A	ACSS2	Z82214.4-003(hsa)	GAS8	CASP6	TRPM7	CCND3
CDT1	ING3	POLD1	HUS1B	ARL3	YWHAZ	FNDC4	LUZP1	MED8
RP4-695O20__B.10	GALT	STRIP2	LRPPRC	MXI1	CDAN1	DAMA-236L13.16-002(hsa)	PRKDC	SYT16
BBS2	NA	SPHAR	EVL	C1orf131	YKT6	WHSC1	TMEM206	TBRG1
ANO6	MIR1199	FAM86A	LCP1	AXL	DKFZp781H1755(hsa)	COX15	RP11-20I23.8	NR2F2
ZDHHC21	AFF1	TJP2	CXCL3	WWC3	EIF4ENIF1	ZNF319	KCTD15	ZMYM3
ARMCX5	COPS7B	TNFSF15	ZNF467	SAMD4A	ERAL1	LOXL1	AP1G1	RRAS
SDR39U1	MCM6	KIAA0332(hsa)	ZNF548	KNS2(hsa)	FBXL19	PHF12	SMAD7	CARHSP1
CCNB1	MED21	FAM83A	SREBF2	NUFIP2	CRELD1	DNAL4	ORC6	KIF5B
TOR1A	DCAKD	TNFAIP3	ZNF358	ADNP1(hsa)	C1orf159	TOM1	PRKAR2B	NCOA1
ATP2C1	PLAGL2	KIAA0061(hsa)	ENKD1	POGK	PRKX	EIF4E2	SURF2	LIN37
TTL	TFEB	ELP5	FAM188B	RUNX3	CLAPA2(hsa)	CHMP3	CDC42GAP(hsa)	ST3GAL4
VPS45	PME1(hsa)	DTNB	ZDHHC12	TP73	SLC25A36	HNRNPH2	CIZ1	ACOT8
PINK1	C19orf84	PLEKHH2	TAF12	RAPGEF2	CUL7	TMED6	BVES	SLC27A4
UCK(hsa)	LPXN	THRA	HIF1AN	UCP2	TIGAR(hsa)	DMPK	CYCS	HIST2H2AB
UMPK(hsa)	MTX3	CCSER2	XXbac-BPG254B15.2-002(hsa)	RAB1B	NTPCR	CREB3L2	MEF2C	SOBP
RIN1	NF2	PI4KB	BTF(hsa)	NFATC2IP	RFX3	ATXN7	GLRX5	GPATCH4
GDAP1	ZNF33B	RAB21	RHOC	MSL1	ZNF708	EPS8R2(hsa)	CEP55	H2AFQ(hsa)
HOXC9	USP1	SIRT5	SENP6	H4FH(hsa)	CPT1(hsa)	TSPYL1	UBE2W	DAXX
NUP62	ACVR1B	IMPD1(hsa)	IL22RA1	MARCH2	REXO2	TFB1M	RBP5	HSJ2(hsa)
UBR1	KIF15	LETM2	SIRT2	TMEM243	SPFH2(hsa)	DHRS13	C8orf2(hsa)	ABCF1
SOCS4	KIF2C	CDK1	ZMIZ2	TRAK1	MBD6	CYB561	TMEM127	PDGFRB
MED13L	LRRC8A	GON4L	ZNF385C	C1orf198	GDI1	EPM2AIP1	hCG_41525(hsa)	PIN4
IL18BP	TNS4	AIM1	ECM1	DYNC1LI2	MFSD3	GINS2	MKNK2	ARHGEF3
MRPS26	p16Ink4a(hsa)	MLLT3	HPSE	ZNF503-AS1	ZNF592	SIX5	PSRC1	ZNF689
BCL9	DDX56	SFT2D2	RECK	CISH	VPS45B(hsa)	CNTNAP3	FANCA	PTPN18
TENC1	UBQLN1	GPR158	CTTN	FABP5	CYP51(hsa)	TCTA	SIGIRR	PHLDA1
NOGOC(hsa)	TMEM161A	TBC1D2	GABPB1	PRO0750(hsa)	PSMB2	PARP1	ZIC5	HOXA13
LPPR4	NR6A1	LAMP3	UBA2	IFRD2	STIM1	GK5	DCLRE1B	WBP11
SRR	CDKN3	TMBIM1	EEF1D	PTPLA	MTA2	SLC9A3R2	HERPUD2	RP11-65F13.2.1
FIGF	TGFA	JOSD2	OSR2	CNNM3	SEMA4C	CTSB	PDE7A	C19orf1(hsa)
CYFIP2	C10orf10	NUP160	CCDC150	OMB5(hsa)	TAF4B	H3F3B	PLK1	ARRB1
RPF1	AP2B1	SLC37A4	VICKZ3(hsa)	PALM3	FAM81A	FOXP2	ZNF367	C21orf33
FGD6	DLG1	SHCBP1	ICAM1	HIF1A	hCG_15646(hsa)	RPL37A	RP11-6D1.6	CYB5R2
CIC	SORT1	SPCS2	NLRX1	PPP1R10	TRIP5(hsa)	WWOX	TSN	ADAM15
COMMD9	HIST1H2BO	TEAD3	RNF40	CLN3	H4/O(hsa)	HNRPUL1(hsa)	GYS1	CD44
RP11-569G9.2-002(hsa)	PHGDH	TMEM109	ZNF207	TREX1	TMPO	SEC24B	ARPC4	AP2M1
C15orf26	DDX38(hsa)	MTFMT	H4/M(hsa)	E2F3	TNK2	SERPINB5	ZFP36L2	TFAM
PAQR5	RSF1	EPN2	ITSN2	POGZ	SRRM2	CLPTM1	RWDD2B	PRKACB
GNA12	TBCK	MSL2	HNRNPK	ARFRP1	RNFT1	MPHOSPH9	SUFU	ARHGEF28
PSMD5	RTFDC1	EIF4EBP1	TNRC15(hsa)	SUSD2	CAPN10	CAPN1	H4/B(hsa)	SERPINA1
HES2	NEURL1B	GSTM1	GRN	DAMC-157M7.11-002(hsa)	PTPRG	SLC22A17	VEGFB	MAP1B
NUMBL	XRCC1	IFT57	SIPA1L1	CYB5D1	CCDC167	F12	C16orf13	CD47
SNRPD2	E2F8	SLC26A6	PP2593(hsa)	APITD1	H2AFR(hsa)	STAMBP	AURKB	H2AFA(hsa)

NUCB2	FACL2(hsa)	LONRF1	STC1	RP11-176F3.6-007(hsa)	TEX261	DLD	ZNF3	C14orf154(hsa)
EVI5L	CHD8	MINK1	MIS18BP1	RGMB	GEMIN5	SLC5A2	UPP1	SWI5
RP11-321N4.1-004(hsa)	NRDE2	LLPH	AZIN1	H4FB(hsa)	RPAP1	CERS6	MBLAC1	SOX5
SR140(hsa)	USP20	C15orf38	RIMS4	NRIP3	H4FO(hsa)	TMEM150C	USP11	MIER3
TPD52	SMIM1	CENPQ	GLTP	MCTP2	KLHL15	NA	HIST1H1B	VAMP2
VPS45A(hsa)	APEH	TIFA	RALGPS2	C2CD2	PRIM1	GAPVD1	TMEM5	ADH5
EIF2AK1	TIMMDC1	MFN2	GAPEX5(hsa)	VPS37A	FXR1	DDX39B	NASP	MPV17L2
UBE2E3	CYP51A1	DDX17	VAV2	KRI1	CDCL1(hsa)	G3BP2	CMK(hsa)	CAMTA2
ZNF621	DAMC-157M7.11-003(hsa)	RAB12	C19orf82	HJURP	MBLAC2	RBBP7	ADK	XYLT1
RFX2	DBNDD1	RFNG	TXNDC16	TMEM97	MTA1L1(hsa)	SEC22A	KHNYN	C17orf80
ST3GAL3	IST1	DCTPP1	H2AFV	GINS1	TSPAN4	C7orf55	NFYC	tcag7.648(hsa)
AREL1	MBOAT7	DHFR	ZNF317	FUT2	RASA4	SPATA20	UBE2S	ALDH7(hsa)
SH3BGRL	SEPT1	EGFR	DCAF7	OPHN2(hsa)	CDH24	FBXO27	PLEKHG6	ATG4B
C1orf43	OGT	SSR2	PP13181(hsa)	GDAP2	SENP1	CYTH1	ULK1	FAM134B
PACS1	BTD	AFG3L2	SBF1	CEP41	SLC26A2	PPP2R3B	MAGT1	SIAH2
NELFD(hsa)	C2CD2L	SMTNL2	ZNF253	HIST1H2AG	NXPE1	TRAPPC6B	CWC15	DGKH
CD200	RPS15A	KIAA0391	RAB43	ATRIP	HIST1H2AL	FAM214B	C5orf63	NUDT19
ZUBR1(hsa)	OAF	ASPHD2	APOL4	NICN1	H3F3A	ANO10	SRSF3	LHX2
SFN	ZNF804A	SLC2A11	CREB1	MAPK8	H1F3(hsa)	FAM149B1	ARHGEF17	TRIM26
TOMM40	CDC37	DNAJC4	GYG1	ZAP70	LYST	MFF	FOSL1	UBR4
TRIP(hsa)	HIST1H1E	TYMK(hsa)	ETV5	MPEG1	EPHA2	ADTAB(hsa)	FUT8	NLGN3
CDS2	PRPF40A	PPP3CC	CDH10	TMLHE	DNM1L	TP53I11	SNX25	NA
ANGEL2	SIAE	P4HB	E1BAP5(hsa)	MIF4GD	My043(hsa)	LRRC23	MOAP1	SENP5
NUDT12	ALDR1(hsa)	RP11-511I2.1-004(hsa)	ZNF436	TMEM116	CERS5	ZFR	WIPI2	XPO5
RNF38	AKR7(hsa)	KIAA0097(hsa)	SORBS3	ATL2	LARP4	NUTF2	TRAPPC13	NA
KIN	ALDH3A2	PRAF2	PKP2	RP11-517O1.1-006(hsa)	CUX1	VPS33A	H3FH(hsa)	HYPA(hsa)
KIAA0060(hsa)	TTC38	CSTB	FBRSL1	RQCD1	RIMKLB	DNAH14	AIMP2	CDC25A
FER1L3(hsa)	ARHGAP24	RDH13	COL6A1	ZC3H7B	SEC31A	HNRNPH1	ALDOA	IRF2BP1
MICALL1	NAP1L1	KPNA2	KIAA1558(hsa)	C8orf37	BACE2	PYGB	FAM65A	FNIP2
MRS2	BLOC1S1	ZYX	SEC63	KTN1	TIGD7	PELI3	PTHLH	EREG
RPS28	LTBR	IER5L	OK/SW-cl.35(hsa)	POLR2F	FAM219A	GSTK1	BCAR3	ENTPD3-AS1
ZBTB38	STX1A	KLF7	MED4	MRJ(hsa)	PSMG1	CD22	LERK2(hsa)	RP3-467K16.3-002(hsa)
ZNF93	GPIP137(hsa)	EVI5	SURF-4(hsa)	PKMYT1	ZC3H14	OAZ2	NR1D2	ICBP90(hsa)
CAMKMT	PVR	TPX2	HSPA1A	DNAJB9	RAB40A	RABGDIA(hsa)	SUSD3	C2orf43
NAT16	PIK3R2	GGH	SEMA4G	HERC2	RNASE4	FANCG	LAMP1	RP11-1149O23.3
RIPK2	PLEKHF1	FAM174A	FAM76B	PAG1	LDHA	DCP1B	NIPSNAP1	TGFB3
MED25	ARL2BP	TMOD3	HEATR3	CSNK1A1	CTNNB(hsa)	RDR(hsa)	HOXC4	TIMP1
GLIS2	SH3RF2	ELOVL3	FRG1	CREBRF	LZTS2	FAM210A	NT5B(hsa)	TRIM28
AIM1L	ARHGEF12	KRTDAP	SZT2	CGGBP1	GLCE	IRGQ1(hsa)	FGFR2	MSX2
PKN3	SGPL1	TRAPPC4	TRIM41	HME1(hsa)	FAM3C	EPHB6	METTL3	YY1AP1
C1orf27	REEP3	KIAA1578(hsa)	BM28(hsa)	CEP68	FBXO9	DGAT1	UBXN2B	RXRA
PEX19	SLAMF1	RPUSD1	EZH1	hCG_2040048(hsa)	KIAA0642(hsa)	AKAP13	THBD	WHAMM
KSR1	IL2RB	NACC1	FAM122B	CAPN3	RGCC	ESYT2	AK1	ZNF213
FUS	TP53I13	FAM64A	C7orf73	NKD2	CRYL1	GNAI2	HSPA1B	SPRY2
THEM6	FBXO42	SERPINH1	RGMA	CYP4F3	PDRG1	CPA4	PRR19	RPL7
FITM2	ALG1	ZNF226	APOA1BP	RNASEK-C17orf49	GRINA	IGFBP3	STEAP3	MEX3C
KIAA0886(hsa)	SZRD1	ZNF580	ARHGEF10L	RNF24	TMEM59L	PIK3CA	TRIM58	CACNB3
ANXA4	B9D1	NOTCH2	ATG14	MRE11(hsa)	PIP5K1A	H2AFI(hsa)	IQSEC2	ZNF581
NELFCD	TRAF7	ZNF519	AGPAT2	SLC29A1	FAM120C	SLMO1	CDC42SE1	YTHDC1
ITGB8	LMBR1L	GABARAP	ZNF576	YPEL3	TNFRSF25	GNB2L1	SERF1B	SCMH1
PGR	YWHAG	CHPT1	ASAP1	RABIF	CHRAC1	SLC1A4	C11orf68	USH1C
CADM2	NPW	HEATR1	GOSR2	CHCHD10	C6orf62	KIAA1011(hsa)	MGAT5B	R3HDM1
FXYD1	PORCN	CARS	ZDHHC16	LNPEP	SRC	NEU1	ATOH8	H4FA(hsa)

FAM126B	H4/I(hsa)	PTMS	ZBTB46	AFF3	TCOF1	RALGDS	SLC27A2	DSP
PWWP2A	C12orf5	HIST1H3E	RNF10	CXorf38	KLHL23	FASTKD5	ITGA6	NPWBP(hsa)
RHOBTB3	PPP1CC	TSPAN12	KIAA1524	DLC1	DNER	RP11-30H12.3-002(hsa)	IGDCC4	SCRIB
WDR7	MCL1	CHMP2A	MAPK1IP1L	TIRAP	DUSP8	ZNF211	hCG_2017814(hsa)	RPL10
RTN4R	PTPN4	TMEM132E	GREB1L	POLD3	STC2	LGALS8	NUA(hsa)	RIN3
ASIC1	KIAA1398(hsa)	GLS	EFHD2	RABGGTA	PPAPDC1B	ZNF81	CEP290	H2BFN(hsa)
CDK4	POMT2	PLOD3	EDEM3	CLEC11A	PLAUR	FBXO44	TMEM200B	MECR
PROM2	TMEM37	ID4	cyclinE2(hsa)	LINC00472	SOX6	GRIN2D	CLAPA1(hsa)	ZFAT
JMJD4	SAZD(hsa)	VCL	STK32C	SUPT16H	ICAM3	LSM2	NAPA	IL6R
EPB41L2	BBS1	DHX40	CHP1	DAQB-147D11.2-002(hsa)	CYBB	KIAA0355	ZNF618	ABRACL
COG2	TBL3	ARHGAP26	PPP3CA	FOXN3	hCG_29955(hsa)	LONRF2	TANC2	UBOX5
GPHN	METTL1	USP47	ZBED1	FAM120AOS	ANKFY1	SHMT1	METTL7B	C9orf142
BSDC1	NFKBID	MAG	MLLT4	SNAPC5	DDOST	TAF1C	SERINC2	CYB561D2
CXCL8	PTPRR	TLE2	AKT2	MYC	CORO2B	KRT222	EG5(hsa)	COA7
RAC2	FBXL12	IMPDH2	GLUT14(hsa)	PPP6R3	ATF1	GPSM2	EPHA4	KCND3
ZNF879	MRPL34	CTSC	TRIM14	ADAR2(hsa)	PTP4A1	PLXNC1	RRBP1	EFNA1
NUCKS1	H3FL(hsa)	AMOT	CDA	DEDD	SDAD1	C5orf24	DECR2	ELF1
BAIAP3	FAM35A	VAPA	RANBP3	ZDHHC6	PAK4	VPS39	DIXDC1	PRR25
BLM	ZNF639	RHOU	CD320	SMU1	LRRC28	DDA1	FAM171A1	ZCCHC3
METAP1	RP11-321N4.1-003(hsa)	H4FD(hsa)	AGPAT1	MRPL16	H3FF(hsa)	SPINT1	PDLIM7	RBM15B
RDH11	TAOK2	PLEKHG4B	ATP6V1B2	ACAP3	FGFR1OP2	ASS1	SCNN1A	TXNIP
RAPGEFL1	ZDHHC8	DPAGT1	PCGF5	TATDN2	MESDC2	IL17RE	ST5	LUC7L2
SCD5	UHRF2	C9orf117	UBQLN2	UBE3C	SHOC2	NCDN	VEGFA	HIST1H4J
HIST1H1D	ACTG1	GM2A	RBM8A	KIF4A	DNMT3A	FAM24B	TBX18	KMT2B
TPM1	PCK2	RPP25	SLC7A1	PARVA	TOLLIP	CYP1B1	DAMA-236L13.16-005(hsa)	HIST1H2BO
LSR	TSPAN3	NHLRC2	FOXP4	CENPF	KMT2D	SCD	SYT15	HIST1H4E
RP3-339A18.4-002(hsa)	FOXM1	IKZF4	D7SR(hsa)	IKZF5	SEPN1	HOMER1	PTEN	NEFL
KCTD12	FOXH1	ACTR8	MAWD(hsa)	SON	VPS26B	ACSL4	KIAA0660(hsa)	GGNBP2
PREX1	DISP2	RP3-511B24.2-013(hsa)	MYO1D	HCFC1	LY6G5C	KLHL9	AP2S1	HIST1H4B
ATN1	SMARCA2	ARFGEF2	ARHGAP35	GFAT(hsa)	FSCN1	GRHPR	NUB1	ZNF785
UBAC2	ETS2	KCTD5	SLC2A3	HIST1H2AK	SNF2L4(hsa)	MVD	NFATC3	SOCS7
NOTCH2NL	RR2(hsa)	XBP1	FBXO34	TRIM3	GPALPP1	SGT1(hsa)	APBB2	F8A1
CCDC92	KIAA0324(hsa)	PPP2R5C	ASAP3	EPLG2(hsa)	KIAA0226	CYB5B	PSDR1(hsa)	C2ORF15
WIPI1	RABGAP1	ITM2B	TTC28	KOC1(hsa)	SLC25A22	TFIP11	ZC3H4	CT45A1
CTNND1	WDR70	hCG_39606(hsa)	SLC9A3R1	CD46	C1orf56	LGALS1	TTC19	CYFIP1
FAR2	SATB2	NDRG1	C8orf58	NOS3	MPZL2	CHAF1A	CDPF1	ZNF230
LMAN2L	HNRNPDL	HSBP1	PTPN13	CTDSPL	LDLRAD3	C14orf28	FGFR1	ZNF2
RP11-517O1.1-002(hsa)	PXN	SNAI1	IDI1	MYO18B	POP1	DNA2	SUPT7L	CFHR2
KIAA1324	TM9SF3	B3GALNT2	FOXP1	FBXL17	KDM6A	EDC3	SCARB1	NUDT3
NXPE3	GSPT2	SPATS2L	hCG_2031635(hsa)	FBXO3	NMRAL1	TNFRSF1A	DUSP16	C5orf22
HSPC130(hsa)	DBT	POLR1D	ATM	CD68	EPS8L2	KIF18B	MARCH6	HIST1H2BM
RAD23B	RC3H2	ZNF768	SOGA1	RHOGAP1(hsa)	UBE2K	SREK1IP1	PALLD	S1PR2
SH3PXD2A	SLC2A13	ATAD5	RABGGTB	B3GNT4	H2AFO(hsa)	GUF1	ARC	HIST1H4K
HIST1H2AI	ORC1	SEC61A1	ZSWIM1	MISP	CAMK1	AURKA	RP11-90D4.2	NR4A1
ADCK2	EFL3(hsa)	SPTBN2	CAPZB	MARCH8	KIAA0462(hsa)	ESPN	SIX4	SNX12
H4/H(hsa)	KIAA1321(hsa)	FUT1	LAMC3	MKLN1	hCG_1992160(hsa)	ZNF644	SNED1	CT45A2
ZNF513	NUP210	MGLL	SHKBP1	SEC14L2	NUP50	TATDN3	SGK3	PCGF2
hCG_1994842(hsa)	ZZZ3	CDKN2A	DUSP10	SMIM15	THAP2	KIF2(hsa)	TEX264	CISD3

FRS3	TUBB1	RTKN2	TAGLN2	TCTN1	BCORL1	ZNF512B	SPTBN1	C17orf100
USP25	RNF39	MIEF1	DYX1C1	TXNRD1	MMAB	DKFZp666B209(hsa)	HNRNPUL1	SERPINA3
ZGRF1	ZNF16	KIAA0513	MINOS1	LCN2	H4FE(hsa)	PCDHGA4	CHD6	KLRK1
H2AFJ	MAPKAP1	FAM175B	RFXANK	BCKDK	H4FJ(hsa)	LEPREL2	EIF3G	GTF2I
ACSS1	X104(hsa)	PPP1R12B	TIMM10	NIPSNAP3A	PEA15	DNAJB12	CCDC85C	RASL10B
TKT	ACOX3	RPP30	FURIN	PPM1A	IPO5	KDM7A	HSPA1(hsa)	MYO19
BRD2	TRIM35	H6PD	FLAF1(hsa)	PARP16	CACNA2D4	THOC2	TMEM55A	CXorf65
USP36	PERP	TJP1	ACP6	RAB36	AIF1L	GPN2	LATS1	VAMP3
USP38	GMFB	LOXL3	TES	TOM40(hsa)	RARA	NUP43	HOXB13	TSACC
PXDC1	PMEPA1	KDM2A	HECTD2	ABHD16A	SNAPC1	PLEKHA1	PPP1R1C	TMEM189-UBE2V1
RNASEL	SLC35E4	PGM3	CECR2	HSPA14	CASP9	D4S234E	SLC35F6	SYNRG
ELAVL2	MYOF	CDK5R1	ANKS1A	COPS3	CTC-281B15.1	FAM102A	CENPU	FCGBP
FEM1A	TCF7L1	C1QBP	CEBPB	PRELID1	SH3RF1	LEF1	STYXL1	HIST1H4D
M6PR	PAST1(hsa)	PLGLB2	PERQ2(hsa)	PIGQ	CPLX2	OCRL	AP1B1	HIST2H2AA3
HNRNPA1	SMC6	UBE2J1	TICRR	hCG_23463(hsa)	ETHE1	PVT1	DMRT1	ACACA
DICER1	ARPP19	UBC	STK38L	RAD23A	RYR3	MRPL28	KIAA1841	NATD1
TMEM205	WDR27	RRP36	IDH3A	CRBN	STX18	LIN7A	IRAK1	TAF15
FRMD4A	GDF11	NAAA	GNA11	PLEKHO2	SESN2	KMT2A	C17orf97	TFDP2
ACOX1	NABP1	H2AFE(hsa)	ISOC1	DOK3	NARF	ROGDI	UBALD2	TPTE
TMEM141	SERGEF	ZEB1	LLGL2	DEPDC1	RTP(hsa)	FAM222A	IRAK4	NPHP3
ARSDR1(hsa)	TESK2	S100L(hsa)	STRIP1	SFXN5	VPS52	RARG	NFIC	MLLT6
CAP43(hsa)	NKTR	H2AFN(hsa)	LYPD3	SPFH1(hsa)	CDC7	LRRFIP1	DAAM1	CTDSP2
DRG1	NPTXR	TMC8	GTF3C2	SFXN3	RNMTL1	KIAA0368	SPSB3	CWC25
DNM2	FAM127A	ZNF280C	SOSTDC1	RPGRIP1L	MNEI(hsa)	CHCHD6	PLK4	IKBKE
SDC1	CMAS	TMEM164	PNT5(hsa)	CNFN	RAP1GDS1	CDK20	CAP1	ORAI1
E2F2	ETS1	MORN4	ETNK1	KHK	FUK	AKR1A1	H2AFL(hsa)	WBP1
MLF2	RILP	NARS	PEG10	USP46	PPDPF	HSPC272(hsa)	BMP7	ZNF280B
CASP7	BRIP1	PPIP5K1	MOB3B	SCAF6(hsa)	C6orf89	ALDH4(hsa)	RP11-316M21.1-003(hsa)	HIST1H4A
ATXN7L2	SH3BP2	HEBP1	FAH	TP53INP1	FGF18	C8orf22	ABCA3	BRD4
TPD52L1	PTPN14	OK/SW-cl.73(hsa)	SRRD	LRP8	HIST1H3H	GOLGB1	EMC1	
RCN1	EIF2AK4	HMP(hsa)	GHDC	ELMSAN1	FAM76A	PPP1R21	LRRC49	
RBM12	HS2ST1	TMEM179B	CUEDC1	HIST1H3D	RP5-824I19__A.1-011(hsa)	PDE5A	BBX	
SCAI	POLM	BRG1(hsa)	LCORL	MRPS15	RP11-874J12.2	RHBDD2	INF2	
ZBTB10	SYPL1	MTMR2	AKT1S1	ARL17A	CCPG1	UBAP2L	BBS4	
MAST4	PSME3	GAS2L3	ZNF397	EQTN	AUNIP	MED19	DCAF16	
FAM32A	LRRC8D	OSBPL7	GFPT(hsa)	SLC4A2	KCNC4	ZBTB48	GRHL1	
CAMK4	LINS	QSER1	WSB2	C12orf29	TMED4	NAGS	TAF5	

miR-345-5p

TERF2	CPSF7	ZNF664	LSM5	SPEN	FIGNL1	RSBN1L	DGKD	DVL2
C17orf85	PTRF	NT5E	DHODH	GFPT1	SRRM2	STARD9	ZNF106	PSMG4
AK2	CDC25B	BCL2	LIMS1	MYLIP	METTL21A	DMPK	NCKAP1	ZBTB25
NEK7	DHCR24	RNF111	TOP2B	GCN1L1	SEC31A	G3BP2	PSMA4	RNF139
ZEB2	LENG8	BOD1L1	PANK3	CLTA	AQR	FASTKD5	MAN1A1	EIF1
TMED8	PRRC2B	ALDH1L2	EEF1D	SLC19A2	SHOC2	RWDD2A	ACIN1	MGAT4B
DAZAP2	SRP19	DAP	GPR35	RYR2	KMT2D	SCD	XRCC5	AMN1
GUCD1	DLG1	AIM1	MIS18BP1	PPP1R10	IPO4	DNAJB12	RND3	RBM15
LARS2	SKIV2L2	IGF1R	EDEM3	ARFRP1	SESN2	SCAF1	TCF7	AUP1
S100A16	IDS	FMR1	AKT2	F2R	TNFRSF10B	NDC1	PHIP	RTN3
SENP3	NELL2	PPM1E	PCGF5	PPP6R3	CYSLTR1	CTCF	FBXO33	TMBIM6
ADCY1	HTT	GGH	SLC7A1	SON	C16orf72	PKN2	ZNF652	OR51E2
TTC39C	CCND2	PTGR2	RABGGTB	STRN4	DYNLL2	CDKN1A	TTC19	EIF4G2
DNAJC6	TAOK2	GNB1	RPLP2	TIAL1	NF1	UHMK1	FGFR1	AP3M2
ATN1	MOCOS	BMP6	ZBTB37	ZC3H12C	FAIM	ANKHD1-EIF4EBP3	MARCH6	MIER3
GARS	DBT	RHOA	MLEC	ECT2	MALSU1	CNEP1R1	SPTBN1	VAMP2
FAM32A	BACH1	EIF1AX	LARP1	PSMD14	KIAA0100	KIAA0930	MUM1	UBR4
MTHFD2	KIAA0319L	ATP6V1G1	ZC4H2	HIC1	MED28	DONSON	MAPK1	INO80D
STK4	E4F1	GBF1	HDGF	ZC3H13	EIF2S1	VPS28	MTMR3	TXNIP
MMS22L	GSE1	GSTO1	KLHDC10	RAB18	JADE2	TNPO2	FBXW11	GREM1
KIF1B	NFIL3	CHD4	AGO1	GSTP1	GSK3A	BTBD7	RPL27A	SYS1
UBQLN4	GPSM1	SMARCD1	BTG2	PPM1L	SLC39A6	CKS2	TET2	MARCKS
ZNFX1	PERM1	SLCO3A1	FBN2	EIF4ENIF1	LMAN1	MPPED2	SYNGR1	PCGF2
BRD8	NHLRC3	CHMP4B	TOP2A	NRP1	PRDM2	CNBP	SART3	MAGIX
UBR5	FBXO4	TMEM145	LPAR1	IPMK	ZNF35	SASH1	RAD9B	MYO18A

miR-200c-3p

TERF2	COPS8	ORMDL3	QSER1	EDEM3	MCTP2	KBTBD11	FBXO22	LATS1
RTN4	MSMO1	LMO7	HNRNPU	FAM46A	GDAP2	EIF3J	FN1	THNSL1
RPRD2	ZNF605	FZD6	HSDL1	PPP3CA	ANP32B	C6orf62	MYCN	ALDH1A1
SLC9A2	CCNL2	SOAT1	PPP1R18	DPY19L4	CEP41	MAPRE1	UBN2	CAB39
HSPA13	MCM8	RDX	SPRED2	UBA6	YEATS4	PTP4A1	C22orf39	ATP13A3
ANP32E	GNS	TOB1	VLDLR	CHURC1	DNAJB9	SHOC2	SPTSSA	PHOSPHO2
ADAM9	HNRNPD	INPP5F	UBE2I	PTK2	PPP2R2A	DNMT3A	RAB22A	NUPL1
CCNG1	VAT1L	CTBP2	VEPH1	C9orf41	RNF24	SESN1	LRIF1	RNF168
DUSP4	DENND4B	TMEM167B	KRAS	TTC8	KIDINS220	SIAH1	BAG6	AMOTL1
RAB4A	FBXO30	GAN	EPDR1	NLK	LNPEP	ANKRD46	RBM12B	LATS2
C22orf46	TCF4	WNK1	HEG1	CEBPB	GPR98	PPP1R9B	ZNF92	FBXW11
NEK7	TMED7	GSE1	IPO7	MGAT3	QKI	SDE2	UHMK1	PPP2R1B
ZEB2	APC	HOXB5	SEPT7	SEPSECS	ASXL3	SCO1	LHFP	GREB1
MLK4	CRK	PBX1	SETD2	ISOC1	SON	KIF11	BIRC3	ABL2
ZMAT3	FBXW7	B4GALT6	SLC35A4	ARHGAP18	LPL	TNFRSF10B	ATG12	KDR
JARID2	DHX29	NBR1	SERINC1	TUBB4A	DOCK5	SURF4	BDP1	USP33
CASP8AP2	UHRF1BP1L	TRPS1	FSTL1	GART	MARCH8	FAM220A	SYNE2	RANBP9
CNOT6	SHC1	MGA	FBXL16	CCDC127	MKLN1	CHD2	NAA16	RAD21
GSKIP	SIRT1	GAL3ST1	SLC16A9	BRCA1	SMIM15	DZIP1	ARF6	KMT2C
FBXL3	NLN	CLCC1	PHF10	ZNF217	PLK2	C16orf72	USP9X	CDK16
NTRK2	RUNX1	NET1	TCP1	ITGAV	MYO9A	FAM199X	MBNL1	ADCY9
PKIA	ABCC1	TMEM64	CPD	STK17B	WSB1	ST7	PRC1	HELLS
UBFD1	CLCN3	KLHL20	MSN	ZNF697	TP53INP1	TSC22D2	NFIA	BTG1
COA4	PAK2	SLC22A25	ZCCHC2	TRIM59	ELMSAN1	PPM1F	TBP	PCDH9
SLMAP	ATMIN	SMARCAD1	NUDCD1	TMX4	SERPINI1	BTAF1	GPM6A	CAPRIN2
TSC1	AKIRIN2	PTDSS1	NAP1L5	PSMB4	SNX1	CLASP1	CNEP1R1	BAP1
PIM2	ANKRD52	BCL2	LBR	CCDC43	TMBIM4	ZC3H11A	EDNRA	MRPL15
RBM27	COX11	CDKN1B	TAOK1	PMAIP1	GLI3	ADNP	ALG13	PARD6B
PEAK1	SPATA2	SRGAP1	RBAK	CDH1	ARPC2	HSPA8	EIF4EBP2	SERBP1
ARL6IP5	MTUS1	DDB2	KIAA0101	RAB11FIP2	RELN	FIGN	PPM1B	FNBP4
UGCG	PPP2CA	STYX	CANX	PITPNM3	BLOC1S5-TXNDC5	CCNY	MORC3	RAP1B
ERBB2IP	SETD7	MED13	CDR2	KIAA1432	NMD3	KDSR	BCL11B	KLF6
TNPO1	LRRC8B	TUBB	SMG8	DDIT4	PCBP1	CDK13	NOL8	FSD1L
UBE2Z	ZNF431	SLC31A1	KBTBD6	SFXN1	ANKRD42	RANBP2	HMGCS1	RANBP6
GLG1	CPEB4	XPOT	TIPARP	NFAT5	KANK1	SRP72	PTPN12	KIAA1279
POMZP3	GAPDH	DDX6	SIKE1	MFSD9	STX17	FAM13B	EXOSC6	PDK2
ZNF638	ADAR	RALBP1	CWC27	CHMP2B	VIM	MDM4	BCOR	RNF139
PDCD4	ZNF131	SH3GL1	ALDH6A1	PIKFYVE	TMED10	OSTM1	CREG1	CBX5
UBE2R2	SUGP2	HIPK3	FAM60A	DPPA4	REL	B3GALTL	LIMA1	KLF4
TM4SF1	USP31	DLG5	NCOA3	TET3	FRMD6	KRT80	CUL3	TBL1XR1
OTUD3	TCF7L2	PLD1	DDX60	ICA1	ANKRD44	TAF7	BTBD7	WDR37
SPPL3	PLAGL2	MYZAP	LPP	CHN2	EIF2S3L	USP42	TMEM263	C14orf132
PPP2R5E	ZYG11B	CAND1	TMEM123	AP1S2	LPIN2	NMT1	ATP2A2	DUSP1
FEM1B	USP1	PXYLP1	GIGYF1	YIF1B	TTC3	SLC20A1	HIPK1	SPOPL
SMC1A	UBQLN1	AGO2	SKP1	NAV2	E-cadherin(hsa)	LAMC1	AHNAK	SECISBP2L
TROVE2	DLG1	KIAA2018	ANLN	PAN2	DBR1	RPS6KA3	YES1	MTSS1L
ZBTB11	HOXD13	EPG5	NSD1	LARP1B	ZMYM2	ERC1	NRIP1	PRRC2A
CNOT11	SMURF2	MCFD2	RASSF2	RSU1	FLT1	LMAN1	ZBTB34	RAB5B
ID2	RNF103	LMNB1	PLEKHA8	PLCG1	KANK2	VMA21	KCTD15	AKAP11
SLC30A5	RSF1	ZNF302	CMTR2	PHC3	ZBTB5	PROSER1	RGL1	MAP7D1
PACRGL	VAC14	TNKS2	RPS6KA5	TOMM5	SLC38A2	DESI1	LEPR	BAG5
RNF2	CSNK1G3	GOLT1B	MFHAS1	NCOR2	EIF4B	PRDM2	GLRX5	GOLIM4
MED13L	PURA	ATP1B1	RHOT1	SLC16A14	RNF19A	GAB1	PLEKHH1	NIN
WNK3	OGT	ELMO2	TUBA1A	STX16	ARMC9	ZCCHC14	UBE2W	ZNF669
DDX3Y	ZNF597	ARID4A	MITF	PDCD10	ABHD5	DCUN1D1	VMP1	SYVN1
NDFIP2	SOWAHC	CCDC117	NUAK1	JUN	ZC3H13	S100PBP	RAN	KIAA1429
TPD52	ZNF43	EFR3B	DDX21	ZNF711	DENND6A	TWISTNB	MKNK2	VKORC1L1
MXD3	HTT	FBLN5	EXOSC3	CRTAP	NOL4L	ATF7IP	KRR1	ARHGAP12
SSR1	ARL2BP	WEE1	RPP14	ABHD17B	SMAD5	YIPF5	GDI2	BCL2L11
GATM	NEDD4L	NUP160	HDAC4	ATP6AP2	GALK2	ZNF281	TSN	GTF2E1
ZBTB38	YWHAG	NFYA	ASH1L	FANCM	TRIM33	WHSC1	ZFP36L2	NAP1L2
KSR1	SEC23A	SHCBP1	TOP2B	TXLNA	ATP11A	PHF12	C9orf78	HCCS
FUS	MCL1	IGIP	PANK3	CCND1	MMD	EIF4E2	OSBPL8	GPATCH4
SAR1B	PTPN4	CD164	SRSF1	WDR82	CPT2	CFH	NEDD1	XIAP
NUCKS1	TGS1	TIFA	MYNN	FNDC3B	ZFPM1	ILDR1	CWC15	ERI1
MGST1	ZDHHC20	DNAJC5	PURB	EFR3A	SFPQ	RIN2	CMTM6	MIER3
PREX1	CDK2	UGP2	LRPPRC	ZNF678	ITPR1	BRD1	CTDSPL2	RAB34
SMAD2	RAB7A	KPNA2	TAF12	LPAR1	CDYL	DLD	UHRF1BP1	PPP1R3D
SH3PXD2A	ALDH1A3	EVI5	SENP6	B3GNT1	CFL2	CERS6	SLC35B4	KIAA1430
BNIP3	MKL2	TMOD3	MTRNR2L4	TUBA1C	CDCA7	DDX39B	PTAR1	PPP1CB
USP25	EEA1	TARDBP	RECK	PVRL4	BAZ2A	UTP23	MOB1A	C2orf69
CBL	TM9SF3	SERPINH1	GABPB1	CTNNA1	STRN3	KIAA0040	CDC42SE1	SYNCRIP
DICER1	TOMM20	KPNA4	ARIH1	CCNE2	CDKN2AIP	SUZ12	SPIN1	GRIA3
TPD52L1	SNX30	NOTCH2	IARS2	GFPT1	TMEM185B	HNRNPH1	PRKRA	MEX3C
RBM12	LMBR1	SQSTM1	IFIH1	SMARCC2	ZFAND6	SNX13	SUCO	TRAPPC8
PAM	GMFB	TAF4	MED1	MRPS30	NPNT	LGALS8	ZFPM2	YTHDC1
MAST4	TCF7L1	KLF11	ITSN2	SBNO1	ERICH1	GALNT2	ZNF618	INO80D
AMFR	ARPP19	CNIH4	SLC25A38	FEZ2	SLC25A36	CEP57	XPO1	ZFAT
REPS1	BDKRB2	TIMP2	AZIN1	ARL6IP1	ARL5A	C5orf24	TRIM71	FHOD1
SCAF4	PTTG1IP	ITM2B	TXNDC16	USP5	BHLHE22	SCD	DIXDC1	GSPT1
IP6K1	FAM127A	NDRG1	H2AFV	NPEPPS	ZNF280D	UBE2G1	GPBP1	EFNA1
CRKL	ETS1	ASPH	NFYB	SKP2	RAP2C	NGRN	CPEB3	SOCS7
ARHGAP19	NRBP1	RHOA	TEX2	WWC3	KLHL15	SREK1IP1	ZNF652	DACH1
CD59	HUWE1	SPATS2L	SBF1	BLOC1S6	NR5A2	ZNF512B	ZC3H4	MARCKS
STK4	FRS2	ENSA	YME1L1	LRRC58	PPIL4	KDM7A	FGFR1	SYNRG
TET1	CBX4	MIEF1	RYK	FAM8A1	RCHY1	ENPP4	MARCH6	ACACA
TPCN1	NRBF2	CHD7	DAP3	PPP1R10	SLC26A2	KMT2A	FDPS	
RASA2	BCL9L	CDK5R1	ANKRD40	HELZ	TBK1	VEZF1	DCTN6	

ZNF673	ALG8	ZEB1	KLF13	PM20D2	JAZF1	LIMK1	HPS5	
TRIM52	PTPRD	PSAP	ASAP1	E2F3	EPHA2	GEMIN2	PI4K2B	
ELK4	CAMSAP2	COPS2	GREB1L	MXD4	ASCC3	DHX9	ERRFI1	
CLIC4	ZNF121	HIC2	ACVR2B	PAIP2	CUX1	DERL1	ATL3	

miR-10a-3p

CCNT2	SMAD2	KATNBL1	RAB1A	CREB1	CYB5D1	LYST	ICK	USP3
B2M	DICER1	GAB3	QSER1	MED4	ATL2	DDX5	PLAG1	HOXA11
KIAA0754	RBM12	BRIP1	TCF19	GREB1L	PPAP2B	DNM1L	PUM2	LSM14B
C16orf52	TMEM33	SYPL1	ZBTB6	EIF5	DNAJC13	C6orf62	AHR	NME1
PFN1	POLI	BACH1	CDK6	TES	LNPEP	PRKAA1	PIAS4	VEGFA
AK2	PDK4	NOD2	KIF23	STK38L	STAU1	ANKFY1	RSBN1L	SPTBN1
RBM14	DIEXF	GLMN	DDX3X	MOB3B	POLD3	ATF1	SPATA6	FOXO1
HOXB7	HP1BP3	PCTP	COPB1	PISD	SRPK1	IPO5	FREM3	MRPS36
ADAM9	THUMPD3	CAMSAP2	INIP	WSB2	QKI	SESN1	MZT1	RNF168
SLC38A1	CLIC4	FICD	NID1	GART	FBXO3	MCUR1	AKAP5	MAPK6
MED29	HOXB4	FZD6	C12orf75	TFDP2	MKLN1	CDC7	ZNF564	FBXW11
EGR1	MSMO1	UBE2E2	PHF10	BRCA1	PPM1A	EBAG9	G3BP2	ABL2
FOS	KIF1B	TRIP11	ADD3	LARP1	SRSF6	TRIO	BRPF1	GAS1
MTA3	CCNL2	CHUK	CCDC71L	PLCB1	C21orf91	RDH10	SH3BP5L	DNAJC27
DARS2	ZNF83	TMEM64	HPS3	DOCK10	PPP1R14C	INTS6	ATP6V1C1	MGAT4A
BRWD3	HNRNPA3	ZNF664	SLCO3A1	CNIH1	ITPRIPL2	TSC22D2	SCD	PARD6B
PKHD1L1	C5orf42	TNFRSF11A	APP	DDX18	PLEKHA3	TXLNG	SDC2	MRPL49
ARL4A	VANGL1	FOXK1	MAT2B	PMAIP1	PCBP1	SSX2IP	SREK1IP1	TOP1
PAGR1	PTCH1	HBP1	KIAA2026	ANP32A	SLC16A1	CCDC50	ZNF644	BRWD1
ZBED5	DBN1	SLC31A1	KIAA0101	KIAA1432	CLPX	HMG20A	OCRL	KLF4
PEAK1	FCHSD2	MACF1	PLAGL1	NFAT5	SKIL	RANBP2	MALT1	SLAIN1
ENC1	HLTF	SQLE	PDS5A	PIKFYVE	ATRX	HINT3	XPO4	PIAS1
TNPO1	ANKRD52	COL19A1	TMEM38B	PPP1R12A	VCP	SSR3	ZNRF2	ZNF91
ZNF638	SAMD8	GLUD1	KIAA1109	AGO1	TTC3	OSTM1	SLC7A2	IREB2
TGFBR1	CYTH2	AFF4	RFX7	SND1	IRF3	FN3KRP	TMEM45A	PTGES3
MACC1	SETD7	IBTK	FJX1	MAPK9	PSMD14	FNIP1	UHMK1	TSPAN13
KIAA1731	CEBPZ	CSDE1	HK2	PCNX	GDE1	BTF3L4	BIRC3	ARHGAP5
IRF4	HABP4	STAG1	AGPAT5	FMNL2	RNF19A	C19orf10	HIAT1	AKAP11
PDHX	ING3	RNF217	LPP	BTG2	EIF4G1	AKIRIN1	TNRC6A	MDM2
FEM1B	USP31	SETD5	GIGYF1	PHC3	RAB18	ZNF451	RNF170	SLC39A14
ARID3A	FKTN	SFT2D2	RCN2	B4GALT1	DENND6A	ARHGAP29	NFIA	SLC30A6
MET	UBQLN1	ZNF280A	CMTR2	PNN	CSRNP3	GFM1	PAN3	NIPA1
ATP6V0E1	ATG2B	AGFG1	UNG	PPP1R15B	FOSL2	NDUFB2	SESN3	HMSD
DENND4A	PID1	RAB12	ZNF680	ABHD17B	ELOVL5	LMAN1	ACTR3B	BMPR2
ZNF616	TAF2	PABPC4	INSIG1	FBXO28	STAM2	PRKAR1A	ANKEF1	BCL2L11
PHF20L1	TEFM	DSE	ZNF354B	CCND1	HIST1H4H	VMA21	ZNF483	KIF5B
ANO6	ARHGEF12	PRAF2	LRIG2	SMAD4	ATIC	SUMO3	ZBTB1	LMO4
CUTC	MCL1	UGP2	PDE4D	CTNNB1	PCMTD2	PROSER1	PHF13	PLS3
PSMD5	RAP1A	TPX2	DDX21	ANXA7	TEAD1	EP300	RASSF8	DGKH
CHRM5	CPS1	LY75	ACBD3	WDR44	AEN	MIDN	AHNAK	PPP1CB
SSR1	ZDHHC5	PRR5L	SREBF2	GFPT1	USO1	PRDM2	YES1	NDEL1
CAMKMT	SMARCA2	TARDBP	ZXDB	ASPM	DOCK4	DDAH1	PRKDC	F11R
RIPK2	ETS2	TXNDC11	RPS8	PBX3	IPMK	SLC10A7	JADE1	
PGAP1	HOXA9	RDH14	CTTN	AXL	SLC9A9	BTG3	TRRAP	
LAMB1	RC3H2	RHOA	RNF149	NUFIP2	DLGAP5	TMEM245	PDZD8	
CXCL8	CFDP1	POLR1D	TRIP12	ABCC9	ADAM22	THTPA	U2AF1	
CCT2	KAT6B	FEM1C	ZNF317	TRAK1	TFPI2	ARL5B	SNX5	
KDM5C	PDE10A	H6PD	DCAF7	PAICS	RCHY1	MPRIP	CLSTN1	

Table A4. Complete output from DIANA-miRPath v.3

Table A4. Complete output from DIANA-miRPath v.3 when using the miRNA signature (miR-34a-5p, miR-345-5p, miR-200c-3p, miR-10a-3p) as input, from both approaches: MicroT-CDS (*in silico* miRNA target prediction algorithm) and TarBase (experimentally supported approach). The columns show the influenced KEGG pathways, *p*-values after Benjamini-Hochberg correction, number of targeted genes and number of associated miRNAs from the signature. Significant *p*-values are highlighted in bold.

KEGG pathway	*p*-value	#genes	#miRNAs
MicroT-CDS			
Mucin type O-Glycan biosynthesis	**1.184275e-12**	5	2
Glycosphingolipid biosynthesis - lacto and neolacto series	**3.884874e-10**	4	2
Biotin metabolism	**0.0006064075**	1	1
Proteoglycans in cancer	**0.0007940732**	22	1
ErbB signaling pathway	**0.0008384294**	16	1
MicroRNAs in cancer	**0.00138567**	20	1
Thyroid hormone signaling pathway	**0.002132528**	16	2
Phosphatidylinositol signaling system	**0.00647907**	12	2
Neurotrophin signaling pathway	**0.009801489**	22	1
Renal cell carcinoma	**0.01097907**	13	1
Glycosaminoglycan biosynthesis - heparan sulfate / heparin	**0.01113265**	4	1
Lysine degradation	**0.01605634**	5	2
Hippo signaling pathway	**0.04556428**	9	1
Axon guidance	**0.04746173**	13	1
TGF-beta signaling pathway	0.05008158	8	1
Glioma	0.06553687	7	1
Circadian rhythm	0.06948476	5	1
Sphingolipid signaling pathway	0.07745415	12	1
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	0.1093001	2	1
Choline metabolism in cancer	0.1191737	12	1
MAPK signaling pathway	0.1518697	28	1
Sphingolipid metabolism	0.1821781	6	1
ECM-receptor interaction	0.2101148	3	1
FoxO signaling pathway	0.236839	17	1
Oocyte meiosis	0.248682	12	1
Glycosphingolipid biosynthesis - globo series	0.2859095	3	1
Small cell lung cancer	0.3219701	12	1
Carbohydrate digestion and absorption	0.3981785	1	1
SNARE interactions in vesicular transport	0.4719691	5	1
Base excision repair	0.5364044	1	1
Inflammatory mediator regulation of TRP channels	0.5608684	9	1
TarBase			
Pathways in cancer	**7.806205e-09**	141	3
MicroRNAs in cancer	**3.356072e-08**	36	1
Proteoglycans in cancer	**4.259346e-08**	82	4
Adherens junction	**4.849975e-08**	41	3
Colorectal cancer	**6.624356e-06**	37	3
Fatty acid biosynthesis	**1.571663e-05**	4	1
p53 signaling pathway	**8.629925e-05**	40	2
Endocytosis	**0.000155911**	80	2
Cell cycle	**0.0001862279**	52	1
Chronic myeloid leukemia	**0.0003003506**	37	2
Hippo signaling pathway	**0.0005622035**	26	2
Metabolism of xenobiotics by cytochrome P450	**0.0007820423**	2	1
Lysine degradation	**0.0007881832**	7	1
TGF-beta signaling pathway	**0.0009288388**	13	2
Prostate cancer	**0.001594596**	39	2
Hepatitis B	**0.002008518**	54	2
Oocyte meiosis	**0.002446339**	16	1
Bladder cancer	**0.002960341**	22	1
Thyroid cancer	**0.003013095**	16	1
Pancreatic cancer	**0.003759289**	33	1
Transcriptional misregulation in cancer	**0.004338518**	62	1

Viral carcinogenesis	**0.004834794**	74	2
Bacterial invasion of epithelial cells	**0.00518036**	36	2
Ubiquitin mediated proteolysis	**0.006448268**	22	1
Melanoma	**0.007508105**	30	1
Non-small cell lung cancer	**0.007831112**	26	1
Fatty acid metabolism	**0.008055732**	14	1
Steroid biosynthesis	**0.008281795**	1	1
Other types of O-glycan biosynthesis	**0.01126346**	13	1
Protein processing in endoplasmic reticulum	**0.01302932**	17	1
Regulation of actin cytoskeleton	**0.014233**	71	1
Glioma	**0.01423311**	26	1
FoxO signaling pathway	**0.01449044**	12	1
Thyroid hormone signaling pathway	**0.01501771**	44	1
Endometrial cancer	**0.01737389**	22	1
Small cell lung cancer	**0.03340758**	13	1
HTLV-I infection	**0.04337413**	81	1
Glycosphingolipid biosynthesis - lacto and neolacto series	**0.04423077**	9	1
Shigellosis	0.05387291	29	2
DNA replication	0.08411226	17	1
Renal cell carcinoma	0.09253995	26	1
Ras signaling pathway	0.1004746	24	1
Central carbon metabolism in cancer	0.1017866	27	1
PI3K-Akt signaling pathway	0.1389146	29	1
MAPK signaling pathway	0.140083	76	1
Huntington's disease	0.1587988	2	1
Wnt signaling pathway	0.1749672	14	1
Alcoholism	0.2882482	60	1
RNA transport	0.2884383	16	1
Neurotrophin signaling pathway	0.3112592	15	1
Progesterone-mediated oocyte maturation	0.4438475	31	1
Epithelial cell signaling in Helicobacter pylori infection	0.4488572	7	1
HIF-1 signaling pathway	0.5095467	35	1
Adrenergic signaling in cardiomyocytes	0.52571	10	1

Table A5. Comparison of studies investigating miRNAs from brain samples

Table A5. Comparison of studies investigating miRNAs from brain samples of FTD and/or ALS patients.

	Kocerha et al. 2011[1]	Chen-Plotkin et al. 2012[2]	Hébert et al., 2013[3]	Gascon et al., 2014[4]	Helferich et al. 2018[5]	Jawaid et al. 2019[6]
Disease	FTD	FTD	FTD	FTD	ALS	FTD, ALS
Cohort	Mixed sporadic/genetic	Mixed sporadic/genetic	Not mentioned	Sporadic	Mixed sporadic/genetic	Sporadic
Patients, n= Discovery/replication	32 PGRN+ 8 PGRN-	5 PGRN+ 7 PGRN-	5/14	5	5 C9orf72+ 1 SOD1+ 16 sporadic	10 ALS 9/12 FTD
Methods of analysis	Microarray	Microarray	Deep sequencing, qRT-PCR	qRT-PCR	qRT-PCR	qRT-PCR
Major deregulated miRNAs	miR-922 miR-516a-3p miR-571 miR-548b-5p miR-548c-5p	miR-132 miR-212	miR-132-3p	miR-124	miR-1825	miR-183/96/182

Figure A1. Stratified nested cross-validation

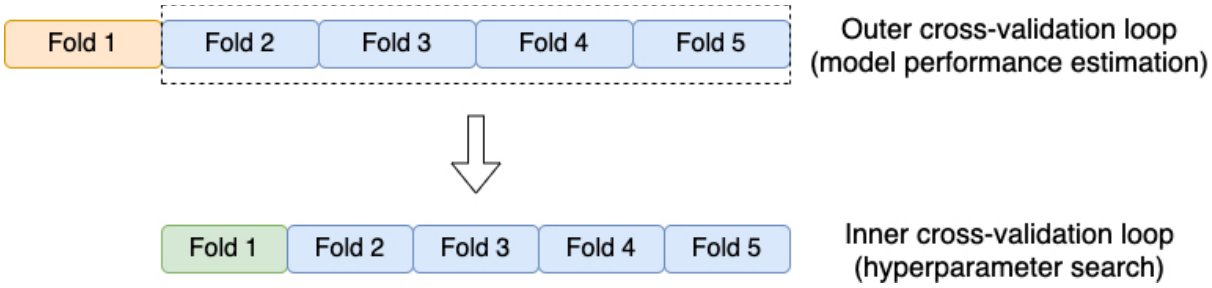


Figure A1. Stratified nested cross-validation scheme, with 5-fold outer and inner cross-validation. The outer cross-validation loop splits the whole dataset into five folds, using four folds as training data and one as test data at each of its iterations. Model performance is estimated averaging the ROC AUC using all five folds as test data. The inner 5-fold cross-validation loop splits the training data into four training folds and one validation fold at each of its iterations. For every outer loop iteration, one hyperparameter is chosen based on the highest average ROC AUC over all five validation folds.

Figure A2. Expression heatmap of miRNA signature

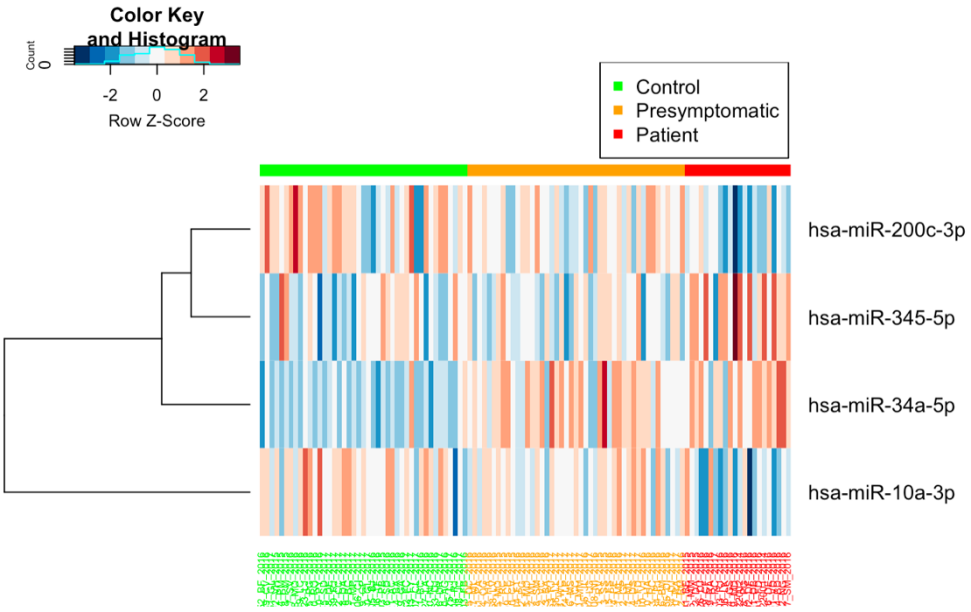


Figure A2. Expression heatmap and hierarchical clustering of the four miRNAs identified as differentially expressed. Rows represent miRNAs and columns represent individuals ordered by clinical status (control, presymptomatic and patient from left to right). The $\log_2$ expression levels of each miRNA are rescaled to have a mean of 0 and a standard deviation of 1, and z-scores are indicated by color: shades of blue indicate low-expression values, white indicates mean expression and shades of red indicate high-expression values. Dendrogram added to the left size clusters the rows according to the miRNA expression means.

Figure A3. Bootstrapped ROC AUC scores

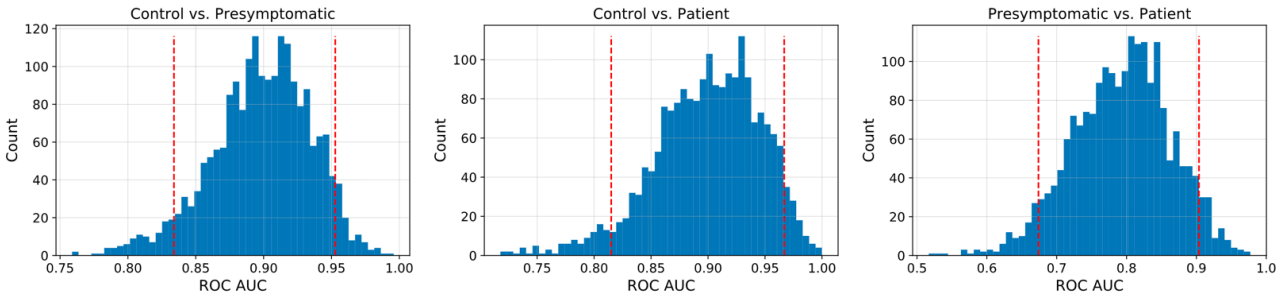


Figure A3. Logistic regression bootstrapped ROC AUC scores obtained with 2000 bootstrap samples for each pairwise comparison, using as features the expression levels of the miRNA signature (miR-34a-5p, miR-345-5p, miR-200c-3p and miR-10a-3p). The red dashed lines indicate the 5th and 95th percentiles (empirical 90% confidence intervals): [0.83, 0.95] for control vs. presymptomatic, [0.82, 0.97] for control vs. patient and [0.67, 0.90] for presymptomatic vs. patient.

Figure A4. ROC AUC scores with 100 different fold splits

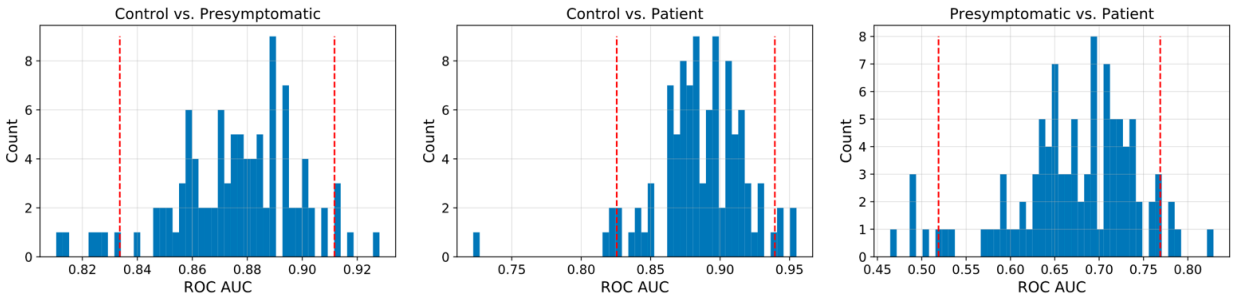


Figure A4. Logistic regression ROC AUC scores obtained with 5-fold cross-validation with 100 different fold splits, using as features the expression levels of differentially expressed miRNAs computed with only the outer cross-validation loop training data (four out of five folds) at each iteration. The red dashed lines indicate the 5th and 95th percentiles (empirical 90% confidence intervals): [0.83, 0.91] for control vs. presymptomatic, [0.83, 0.94] for control vs. patient and [0.52, 0.77] for presymptomatic vs. patient.

Figure A5. Presymptomatic subjects probability scores

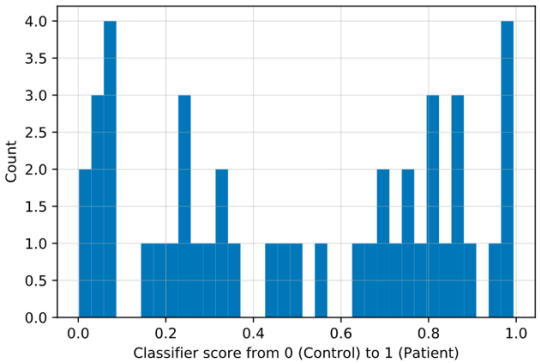


Figure A5. Logistic regression probability scores for all presymptomatic subjects (N=45), when model was trained with the expression levels of differentially expressed miRNAs in controls (N=43) and patients (N=22). Scores near 0 indicate that the subject has a miRNA profile similar to controls; scores close to 1 mean that the individual has a miRNA profile similar to patients.

Figure A6. Heatmap of the level of enrichment in KEGG pathways

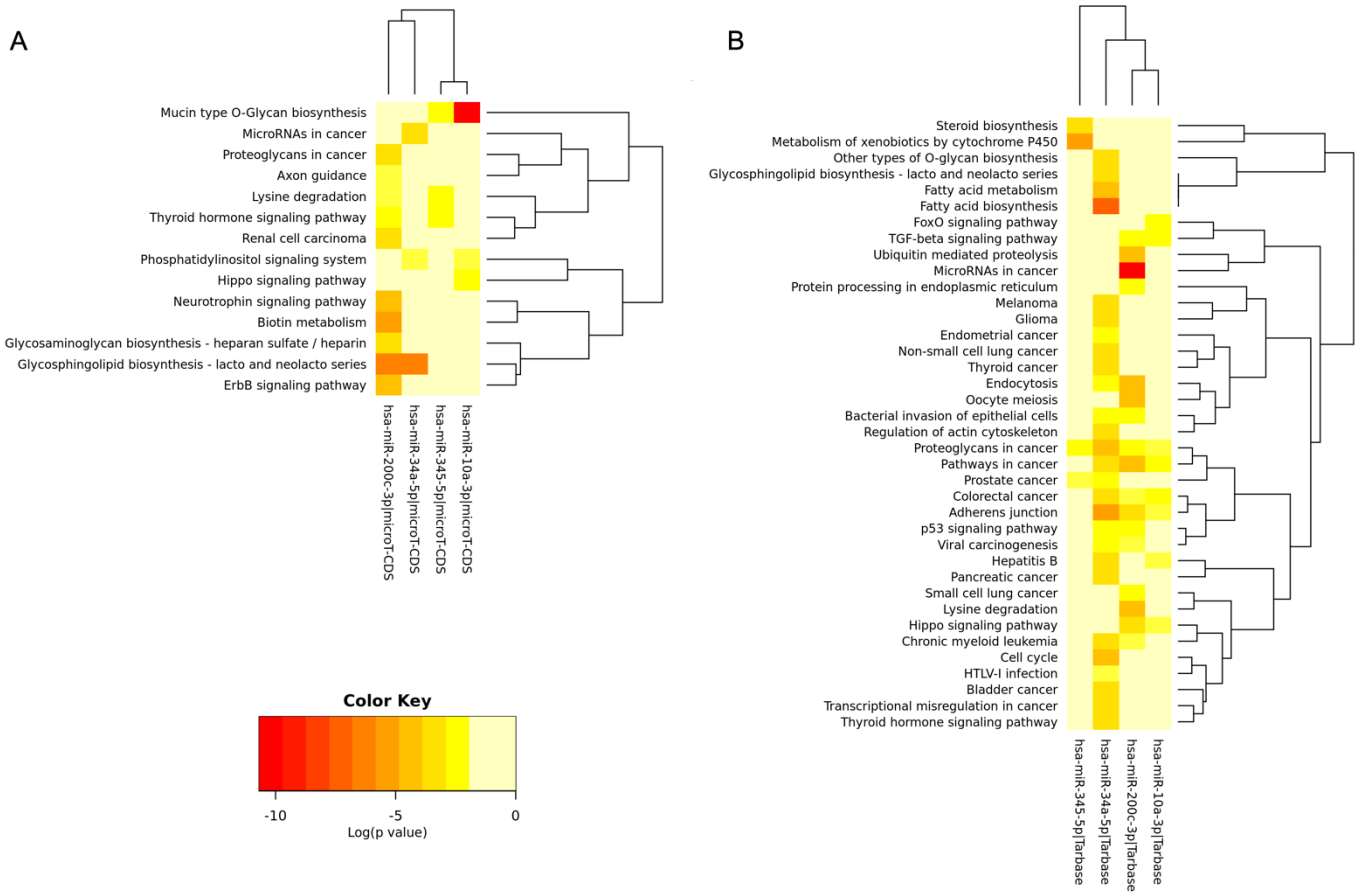


Figure A6. Heatmaps depicting the level of enrichment in KEGG pathways for the four differentially expressed miRNAs, as computed by the *in silico* target prediction algorithm (A) and the experimentally supported approach (B). Pathways with similar enrichment patterns are clustered together.

SUPPLEMENTARY REFERENCES

1 Kocerha J, Kouri N, Baker M, *et al.* Altered microRNA expression in frontotemporal lobar degeneration with TDP-43 pathology caused by progranulin mutations. *BMC Genomics* 2011;**12**:527. doi:10.1186/1471-2164-12-527

2 Chen-Plotkin AS, Unger TL, Gallagher MD, *et al.* TMEM106B, the Risk Gene for Frontotemporal Dementia, Is Regulated by the microRNA-132/212 Cluster and Affects Progranulin Pathways. *J Neurosci* 2012;**32**:11213–27. doi:10.1523/JNEUROSCI.0521-12.2012

3 Hébert SS, Wang W-X, Zhu Q, *et al.* A Study of Small RNAs from Cerebral Neocortex of Pathology-Verified Alzheimer's Disease, Dementia with Lewy Bodies, Hippocampal Sclerosis, Frontotemporal Lobar Dementia, and Non-Demented Human Controls. *J Alzheimers Dis JAD* 2013;**35**:335–48. doi:10.3233/JAD-122350

4 Gascon E, Lynch K, Ruan H, *et al.* Alterations in microRNA-124 and AMPA receptors contribute to social behavioral deficits in frontotemporal dementia. *Nat Med* 2014;**20**:1444–51. doi:10.1038/nm.3717

5 Helferich AM, Brockmann SJ, Reinders J, *et al.* Dysregulation of a novel miR-1825/TBCB/TUBA4A pathway in sporadic and familial ALS. *Cell Mol Life Sci* 2018;**75**:4301–19. doi:10.1007/s00018-018-2873-1

6 Jawaid A, Woldemichael BT, Kremer EA, *et al.* Memory Decline and Its Reversal in Aging and Neurodegeneration Involve miR-183/96/182 Biogenesis. *Mol Neurobiol* 2019;**56**:3451–62. doi:10.1007/s12035-018-1314-3

Appendix B

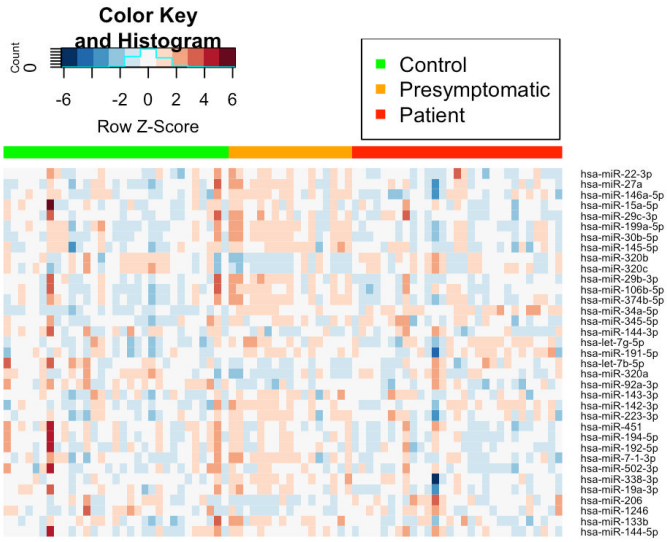
Circulating microRNA signatures as potential biomarkers in genetic frontotemporal dementia and amyotrophic lateral sclerosis

This appendix is the supplementary material of the Chapter 5:

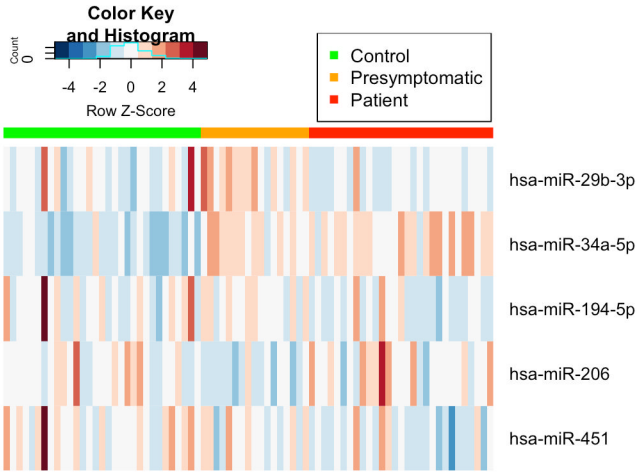
- **Kmetzsch, V.**, Latouche, M., Saracino, D., Rinaldi, D., Camuzat, A., Gareau, T., the French research network on FTD/ALS, Le Ber, I., Colliot, O., Becker, E, "Circulating microRNA signatures as potential biomarkers in genetic frontotemporal dementia and amyotrophic lateral sclerosis". Submitted to *Molecular Psychiatry*.

Supplementary Figure 1. Expression heatmaps of differentially expressed microRNAs. Rows represent microRNAs and columns represent individuals ordered by clinical status (control, presymptomatic, and patient). The log_2 expression values of microRNAs are standardized (mean of 0 and standard deviation of 1), and z-scores are indicated by colors (blue indicates underexpression and red indicates overexpression). A) All 35 differentially expressed microRNAs identified in the *C9orf72* cohort. B) Zoom over five of the most differentially expressed microRNAs identified in the *C9orf72* cohort. C) All five differentially expressed microRNAs identified in the *GRN* cohort.

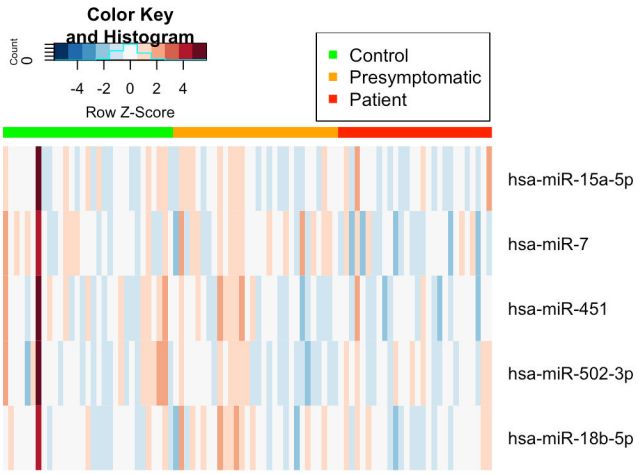
A)



B)



C)



Supplementary Table 1. Complete output from differential expression analyses in the *C9orf72* cohort, for each pairwise comparison between the clinical groups. The columns show the 30 studied miRNAs, the log-fold change when comparing the clinical groups, the unadjusted *P* values, and finally the adjusted *P* values after Benjamini-Hochberg. For each pairwise comparison, a positive log-fold change means that the miRNA is overexpressed in the first group. Controls (n=31), *C9orf72* presymptomatic subjects (n=17), and *C9orf72* patients (n=29). Adjusted *P* values lower than 0.05 are shown in bold.

miRNA	log-fold change	*P* value	adjusted *P* value
Controls vs. *C9orf72* presymptomatic subjects			
miR-34a-5p	-1.58	3.72E-10	**2.42E-08**
miR-338-3p	-0.79	3.48E-04	**9.53E-03**
miR-142-3p	-0.82	4.90E-04	**9.53E-03**
miR-320a	0.74	5.87E-04	**9.53E-03**
miR-145-5p	-0.94	2.29E-03	**2.53E-02**
miR-92a-3p	0.75	2.63E-03	**2.53E-02**
let-7g-5p	-0.46	2.73E-03	**2.53E-02**
miR-199a-5p	-1.13	3.62E-03	**2.62E-02**
miR-206	2.04	3.62E-03	**2.62E-02**
miR-30b-5p	-1.17	4.45E-03	**2.89E-02**
miR-191-5p	-0.44	5.34E-03	**3.00E-02**
miR-27a	-0.89	5.53E-03	**3.00E-02**
miR-320b	0.76	7.88E-03	**3.94E-02**
miR-143-3p	-0.67	9.46E-03	**4.22E-02**
miR-1246	1.10	9.73E-03	**4.22E-02**
miR-223-3p	-0.70	1.08E-02	**4.38E-02**
miR-144-3p	0.87	1.17E-02	**4.46E-02**
let-7b-5p	0.39	1.52E-02	5.50E-02
miR-151a-5p	-0.50	1.83E-02	6.25E-02
miR-1234-3p	1.40	2.01E-02	6.52E-02
miR-26a-5p	-0.49	2.11E-02	6.52E-02
miR-374b-5p	-0.80	2.45E-02	7.01E-02
miR-146a-5p	-0.68	2.48E-02	7.01E-02
miR-320c	0.58	3.60E-02	9.37E-02
miR-301a-3p	-0.46	3.60E-02	9.37E-02
miR-144-5p	1.18	4.82E-02	1.20E-01
miR-7-1-3p	-0.55	4.99E-02	1.20E-01
miR-425-5p	0.29	5.39E-02	1.23E-01
miR-9	-0.37	5.51E-02	1.23E-01
miR-345-5p	-0.40	5.72E-02	1.24E-01
miR-451	1.02	5.90E-02	1.24E-01
miR-423-3p	-0.37	7.56E-02	1.53E-01
miR-10a-3p	-0.38	1.05E-01	2.08E-01
miR-29b-3p	-0.41	1.23E-01	2.34E-01
miR-194-5p	0.55	1.27E-01	2.36E-01
miR-7	-0.28	1.44E-01	2.60E-01
miR-192-5p	0.39	1.67E-01	2.93E-01
miR-200c-3p	-0.26	1.79E-01	3.01E-01
let-7e	-0.35	1.82E-01	3.01E-01
miR-26b-5p	-0.20	1.88E-01	3.01E-01
miR-127-3p	-0.37	1.90E-01	3.01E-01
miR-1	-0.57	2.13E-01	3.28E-01

miR-22-3p	-0.17	2.17E-01	3.28E-01
miR-106b-5p	-0.27	2.34E-01	3.45E-01
miR-4649-5p	0.35	2.52E-01	3.61E-01
miR-502-3p	0.30	2.56E-01	3.61E-01
miR-18b-5p	0.26	2.84E-01	3.93E-01
miR-129-1-3p	-0.72	3.13E-01	4.24E-01
miR-181a-5p	0.16	3.58E-01	4.74E-01
miR-15a-5p	0.16	4.72E-01	6.14E-01
miR-106a-5p	0.20	4.87E-01	6.21E-01
miR-181b-5p	0.11	5.06E-01	6.33E-01
miR-133a	-0.23	5.36E-01	6.49E-01
miR-133a-3p	-0.23	5.39E-01	6.49E-01
miR-29c-3p	-0.13	5.71E-01	6.74E-01
miR-133b	-0.20	6.39E-01	7.41E-01
miR-4745-5p	-0.28	7.24E-01	8.24E-01
miR-19a-3p	-0.12	7.35E-01	8.24E-01
miR-663a	-0.10	7.51E-01	8.28E-01
miR-1915-3p	0.85	9.10E-01	9.86E-01
miR-335-5p	-0.02	9.29E-01	9.90E-01
miR-4530	0.00	1.00E+00	1.00E+00
miR-1825	0.00	1.00E+00	1.00E+00
miR-4299	0.00	1.00E+00	1.00E+00
miR-3665	0.00	1.00E+00	1.00E+00
Controls vs. *C9orf72* patients			
miR-34a-5p	-1.49	7.78E-08	**5.06E-06**
miR-451	2.20	2.87E-05	**6.80E-04**
miR-194-5p	1.55	3.21E-05	**6.80E-04**
miR-144-5p	2.68	4.19E-05	**6.80E-04**
miR-29b-3p	1.01	2.86E-04	**3.72E-03**
miR-29c-3p	0.80	6.61E-04	**7.16E-03**
miR-192-5p	0.99	9.39E-04	**8.72E-03**
miR-19a-3p	1.25	1.32E-03	**1.08E-02**
miR-502-3p	0.80	4.84E-03	**3.50E-02**
miR-15a-5p	0.62	6.70E-03	**4.36E-02**
miR-206	-1.81	7.92E-03	**4.68E-02**
let-7e	-0.74	9.91E-03	5.37E-02
miR-133b	1.11	1.45E-02	7.26E-02
miR-18b-5p	0.62	1.59E-02	7.37E-02
miR-106a-5p	0.69	1.88E-02	8.17E-02
miR-1915-3p	1.44	2.04E-02	8.29E-02
miR-129-1-3p	-1.54	3.45E-02	1.32E-01
miR-191-5p	-0.33	4.58E-02	1.64E-01
miR-106b-5p	0.46	4.79E-02	1.64E-01
miR-144-3p	0.65	5.17E-02	1.68E-01
miR-7-1-3p	0.53	6.26E-02	1.94E-01
miR-30b-5p	0.73	8.19E-02	2.40E-01
miR-374b-5p	0.62	8.50E-02	2.40E-01
miR-4745-5p	-1.23	9.18E-02	2.49E-01
miR-27a	0.51	1.19E-01	3.11E-01
miR-425-5p	0.24	1.27E-01	3.18E-01
miR-199a-5p	0.57	1.45E-01	3.39E-01
miR-200c-3p	-0.29	1.51E-01	3.39E-01
let-7g-5p	-0.22	1.56E-01	3.39E-01
miR-133a-3p	0.53	1.62E-01	3.39E-01
miR-133a	0.53	1.62E-01	3.39E-01
miR-92a-3p	0.33	1.74E-01	3.53E-01
miR-151a-5p	-0.28	1.97E-01	3.71E-01

miR-10a-3p	-0.31	1.98E-01	3.71E-01
miR-22-3p	0.19	2.00E-01	3.71E-01
miR-320c	-0.33	2.47E-01	4.46E-01
miR-345-5p	0.22	3.25E-01	5.70E-01
miR-4649-5p	-0.60	3.47E-01	5.91E-01
miR-146a-5p	0.28	3.55E-01	5.91E-01
miR-335-5p	0.19	3.70E-01	6.02E-01
miR-1	-0.37	3.91E-01	6.20E-01
miR-26b-5p	0.12	4.43E-01	6.86E-01
miR-26a-5p	-0.16	4.76E-01	7.20E-01
miR-1246	0.28	4.94E-01	7.30E-01
miR-1234-3p	0.41	5.06E-01	7.32E-01
miR-320b	-0.18	5.27E-01	7.45E-01
miR-320a	-0.11	6.05E-01	8.36E-01
miR-9	0.09	6.40E-01	8.66E-01
let-7b-5p	-0.07	6.76E-01	8.78E-01
miR-181b-5p	0.07	6.77E-01	8.78E-01
miR-223-3p	0.12	6.89E-01	8.78E-01
miR-127-3p	-0.11	7.08E-01	8.85E-01
miR-301a-3p	0.08	7.33E-01	8.99E-01
miR-145-5p	0.10	7.48E-01	9.01E-01
miR-423-3p	-0.06	7.90E-01	9.33E-01
miR-181a-5p	0.04	8.21E-01	9.53E-01
miR-142-3p	-0.03	8.89E-01	9.97E-01
miR-338-3p	-0.03	8.90E-01	9.97E-01
miR-7	-0.01	9.42E-01	1.00E+00
miR-663a	-0.18	9.51E-01	1.00E+00
miR-143-3p	0.01	9.56E-01	1.00E+00
miR-4530	0.00	1.00E+00	1.00E+00
miR-1825	0.00	1.00E+00	1.00E+00
miR-4299	0.00	1.00E+00	1.00E+00
miR-3665	0.00	1.00E+00	1.00E+00
***C9orf72* presymptomatic subjects vs. *C9orf72* patients**			
miR-206	-3.85	2.35E-06	**9.04E-05**
miR-29b-3p	1.42	2.78E-06	**9.04E-05**
miR-30b-5p	1.90	3.15E-05	**6.83E-04**
miR-199a-5p	1.70	6.58E-05	**8.93E-04**
miR-27a	1.41	6.87E-05	**8.93E-04**
miR-29c-3p	0.92	1.86E-04	**2.01E-03**
miR-320a	-0.85	2.17E-04	**2.02E-03**
miR-374b-5p	1.42	2.87E-04	**2.33E-03**
miR-7-1-3p	1.08	4.35E-04	**3.14E-03**
miR-19a-3p	1.37	6.33E-04	**4.11E-03**
miR-338-3p	0.76	1.26E-03	**7.43E-03**
miR-145-5p	1.04	1.41E-03	**7.61E-03**
miR-142-3p	0.79	1.57E-03	**7.83E-03**
miR-320b	-0.94	1.86E-03	**8.61E-03**
miR-320c	-0.92	2.09E-03	**9.06E-03**
miR-106b-5p	0.72	2.97E-03	**1.21E-02**
miR-146a-5p	0.96	3.36E-03	**1.28E-02**
miR-133b	1.31	6.20E-03	**2.23E-02**
miR-223-3p	0.82	6.53E-03	**2.23E-02**
let-7b-5p	-0.46	7.16E-03	**2.33E-02**
miR-345-5p	0.61	7.69E-03	**2.38E-02**
miR-194-5p	1.01	8.56E-03	**2.53E-02**
miR-143-3p	0.68	1.13E-02	**3.19E-02**
miR-22-3p	0.36	1.77E-02	**4.80E-02**

miR-301a-3p	0.54	2.24E-02	5.83E-02
miR-144-5p	1.50	2.38E-02	5.91E-02
miR-9	0.46	2.45E-02	5.91E-02
miR-451	1.18	2.76E-02	6.40E-02
miR-26b-5p	0.33	4.62E-02	1.03E-01
miR-15a-5p	0.46	4.75E-02	1.03E-01
miR-1246	-0.82	5.16E-02	1.06E-01
miR-192-5p	0.59	5.33E-02	1.06E-01
miR-133a	0.76	5.52E-02	1.06E-01
miR-133a-3p	0.76	5.57E-02	1.06E-01
miR-4649-5p	-0.95	8.39E-02	1.55E-01
miR-502-3p	0.50	8.58E-02	1.55E-01
miR-92a-3p	-0.42	9.76E-02	1.71E-01
miR-106a-5p	0.49	1.02E-01	1.74E-01
miR-1234-3p	-0.99	1.12E-01	1.86E-01
miR-26a-5p	0.33	1.38E-01	2.22E-01
let-7g-5p	0.23	1.40E-01	2.22E-01
miR-4745-5p	-0.95	1.44E-01	2.23E-01
miR-423-3p	0.31	1.58E-01	2.39E-01
miR-18b-5p	0.36	1.68E-01	2.43E-01
let-7e	-0.38	1.71E-01	2.43E-01
miR-129-1-3p	-0.83	1.72E-01	2.43E-01
miR-7	0.26	1.86E-01	2.57E-01
miR-151a-5p	0.21	3.40E-01	4.53E-01
miR-335-5p	0.21	3.41E-01	4.53E-01
miR-127-3p	0.26	3.94E-01	5.12E-01
miR-191-5p	0.10	5.21E-01	6.46E-01
miR-181a-5p	-0.12	5.22E-01	6.46E-01
miR-144-3p	-0.22	5.27E-01	6.46E-01
miR-1	0.20	6.86E-01	8.21E-01
miR-34a-5p	0.09	6.95E-01	8.21E-01
miR-425-5p	-0.05	7.35E-01	8.54E-01
miR-10a-3p	0.06	7.93E-01	9.04E-01
miR-181b-5p	-0.04	8.34E-01	9.35E-01
miR-200c-3p	-0.03	8.72E-01	9.61E-01
miR-663a	-0.07	9.82E-01	1.00E+00
miR-1915-3p	0.59	1.00E+00	1.00E+00
miR-4530	0.00	1.00E+00	1.00E+00
miR-1825	0.00	1.00E+00	1.00E+00
miR-4299	0.00	1.00E+00	1.00E+00
miR-3665	0.00	1.00E+00	1.00E+00

Supplementary Table 2. Complete output from differential expression analyses in the *GRN* cohort, for each pairwise comparison between clinical groups. The columns show the 30 studied miRNAs, the log-fold change when comparing the clinical groups, the unadjusted *P* values, and finally the adjusted *P* values after Benjamini-Hochberg. For each pairwise comparison, a positive log-fold change means that the miRNA is overexpressed in the first group. Controls (n=31), *GRN* presymptomatic subjects (n=30), *GRN* patients (n=28). Adjusted *P* values lower than 0.05 are shown in bold.

miRNA	log-fold change	*P* value	adjusted *P* value
Controls vs. *GRN* presymptomatic subjects			
miR-502-3p	0.53	1.58E-02	3.71E-01
miR-451	0.78	3.88E-02	3.71E-01
let-7e	-0.39	4.87E-02	3.71E-01
miR-206	1.01	4.95E-02	3.71E-01
miR-92a-3p	0.27	1.15E-01	6.04E-01
miR-345-5p	0.26	1.21E-01	6.04E-01
miR-15a-5p	0.22	2.23E-01	7.11E-01
let-7g-5p	-0.13	2.40E-01	7.11E-01
miR-320a	0.18	2.55E-01	7.11E-01
miR-30b-5p	0.33	2.65E-01	7.11E-01
miR-22-3p	0.11	2.70E-01	7.11E-01
miR-223-3p	0.21	2.87E-01	7.11E-01
miR-301a-3p	-0.17	3.25E-01	7.11E-01
miR-10a-3p	-0.19	3.32E-01	7.11E-01
miR-320b	0.20	3.62E-01	7.20E-01
miR-1246	0.32	3.91E-01	7.20E-01
miR-34a-5p	-0.13	4.21E-01	7.20E-01
let-7b-5p	0.10	4.32E-01	7.20E-01
miR-200c-3p	-0.09	4.60E-01	7.26E-01
miR-146a-5p	0.13	5.84E-01	8.76E-01
miR-26a-5p	-0.07	6.22E-01	8.89E-01
miR-335-5p	-0.07	6.83E-01	9.31E-01
miR-663a	0.13	7.30E-01	9.31E-01
miR-7	-0.05	7.45E-01	9.31E-01
miR-9	-0.03	8.77E-01	9.79E-01
miR-106b-5p	0.02	8.92E-01	9.79E-01
miR-106a-5p	0.03	9.05E-01	9.79E-01
miR-18b-5p	-0.02	9.16E-01	9.79E-01
miR-127-3p	0.01	9.60E-01	9.79E-01
miR-26b-5p	0.00	9.79E-01	9.79E-01
Controls vs. *GRN* patients			
miR-451	2.23	2.65E-06	**7.96E-05**
miR-15a-5p	0.77	3.03E-04	**4.54E-03**
miR-502-3p	0.82	1.73E-03	**1.73E-02**
miR-7	0.56	4.56E-03	**3.42E-02**
miR-18b-5p	0.64	7.39E-03	**4.44E-02**
miR-106a-5p	0.68	1.14E-02	5.72E-02
miR-92a-3p	0.51	1.57E-02	6.72E-02
miR-106b-5p	0.47	2.67E-02	9.44E-02
let-7b-5p	0.33	2.83E-02	9.44E-02
miR-223-3p	-0.35	1.24E-01	3.72E-01
miR-320a	0.24	2.17E-01	5.93E-01

miR-22-3p	0.14	2.53E-01	6.33E-01
miR-26b-5p	0.15	3.31E-01	7.53E-01
miR-26a-5p	-0.17	3.51E-01	7.53E-01
miR-663a	-0.60	3.93E-01	7.85E-01
miR-345-5p	-0.15	4.43E-01	8.30E-01
miR-1246	0.29	5.11E-01	8.51E-01
miR-200c-3p	-0.10	5.11E-01	8.51E-01
miR-335-5p	0.12	5.43E-01	8.58E-01
miR-30b-5p	0.17	6.22E-01	8.65E-01
miR-320b	0.12	6.52E-01	8.65E-01
let-7e	-0.11	6.72E-01	8.65E-01
miR-9	-0.07	7.24E-01	8.65E-01
miR-146a-5p	0.09	7.42E-01	8.65E-01
miR-206	-0.20	7.47E-01	8.65E-01
miR-34a-5p	-0.06	7.56E-01	8.65E-01
let-7g-5p	0.04	7.98E-01	8.65E-01
miR-301a-3p	0.05	8.08E-01	8.65E-01
miR-127-3p	-0.05	8.65E-01	8.88E-01
miR-10a-3p	-0.03	8.88E-01	8.88E-01
***GRN* presymptomatic subjects vs. *GRN* patients**			
miR-451	1.45	3.35E-03	6.92E-02
miR-7	0.61	4.61E-03	6.92E-02
miR-18b-5p	0.66	1.31E-02	1.19E-01
miR-15a-5p	0.55	1.59E-02	1.19E-01
miR-106a-5p	0.65	2.52E-02	1.51E-01
miR-223-3p	-0.55	3.11E-02	1.56E-01
miR-206	-1.22	4.58E-02	1.96E-01
miR-106b-5p	0.45	5.79E-02	2.00E-01
miR-345-5p	-0.41	6.01E-02	2.00E-01
let-7b-5p	0.23	1.44E-01	4.32E-01
let-7g-5p	0.17	2.59E-01	5.96E-01
let-7e	0.29	2.62E-01	5.96E-01
miR-92a-3p	0.24	2.83E-01	5.96E-01
miR-663a	-0.74	2.87E-01	5.96E-01
miR-502-3p	0.29	2.98E-01	5.96E-01
miR-301a-3p	0.22	3.28E-01	6.15E-01
miR-26b-5p	0.14	3.72E-01	6.49E-01
miR-335-5p	0.18	3.90E-01	6.49E-01
miR-10a-3p	0.16	5.34E-01	8.44E-01
miR-26a-5p	-0.10	6.14E-01	9.20E-01
miR-30b-5p	-0.16	6.67E-01	9.41E-01
miR-34a-5p	0.07	7.38E-01	9.41E-01
miR-320b	-0.08	7.64E-01	9.41E-01
miR-320a	0.05	7.88E-01	9.41E-01
miR-127-3p	-0.06	8.40E-01	9.41E-01
miR-9	-0.04	8.42E-01	9.41E-01
miR-22-3p	0.03	8.47E-01	9.41E-01
miR-146a-5p	-0.04	9.07E-01	9.63E-01
miR-1246	-0.03	9.49E-01	9.63E-01
miR-200c-3p	-0.01	9.63E-01	9.63E-01

Appendix C

Disease progression score estimation from multimodal imaging and microRNA data using supervised variational autoencoders

This appendix is the supplementary material of the Chapter 6:

- **Kmetzsch, V.**, Becker, E., Saracino, D., Rinaldi, D., Camuzat, A., Le Ber, I., Colliot, O., "Disease progression score estimation from multimodal imaging and microRNA data using supervised variational autoencoders". Under review at the *IEEE Journal of Biomedical and Health Informatics*.

Algorithm 1 Synthetic datasets generation

Input: number of subjects n, number of features m, disease progression scores values v_i (i=1,...,n).
Output: set L containing the datasets $\mathcal{D} \in \mathbb{R}^{n \times m}$
$L = \{\}$
for f in $\{0, 2, 5, 10, 15, 20, 25, 30, 35, 40\}$ **do**
  for s in $\{0.001, 0.2, 0.5, 0.8, 1, 5\}$ **do**
    $C_{1,...,\frac{m}{2}} \leftarrow \text{NB}(r = 3000,\ p = 0.75,\ size = (n, \frac{m}{2}))$
    $C_{\frac{m}{2}+1,...,m} \leftarrow \mathcal{N}(\mu\text{=}1000,\ \sigma\text{=}200,\ size\text{=}(n, \frac{m}{2}))$
    /* f features from each modality positively correlate with disease progression */
    for j in $\{1, ..., f\} \cup \{\frac{m}{2} + 1, ..., \frac{m}{2} + 1 + f\}$ **do**
      $C_j \leftarrow v \odot C_j$
    end for
    /* the next f features are negatively correlated */
    for j in $\{f, ..., 2f\} \cup \{\frac{m}{2} + 1 + f, ..., \frac{m}{2} + 1 + 2f\}$ **do**
      $C_j \leftarrow \frac{1}{v} \odot C_j$
    end for
    /* normalize and add zero-mean Gaussian noise */
    for j in {1,...,m} **do**
      $C_j \leftarrow \frac{C_j - min(C_j)}{max(C_j) - min(C_j)}$
    end for
    $\mathcal{D} \leftarrow \mathcal{D} + \mathcal{N}(\mu\text{=}0,\ \sigma\text{=s},\ size\text{=}(n, m))$
    $L = L \cup \{D\}$
  end for
end for
return L

Bibliography

Abramzon, Y. A., P. Fratta, B. J. Traynor, and R. Chia (2020). "The Overlapping Genetics of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia". In: *Frontiers in Neuroscience* 14.

Aksman, L. M., M. A. Scelsi, A. F. Marquand, D. C. Alexander, S. Ourselin, A. Altmann, and for ADNI (Sept. 2019). "Modeling longitudinal imaging biomarkers with parametric Bayesian multi-task learning". In: *Human Brain Mapping* 40.13, pp. 3982–4000. DOI: 10.1002/hbm.24682.

Andersson, J., M Jenkinson, and S Smith (2010). "Non-linear registration, aka spatial normalisation". In: *FMRIB Technical Report TR07JA2*.

Antelmi, L., N. Ayache, P. Robert, and M. Lorenzi (May 2019). "Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data". In: *International Conference on Machine Learning*. PMLR, pp. 302–311.

Antonelli, L., M. R. Guarracino, L. Maddalena, and M. Sangiovanni (July 2019). "Integrating imaging and omics data: A review". In: *Biomedical Signal Processing and Control* 52, pp. 264–280. DOI: 10.1016/j.bspc.2019.04.032.

Archetti, D., S. Ingala, V. Venkatraghavan, V. Wottschel, A. L. Young, M. Bellio, E. E. Bron, S. Klein, F. Barkhof, D. C. Alexander, N. P. Oxtoby, G. B. Frisoni, and A. Redolfi (July 2019). "Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease". In: *NeuroImage : Clinical* 24, p. 101954. DOI: 10.1016/j.nicl.2019.101954.

Arnatkeviciute, A., B. D. Fulcher, and A. Fornito (Apr. 2019). "A practical guide to linking brain-wide gene expression and neuroimaging data". In: *NeuroImage* 189, pp. 353–367. DOI: 10.1016/j.neuroimage.2019.01.011.

Ashburner, J. (Aug. 2012). "SPM: A history". In: *NeuroImage*. 20 YEARS OF fMRI 62.2, pp. 791–800. DOI: 10.1016/j.neuroimage.2011.10.025.

Atkinson, R. A. K., C. M. Fernandez-Martos, J. D. Atkin, J. C. Vickers, and A. E. King (Sept. 2015). "C9ORF72 expression and cellular localization over mouse development". In: *Acta Neuropathologica Communications* 3, p. 59. DOI: 10.1186/s40478-015-0238-7.

Azuaje, F. (2010). "Integrative Data Analysis for Biomarker Discovery". In: *Bioinformatics and Biomarker Discovery*. John Wiley & Sons, Ltd, pp. 137–154. DOI: 10.1002/9780470686423.ch8.

Becker-André, M and K Hahlbrock (Nov. 1989). "Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY)." In: *Nucleic Acids Research* 17.22, pp. 9437–9446.

Behrad, F. and M. Saniee Abadeh (Aug. 2022). "An overview of deep learning methods for multimodal medical data mining". In: *Expert Systems with Applications* 200, p. 117006. DOI: 10.1016/j.eswa.2022.117006.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1, pp. 289–300.

Berg, M. M. J. van den, J. Krauskopf, J. G. Ramaekers, J. C. S. Kleinjans, J. Prickaerts, and J. J. Briedé (Feb. 2020). "Circulating microRNAs as potential biomarkers for psychiatric and neurodegenerative disorders". In: *Progress in Neurobiology* 185, p. 101732. DOI: 10.1016/j.pneurobio.2019.101732.

Berkhahn, F., R. Keys, W. Ouertani, N. Shetty, and D. Geißler (Nov. 2019). "Augmenting Variational Autoencoders with Sparse Labels: A Unified Framework for Unsupervised, Semi-(un)supervised, and Supervised Learning". In: *arXiv:1908.03015 [cs, stat]*.

Bernstein, M. A., K. F. King, and X. J. Zhou (Sept. 2004). *Handbook of MRI Pulse Sequences*. Elsevier.

Bertrand, A., J. Wen, D. Rinaldi, M. Houot, S. Sayah, A. Camuzat, et al. (2018). "Early Cognitive, Structural, and Microstructural Changes in Presymptomatic C9orf72 Carriers Younger Than 40 Years". In: *JAMA neurology* 75.2, pp. 236–245. DOI: 10.1001/jamaneurol.2017.4266.

Biomarkers Definitions Working Group. (Mar. 2001). "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework". In: *Clinical Pharmacology and Therapeutics* 69.3, pp. 89–95. DOI: 10.1067/mcp.2001.113989.

Bitar, R., G. Leung, R. Perng, S. Tadros, A. R. Moody, J. Sarrazin, C. McGregor, M. Christakis, S. Symons, A. Nelson, and T. P. Roberts (Mar. 2006). "MR Pulse Sequences: What Every Radiologist Wants to Know but Is Afraid to Ask". In: *RadioGraphics* 26.2, pp. 513–537. DOI: 10.1148/rg.262055063.

Boese, A. S., R. Saba, K. Campbell, A. Majer, S. Medina, L. Burton, T. F. Booth, P. Chong, G. Westmacott, S. M. Dutta, J. A. Saba, and S. A. Booth (Mar. 2016). "MicroRNA abundance is altered in synaptoneurosomes during prion disease". In: *Molecular and Cellular Neuroscience* 71, pp. 13–24. DOI: 10.1016/j.mcn.2015.12.001.

Brennan, S., M. Keon, B. Liu, Z. Su, and N. K. Saksena (Nov. 2019). "Panoramic Visualization of Circulating MicroRNAs Across Neurodegenerative Diseases in Humans". In: *Molecular Neurobiology* 56.11, pp. 7380–7407. DOI: 10.1007/s12035-019-1615-1.

Brown, M. A. and R. C. Semelka (Dec. 1999). "MR Imaging Abbreviations, Definitions, and Descriptions: A Review". In: *Radiology* 213.3, pp. 647–662. DOI: 10.1148/radiology.213.3.r99dc18647.

Brown, P. O. and D. Botstein (Jan. 1999). "Exploring the new world of the genome with DNA microarrays". In: *Nature Genetics* 21.1, pp. 33–37. DOI: 10.1038/4462.

Buratti, E. and F. E. Baralle (Aug. 2010). "The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation". In: *RNA biology* 7.4, pp. 420–429. DOI: 10.4161/rna.7.4.12205.

Burberry, A., M. F. Wells, F. Limone, A. Couto, K. S. Smith, J. Keaney, G. Gillet, N. van Gastel, J.-Y. Wang, O. Pietilainen, M. Qian, P. Eggan, C. Cantrell, J. Mok, I. Kadiu, D. T. Scadden, and K. Eggan (2020). "C9orf72 suppresses systemic and neural inflammation induced by gut bacteria". In: *Nature* 582.7810, pp. 89–94. DOI: 10.1038/s41586-020-2288-7.

Burrell, J. R., M. C. Kiernan, S. Vucic, and J. R. Hodges (Sept. 2011). "Motor neuron dysfunction in frontotemporal dementia". In: *Brain: A Journal of Neurology* 134.Pt 9, pp. 2582–2594. DOI: 10.1093/brain/awr195.

Bzdok, D., D. Engemann, and B. Thirion (Oct. 2020). "Inference and Prediction Diverge in Biomedicine". In: *Patterns*, p. 100119. DOI: 10.1016/j.patter.2020.100119.

Carreiro, A. V., A. Mendonça, M. d. Carvalho, and S. C. Madeira (2015). "Integrative biomarker discovery in neurodegenerative diseases". In: *WIREs Systems Biology and Medicine* 7.6, pp. 357–379. DOI: 10.1002/wsbm.1310.

Cedazo-Minguez, A. and B. Winblad (Jan. 2010). "Biomarkers for Alzheimer's disease and other forms of dementia: clinical needs, limitations and future aspects". In: *Experimental Gerontology* 45.1, pp. 5–14. DOI: 10.1016/j.exger.2009.09.008.

Chen-Plotkin, A. S., T. L. Unger, M. D. Gallagher, E. Bill, L. K. Kwong, L. Volpicelli-Daley, J. I. Busch, S. Akle, M. Grossman, V. Van Deerlin, J. Q. Trojanowski, and V. M.-Y. Lee (Aug. 2012). "TMEM106B, the Risk Gene for Frontotemporal Dementia, Is Regulated by the microRNA-132/212 Cluster and Affects Progranulin Pathways". In: *The Journal of Neuroscience* 32.33, pp. 11213–11227. DOI: 10.1523/JNEUROSCI.0521-12.2012.

Cheng, J., M. Gao, J. Liu, H. Yue, H. Kuang, J. Liu, and J. Wang (Feb. 2022). "Multimodal Disentangled Variational Autoencoder With Game Theoretic Interpretability for Glioma Grading". In: *IEEE Journal of Biomedical and Health Informatics* 26.2, pp. 673–684. DOI: 10.1109/JBHI.2021.3095476.

Chiò, A., G. Logroscino, B. J. Traynor, J. Collins, J. C. Simeone, L. A. Goldstein, and L. A. White (2013). "Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature". In: *Neuroepidemiology* 41.2, pp. 118–130. DOI: 10.1159/000351153.

Chua, C. E. L. and B. L. Tang (Feb. 2019). "miR-34a in Neurophysiology and Neuropathology". In: *Journal of molecular neuroscience: MN* 67.2, pp. 235–246. DOI: 10.1007/s12031-018-1231-y.

Cock, P. J. A., C. J. Fields, N. Goto, M. L. Heuer, and P. M. Rice (Apr. 2010). "The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants". In: *Nucleic Acids Research* 38.6, pp. 1767–1771. DOI: 10.1093/nar/gkp1137.

Cosín-Tomás, M., A. Antonell, A. Lladó, D. Alcolea, J. Fortea, M. Ezquerra, A. Lleó, M. J. Martí, M. Pallàs, R. Sanchez-Valle, J. L. Molinuevo, C. Sanfeliu, and P. Kaliman (2017). "Plasma miR-34a-5p and miR-545-3p as Early Biomarkers of Alzheimer's Disease: Potential and Limitations". In: *Molecular Neurobiology* 54.7, pp. 5550–5562. DOI: 10.1007/s12035-016-0088-8.

Crick, F. (Aug. 1970). "Central Dogma of Molecular Biology". In: *Nature* 227.5258, pp. 561–563. DOI: 10.1038/227561a0.

Cui, Y., B. Liu, S. Luo, X. Zhen, M. Fan, T. Liu, W. Zhu, M. Park, T. Jiang, J. S. Jin, and Alzheimer's Disease Neuroimaging Initiative (2011). "Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors". In: *PloS One* 6.7, e21896. DOI: 10.1371/journal.pone.0021896.

De Felice, B., A. Annunziata, G. Fiorentino, M. Borra, E. Biffali, C. Coppola, R. Cotrufo, J. Brettschneider, M. L. Giordana, T. Dalmay, G. Wheeler, and R. D'Alessandro (Oct. 2014). "miR-338-3p is over-expressed in blood, CFS, serum and spinal cord from sporadic amyotrophic lateral sclerosis patients". In: *neurogenetics* 15.4, pp. 243–253. DOI: 10.1007/s10048-014-0420-2.

DeJesus Hernandez, M., I. R. Mackenzie, B. F. Boeve, A. L. Boxer, M. Baker, N. J. Rutherford, et al. (Oct. 2011). "Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS". In: *Neuron* 72.2, pp. 245–256. DOI: 10.1016/j.neuron.2011.09.011.

Dekker, I., M. M. Schoonheim, V. Venkatraghavan, A. J. Eijlers, I. Brouwer, E. E. Bron, S. Klein, M. P. Wattjes, A. M. Wink, J. J. Geurts, B. M. Uitdehaag, N. P. Oxtoby, D. C. Alexander, H. Vrenken, J. Killestein, F. Barkhof, and V. Wottschel (Dec. 2020). "The sequence of structural, functional and cognitive changes in multiple sclerosis". In: *NeuroImage : Clinical* 29, p. 102550. DOI: 10.1016/j.nicl.2020.102550.

Denk, J., F. Oberhauser, J. Kornhuber, J. Wiltfang, K. Fassbender, M. L. Schroeter, A. E. Volk, J. Diehl-Schmid, J. Prudlo, A. Danek, B. Landwehrmeyer, M. Lauer, M. Otto, H. Jahn, and FTLDc study group (2018). "Specific serum and CSF microRNA profiles distinguish sporadic behavioural variant of frontotemporal dementia compared with Alzheimer patients and cognitively healthy controls". In: *PloS One* 13.5, e0197329. DOI: 10.1371/journal.pone.0197329.

Desikan, R. S., F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany (July 2006). "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest". In: *NeuroImage* 31.3, pp. 968–980. DOI: 10.1016/j.neuroimage.2006.01.021.

Despotović, I., B. Goossens, and W. Philips (Mar. 2015). "MRI Segmentation of the Human Brain: Challenges, Methods, and Applications". In: *Computational and Mathematical Methods in Medicine* 2015, e450341. DOI: 10.1155/2015/450341.

Dobrowolny, G., J. Martone, E. Lepore, I. Casola, A. Petrucci, M. Inghilleri, M. Morlando, A. Colantoni, B. M. Scicchitano, A. Calvo, G. Bisogni, A. Chiò, M. Sabatelli,

I. Bozzoni, and A. Musarò (Jan. 2021). "A longitudinal study defined circulating microRNAs as reliable biomarkers for disease prognosis and progression in ALS human patients". In: *Cell Death Discovery* 7, p. 4. DOI: 10.1038/s41420-020-00397-6.

Eisen, A., M. Kiernan, H. Mitsumoto, and M. Swash (Nov. 2014). "Amyotrophic lateral sclerosis: a long preclinical period?" In: *Journal of Neurology, Neurosurgery & Psychiatry* 85.11, pp. 1232–1238. DOI: 10.1136/jnnp-2013-307135.

Eitan, C. and E. Hornstein (Sept. 2016). "Vulnerability of microRNA biogenesis in FTD–ALS". In: *Brain Research*. RNA Metabolism in Neurological Disease 1647, pp. 105–111. DOI: 10.1016/j.brainres.2015.12.063.

Ende, E. L. van der, E. E. Bron, J. M. Poos, L. C. Jiskoot, J. L. Panman, J. M. Papma, et al. (Oct. 2021). "A data-driven disease progression model of fluid biomarkers in genetic frontotemporal dementia". In: *Brain: A Journal of Neurology*, awab382. DOI: 10.1093/brain/awab382.

Eshaghi, A., R. V. Marinescu, A. L. Young, N. C. Firth, F. Prados, M Jorge Cardoso, C. Tur, F. De Angelis, N. Cawley, W. J. Brownlee, N. De Stefano, M Laura Stromillo, M. Battaglini, S. Ruggieri, C. Gasperini, et al. (June 2018). "Progression of regional grey matter atrophy in multiple sclerosis". In: *Brain* 141.6, pp. 1665–1677. DOI: 10.1093/brain/awy088.

Evans, A., D. Collins, S. Mills, E. Brown, R. Kelly, and T. Peters (Oct. 1993). "3D statistical neuroanatomical models from 305 MRI volumes". In: *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, 1813–1817 vol.3. DOI: 10.1109/NSSMIC.1993.373602.

Farg, M. A., A. Konopka, K. Y. Soo, D. Ito, and J. D. Atkin (2017). "The DNA damage response (DDR) is induced by the C9orf72 repeat expansion in amyotrophic lateral sclerosis". In: *Human Molecular Genetics* 26.15, pp. 2882–2896. DOI: 10.1093/hmg/ddx170.

Feneberg, E., P. Steinacker, A. E. Volk, J. H. Weishaupt, M. A. Wollmer, A. Boxer, H. Tumani, A. C. Ludolph, and M. Otto (Mar. 2016). "Progranulin as a candidate biomarker for therapeutic trial in patients with ALS and FTLD". In: *Journal of Neural Transmission* 123.3, pp. 289–296. DOI: 10.1007/s00702-015-1486-1.

Firth, N. C., S. Primativo, E. Brotherhood, A. L. Young, K. X. Yong, S. J. Crutch, D. C. Alexander, and N. P. Oxtoby (July 2020). "Sequences of cognitive decline in typical Alzheimer's disease and posterior cortical atrophy estimated using a novel event-based model of disease progression". In: *Alzheimer's & Dementia* 16.7, pp. 965–973. DOI: 10.1002/alz.12083.

Fischl, B. (Aug. 2012). "FreeSurfer". In: *NeuroImage* 62.2, pp. 774–781. DOI: 10.1016/j.neuroimage.2012.01.021.

Fischl, B., D. H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness, A. Montillo, N. Makris, B. Rosen, and A. M. Dale (Jan. 2002). "Whole Brain Segmentation". In: *Neuron* 33.3, pp. 341–355. DOI: 10.1016/S0896-6273(02)00569-X.

Fonteijn, H. M., M. Modat, M. J. Clarkson, J. Barnes, M. Lehmann, N. Z. Hobbs, R. I. Scahill, S. J. Tabrizi, S. Ourselin, N. C. Fox, and D. C. Alexander (Apr. 2012). "An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease". In: *NeuroImage* 60.3, pp. 1880–1889. DOI: 10.1016/j.neuroimage.2012.01.062.

Freischmidt, A., A. Goswami, K. Limm, V. L. Zimyanin, M. Demestre, H. Glaß, K. Holzmann, A. M. Helferich, S. J. Brockmann, P. Tripathi, A. Yamoah, I. Poser, P. J. Oefner, T. M. Böckers, E. Aronica, A. C. Ludolph, P. M. Andersen, A. Hermann, J. Weis, J. Reinders, K. M. Danzer, and J. H. Weishaupt (2021). "A serum microRNA sequence reveals fragile X protein pathology in amyotrophic lateral sclerosis". In: *Brain* 144.4, pp. 1214–1229. DOI: 10.1093/brain/awab018.

Freischmidt, A., K. Müller, L. Zondler, P. Weydt, B. Mayer, C. A. F. von Arnim, A. Hübers, J. Dorst, M. Otto, K. Holzmann, A. C. Ludolph, K. M. Danzer, and J. H. Weishaupt (Sept. 2015). "Serum microRNAs in sporadic amyotrophic lateral sclerosis". In: *Neurobiology of Aging* 36.9, 2660.e15–2660.e20. DOI: 10.1016/j.neurobiolaging.2015.06.003.

Freischmidt, A., K. Müller, L. Zondler, P. Weydt, A. E. Volk, A. L. Božič, M. Walter, M. Bonin, B. Mayer, C. A. F. von Arnim, M. Otto, C. Dieterich, K. Holzmann, P. M. Andersen, A. C. Ludolph, K. M. Danzer, and J. H. Weishaupt (Nov. 2014). "Serum microRNAs in patients with genetic amyotrophic lateral sclerosis and pre-manifest mutation carriers". In: *Brain: A Journal of Neurology* 137.Pt 11, pp. 2938–2950. DOI: 10.1093/brain/awu249.

Friedländer, M. R., S. D. Mackowiak, N. Li, W. Chen, and N. Rajewsky (Jan. 2012). "miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades". In: *Nucleic Acids Research* 40.1, pp. 37–52. DOI: 10.1093/nar/gkr688.

Fu, J., L. Peng, T. Tao, Y. Chen, Z. Li, and J. Li (Nov. 2019). "Regulatory roles of the miR-200 family in neurodegenerative diseases". In: *Biomedicine & Pharmacotherapy* 119, p. 109409. DOI: 10.1016/j.biopha.2019.109409.

Gabel, M. C., R. J. Broad, A. L. Young, S. Abrahams, M. E. Bastin, R. A. L. Menke, A. Al-Chalabi, L. H. Goldstein, S. Tsermentseli, D. C. Alexander, M. R. Turner, P. N. Leigh, and M. Cercignani (May 2020). "Evolution of white matter damage in amyotrophic lateral sclerosis". In: *Annals of Clinical and Translational Neurology* 7.5, pp. 722–732. DOI: 10.1002/acn3.51035.

Gallivanone, F., C. Cava, F. Corsi, G. Bertoli, and I. Castiglioni (Nov. 2019). "In Silico Approach for the Definition of radiomiRNomic Signatures for Breast Cancer Differential Diagnosis". In: *International Journal of Molecular Sciences* 20.23. DOI: 10.3390/ijms20235825.

Garbarino, S., M. Lorenzi, N. P. Oxtoby, E. J. Vinke, R. V. Marinescu, A. Eshaghi, M. A. Ikram, W. J. Niessen, O. Ciccarelli, F. Barkhof, J. M. Schott, M. W. Vernooij, and D. C. Alexander (2019). "Differences in topological progression profile among

neurodegenerative diseases from imaging data". In: *eLife* 8. Ed. by F. P. de Lange, e49298. DOI: 10.7554/eLife.49298.

Gascon, E. and F.-B. Gao (June 2014). "The emerging roles of microRNAs in the pathogenesis of frontotemporal dementia-amyotrophic lateral sclerosis (FTD-ALS) spectrum disorders". In: *Journal of Neurogenetics* 28.1-2, pp. 30–40. DOI: 10.3109/01677063.2013.876021.

Gascon, E., K. Lynch, H. Ruan, S. Almeida, J. Verheyden, W. W. Seeley, D. W. Dickson, L. Petrucelli, D. Sun, J. Jiao, H. Zhou, M. Jakovcevski, S. Akbarian, W.-D. Yao, and F.-B. Gao (Dec. 2014). "Alterations in microRNA-124 and AMPA receptors contribute to social behavioral deficits in frontotemporal dementia". In: *Nature medicine* 20.12, pp. 1444–1451. DOI: 10.1038/nm.3717.

Gorno-Tempini, M. L., A. E. Hillis, S. Weintraub, A. Kertesz, M. Mendez, S. F. Cappa, J. M. Ogar, J. D. Rohrer, S. Black, B. F. Boeve, F. Manes, N. F. Dronkers, R. Vandenberghe, K. Rascovsky, K. Patterson, B. L. Miller, D. S. Knopman, J. R. Hodges, M. M. Mesulam, and M. Grossman (Mar. 2011). "Classification of primary progressive aphasia and its variants". In: *Neurology* 76.11, pp. 1006–1014. DOI: 10.1212/WNL.0b013e31821103e6.

Grasso, M., P. Piscopo, A. Confaloni, and M. A. Denti (May 2014). "Circulating miRNAs as biomarkers for neurodegenerative disorders". In: *Molecules (Basel, Switzerland)* 19.5, pp. 6891–6910. DOI: 10.3390/molecules19056891.

Grasso, M., P. Piscopo, A. Crestini, A. Confaloni, and M. A. Denti (2015). "Circulating microRNAs in Neurodegenerative Diseases". In: *Circulating microRNAs in Disease Diagnostics and their Potential Biological Relevance*. Ed. by P. Igaz. Experientia Supplementum. Basel: Springer, pp. 151–169. DOI: 10.1007/978-3-0348-0955-9_7.

Grasso, M., P. Piscopo, G. Talarico, L. Ricci, A. Crestini, G. Tosto, M. Gasparini, G. Bruno, M. A. Denti, and A. Confaloni (Dec. 2019). "Plasma microRNA profiling distinguishes patients with frontotemporal dementia from healthy subjects". In: *Neurobiology of Aging* 84, 240.e1–240.e12. DOI: 10.1016/j.neurobiolaging.2019.01.024.

Greve, D. N. and B. Fischl (Oct. 2009). "Accurate and robust brain image alignment using boundary-based registration". In: *NeuroImage* 48.1, pp. 63–72. DOI: 10.1016/j.neuroimage.2009.06.060.

Gupta, H., P. K. Sahu, R. Pattnaik, A. Mohanty, M. Majhi, A. K. Mohanty, L. Pirpamer, A. Hoffmann, S. Mohanty, and S. C. Wassmer (2021). "Plasma levels of hsa-miR-3158-3p microRNA on admission correlate with MRI findings and predict outcome in cerebral malaria". In: *Clinical and Translational Medicine* 11.6, e396. DOI: 10.1002/ctm2.396.

Haacke, E. M., R. W. Brown, M. R. Thompson, and R. Venkatesan (June 1999). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley.

Hastie, T. and W. Stuetzle (1989). "Principal Curves". In: *Journal of the American Statistical Association* 84.406, pp. 502–516. DOI: 10.1080/01621459.1989.10478797.

Helferich, A. M., S. J. Brockmann, J. Reinders, D. Deshpande, K. Holzmann, D. Brenner, P. M. Andersen, S. Petri, D. R. Thal, J. Michaelis, M. Otto, S. Just, A. C. Ludolph, K. M. Danzer, A. Freischmidt, and J. H. Weishaupt (Dec. 2018). "Dysregulation of a novel miR-1825/TBCB/TUBA4A pathway in sporadic and familial ALS". In: *Cellular and Molecular Life Sciences* 75.23, pp. 4301–4319. DOI: 10.1007/s00018-018-2873-1.

Hemond, C. C., B. C. Healy, S. Tauhid, M. A. Mazzola, F. J. Quintana, R. Gandhi, H. L. Weiner, and R. Bakshi (Feb. 2019). "MRI phenotypes in MS". In: *Neurology® Neuroimmunology & Neuroinflammation* 6.2. DOI: 10.1212/NXI.0000000000000530.

Huntzinger, E. and E. Izaurralde (Feb. 2011). "Gene silencing by microRNAs: contributions of translational repression and mRNA decay". In: *Nature Reviews Genetics* 12.2, pp. 99–110. DOI: 10.1038/nrg2936.

Hébert, S. S., W.-X. Wang, Q. Zhu, and P. T. Nelson (2013). "A Study of Small RNAs from Cerebral Neocortex of Pathology-Verified Alzheimer's Disease, Dementia with Lewy Bodies, Hippocampal Sclerosis, Frontotemporal Lobar Dementia, and Non-Demented Human Controls". In: *Journal of Alzheimer's disease : JAD* 35.2, pp. 335–348. DOI: 10.3233/JAD-122350.

Jack, C. R., D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, and J. Q. Trojanowski (Feb. 2013). "Update on hypothetical model of Alzheimer's disease biomarkers". In: *Lancet neurology* 12.2, pp. 207–216. DOI: 10.1016/S1474-4422(12)70291-0.

Jack, C. R., D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski (Jan. 2010). "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade". In: *The Lancet. Neurology* 9.1, pp. 119–128. DOI: 10.1016/S1474-4422(09)70299-6.

Jauhari, A., T. Singh, P. Singh, D. Parmar, and S. Yadav (2018). "Regulation of miR-34 Family in Neuronal Development". In: *Molecular Neurobiology* 55.2, pp. 936–945. DOI: 10.1007/s12035-016-0359-4.

Jawaid, A., B. T. Woldemichael, E. A. Kremer, F. Laferriere, N. Gaur, T. Afroz, M. Polymenidou, and I. M. Mansuy (May 2019). "Memory Decline and Its Reversal in Aging and Neurodegeneration Involve miR-183/96/182 Biogenesis". In: *Molecular Neurobiology* 56.5, pp. 3451–3462. DOI: 10.1007/s12035-018-1314-3.

Jedynak, B. M., A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jedynak, B. Caffo, and J. L. Prince (Nov. 2012). "A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort". In: *NeuroImage* 63.3, pp. 1478–1486. DOI: 10.1016/j.neuroimage.2012.07.059.

Jenkinson, M., C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith (Aug. 2012). "FSL". In: *NeuroImage* 62.2, pp. 782–790. DOI: 10.1016/j.neuroimage.2011.09.015.

Ji, T., S. T. Vuppala, G. Chowdhary, and K. Driggs-Campbell (Oct. 2021). "Multi-Modal Anomaly Detection for Unstructured and Uncertain Environments". In: *Proceedings of the 2020 Conference on Robot Learning*. PMLR, pp. 1443–1455.

Jin, J., Y. Cheng, Y. Zhang, W. Wood, Q. Peng, E. Hutchison, M. P. Mattson, K. G. Becker, and W. Duan (Nov. 2012). "Interrogation of brain miRNA and mRNA expression profiles reveals a molecular regulatory network that is perturbed by mutant huntingtin". In: *Journal of Neurochemistry* 123.4, pp. 477–490. DOI: 10.1111/j.1471-4159.2012.07925.x.

Kalavathi, P. and V. B. S. Prasath (June 2016). "Methods on Skull Stripping of MRI Head Scan Images-a Review". In: *Journal of Digital Imaging* 29.3, pp. 365–379. DOI: 10.1007/s10278-015-9847-8.

Katrib, A., W. Hsu, A. Bui, and Y. Xing (Mar. 2016). "'RADIOTRANSCRIPTOMICS": A synergy of imaging and transcriptomics in clinical assessment". In: *Quantitative Biology (Beijing, China)* 4.1, pp. 1–12. DOI: 10.1007/s40484-016-0061-6.

Keil, J. M., A. Qalieh, and K. Y. Kwan (Mar. 2018). "Brain Transcriptome Databases: A User's Guide". In: *Journal of Neuroscience* 38.10, pp. 2399–2412. DOI: 10.1523/JNEUROSCI.1930-17.2018.

Kingma, D. P. and J. Ba (Jan. 2017). "Adam: A Method for Stochastic Optimization". In: *arXiv:1412.6980 [cs]*.

Kingma, D. P., D. J. Rezende, S. Mohamed, and M. Welling (Oct. 2014). "Semi-Supervised Learning with Deep Generative Models". In: *arXiv:1406.5298 [cs, stat]*.

Kingma, D. P. and M. Welling (May 2014). "Auto-Encoding Variational Bayes". In: *arXiv:1312.6114 [cs, stat]*.

Kmetzsch, V., V. Anquetil, D. Saracino, D. Rinaldi, A. Camuzat, T. Gareau, L. Jornea, S. Forlani, P. Couratier, D. Wallon, F. Pasquier, N. Robil, P. d. l. Grange, I. Moszer, I. L. Ber, O. Colliot, and E. Becker (May 2021). "Plasma microRNA signature in presymptomatic and symptomatic subjects with C9orf72-associated frontotemporal dementia and amyotrophic lateral sclerosis". In: *Journal of Neurology, Neurosurgery & Psychiatry* 92.5, pp. 485–493. DOI: 10.1136/jnnp-2020-324647.

Kmetzsch, V., E. Becker, D. Saracino, V. Anquetil, D. Rinaldi, A. Camuzat, T. Gareau, I. L. Ber, and O. Colliot (Apr. 2022). "A multimodal variational autoencoder for estimating progression scores from imaging and microRNA data in rare neurodegenerative diseases". In: *Medical Imaging 2022: Image Processing*. Vol. 12032. SPIE, pp. 376–382. DOI: 10.1117/12.2607250.

Kocerha, J., N. Kouri, M. Baker, N. Finch, M. DeJesus-Hernandez, J. Gonzalez, K. Chidamparam, K. A. Josephs, B. F. Boeve, N. R. Graff-Radford, J. Crook, D. W. Dickson, and R. Rademakers (Oct. 2011). "Altered microRNA expression in frontotemporal lobar degeneration with TDP-43 pathology caused by progranulin mutations". In: *BMC genomics* 12, p. 527. DOI: 10.1186/1471-2164-12-527.

Koval, I., A. Bône, M. Louis, T. Lartigue, S. Bottani, A. Marcoux, J. Samper-González, N. Burgos, B. Charlier, A. Bertrand, S. Epelbaum, O. Colliot, S. Allassonnière, and

S. Durrleman (Apr. 2021). "AD Course Map charts Alzheimer's disease progression". In: *Scientific Reports* 11.1, p. 8020. DOI: 10.1038/s41598-021-87434-1.

Kovanda, A., L. Leonardis, J. Zidar, B. Koritnik, L. Dolenc-Groselj, S. Ristic Kovacic, T. Curk, and B. Rogelj (Apr. 2018). "Differential expression of microRNAs and other small RNAs in muscle tissue of patients with ALS and healthy age-matched controls". In: *Scientific Reports* 8.1, p. 5609. DOI: 10.1038/s41598-018-23139-2.

Kozomara, A. and S. Griffiths-Jones (Jan. 2011). "miRBase: integrating microRNA annotation and deep-sequencing data". In: *Nucleic Acids Research* 39.Database issue, pp. D152–D157. DOI: 10.1093/nar/gkq1027.

Kudo, L. C., L. Parfenova, N. Vi, K. Lau, J. Pomakian, P. Valdmanis, G. A. Rouleau, H. V. Vinters, M. Wiedau-Pazos, and S. L. Karsten (Aug. 2010). "Integrative gene-tissue microarray-based approach for identification of human disease biomarkers: application to amyotrophic lateral sclerosis". In: *Human Molecular Genetics* 19.16, pp. 3233–3253. DOI: 10.1093/hmg/ddq232.

Kukurba, K. R. and S. B. Montgomery (Apr. 2015). "RNA Sequencing and Analysis". In: *Cold Spring Harbor protocols* 2015.11, pp. 951–969. DOI: 10.1101/pdb.top084970.

Kuo, M. D. and N. Jamshidi (Feb. 2014). "Behind the numbers: Decoding molecular phenotypes with radiogenomics–guiding principles and technical considerations". In: *Radiology* 270.2, pp. 320–325. DOI: 10.1148/radiol.13132195.

Lal, A., M. P. Thomas, G. Altschuler, F. Navarro, E. O'Day, X. L. Li, C. Concepcion, Y.-C. Han, J. Thiery, D. K. Rajani, A. Deutsch, O. Hofmann, A. Ventura, W. Hide, and J. Lieberman (Nov. 2011). "Capture of MicroRNA–Bound mRNAs Identifies the Tumor Suppressor miR-34a as a Regulator of Growth Factor Signaling". In: *PLoS Genetics* 7.11. DOI: 10.1371/journal.pgen.1002363.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg (Mar. 2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome". In: *Genome Biology* 10.3, R25. DOI: 10.1186/gb-2009-10-3-r25.

Lashley, T., J. D. Rohrer, S. Mead, and T. Revesz (2015). "Review: An update on clinical, genetic and pathological aspects of frontotemporal lobar degenerations". In: *Neuropathology and Applied Neurobiology* 41.7, pp. 858–881. DOI: 10.1111/nan.12250.

Le Ber, I., E. Guedj, A. Gabelle, P. Verpillat, M. Volteau, C. Thomas-Anterion, M. Decousus, D. Hannequin, P. Véra, L. Lacomblez, A. Camuzat, M. Didic, M. Puel, J.-A. Lotterie, V. Golfier, A.-M. Bernard, M. Vercelletto, C. Magne, F. Sellal, I. Namer, B.-F. Michel, J. Pasquier, F. Salachas, J. Bochet, French research network on FTD/FTD-MND, A. Brice, M.-O. Habert, and B. Dubois (Nov. 2006). "Demographic, neurological and behavioural characteristics and brain perfusion SPECT in frontal variant of frontotemporal dementia". In: *Brain: A Journal of Neurology* 129.Pt 11, pp. 3051–3065. DOI: 10.1093/brain/awl288.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup (Aug. 2009).

"The Sequence Alignment/Map format and SAMtools". In: *Bioinformatics (Oxford, England)* 25.16, pp. 2078–2079. DOI: 10.1093/bioinformatics/btp352.

Li, L. and J. Wang (Mar. 2019). "Roles of extracellular microRNAs in central nervous system". In: *ExRNA* 1.1, p. 13. DOI: 10.1186/s41544-019-0011-3.

Li, T., A. K. Sahu, A. Talwalkar, and V. Smith (May 2020). "Federated Learning: Challenges, Methods, and Future Directions". In: *IEEE Signal Processing Magazine* 37.3, pp. 50–60. DOI: 10.1109/MSP.2020.2975749.

Liao, Y., G. K. Smyth, and W. Shi (Apr. 2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". In: *Bioinformatics (Oxford, England)* 30.7, pp. 923–930. DOI: 10.1093/bioinformatics/btt656.

Lorenzi, M., M. Filippone, D. C. Alexander, and S. Ourselin (Apr. 2019a). "Disease Progression Modeling and Prediction through Random Effect Gaussian Processes and Time Transformation". In: *NeuroImage* 190, pp. 56–68. DOI: 10.1016/j.neuroimage.2017.08.059.

Lorenzi, M., M. Filippone, G. B. Frisoni, D. C. Alexander, and S. Ourselin (Apr. 2019b). "Probabilistic disease progression modeling to characterize diagnostic uncertainty: Application to staging and prediction in Alzheimer's disease". In: *NeuroImage* 190, pp. 56–68. DOI: 10.1016/j.neuroimage.2017.08.059.

Lowe, R., N. Shirley, M. Bleackley, S. Dolan, and T. Shafee (2017). "Transcriptomics technologies". In: *PLoS computational biology* 13.5, e1005457. DOI: 10.1371/journal.pcbi.1005457.

Mackenzie, I. R. A., P. Frick, and M. Neumann (Mar. 2014). "The neuropathology associated with repeat expansions in the C9ORF72 gene". In: *Acta Neuropathologica* 127.3, pp. 347–357. DOI: 10.1007/s00401-013-1232-4.

Mackenzie, I., R. Rademakers, and M. Neumann (2010). "TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia". In: *The Lancet Neurology* 9.10, pp. 995–1007. DOI: 10.1016/S1474-4422(10)70195-2.

Madabhushi, A. and J. Udupa (May 2005). "Interplay between intensity standardization and inhomogeneity correction in MR image processing". In: *IEEE Transactions on Medical Imaging* 24.5, pp. 561–576. DOI: 10.1109/TMI.2004.843256.

Magen, I., N. Yacovzada, J. D. Warren, C. Heller, I. Swift, Y. Bobeva, A. Malaspina, J. D. Rohrer, P. Fratta, and E. Hornstein (Jan. 2020). "Classification and prediction of frontotemporal dementia based on plasma microRNAs". In: *medRxiv*, p. 2020.01.22.20018408. DOI: 10.1101/2020.01.22.20018408.

Magen, I., N. S. Yacovzada, E. Yanowski, A. Coenen-Stass, J. Grosskreutz, C.-H. Lu, L. Greensmith, A. Malaspina, P. Fratta, and E. Hornstein (Nov. 2021). "Circulating miR-181 is a prognostic biomarker for amyotrophic lateral sclerosis". In: *Nature Neuroscience* 24.11, pp. 1534–1541. DOI: 10.1038/s41593-021-00936-z.

Marinescu, R. V., A. Eshaghi, M. Lorenzi, A. L. Young, N. P. Oxtoby, S. Garbarino, S. J. Crutch, D. C. Alexander, and Alzheimer's Disease Neuroimaging Initiative (May 2019). "DIVE: A spatiotemporal progression model of brain pathology in

neurodegenerative disorders". In: *NeuroImage* 192, pp. 166–177. DOI: 10.1016/j.neuroimage.2019.02.053.

Martin, M. (May 2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads". In: *EMBnet.journal* 17.1, pp. 10–12. DOI: 10.14806/ej.17.1.200.

Martins, D., A. Giacomel, S. C. Williams, F. Turkheimer, O. Dipasquale, and M. Veronese (Dec. 2021). "Imaging transcriptomics: Convergent cellular, transcriptomic, and molecular neuroimaging signatures in the healthy adult human brain". In: *Cell Reports* 37.13, p. 110173. DOI: 10.1016/j.celrep.2021.110173.

McCormick, M., X. Liu, L. Ibanez, J. Jomier, and C. Marion (2014). "ITK: enabling reproducible research and open science". In: *Frontiers in Neuroinformatics* 8.

McRobbie, D. W., E. A. Moore, M. J. Graves, and M. R. Prince (2006). *MRI from Picture to Proton*. 2nd ed. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511545405.

Meeter, L. H., L. D. Kaat, J. D. Rohrer, and J. C. van Swieten (July 2017). "Imaging and fluid biomarkers in frontotemporal dementia". In: *Nature Reviews Neurology* 13.7, pp. 406–419. DOI: 10.1038/nrneurol.2017.75.

Mehdipour Ghazi, M., M. Nielsen, A. Pai, M. J. Cardoso, M. Modat, S. Ourselin, and L. Sørensen (Apr. 2019). "Training recurrent neural networks robust to incomplete data: Application to Alzheimer's disease progression modeling". In: *Medical Image Analysis* 53, pp. 39–46. DOI: 10.1016/j.media.2019.01.004.

Mehdipour Ghazi, M., M. Nielsen, A. Pai, M. Modat, M. Jorge Cardoso, S. Ourselin, and L. Sørensen (Jan. 2021). "Robust parametric modeling of Alzheimer's disease progression". In: *NeuroImage* 225, p. 117460. DOI: 10.1016/j.neuroimage.2020.117460.

Modi, P. K., S. Jaiswal, and P. Sharma (Jan. 2016). "Regulation of Neuronal Cell Cycle and Apoptosis by MicroRNA 34a". In: *Molecular and Cellular Biology* 36.1, pp. 84–94. DOI: 10.1128/MCB.00589-15.

Montembeault, M., S. Sayah, D. Rinaldi, B. Le Toullec, A. Bertrand, A. Funkiewiez, D. Saracino, A. Camuzat, P. Couratier, M. Chouly, D. Hannequin, C. Aubier-Girard, F. Pasquier, X. Delbeuck, O. Colliot, B. Batrancourt, C. Azuar, R. Lévy, B. Dubois, I. Le Ber, R. Migliaccio, and PrevDemAls study group (Apr. 2020). "Cognitive inhibition impairments in presymptomatic C9orf72 carriers". In: *Journal of Neurology, Neurosurgery, and Psychiatry* 91.4, pp. 366–372. DOI: 10.1136/jnnp-2019-322242.

N, S., B. Paige, J.-W. van de Meent, A. Desmaison, N. Goodman, P. Kohli, F. Wood, and P. Torr (2017). "Learning Disentangled Representations with Semi-Supervised Deep Generative Models". In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc.

Neary, D., J. Snowden, and D. Mann (Nov. 2005). "Frontotemporal dementia". In: *The Lancet. Neurology* 4.11, pp. 771–780. DOI: 10.1016/S1474-4422(05)70223-4.

Nosek, B. A., C. R. Ebersole, A. C. DeHaven, and D. T. Mellor (Mar. 2018). "The preregistration revolution". In: *Proceedings of the National Academy of Sciences of the United States of America* 115.11, pp. 2600–2606. DOI: 10.1073/pnas.1708274114.

Nyul, L., J. Udupa, and X. Zhang (Feb. 2000). "New variants of a method of MRI scale standardization". In: *IEEE Transactions on Medical Imaging* 19.2, pp. 143–150. DOI: 10.1109/42.836373.

Oliveira, F. P. and J. M. R. Tavares (Jan. 2014). "Medical image registration: a review". In: *Computer Methods in Biomechanics and Biomedical Engineering* 17.2, pp. 73–93. DOI: 10.1080/10255842.2012.670855.

Orrù, G., W. Pettersson-Yeo, A. F. Marquand, G. Sartori, and A. Mechelli (Apr. 2012). "Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review". In: *Neuroscience and Biobehavioral Reviews* 36.4, pp. 1140–1152. DOI: 10.1016/j.neubiorev.2012.01.004.

Oxtoby, N. P., L.-A. Leyland, L. M. Aksman, G. E. C. Thomas, E. L. Bunting, P. A. Wijeratne, A. L. Young, A. Zarkali, M. M. X. Tan, F. D. Bremner, P. A. Keane, H. R. Morris, A. E. Schrag, D. C. Alexander, and R. S. Weil (Feb. 2021). "Sequence of clinical and neurodegeneration events in Parkinson's disease progression". In: *Brain* 144.3, pp. 975–988. DOI: 10.1093/brain/awaa461.

Oxtoby, N. P., A. L. Young, D. M. Cash, T. L. S. Benzinger, A. M. Fagan, J. C. Morris, R. J. Bateman, N. C. Fox, J. M. Schott, and D. C. Alexander (May 2018). "Data-driven models of dominantly-inherited Alzheimer's disease progression". In: *Brain* 141.5, pp. 1529–1544. DOI: 10.1093/brain/awy050.

Panman, J. L., V. Venkatraghavan, E. L. v. d. Ende, R. M. E. Steketee, L. C. Jiskoot, J. M. Poos, E. G. P. Dopper, et al. (Jan. 2021). "Modelling the cascade of biomarker changes in GRN-related frontotemporal dementia". In: *Journal of Neurology, Neurosurgery & Psychiatry*. DOI: 10.1136/jnnp-2020-323541.

Pasinelli, P. and R. H. Brown (Sept. 2006). "Molecular biology of amyotrophic lateral sclerosis: insights from genetics". In: *Nature Reviews Neuroscience* 7.9, pp. 710–723. DOI: 10.1038/nrn1971.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12.Oct, pp. 2825–2830.

Piscopo, P., M. Grasso, M. Puopolo, E. D'Acunto, G. Talarico, A. Crestini, M. Gasparini, R. Campopiano, S. Gambardella, A. E. Castellano, G. Bruno, M. A. Denti, and A. Confaloni (Jan. 2018). "Circulating miR-127-3p as a Potential Biomarker for Differential Diagnosis in Frontotemporal Dementia". In: *Journal of Alzheimer's Disease* 65.2, pp. 455–464. DOI: 10.3233/JAD-180364.

Popuri, K., E. Dowds, M. F. Beg, R. Balachandar, M. Bhalla, C. Jacova, A. Buller, P. Slack, P. Sengdy, R. Rademakers, D. Wittenberg, H. H. Feldman, I. R. Mackenzie, and G.-Y. R. Hsiung (2018). "Gray matter changes in asymptomatic C9orf72 and GRN mutation carriers". In: *NeuroImage. Clinical* 18, pp. 591–598. DOI: 10.1016/j.nicl.2018.02.017.

Potla, P., S. A. Ali, and M. Kapoor (Mar. 2021). "A bioinformatics approach to microRNA-sequencing analysis". In: *Osteoarthritis and Cartilage Open* 3.1, p. 100131. DOI: 10.1016/j.ocarto.2020.100131.

Pottier, C., T. A. Ravenscroft, M. Sanchez-Contreras, and R. Rademakers (2016). "Genetics of FTLD: overview and what else we can expect from genetic studies". In: *Journal of Neurochemistry* 138.S1, pp. 32–53. DOI: 10.1111/jnc.13622.

Prayitno, C.-R. Shyu, K. T. Putra, H.-C. Chen, Y.-Y. Tsai, K. S. M. T. Hossain, W. Jiang, and Z.-Y. Shae (Jan. 2021). "A Systematic Review of Federated Learning in the Healthcare Area: From the Perspective of Data Properties and Applications". In: *Applied Sciences* 11.23, p. 11191. DOI: 10.3390/app112311191.

Raheja, R., K. Regev, B. C. Healy, M. A. Mazzola, V. Beynon, F. von Glehn, A. Paul, C. Diaz-Cruz, T. Gholipour, B. I. Glanz, P. Kivisakk, T. Chitnis, H. L. Weiner, J. D. Berry, and R. Gandhi (Aug. 2018). "Correlating serum microRNAs and clinical parameters in Amyotrophic lateral sclerosis". In: *Muscle & nerve* 58.2, pp. 261–269. DOI: 10.1002/mus.26106.

Ramirez, J., M. F. Holmes, C. J. M. Scott, M. Ozzoude, S. Adamo, G. M. Szilagyi, M. Goubran, F. Gao, S. R. Arnott, J. M. Lawrence-Dewar, D. Beaton, S. C. Strother, D. P. Munoz, M. Masellis, R. H. Swartz, R. Bartha, S. Symons, S. E. Black, and T. O. I. (2020). "Ontario Neurodegenerative Disease Research Initiative (ONDRI): Structural MRI Methods and Outcome Measures". In: *Frontiers in Neurology* 11.

Rascovsky, K., J. R. Hodges, D. Knopman, M. F. Mendez, J. H. Kramer, J. Neuhaus, J. C. van Swieten, et al. (Sept. 2011). "Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia". In: *Brain* 134.9, pp. 2456–2477. DOI: 10.1093/brain/awr179.

Ratnavalli, E., C. Brayne, K. Dawson, and J. R. Hodges (June 2002). "The prevalence of frontotemporal dementia". In: *Neurology* 58.11, pp. 1615–1621. DOI: 10.1212/wnl.58.11.1615.

Regev, K., B. C. Healy, F. Khalid, A. Paul, R. Chu, S. Tauhid, S. Tummala, C. Diaz-Cruz, R. Raheja, M. A. Mazzola, F. v. Glehn, P. Kivisakk, S. L. Dupuy, G. Kim, T. Chitnis, H. L. Weiner, R. Gandhi, and R. Bakshi (Mar. 2017). "Association Between Serum MicroRNAs and Magnetic Resonance Imaging Measures of Multiple Sclerosis Severity". In: *JAMA Neurology* 74.3, pp. 275–285. DOI: 10.1001/jamaneurol.2016.5197.

Renton, A. E., E. Majounie, A. Waite, J. Simón-Sánchez, S. Rollinson, J. R. Gibbs, et al. (Oct. 2011). "A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD". In: *Neuron* 72.2, pp. 257–268. DOI: 10.1016/j.neuron.2011.09.010.

Rieke, N., J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso (Sept. 2020). "The future of digital health with federated learning". In: *NPJ Digital Medicine* 3, p. 119. DOI: 10.1038/s41746-020-00323-1.

Rivero-Santana, A., D. Ferreira, L. Perestelo-Pérez, E. Westman, L.-O. Wahlund, A. Sarría, and P. Serrano-Aguilar (2017). "Cerebrospinal Fluid Biomarkers for the Differential Diagnosis between Alzheimer's Disease and Frontotemporal Lobar Degeneration: Systematic Review, HSROC Analysis, and Confounding Factors". In: *Journal of Alzheimer's disease: JAD* 55.2, pp. 625–644. DOI: 10.3233/JAD-160366.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth (Jan. 2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". In: *Bioinformatics (Oxford, England)* 26.1, pp. 139–140. DOI: 10.1093/bioinformatics/btp616.

Robinson, M. D. and A. Oshlack (Mar. 2010). "A scaling normalization method for differential expression analysis of RNA-seq data". In: *Genome Biology* 11.3, R25. DOI: 10.1186/gb-2010-11-3-r25.

Rohrer, J. D., J. M. Nicholas, D. M. Cash, J. van Swieten, E. Dopper, L. Jiskoot, R. van Minkelen, S. A. Rombouts, M. J. Cardoso, S. Clegg, M. Espak, S. Mead, D. L. Thomas, E. De Vita, M. Masellis, et al. (Mar. 2015). "Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: a cross-sectional analysis". In: *The Lancet. Neurology* 14.3, pp. 253–262. DOI: 10.1016/S1474-4422(14)70324-2.

Rohrer, J. D., I. O. Woollacott, K. M. Dick, E. Brotherhood, E. Gordon, A. Fellows, J. Toombs, R. Druyeh, M. J. Cardoso, S. Ourselin, J. M. Nicholas, N. Norgren, S. Mead, U. Andreasson, K. Blennow, J. M. Schott, N. C. Fox, J. D. Warren, and H. Zetterberg (Sept. 2016). "Serum neurofilament light chain protein is a measure of disease intensity in frontotemporal dementia". In: *Neurology* 87.13, pp. 1329–1336. DOI: 10.1212/WNL.0000000000003154.

Routier, A., N. Burgos, M. Díaz, M. Bacci, S. Bottani, O. El-Rifai, S. Fontanella, P. Gori, J. Guillon, A. Guyot, R. Hassanaly, T. Jacquemont, P. Lu, A. Marcoux, T. Moreau, J. Samper-González, M. Teichmann, E. Thibeau-Sutre, G. Vaillant, J. Wen, A. Wild, M.-O. Habert, S. Durrleman, and O. Colliot (2021). "Clinica: An Open-Source Software Platform for Reproducible Clinical Neuroscience Studies". In: *Frontiers in Neuroinformatics* 15.

Sadananthan, S. A., W. Zheng, M. W. L. Chee, and V. Zagorodnov (Jan. 2010). "Skull stripping using graph cuts". In: *NeuroImage* 49.1, pp. 225–239. DOI: 10.1016/j.neuroimage.2009.08.050.

Saracino, D., K. Dorgham, A. Camuzat, D. Rinaldi, A. Rametti-Lacroux, M. Houot, F. Clot, P. Martin-Hardy, L. Jornea, C. Azuar, R. Migliaccio, F. Pasquier, P. Couratier, S. Auriacombe, M. Sauvée, C. Boutoleau-Bretonnière, J. Pariente, M. Didic, D. Hannequin, D. Wallon, t. F. R. N. o. Ftd/Ftd-Als, t. P.-D. a. P.-P. s. Groups, O. Colliot, B. Dubois, A. Brice, R. Levy, S. Forlani, and I. L. Ber (Dec. 2021). "Plasma NfL levels and longitudinal change rates in C9orf72 and GRN-associated diseases: from tailored references to clinical applications". In: *Journal of Neurology, Neurosurgery & Psychiatry* 92.12, pp. 1278–1288. DOI: 10.1136/jnnp-2021-326914.

Schiratti, J.-B., S. Allassonnière, O. Colliot, and S. Durrleman (2015). "Learning spatiotemporal trajectories from manifold-valued longitudinal data". In: *Advances in Neural Information Processing Systems*. Vol. 28. Curran Associates, Inc.

— (2017). "A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations". In: *Journal of Machine Learning Research* 18.133, pp. 1–33.

Schneider, R., P. McKeever, T. Kim, C. Graff, J. C. van Swieten, A. Karydas, A. Boxer, H. Rosen, B. L. Miller, R. Laforce, D. Galimberti, M. Masellis, B. Borroni, Z. Zhang, L. Zinman, J. D. Rohrer, M. C. Tartaglia, J. Robertson, and Genetic FTD Initiative (GENFI) (2018). "Downregulation of exosomal miR-204-5p and miR-632 as a biomarker for FTD: a GENFI study". In: *Journal of Neurology, Neurosurgery, and Psychiatry* 89.8, pp. 851–858. DOI: 10.1136/jnnp-2017-317492.

Seelaar, H., J. D. Rohrer, Y. A. L. Pijnenburg, N. C. Fox, and J. C. v. Swieten (May 2011). "Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review". In: *Journal of Neurology, Neurosurgery & Psychiatry* 82.5, pp. 476–486. DOI: 10.1136/jnnp.2010.212225.

Sheinerman, K. S., J. B. Toledo, V. G. Tsivinsky, D. Irwin, M. Grossman, D. Weintraub, H. I. Hurtig, A. Chen-Plotkin, D. A. Wolk, L. F. McCluskey, L. B. Elman, J. Q. Trojanowski, and S. R. Umansky (Nov. 2017). "Circulating brain-enriched microRNAs as novel biomarkers for detection and differentiation of neurodegenerative diseases". In: *Alzheimer's Research & Therapy* 9.1, p. 89. DOI: 10.1186/s13195-017-0316-0.

Sled, J. G., A. P. Zijdenbos, and A. C. Evans (Feb. 1998). "A nonparametric method for automatic correction of intensity nonuniformity in MRI data". In: *IEEE transactions on medical imaging* 17.1, pp. 87–97. DOI: 10.1109/42.668698.

Smith, T., A. Heger, and I. Sudbery (2017). "UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy". In: *Genome Research* 27.3, pp. 491–499. DOI: 10.1101/gr.209601.116.

Sohel, M. H. (Dec. 2016). "Extracellular/Circulating MicroRNAs: Release Mechanisms, Functions and Challenges". In: *Achievements in the Life Sciences* 10.2, pp. 175–186. DOI: 10.1016/j.als.2016.11.007.

Soliman, R., N. O. Mousa, H. R. Rashed, R. R. Moustafa, N. Hamdi, A. Osman, and N. Fahmy (Sept. 2021). "Assessment of diagnostic potential of some circulating microRNAs in Amyotrophic Lateral Sclerosis Patients, an Egyptian study". In: *Clinical Neurology and Neurosurgery* 208, p. 106883. DOI: 10.1016/j.clineuro.2021.106883.

Stark, R., M. Grzelak, and J. Hadfield (Nov. 2019). "RNA sequencing: the teenage years". In: *Nature Reviews Genetics* 20.11, pp. 631–656. DOI: 10.1038/s41576-019-0150-2.

Street, K., D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, and S. Dudoit (June 2018). "Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics". In: *BMC Genomics* 19.1, p. 477. DOI: 10.1186/s12864-018-4772-0.

Su, Z., Y. Zhang, T. F. Gendron, P. O. Bauer, J. Chew, W.-Y. Yang, E. Fostvedt, K. Jansen-West, V. V. Belzil, P. Desaro, A. Johnston, K. Overstreet, S.-Y. Oh, P. K. Todd, J. D. Berry, M. E. Cudkowicz, B. F. Boeve, D. Dickson, M. K. Floeter, B. J. Traynor, C. Morelli, A. Ratti, V. Silani, R. Rademakers, R. H. Brown, J. D. Rothstein, K. B. Boylan, L. Petrucelli, and M. D. Disney (Sept. 2014). "Discovery of a biomarker and lead small molecules to target r(GGGGCC)-associated defects in c9FTD/ALS". In: *Neuron* 83.5, pp. 1043–1050. DOI: 10.1016/j.neuron.2014.07.041.

Sunkin, S. M., L. Ng, C. Lau, T. Dolbeare, T. L. Gilbert, C. L. Thompson, M. Hawrylycz, and C. Dang (Jan. 2013). "Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system". In: *Nucleic Acids Research* 41.Database issue, pp. D996–D1008. DOI: 10.1093/nar/gks1042.

Symms, M., H. R. Jäger, K. Schmierer, and T. A. Yousry (Sept. 2004). "A review of structural magnetic resonance neuroimaging". In: *Journal of Neurology, Neurosurgery & Psychiatry* 75.9, pp. 1235–1244. DOI: 10.1136/jnnp.2003.032714.

Takahashi, I., Y. Hama, M. Matsushima, M. Hirotani, T. Kano, H. Hohzen, I. Yabe, J. Utsumi, and H. Sasaki (Oct. 2015). "Identification of plasma microRNAs as a biomarker of sporadic Amyotrophic Lateral Sclerosis". In: *Molecular Brain* 8.1, p. 67. DOI: 10.1186/s13041-015-0161-7.

Tartaglia, M. C., H. J. Rosen, and B. L. Miller (Jan. 2011). "Neuroimaging in Dementia". In: *Neurotherapeutics* 8.1, pp. 82–92. DOI: 10.1007/s13311-010-0012-2.

Tasca, E., V. Pegoraro, A. Merico, and C. Angelini (Jan. 2016). "Circulating microRNAs as biomarkers of muscle differentiation and atrophy in ALS". In: *Clinical Neuropathology* 35.01, pp. 22–30. DOI: 10.5414/NP300889.

Thompson, P. M., N. G. Martin, and M. J. Wright (Aug. 2010). "Imaging genomics". In: *Current Opinion in Neurology* 23.4, pp. 368–373. DOI: 10.1097/WCO.0b013e32833b764c.

Toivonen, J. M., R. Manzano, S. Oliván, P. Zaragoza, A. García-Redondo, and R. Osta (Feb. 2014). "MicroRNA-206: A Potential Circulating Biomarker Candidate for Amyotrophic Lateral Sclerosis". In: *PLoS ONE* 9.2. Ed. by H. Cai, e89065. DOI: 10.1371/journal.pone.0089065.

Turner, M. R., A. Al-Chalabi, A. Chio, O. Hardiman, M. C. Kiernan, J. D. Rohrer, J. Rowe, W. Seeley, and K. Talbot (Dec. 2017). "Genetic screening in sporadic ALS and FTD". In: *Journal of Neurology, Neurosurgery & Psychiatry* 88.12, pp. 1042–1044. DOI: 10.1136/jnnp-2017-315995.

Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee (June 2010). "N4ITK: Improved N3 Bias Correction". In: *IEEE Transactions on Medical Imaging* 29.6, pp. 1310–1320. DOI: 10.1109/TMI.2010.2046908.

Varcianna, A., M. A. Myszczynska, L. M. Castelli, B. O'Neill, Y. Kim, J. Talbot, S. Nyberg, I. Nyamali, P. R. Heath, M. J. Stopford, G. M. Hautbergue, and L. Ferraiuolo (Feb. 2019). "Micro-RNAs secreted through astrocyte-derived extracellular vesicles cause neuronal network degeneration in C9orf72 ALS". In: *EBioMedicine* 40, pp. 626–635. DOI: 10.1016/j.ebiom.2018.11.067.

Varoquaux, G. (Oct. 2018). "Cross-validation failure: Small sample sizes lead to large error bars". In: *NeuroImage*. New advances in encoding and decoding of brain signals 180, pp. 68–77. DOI: 10.1016/j.neuroimage.2017.06.061.

Venkatraghavan, V., E. E. Bron, W. J. Niessen, and S. Klein (Feb. 2019). "Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling". In: *NeuroImage* 186, pp. 518–532. DOI: 10.1016/j.neuroimage.2018.11.024.

Vergallo, A., S. Lista, Y. Zhao, P. Lemercier, S. J. Teipel, M.-C. Potier, M.-O. Habert, B. Dubois, W. J. Lukiw, and H. Hampel (Jan. 2021). "MiRNA-15b and miRNA-125b are associated with regional Abeta-PET and FDG-PET uptake in cognitively normal individuals with subjective memory complaints". In: *Translational Psychiatry* 11.1, pp. 1–11. DOI: 10.1038/s41398-020-01184-8.

Vlachos, I. S., K. Zagganas, M. D. Paraskevopoulou, G. Georgakilas, D. Karagkouni, T. Vergoulis, T. Dalamagas, and A. G. Hatzigeorgiou (July 2015). "DIANA-miRPath v3.0: deciphering microRNA function with experimental support". In: *Nucleic Acids Research* 43.Web Server issue, W460–W466. DOI: 10.1093/nar/gkv403.

Vovk, U., F. Pernus, and B. Likar (Mar. 2007). "A review of methods for correction of intensity inhomogeneity in MRI". In: *IEEE transactions on medical imaging* 26.3, pp. 405–421. DOI: 10.1109/TMI.2006.891486.

Waller, R., E. F. Goodall, M. Milo, J. Cooper-Knock, M. Da Costa, E. Hobson, M. Kazoka, H. Wollff, P. R. Heath, P. J. Shaw, and J. Kirby (July 2017). "Serum miRNAs miR-206, 143-3p and 374b-5p as potential biomarkers for amyotrophic lateral sclerosis (ALS)". In: *Neurobiology of Aging* 55, pp. 123–131. DOI: 10.1016/j.neurobiolaging.2017.03.027.

Wang, L.-J., S.-C. Li, H.-C. Kuo, W.-J. Chou, M.-J. Lee, M.-C. Chou, H.-H. Tseng, C.-F. Hsu, S.-Y. Lee, and W.-C. Lin (Dec. 2020). "Gray matter volume and microRNA levels in patients with attention-deficit/hyperactivity disorder". In: *European Archives of Psychiatry and Clinical Neuroscience* 270.8, pp. 1037–1045. DOI: 10.1007/s00406-019-01032-x.

Wang, Z., M. Gerstein, and M. Snyder (Jan. 2009). "RNA-Seq: a revolutionary tool for transcriptomics". In: *Nature Reviews Genetics* 10.1, pp. 57–63. DOI: 10.1038/nrg2484.

Watson, C. N., A. Belli, and V. Di Pietro (2019). "Small Non-coding RNAs: New Class of Biomarkers and Potential Therapeutic Targets in Neurodegenerative Disease". In: *Frontiers in Genetics* 10, p. 364. DOI: 10.3389/fgene.2019.00364.

Wen, J., H. Zhang, D. C. Alexander, S. Durrleman, A. Routier, D. Rinaldi, M. Houot, P. Couratier, D. Hannequin, F. Pasquier, J. Zhang, O. Colliot, I. Le Ber, A. Bertrand, and Predict to Prevent Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis (PREV-DEMALS) Study Group (Apr. 2019). "Neurite density is

reduced in the presymptomatic phase of C9orf72 disease". In: *Journal of Neurology, Neurosurgery, and Psychiatry* 90.4, pp. 387–394. DOI: 10.1136/jnnp-2018-318994.

Whitwell, J. L., B. F. Boeve, S. D. Weigand, M. L. Senjem, J. L. Gunter, M. C. Baker, M. DeJesus-Hernandez, D. S. Knopman, Z. K. Wszolek, R. C. Petersen, R. Rademakers, C. R. Jack, and K. A. Josephs (May 2015). "Brain atrophy over time in genetic and sporadic frontotemporal dementia: a study of 198 serial magnetic resonance images". In: *European Journal of Neurology* 22.5, pp. 745–752. DOI: 10.1111/ene.12675.

Whitwell, J. L. and K. A. Josephs (Jan. 2012). "Neuroimaging in frontotemporal lobar degeneration–predicting molecular pathology". In: *Nature Reviews. Neurology* 8.3, pp. 131–142. DOI: 10.1038/nrneurol.2012.7.

Wijeratne, P. A., E. B. Johnson, S. Gregory, N. Georgiou-Karistianis, J. S. Paulsen, R. I. Scahill, S. J. Tabrizi, and D. C. Alexander (Aug. 2021). "A Multi-Study Model-Based Evaluation of the Sequence of Imaging and Clinical Biomarker Changes in Huntington's Disease". In: *Frontiers in Big Data* 4, p. 662200. DOI: 10.3389/fdata.2021.662200.

Xiang, S., L. Yuan, W. Fan, Y. Wang, P. M. Thompson, and J. Ye (Nov. 2014). "Bi-level Multi-Source Learning for Heterogeneous Block-wise Missing Data". In: *NeuroImage* 102 Pt 1, pp. 192–206. DOI: 10.1016/j.neuroimage.2013.08.015.

Xu, Q., Y. Zhao, X. Zhou, J. Luan, Y. Cui, and J. Han (Feb. 2018). "Comparison of the extraction and determination of serum exosome and miRNA in serum and the detection of miR-27a-3p in serum exosome of ALS patients". In: *Intractable & Rare Diseases Research* 7.1, pp. 13–18. DOI: 10.5582/irdr.2017.01091.

Xu, Y., X. Liu, L. Pan, X. Mao, H. Liang, G. Wang, and T. Chen (2021). "Explainable Dynamic Multimodal Variational Autoencoder for the Prediction of Patients with Suspected Central Precocious Puberty". In: *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1. DOI: 10.1109/JBHI.2021.3103271.

Young, A. L., N. P. Oxtoby, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, D. C. Alexander, and on behalf of the Alzheimer's Disease Neuroimaging Initiative (Sept. 2014). "A data-driven model of biomarker changes in sporadic Alzheimer's disease". In: *Brain* 137.9, pp. 2564–2577. DOI: 10.1093/brain/awu176.

Zhang, X., Y. Feng, W. Chen, X. Li, A. V. Faria, Q. Feng, and S. Mori (2019). "Linear Registration of Brain MRI Using Knowledge-Based Multiple Intermediator Libraries". In: *Frontiers in Neuroscience* 13.

Zhuo, J. and R. P. Gullapalli (Jan. 2006). "MR Artifacts, Safety, and Quality Control". In: *RadioGraphics* 26.1, pp. 275–297. DOI: 10.1148/rg.261055134.

Zinn, P. O., B. Mahajan, B. Majadan, P. Sathyan, S. K. Singh, S. Majumder, F. A. Jolesz, and R. R. Colen (2011). "Radiogenomic mapping of edema/cellular invasion MRI-phenotypes in glioblastoma multiforme". In: *PloS One* 6.10, e25451. DOI: 10.1371/journal.pone.0025451.

Zmyslowska, A., M. Stanczak, Z. Nowicka, A. Waszczykowska, D. Baranska, W. Fendler, M. Borowiec, and W. Młynarski (Nov. 2020). "Serum microRNA as indicators of Wolfram syndrome's progression in neuroimaging studies". In: *BMJ Open Diabetes Research and Care* 8.2, e001379. DOI: 10.1136/bmjdrc-2020-001379.