



Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Estatística

Dissertação de Mestrado

Aplicação de Modelos Gráficos em Dados Genômicos da Doença de Alzheimer

por

Rafael da Silva Lins

Brasília

2021

Aplicação de Modelos Gráficos em Dados Genômicos da Doença de Alzheimer

por

Rafael da Silva Lins

Dissertação apresentada ao Departamento de Estatística da Universidade de Brasília, como requisito parcial para obtenção do título de Mestre em Estatística.

Orientadora: Prof^a. Joanlise M. de L. Andrade

Coorientador: Prof. George F. von Borries

Brasília

2021

Dissertação submetida ao Programa de Pós-Graduação em Estatística do Departamento de Estatística da Universidade de Brasília como parte dos requisitos para a obtenção do grau de Mestre em Estatística.

Texto aprovado, em 11 de novembro de 2021, por:

Prof^a. Joanlise Marco de Leon Andrade

Orientadora, EST/UnB

Prof. George Freitas von Borries

Coorientador, EST/UnB

Prof. Eduardo Yoshio Nakano

Membro interno, EST/UnB

Prof^a. Suely Ruiz Giolo

Membro externo, EST/UFPR

Prof. Raul Yukihiro Matsushita

Suplente, EST/UnB

A Deus
A meus pais
A meus irmãos

Resumo

A doença de Alzheimer (DA) é uma doença neurodegenerativa e progressiva que muitas vezes causa danos cognitivos e funcionais irreversíveis ao cérebro, envolvendo a perda progressiva de memória, declínio cognitivo e eventual perda de funções corporais. Um caminho para compreender melhor a doença envolve o estudo das relações entre os mecanismos moleculares do cérebro e seus respectivos resultados funcionais a partir do contraste entre os perfis de expressão gênica de cérebros saudáveis e de cérebros doentes. Nesse contexto, diversos estudos já foram realizados para a análise de genes diferencialmente expressos, contudo, tal tipo de análise não considera em si que os genes podem operar em conjunto na presença da DA. Nesse sentido, mais recentemente a análise de redes gênicas também têm sido utilizada para analisar as relações de dependência estatística entre os genes e, a partir daí, compreender melhor a DA e desenvolver novos métodos de diagnóstico ou intervenções na doença. Desse modo, o presente trabalho busca explorar a estrutura de dependência estatística em dados de expressão gênica da DA. Para tanto, identificam-se os padrões alterados entre as redes de duas condições (ou grupos), a saber, amostras de cérebros saudáveis (ditos controles) e de cérebros com a DA. Por fim, estuda-se o potencial dessa abordagem para identificar os grupos de genes relacionamente afetados pela DA e, portanto, fornecer informações valiosas sobre os mecanismos biológicos alterados na presença da doença.

Palavras-chave: doença de alzheimer, expressão gênica, modelo gráfico probabilístico, co-expressão diferencial, vias biológicas

Abstract

Alzheimer's disease (AD) is a neurodegenerative and progressive disease, which often causes irreversible cognitive and functional damage to the brain, involving progressive memory loss, cognitive decline and eventual loss of bodily functions. One way to better understand the disease involves studying the relationships between the molecular mechanisms of the brain and their respective functional outcomes by contrasting the gene expression profiles of healthy brains and diseased brains. In this context, several studies have been carried out to analyze differentially expressed genes, however, this type of analysis does not consider that genes can operate in pathways, or networks, in the presence of AD. In this sense, more recently, gene network analysis has also been used to analyze the statistical dependence relationships between genes and, from there, to better understand AD and develop new diagnostic methods or interventions in the disease. Thus, this work aims to explore the structure of statistical dependence on AD gene expression data. For this purpose, the altered patterns between the networks of two conditions (or groups) are identified, namely, samples of healthy brains (controls) and brains with AD. Finally, we study the potential of this approach to identify groups of genes relationally affected by AD and, therefore, provide valuable information about the altered biological mechanisms in the presence of the disease.

Keywords: alzheimer's disease, gene expression, probabilistic graphical model, differential coexpression, biological pathways

Sumário

1	Introdução	1
2	Dados de Chips de Microarranjo para Expressão Gênica	4
2.1	Pré-Processamento	6
2.1.1	GC-RMA (do inglês <i>GeneChip Robust Multi-array Average</i>)	6
2.1.2	Normalização Quantílica	7
2.1.3	Mediana Polida	8
3	Redução de Dimensionalidade para Dados de Expressão Gênica	10
3.1	Seleção de Transcritos (ou Genes) Diferencialmente Expressos	10
3.1.1	Filtragem por nível de expressão	11
3.1.2	<i>Fold Change</i> (FC)	11
3.1.3	Teste estatístico	12
3.2	Agrupamento de Transcritos (ou Genes) por Nível de Expressão	13
3.2.1	Agrupamento Particional Baseado em P-valores (PPCLUST)	13
3.2.2	Índice de Rand Ajustado	17
4	Análise de Redes e de Vias Biológicas	19
4.1	Redes de Co-Expressão Diferencial	19
4.2	Redes de Dependência Condicional (Modelos Gráficos Probabilísticos, MGPs)	20
4.2.1	MGPs Não Direcionados	22

4.2.2	MGPs Não Direcionados Gaussianos	23
4.3	Estimação Esparsa da Matriz de Precisão	24
4.3.1	Métodos baseado na log-verossimilhança penalizada	24
4.3.2	Métodos baseados em otimização restrita	26
4.4	Análise de Vias Biológicas (<i>Reactome Pathway Analysis</i>)	28
5	Aplicação em Dados de Expressão Gênica da Doença de Alzheimer (DA)	30
5.1	Estimação das Expressões Gênicas e Análise Exploratória	30
5.1.1	Pré-Processamento	32
5.1.2	Controle de qualidade	32
5.2	Seleção de Transcritos (ou Genes) Diferencialmente Expressos	39
5.2.1	Métodos paramétrico e não paramétrico	39
5.2.2	PPCLUST	44
5.2.3	Normalidade dos Transcritos Selecionados	46
5.3	Análise de Redes e de Vias Biológicas	48
5.3.1	Análise de Co-expressão Diferencial e de Vias Biológicas	48
5.3.2	Análise de Dependência Condicional e de Vias Biológicas	61
6	Discussão e Conclusão	71
A	Apêndice	74
	Referências Bibliográficas	109

Abreviações e Siglas

ARI	<i>Adjusted Rand Index</i>
ASR	análise de super-representação
BH	Benjamini-Hochberg
cDNA	DNA complementar
CE	córtex entorrinal
CLIME	<i>Constrained LI-minimization for Inverse Matrix Estimation</i>
CP	cingulado posterior
CRAN	<i>Comprehensive R Archive Network</i>
CVP	córtex visual primário
DA	doença de Alzheimer
DCGA	<i>Differential Gene Correlation Analysis</i>
DE	diferencialmente expresso
DNA	molécula formadora do genoma (ácido disoxirribonucleico)
FC	<i>Fold-Change</i>
FDR	<i>False Discovery Rate</i>
GC-RMA	<i>GeneChip Robust Multi-array Average</i>
GEO	<i>Gene Expression Omnibus</i>
GFS	giro frontal superior
Glasso	<i>Graphical lasso</i>
GPL	<i>GEO Platform</i>

GTM	giro temporal medial
HDLSS	<i>High Dimensional Low Sample Size</i>
HIP	hipocampo
MAS	<i>Microarray Analysis Suite</i>
MGP	Modelo Gráfico Probabilístico
MM	<i>Mismatch</i>
mRNA	RNA mensageiro
MSE	<i>Mean Squared Error</i>
MTS	<i>Mean Square for Treatments</i>
NSB	<i>Nonspecific Binding</i>
PM	<i>Perfect Match</i>
ReactomePA	<i>Reactome Pathway Analysis</i>
RI	<i>Rand Index</i>
RMA	<i>Robust Multi-array Average</i>
RNA	ácido ribonucleico
SCIO	<i>Sparse Column-Wise Inverse Operator</i>
TIGER	<i>Tuning-Insensitive Graph Estimation and Regression</i>
UG	<i>Undirected Graph</i>

Capítulo 1

Introdução

A doença de Alzheimer (DA) é uma doença neurodegenerativa e progressiva que causa danos cognitivos e funcionais irreversíveis ao cérebro (Goedert e Spillantini, 2006). A DA é caracterizada pela perda progressiva de memória, pelo declínio cognitivo, pela eventual perda de funções corporais e, em um estágio final, à morte. Estima-se que 50 milhões de pessoas estão vivendo com demência em todo o mundo, sendo a DA a principal causa de demência. Prevê-se que este número aumentará nos próximos anos devido ao envelhecimento da população mundial (Association, 2019; Organization, 2019).

Para prevenir uma crise de saúde pública, um grande esforço de pesquisa tem sido dedicado para a melhor compreensão da DA. Embora existam algumas hipóteses populares acerca da geração e do desenvolvimento da doença, a DA está longe de ser totalmente compreendida (Wang e Liu, 2019). As evidências atuais sugerem que o risco de desenvolver a DA é influenciado por fatores genéticos e ambientais. Sendo assim, são necessárias novas hipóteses acerca dos mecanismos biológicos ligados à DA, as quais podem contribuir para o aprimoramento de métodos para o diagnóstico e intervenção contra a doença (Cummings et al., 2020).

Um caminho para compreender as relações entre os mecanismos moleculares do cérebro e seus respectivos resultados funcionais consiste em analisar o perfil de expressão gênica de cérebros saudáveis e de cérebros doentes. Em termos práticos, os genes controlam as funções

de cada célula do nosso organismo. Por exemplo, alguns genes por meio de sua expressão podem determinar características básicas como a cor dos olhos e do cabelo. Outros genes, cujas expressões encontram-se desreguladas, podem estar ligados ao desenvolvimento de certas doenças tal como o Alzheimer. Tais dados de expressão gênica são obtidos pela tecnologia de microarranjo (do inglês *microarray*), a qual permite estimar a expressão de milhares de genes simultaneamente.

O presente trabalho utiliza dados brutos de microarranjo¹ de seis regiões anatomicamente distintas do cérebro de indivíduos *post mortem* portadores da DA, bem como amostras controles de indivíduos saudáveis similares aos casos em idade, sexo e etnia, as quais atendem aos critérios neuropatológicos da DA para indivíduos que estavam cognitivamente intactos no momento da morte. As regiões de interesse do cérebro compreendidas nos dados foram originalmente selecionadas com base em pesquisas sobre a DA (Liang et al., 2007a; Liang et al., 2008).

O objetivo principal deste trabalho consiste na aplicação de procedimentos associados ao preparo de dados de microarranjo e a investigação de redes para explorar a estrutura de dependência estatística em dados de expressão gênica da DA. Nesse contexto, utiliza-se a análise diferencial, a qual examina os padrões alterados entre redes de duas condições (ou grupos), a saber, amostras de cérebros saudáveis (ditos controles) e de cérebros com a DA. Assim sendo, estuda-se o potencial dessa abordagem para identificar os grupos de genes relacionalmente afetados pela DA e, portanto, fornecer informações valiosas sobre os mecanismos biológicos alterados na presença da doença.

O Capítulo 2 revisa os métodos de pré-processamento convencionais para dados brutos de microarranjo Affymetrix. Em seguida, o Capítulo 3 trata da seleção de transcritos (ou genes) diferencialmente expressos e o agrupamento dos transcritos com base no algoritmo PPCLUST. O Capítulo 4 apresenta uma revisão sobre análise de redes, incluindo redes de co-expressão diferencial, modelos gráficos probabilísticos, estimação da matriz de precisão, como também, a análise de vias biológicas (*Reactome Pathway Analysis*). Por fim, o Capítulo 5 apresenta os

¹especificamente, a plataforma Affymetrix Human Genome U133A Array (GPL570)

resultados do pré-processamento dos dados, da seleção dos transcritos diferencialmente expressos, dos transcritos diferencialmente co-expressos, da dependência condicional e da análise das vias biológicas, especificamente das vias mais significativas. Além disso, o Apêndice A dispõe das tabelas com todas as vias biológicas apresentadas no Capítulo 5, especificando os genes identificados a fim de fornecer insumo para estudos futuros.

Capítulo 2

Dados de Chips de Microarranjo para Expressão Gênica

A tecnologia de chips de microarranjo permite a medição simultânea das expressões¹ de milhares de genes em diferentes tecidos de um determinado organismo em um mesmo experimento (Hardin, 2007). Um microarranjo é formado por milhares de pontos microscópicos chamados de *spots*, cada um contendo um grande número de sondas idênticas (segmentos de DNA, cDNA ou oligonucleotídeos) que são fixadas à uma superfície sólida, usualmente uma lâmina de vidro ou de silício (Chaudhuri, 2005).



Figura 2.1: GeneChip Human Genome U133A 2.0 Array. Fonte: www.thermofisher.com.

¹processo pelo qual o mRNA é medido e a quantidade de proteínas produzidas é estimada.

Este trabalho se restringe ao uso de dados obtidos usando um microarranjo de oligonucleotídeos. Alguns dos microarranjos de oligonucleotídeos disponíveis no mercado são, por exemplo, as plataformas Affymetrix (G., 2002) e a Illumina (*Illumina*). A Figura 2.1 mostra a imagem de um chip de microarranjo Affymetrix.

Cada transcrito é representado por um conjunto (tipicamente de 11 a 20 pares) de sondas oligonucleotídicas, sendo que mais de um transcrito pode envolver o mesmo gene. Cada par de sondas consiste em duas cadeias curtas de nucleotídeos com 25 pares de bases, denominadas oligonucleotídeos (Lipshutz et al., 1999). Uma sonda com a sequência nucleotídea idêntica ao transcrito, denominada *perfect match* (PM) e, outra sonda com uma mudança nucleotídea na 13ª base, denominada *mismatch* (MM). As sondas MM são projetadas para estimar o sinal de fundo (do inglês *background signal*) das sondas PM correspondentes. A Figura 2.2 ilustra a composição de um microarranjo Affymetrix.

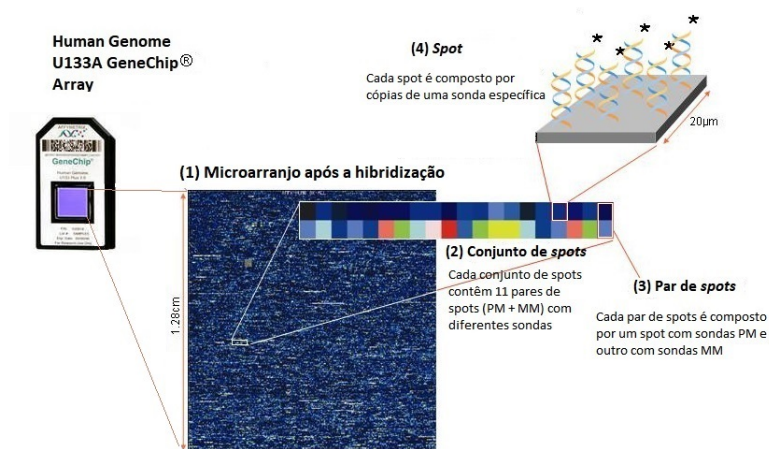


Figura 2.2: Esquema representativo de um microarranjo Affymetrix. Fonte: adaptado de David Givol, Weizmann Institute of Science (www.weizmann.ac.il).

O princípio básico da técnica consiste na hibridização por complementariedade de bases de ácido nucleico, processo no qual cada sonda tende a se ligar apenas à sua sequência complementar de nucleotídeos (Jaluria et al., 2007). Tal processo ocorre entre a sonda e seu DNA complementar (cDNA) correspondente, o qual é obtido a partir do seu respectivo RNA men-

sageiro (mRNA) da amostra, representando assim um único transcrito do genoma. Quando a amostra de interesse contém muitas cópias do mRNA referente a uma dada sequência de DNA, muitas ligações ocorrerão, indicando que o gene correspondente ao mRNA transcrito é possivelmente positivamente expresso, pois assume-se que o número de moléculas de mRNA seja proporcional ao nível de expressão gênica.

Antes do processo de hibridização², as amostras são “marcadas” com químicos fluorescentes sensíveis à luz, possibilitando assim medir a abundância dos transcritos presentes na amostra. Após a hibridização, um *scanner* mede a emissão de fluorescência dos *spots*. As intensidades da fluorescência representam as quantidades de cDNA hibridizado nos *spots* permitindo a estimativa do nível de expressão dos respectivos genes na amostra (Brown e Botstein, 1999).

2.1 Pré-Processamento

Os dados “brutos” obtidos por um experimento com microarranjo Affymetrix são armazenados em um arquivo no formato CEL. Esse arquivo armazena os valores numéricos das intensidades de cores emitidas pelos *spots* a partir do escaneamento do microarranjo. O pré-processamento consiste em transformar os dados “brutos” em estimativas dos níveis de expressão dos transcritos. Para tanto, utiliza-se o pré-processamento denominado GC-RMA, o qual compreende as etapas de correção do ruído de fundo (*Background Correction*), normalização e sumarização.

2.1.1 GC-RMA (do inglês *GeneChip Robust Multi-array Average*)

A interferência causada pela intensidade de fluorescência de *spots* vizinhos adiciona um ruído aleatório positivo que leva à detecção equivocada de fluorescência nos *spots* onde espera-se intensidade nula. Nesse sentido, a correção do ruído de fundo busca subtrair a intensidade de fundo da intensidade observada em cada *spot*. Dentre os métodos para a correção do ruído de fundo, o GC-RMA destaca-se pela sua superioridade conceitual em relação aos demais.

²processo pelo qual as moléculas de cDNA ligam-se às sondas no microarranjo

O método GC-RMA foi desenvolvido por Wu et al. (2004) como uma versão aprimorada do método RMA (Irizarry et al., 2003), o qual considera algumas informações sobre a sequência de sondas para descrever a variação de ligação não específica (NSB, do inglês *nonspecific binding*) e, assim, obter estimativas mais precisas da expressão gênica.

Os autores propõem um modelo estatístico para descrever os sinais observados das sondas PM e MM em um microarray. Dado um par de sondas, o modelo é dado por

$$PM = O_{PM} + N_{PM} + S \text{ e}$$

$$MM = O_{MM} + N_{MM} + \phi S,$$

em que O representa o ruído óptico, N representa o ruído de ligação não específica, e S é uma quantidade proporcional à expressão do RNA mensageiro (quantidade de interesse). O modelo inclui o parâmetro $0 < \phi < 1$ sob a possibilidade de que a sonda MM detecte a quantidade de interesse em alguns pares de *spots*. Assume-se que o termo O siga uma distribuição log-normal e, $\log_2(N_{PM})$ e $\log_2(N_{MM})$ sigam distribuições normais bivariadas com médias μ_{PM} e μ_{MM} , respectivamente, e a mesma variância σ^2 e coeficiente de correlação ρ , constante entre os *spots*. Os termos O e N são assumidos como independentes e as médias μ_{PM} e μ_{MM} são definidas como funções das afinidades das sondas PM e MM, respectivamente. A afinidade é uma medida usada para estimar os efeitos específicos da hibridação com a base que compõe as sondas.

2.1.2 Normalização Quantílica

Tipicamente, um estudo de expressão gênica envolve a utilização de vários microarranjos. Nesse contexto, a normalização permite que os resultados de diferentes microarranjos sejam comparáveis. Tal etapa preocupa-se com a influência de fatores técnicos relacionados às condições amostrais na estimação das expressões gênicas.

A normalização quantílica, introduzida por Bolstad et al. (2003), tem como objetivo fazer com que as leituras de cada microarranjo tenham a mesma distribuição empírica. Tal método

é motivado pela ideia de que se dois vetores de dados possuem a mesma distribuição, então o gráfico quantil-quantil apresentará pontos ao longo de uma linha diagonal e são diferentes se estiverem fora da diagonal. Este conceito é estendido para n dimensões de modo que todos os n vetores de dados tenham a mesma distribuição a partir da projeção dos quantis dos dados na diagonal n -dimensional.

O algoritmo que realiza a normalização quantílica é descrito a seguir.

1. Seja \mathbf{X} uma matriz de dimensão $p \times n$ em que cada coluna contém dados de um microarranjo.
2. Ordene cada coluna de \mathbf{X} e chame-a de \mathbf{X}_{sort} .
3. Tome a média de cada linha de \mathbf{X}_{sort} e atribua a cada elemento para obter o \mathbf{X}'_{sort} .
4. Obtenha a matriz $\mathbf{X}_{normalizada}$ reorganizando cada coluna de \mathbf{X}'_{sort} para ter a mesma ordenação como \mathbf{X} original.

Esse método se sobressai por ser menos custoso computacionalmente e se tratar de um método considerado padrão-ouro para a etapa de normalização (Bolstad et al., 2003).

2.1.3 Mediana Polida

A etapa de sumarização consiste em combinar as intensidades dos pares de *spots* a fim de estimar as expressões gênicas dos transcritos correspondentes.

No método da mediana polida um modelo linear *multichip* é ajustado aos dados para cada conjunto de *spots*. Em particular para o conjunto k com $i = 1, \dots, I_k$ e os dados de J chips indexados por $j = 1, \dots, J$, aplica-se o seguinte modelo

$$\log_2(PM_{ij}^{(k)}) = \alpha_i^{(k)} + \beta_j^{(k)} + \epsilon_{ij}^{(k)},$$

em que α_i é o efeito da sonda i , β_j é o efeito do microarranjo j (quantidade de interesse) e ϵ_{ij} é o ruído aleatório. A mediana polida é um algoritmo para ajustar o modelo de forma robusta. Ressalta-se que os valores de expressão estimados usando este método estarão na escala \log_2 .

Após a etapa de sumarização, os dados de expressão gênica podem ser estruturados em uma matriz $p \times n$, em que p é o número de transcritos e n é o número de observações, tratadas como “amostras” em textos biológicos. Assim sendo, as expressões gênicas dos transcritos são dispostas nas linhas do conjunto de dados e as colunas representam as unidades amostrais para as diferentes condições amostrais (casos e controles, por exemplo). As entradas dos dados correspondem ao valor de expressão de um gene (linha) em um chip (coluna). A notação $X_{i,j}$ corresponde à expressão estimada do transcrito i na amostra (chip de microarranjo) j .

Transcrito	Chip 1	Chip 2	...	Chip n
1	X_{11}	X_{11}	...	X_{1n}
2	X_{21}	X_{22}	...	X_{2n}
\vdots	\vdots	\vdots	\vdots	\vdots
p	X_{p1}	X_{p2}	...	X_{pn}

Tabela 2.1: Estrutura de dados de expressão gênica.

Neste trabalho, utiliza-se o *software* R 3.6.1 (R Core Team, 2017) juntamente com alguns pacotes disponíveis no repositório CRAN e no projeto Bioconductor³. Esse projeto, criado em 2002, engloba atualmente grande parte dos pacotes do *software* usados para análise de dados genômicos de alto rendimento, totalizando mais de mil pacotes. Assim como o *software*, o programa é de código aberto.

³bioconductor.org

Capítulo 3

Redução de Dimensionalidade para Dados de Expressão Gênica

A presença de transcritos não informativos nos dados de microarranjo pode induzir viés e reduzir o poder estatístico em análises posteriores. Um grande desafio na análise de dados de expressão gênica advém da alta dimensionalidade devido ao grande número de variáveis genéticas (por exemplo, transcritos, genes) medidas em um número reduzido de unidades amostrais (Buhlmann et al., 2016). Nesse contexto, a redução de dimensionalidade é frequentemente utilizada para selecionar um subconjunto de transcritos relevantes eliminando os desinteressantes e, assim, diminuir o custo de lidar com um grande número de variáveis.

3.1 Seleção de Transcritos (ou Genes) Diferencialmente Expressos

Uma possível estratégia para a seleção de transcritos envolve a filtragem de transcritos com expressões constantes, visto que esses não podem contribuir com a capacidade de discriminação. Outro critério amplamente utilizado envolve a exclusão de transcritos com *fold change* (FC) abaixo de um limiar pré-estabelecido, o qual indica a razão entre a expressão dos transcritos nos grupos de casos em relação ao grupo de controles.

3.1.1 Filtragem por nível de expressão

Em geral, alguns transcritos não serão expressos em todas as condições amostrais ou pelo menos em uma delas. Os transcritos cujas expressões são muito baixas produzem pouca evidência para avaliar diferenças na expressão gênica entre duas ou mais condições, assim pode ser sensato reduzir o conjunto de transcritos àqueles que foram expressos e/ou identificados em pelo menos uma condição. Além disso, do ponto de vista biológico, um transcrito deve ser expresso em algum nível mínimo antes que seja considerado biologicamente relevante. Nesse sentido, usualmente realiza-se uma filtragem (descarte) de transcritos antes de análises posteriores. Como regra geral, os transcritos são descartados se apresentarem expressões muito baixas ou nulas em todas as observações para qualquer uma das condições amostrais.

3.1.2 Fold Change (FC)

Em algumas situações, há o interesse em se estudar os transcritos cujas razões entre os níveis de expressão das diferentes condições possuam uma magnitude superior a um dado valor de corte. Um gene, por exemplo, pode ser superexpresso ou subexpresso na condição B versus A e, portanto, é observado um aumento ou diminuição no seu nível de expressão. Ao analisar casos e controles, o *fold change* mensura quantas vezes a média (ou mediana) dos casos é maior que a média (ou mediana) dos controles. Tipicamente, toma-se o logaritmo na base 2 do *fold change*. Dessa forma, seus valores positivos denotam um aumento e valores negativos um decréscimo na expressão gênica (0 significa que não há alteração). Sejam B_{ij} com $j = 1, 2, \dots, n_1$ e A_{ik} com $k = 1, 2, \dots, n_2$ as expressões gênicas estimadas para o i -ésimo transcrito nas condições B e A , respectivamente. Assim, o *fold change* é dado por

$$FC_i = \frac{\bar{B}_i}{\bar{A}_i}, \text{ em que } \bar{B}_i = \sum_{j=1}^{n_1} \frac{B_{ij}}{n_1} \text{ e } \bar{A}_i = \sum_{k=1}^{n_2} \frac{A_{ik}}{n_2}.$$

Usando a mediana, tem-se

$$FC_i^* = \frac{\tilde{B}_i}{\tilde{A}_i}, \text{ em que } \tilde{B}_i = \text{mediana do } i\text{-ésimo transcrito na condição B e}$$
$$\tilde{A}_i = \text{mediana do } i\text{-ésimo transcrito na condição A.}$$

3.1.3 Teste estatístico

O teste estatístico é um critério frequentemente utilizado para a identificação de transcritos diferencialmente expressos nas condições amostrais. Para cada transcrito, a estatística do teste e seu p-valor correspondente são calculados para determinar se a diferença entre médias (ou medianas) das expressão dos grupos pode ser considerada significativa. Assim sendo, o p-valor denota a probabilidade de observar um resultado experimental igual ou mais extremo, supondo a igualdade do nível de expressão gênica entre as condições. Um p-valor pequeno indica evidência de expressão diferencial, sendo superexpressão quando um dos grupos (geralmente dos casos) apresenta em média valores de expressão superiores aos do outro grupo (geralmente controles) e subexpressão quando acontece o contrário..

À medida que centenas ou milhares de testes são conduzidos, simplesmente usar o nível de significância sem ajustá-lo para múltiplos testes aumentará a chance de encontrar falsos positivos. Para contornar esse problema introduzimos a seguir o método BH (Benjamini e Hochberg, 1995), amplamente utilizado para o controle da taxa de descoberta falsa.

Controle da Taxa de Descoberta Falsa

A taxa de descoberta falsa (FDR, do inglês *False Discovery Rate*) é definida como a proporção esperada de hipóteses (nulas) rejeitadas incorretamente dentre todas as hipóteses rejeitadas. Considere o problema de testar simultaneamente m hipóteses (nulas) das quais m_0 são verdadeiras. A Tabela 3.1 lista todos os resultados possíveis dos testes de hipótese.

	H_0 não rejeitada	H_0 rejeitada	Total
H_0 verdadeira	U	V	m_0
H_0 falsa	T	S	$m - m_0$
Total	$m - R$	R	m

Tabela 3.1: Possíveis resultados para m testes de hipótese (nula).

Todas as quantidades apresentadas na Tabela 3.1 são variáveis aleatórias não observáveis, exceto m e R . O método Benjamini-Hochberg (BH) é um procedimento que controla a taxa de descoberta falsa de modo que $\text{FDR} \leq \alpha$, ou seja, $\mathbb{E}\left(\frac{V}{R}\right) \leq \alpha$ (Benjamini e Hochberg, 1995). Os p-valores ajustados pelo método recebem o nome de q-valor, os quais são definidos recursivamente começando pelo maior p-valor. Sejam $p_{(1)} < \dots < p_{(m)}$ os p-valores não ajustados ordenados, então

$$q_{(i)} = \begin{cases} p_{(m)}, & i = m \\ \min\{q_{(i+1)}, (m/i)p_{(i)}\}, & i = (m-1), \dots, 1. \end{cases} \quad (3.1)$$

3.2 Agrupamento de Transcritos (ou Genes) por Nível de Expressão

O princípio básico do agrupamento de dados envolve o particionamento de um conjunto de elementos em grupos de tal forma que elementos no mesmo grupo sejam mais semelhantes entre si que elementos de grupos diferentes, segundo algum critério. O agrupamento de dados em si é uma técnica de aprendizado não supervisionado para a identificação de padrões subjacentes dos dados. Na análise de dados de expressão gênica, esses métodos apresentam potencial de identificar “grupos” de transcritos (ou genes) relacionados funcionalmente.

3.2.1 Agrupamento Parcial Baseado em P-valores (PPCLUST)

Um grande desafio no agrupamento de dados genéticos advém da alta dimensionalidade devido ao grande número de variáveis genéticas (por exemplo, transcritos, genes) medidas em um pequeno número de unidades amostrais (Buhlmann et al., 2016).

O algoritmo PPCLUST é capaz de agrupar um grande número de variáveis com poucas observações, cujo formato dos dados segue o apresentado na Tabela 2.1, sendo p grande e n pequeno. A medida de similaridade utilizada no algoritmo consiste no p -valor obtido a partir do teste não paramétrico desenvolvido por Wang e Akritas (2004) que envolve uma hipótese nula de igualdade de funções de distribuição para cada nível de fator quando o número de níveis fator é grande e o número de replicações é pequeno (tal como os dados da Tabela 2.1). Em razão da medida de similaridade usada, tal algoritmo herda a robustez a *outliers* e a invariação a transformações monótonas dos dados (ver von Borries (2008) e von Borries (2009)).

Denote R_{ij} como sendo o posto (ou *rank*) da observação X_{ij} no conjunto de todas as $n_1 + n_2 + \dots + n_a$ observações. Os dados podem ser apresentados como uma matriz com elementos X_{ij} conforme mostra a Tabela 2.1. A hipótese de interesse pode ser escrita como

$$H_0 : F_1(x) = \dots = F_a(x),$$

em que $F_i(x)$ denota uma função de distribuição de probabilidade arbitrária. Ou seja, H_0 estabelece que todas as observações advêm da mesma distribuição. Então, sob H_0 , os postos são distribuídos uniformemente entre 1 e $\sum_{i=1}^a n_i$. Agora, defina a estatística de teste

$$F_R = \frac{MST_R}{MSE_R},$$

em que MST_R é o erro quadrático médio devido aos níveis de fator e o MSE_R é a estimativa da variância amostral, também obtida por meio dos postos. Isso é,

$$MST_R = \frac{1}{a-1} \sum_{i=1}^a (\bar{R}_{i.} - \tilde{R}_{..})^2$$

$$MSE_R = \frac{1}{a} \sum_{i=1}^a \frac{1}{n_i} S_{R,i}^2.$$

Note que $\bar{R}_{i.} = n_i^{-1} \sum_{j=1}^{n_i} R_{ij}$, sendo R_{ij} o posto médio de cada nível do fator, $\tilde{R}_{..} =$

$a^{-1} \sum_{i=1}^a \bar{R}_i$ é a média geral dos níveis de fator e $S_{R,i}^2$ é a variância amostral calculada para cada nível de fator.

Wang e Akritas (2004) mostram que a distribuição assintótica da estatística $\sqrt{a}(F_R - 1)$ sob H_0 converge para uma distribuição normal. A partir desse resultado, o p-valor do teste pode ser facilmente calculado. Um grande p-valor indica que os dados testados não fornecem evidências para concluir que os níveis de fator possuem distribuições diferentes. Nesse caso, conclui-se que tais níveis de fator pertencem ao mesmo grupo. Por outro lado, um pequeno p-valor fornece evidências contra H_0 , indicando que pelo menos duas das distribuições testadas são diferentes.

Fundamentalmente, o algoritmo PPCLUST parte do princípio de que os dados de alta dimensão são originados de uma mistura de distribuições desconhecidas $F_i(x)$, no qual cada distribuição define um grupo. Nesse sentido, o agrupamento é considerado como um problema de detecção de uma diferença significativa na distribuição das observações. Em outras palavras, o objetivo final do algoritmo é concluir que os (*clusters*) obtidos do conjunto de dados provêm de distribuições diferentes. A ideia principal envolve testar iterativamente a hipótese nula do teste não-paramétrico de Wang e Akritas (2004) em uma dada partição (ou totalidade) dos dados de modo que uma partição (ou totalidade) seja dividida em duas menores quando a hipótese nula é rejeitada.

A escolha do nível de significância (limiar) é estabelecido arbitrariamente. Quanto mais próximo de zero for esse limiar, mais conservador o algoritmo será em detectar diferenças entre grupos, podendo identificar menos grupos ao final do procedimento. Por outro lado, quanto maior o limiar escolhido (próximo de 1), menos conservador será o algoritmo, favorecendo a identificação de um número maior de grupos.

Seja c o índice do grupo em que o teste está sendo aplicado e $nv(D)$ o número de variáveis (linhas) em um conjunto de dados D . O pseudo-código do PPCLUST desenvolvido por von Borries (2008) é reproduzido abaixo.

Algoritmo 1: PPCLUST

Entrada: Matriz de observações de todas as variáveis como na Tabela 2.1
Saída: Identificação do grupo ao qual cada variável está classificada

- 1 faça $c = 1$ e $D_1 = \text{Dados}$;
- 2 obtenha o rank de todas as observações em D_1 e armazene-as em D_{1R} no mesmo formato da Tabela 2.1;
- 3 calcule a mediana dos ranks para cada variável (ou seja, cada linha) em D_{1R} ;
- 4 ordene as variáveis em D_{1R} de acordo com suas medianas;
- 5 realize o teste para avaliar se as variáveis em D_1 têm distribuição idêntica;
- 6 **if** H_0 não é rejeitada **then**
- 7 relatar todas as variáveis em D_1 como pertencentes a um único grupo. Vá para o passo 43;
- 8 **else**
- 9 continue para o passo 11;
- 10 **end**
- 11 tome a primeira metade do número (arredondado para inteiro) de variáveis de linhas consecutivas de D_1 e denote os dados neste subconjunto incluindo as observações correspondentes como D_2 ;
- 12 aplique o teste para avaliar se as variáveis em D_2 têm distribuição idêntica;
- 13 **if** H_0 não é rejeitada **then**
- 14 atribua as variáveis de D_2 ao grupo c ;
- 15 faça $c = c + 1$;
- 16 remova as variáveis de D_1 em D_2 ;
- 17 **if** $nv(D_1) = 0$ **then**
- 18 vá para o passo 43;
- 19 **else**
- 20 teste se cada variável em D_1 pertence ao grupo recém atribuído, testando a hipótese de que todas as variáveis envolvidas têm a mesma distribuição. Remova a variável de D_1 quando H_0 não é rejeitada e o coloque-a nesse recém atribuído;
- 21 seja D_2 o conjunto que contém as variáveis restantes em D_1 e vá para o passo 29;
- 22 **end**
- 23 **else**
- 24 pegue a primeira metade do número (arredondado para inteiro) de variáveis de D_2 e denote os dados deste subconjunto como D_3 ;
- 25 atribua todas as variáveis que não estão em D_3 a D_1 ;
- 26 faça $D_2 = D_3$ e remova D_3 ;
- 27 continue para o passo 29;
- 28 **end**
- 29 **if** $nv(D_2) = 1$ **then**
- 30 atribua a variável em D_2 ao grupo 0;
- 31 remova as variáveis em D_2 de D_1 ;
- 32 **if** $nv(D_1) = 0$ **then**
- 33 vá para o passo 43;
- 34 **else**
- 35 **if** $nv(D_1) = 1$ **then**
- 36 faça $D_2 = D_1$ e volte ao passo 29;
- 37 **else**
- 38 faça $D_2 = D_1$ e volte ao passo 12;
- 39 **end**
- 40 **end**
- 41 **else**
- 42 **end**
- 43 finalize o procedimento e reporte os grupos identificados;

Vários trabalhos discutem estudos numéricos e aplicam o PPCLUST em dados de expressão gênica (von Borries, 2008; von Borries, 2009; Silva, 2012; Souza, 2013). Nesse contexto, o algoritmo possibilita a identificação de transcritos (ou genes) relevantes eliminando os desin-

teressantes (grupos intermediários).

3.2.2 Índice de Rand Ajustado

O Índice de Rand Ajustado (ARI, do inglês *Adjusted Rand Index*) é uma correção do Índice de Rand (Rand, 1971, RI) desenvolvido por Hubert e Arabie (1985), o qual mede a concordância entre duas partições de um mesmo conjunto de elementos. Diferente do RI, no caso em que as partições são aleatoriamente dispostas, seu valor esperado é igual a 0, além disso, seu valor máximo não ultrapassa o valor 1.

Considerando, por exemplo, uma partição $R = \{r_1, r_2, \dots, r_k\}$ representando k grupos de referência que são usados para comparar os procedimentos de agrupamento e, sendo $V = \{v_1, v_2, \dots, v_c\}$ uma partição de c grupos obtidos a partir de algum algoritmo de agrupamento. A Tabela 3.2, apresenta os resultados de ambas as partições dos dados, R e V , sendo n_{ij} o número de elementos que estão em ambos os grupos r_i e v_j , com $i = 1, \dots, k, j = 1, \dots, c$. Além disso, $n_{i.} = \sum_{j=1}^c n_{ij}$ e $n_{.j} = \sum_{i=1}^k n_{ij}$.

Grupo	v_1	v_2	...	v_c	Total
r_1	n_{11}	n_{12}	...	n_{1c}	$n_{1.}$
r_2	n_{21}	n_{22}	...	n_{2c}	$n_{2.}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
r_k	n_{k1}	n_{k2}	...	n_{kc}	$n_{k.}$
Total	$n_{.1}$	$n_{.2}$...	$n_{.c}$	n

Tabela 3.2: Tabela cruzada para o agrupamento.

O ARI pode ser calculado por:

$$\frac{\sum_i \sum_j \binom{n_{ij}}{2} - \frac{\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2}}{\binom{n}{2}}}{\frac{1}{2} [\sum_i \binom{n_{i.}}{2} + \sum_j \binom{n_{.j}}{2}] - \frac{\sum_i \binom{n_{i.}}{2} \sum_j \binom{n_{.j}}{2}}{\binom{n}{2}}}$$

Em que, o ARI é igual a 1 quando as duas partições concordam exatamente e 0 quando as duas partições concordam não mais do que o esperado por acaso. Supondo duas partições: $A =$

$\{(1, 2, 3, 4, 5, 6), (7, 8, 9, 10, 11, 12)\}$ e $B = \{(7, 8, 9, 10), (1, 2, 11, 12), (3, 4, 5, 6)\}$. A tabela cruzada entre as partições é apresentada abaixo.

Grupo	b_1	b_2	b_3	Total
a_1	0	2	4	6
a_2	4	2	0	6
Total	4	4	4	12

Tabela 3.3: Tabela cruzada das partições A e B.

Basicamente, essa medida baseia-se no número de objetos que foram alocados juntos e separados tanto em A quanto em B . Nota-se que nem todo o quarteto que está no mesmo grupo de B também está no mesmo grupo de A (nesse caso, $ARI = 0,36$).

Capítulo 4

Análise de Redes e de Vias Biológicas

4.1 Redes de Co-Expressão Diferencial

A rede de co-expressão gênica é um tipo de rede popularmente empregada para descrever as interações entre genes ou transcritos (ver Cox e Wermuth, 1996; Weirauch, 2011; Sulaimanov e Koepl, 2016). Nessa rede, as arestas ausentes denotam correlações nulas entre os respectivos pares de genes que, no caso gaussiano, implicam em independência marginal entre os perfis de expressão gênica. A rede de co-expressão baseia-se no princípio de que os genes envolvidos nos mesmos processos biológicos, provavelmente apresentarão perfis de expressão gênica correlacionados.

O objetivo da análise diferencial é comparar as redes de co-expressão entre diferentes condições amostrais e identificar mudanças nas topologias das redes (padrão de conexões). O método de co-expressão diferencial adotado consiste, resumidamente, em transformar os coeficientes de correlação para z -scores e usar as diferenças entre os z -scores para calcular os p -valores das correlações diferenciais entre os transcritos. Além disso, os p -valores resultantes são corrigidos para múltiplos testes.

No que tange o método de co-expressão diferencial, a transformação z de Fisher é aplicada aos coeficientes de correlação em cada condição:

$$z = atanh(r) = \frac{1}{2} \log_e \left(\frac{1+r}{1-r} \right),$$

em que r é o coeficiente de correlação amostral, \log_e é função logarítmica natural e $atanh$ é a função hiperbólica arco-tangente. A variância do score z resultante depende de qual coeficiente de correlação é usado. No presente trabalho, o coeficiente de correlação de Pearson (r_p). Sob a suposição de normalidade bivariada, a variância de z pode ser calculada por

$$var(z) = \frac{1}{n-3},$$

em que n é o tamanho da amostra. Devido ao denominador, a equação exige que haja ao menos 4 observações em cada condição considerada. Segundo Fieller, Hartley e Pearson (1957), a equação de variância acima é válida para amostras pequenas, tal como amostras de tamanho igual a 10.

Assim, a diferença nos scores z (dz) entre duas condições pode ser calculada por,

$$dz = \frac{(z_1 - z_2)}{\sqrt{(s_{z_1}^2 + s_{z_2}^2)}} = \frac{(z_1 - z_2)}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}.^1$$

Sob a hipótese nula, dz é normalmente distribuída com média nula e variância igual a um (distribuição normal padrão). O p-valor do teste é calculado por $p(|Z| > |dz|)$, em que Z denota uma variável aleatória normal padrão.

4.2 Redes de Dependência Condicional (Modelos Gráficos Probabilísticos, MGPs)

A descoberta dos mecanismos por trás de doenças demanda um entendimento sobre como os diferentes elementos funcionais do genoma (por exemplo, DNA, mRNA, miRNA etc) estão relacionados entre si. Uma maneira de descrever as interações entre os elementos de interesse

¹Mckenzie et al. (2016) apresenta esta fórmula de dz com o denominador $\sqrt{|s_{z_1}^2 - s_{z_2}^2|}$, contudo, o pacote *DGCA* dos autores implementa o denominador correto, $\sqrt{(s_{z_1}^2 + s_{z_2}^2)}$.

envolve conceitos de independência condicional entre variáveis aleatórias. Nesse contexto, os Modelos Gráficos Probabilísticos (MGPs) se destacam como uma ferramenta poderosa para a aprendizagem e descrição de independências condicionais (ver Pearl, 1988; Whittaker, 2009; Lauritzen, 1996). Adicionalmente, os MGPs podem ser interpretados como casos especiais de modelos estatísticos da estrutura de independência condicional de um conjunto de variáveis (Maathuis et al., 2018).

Como tratado na Seção 4.1, a rede de co-expressão diferencial baseia-se na correlação marginal. É sabido que esse método não distingue relacionamentos diretos de relacionamentos indiretos, como por exemplo, quando dois genes que mostram perfis de expressão correlacionados devido a um gene intermediário comum. Nesse contexto, investigar o padrão de dependência condicional dos grafos (ou redes) de duas condições distintas (por exemplo, casos vs. controles) fornece informações valiosas, possibilitando a geração de novas hipóteses biológicas.

Um MGP, denotado por $G = (V, E)$, é representado por um grafo, em que V representa um conjunto de variáveis aleatórias (vértices ou nós) denotando as quantidades de interesse e E é um conjunto de arestas representando a estrutura de dependência entre as variáveis. Os grafos são úteis para representar relações de independência entre variáveis aleatórias. Um grafo pode ser não direcionado ou direcionado e pode ser acíclico ou cíclico. Em um grafo não direcionado (UG, do inglês *Undirected Graph*), o conjunto E contém pares não ordenados de vértices. Por exemplo, $(X, Y) \in E$ significa que X e Y são unidos por uma aresta. No caso direcionado, o conjunto E é formado por pares ordenados de vértices. Assim, se $(X, Y) \in E$, então há uma seta apontando de X para Y . Um caminho direcionado que inicia e termina na mesma variável é chamado de ciclo. Logo, um gráfico direcionado acíclico (DAG, do inglês *Directed Acyclic Graph*) não contém ciclos, por definição.

São outros conceitos importantes o de adjacência e caminho. Dois vértices são adjacentes se houver uma aresta entre eles (notação: $X \sim Y$). Por exemplo, na Figura 4.1-(b), X e Y são adjacentes, mas X e Z não são adjacentes. Com base nisso, pode-se introduzir o conceito de caminho. Uma sequência X_0, \dots, X_n é chamada de caminho se $X_{i-1} \sim X_i$, para todo $i \in$

$\{1, \dots, n\}$. Na Figura 4.1-(a), a sequência X, Y, Z é um caminho.

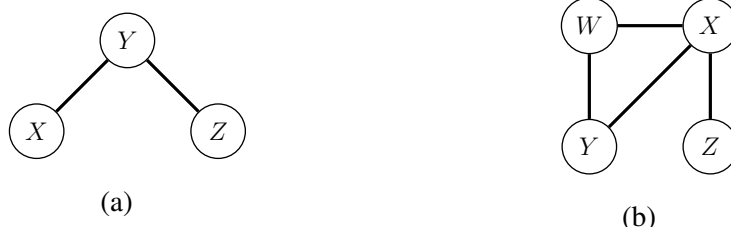


Figura 4.1: (a) um grafo com vértices $V = \{X, Y, Z\}$, cujo conjunto de arestas é $E = \{(X, Y), (Y, Z)\}$ e (b) $\{Y, W\}$ e $\{Z\}$ são separados por $\{X\}$. Além disso, W e Z são separados por $\{X, Y\}$. Fonte: elaboração própria com o uso do pacote *tikZ* do LaTeX.

4.2.1 MGPs Não Direcionados

Em um Modelo Gráfico Probabilístico (MGP) não direcionado, os nós correspondem às variáveis aleatórias e as arestas codificam a independência condicional entre as variáveis. Em uma perspectiva probabilística, tais MGPs também são chamados de campos aleatórios de Markov (do inglês *Markov Random Fields*) ou redes de Markov (do inglês *Markov Networks*) (Pearl, 1988).

Um MGP não direcionado satisfaz as seguintes propriedades:

- **Propriedade de Markov global:** Sejam dois subconjuntos de variáveis A e B . Dizemos que A e B são separados por C em G e escrevem a , se todo caminho em G a partir de um nó em A para um nó em B contém um nó em C .
- **Propriedade de Markov local:** Uma variável é condicionalmente independente de todas as outras variáveis, dados os seus vizinhos.
- **Propriedade de Markov dos pares:** Duas variáveis não adjacentes são condicionalmente independentes, dadas todas as demais variáveis.
- **Fatorização Clique:** As propriedades de fatorização permitem que os modelos gráficos sejam computacionalmente tratáveis.

Para uma exposição mais aprofundada, consulte Pearl (1988), Whittaker (2009), Lauritzen (1996) ou Koller e Friedman (2009).

4.2.2 MGPs Não Direcionados Gaussianos

Assumindo uma amostra aleatória como sendo proveniente de uma distribuição multivariada gaussiana, a independência condicional é tida como uma d -separação, ou seja, quando um conjunto X de variáveis é independente de outro conjunto Y , dado um terceiro conjunto Z . Para distribuições gaussianas multivariadas, a independência condicional é equivalente à uma correlação parcial nula (Lauritzen, 1996, Proposição 5.2).

Considere um vetor aleatório $\mathbf{V} = (X_1, \dots, X_p)$ com distribuição gaussiana p -variada com média $\boldsymbol{\mu}_p$ e matriz de covariância $\boldsymbol{\Sigma}_{p \times p}$, como também a sua inversa, conhecida como matriz de precisão ou de concentração, a qual é denotada por $\mathbf{K} = \boldsymbol{\Sigma}_{p \times p}^{-1}$, sendo k_{ij} a entrada da matriz \mathbf{K} associada a concentração entre as variáveis $X_i, X_j \in \mathbf{V}$. Nesse caso, $k_{ij} = 0$ se e somente se X_i é condicionalmente independente de X_j , dadas todas as variáveis restantes em \mathbf{V} , ou seja,

$$X_i X_j | \mathbf{V} \setminus \{X_i, X_j\} \Leftrightarrow k_{ij} = 0.$$

Em um modelo gráfico, $k_{ij} = 0$ significa que X_i não é conectada por uma aresta a X_j diretamente, a não ser por outras variáveis em \mathbf{V} . Uma aresta ausente entre X_i e X_j informa que a correlação entre esse par de variáveis pode ser explicada por outras variáveis no modelo. A correlação parcial pode ser definida como

$$\rho_{X_i X_j | \mathbf{V} \setminus \{X_i, X_j\}} = \frac{-k_{ij}}{\sqrt{k_{ii} k_{jj}}}.$$

Ao realizar uma análise de dependência condicional, uma abordagem possível consiste em estimar a matriz de precisão de ambos os grupos (casos e controles) separadamente utilizando um dos métodos apresentados na Seção 4.3. Uma observação importante nesse contexto é que há cada vez mais evidências biológicas sugerindo que as redes biomoleculares possuem uma

estrutura esparsa. Essa característica tem sido explorada por métodos esparsos para permitir a estimação de modelos gráficos gaussianos a partir de dados de microarranjo.

4.3 Estimação Esparsa da Matriz de Precisão

A matriz de covariância (Σ) e sua inversa, a matriz de precisão (Σ^{-1}), desempenham um papel importante em métodos de análise multivariada, como no estudo de MGPs gaussianos. Em contraste a muitos conjuntos de dados comuns sobre os quais modelos gráficos gaussianos são aplicados, os dados de microarranjo constituem um problema desafiador, pois os experimentos de microarranjo tipicamente produzem dados com um grande número p de variáveis e um número n substancialmente menor de observações, o que dificulta a estimação da matriz inversa da matriz de covariância, etapa essencial para a estimação de MGPs.

Dada uma amostra de tamanho n , podemos obter um estimador para Σ via método de máxima verossimilhança, sendo $\hat{\Sigma} = n^{-1}XX^T$, em que substituindo o termo n por $n-1$ obtemos o estimador não enviesado de Σ . Intuitivamente, sabe-se que quanto maior o tamanho da amostra, melhor e mais confiáveis serão nossas estimativas. No entanto, em alguns campos de estudo, como na genômica, o número de observações disponível pode ser bem reduzido em comparação com o número de variáveis de interesse. Em contextos como esse, a dimensão p dos dados é bem superior ao tamanho da amostra n , impossibilitando que a matriz $\hat{\Sigma}$ seja invertida. Isso se deve ao fato de que o número de autovalores não nulos de $\hat{\Sigma}$ é igual a $\min(n, p) = n$, porém para ser não singular esse número teria de ser igual a p . Este caso é conhecido como “big p , small n ” ou HDLSS (do inglês *High Dimensional Low Sample Size*) na literatura sobre dados de alta dimensão, e geralmente denotado por $n \ll p$.

4.3.1 Métodos baseado na log-verossimilhança penalizada

O método Glasso (do inglês *Graphical lasso*) introduzido por Friedman, Hastie e Tibshirani (2008) é um dos métodos de estimação da matriz de precisão mais conhecidos. Em tal método,

a estimativa da matriz de precisão pode ser expressa como um problema de maximização da log-verossimilhança penalizada na forma

$$\log |\Theta| - \text{tr}(\hat{\Sigma}\Theta) - \lambda \|\Theta\|_1,$$

sob matrizes definidas positivas Θ , em que $\|\Theta\|_1$ é a norma L_1 da matriz de precisão, computada como a soma dos valores absolutos dos elementos de Θ ; λ é o parâmetro encolhimento dos elementos de Θ , e controla o número de zeros na estimativa final da matriz de precisão, ou seja, a esparsidade.

A Figura 4.2 apresenta os gráficos de calor das matrizes de covariância de dados simulados ($p=100$ e $n=50$). O primeiro gráfico de calor (a) representa a matriz de covariância verdadeira, o segundo (b), a matriz de covariância amostral e, o terceiro (c), a matriz estimada pelo método Glasso.

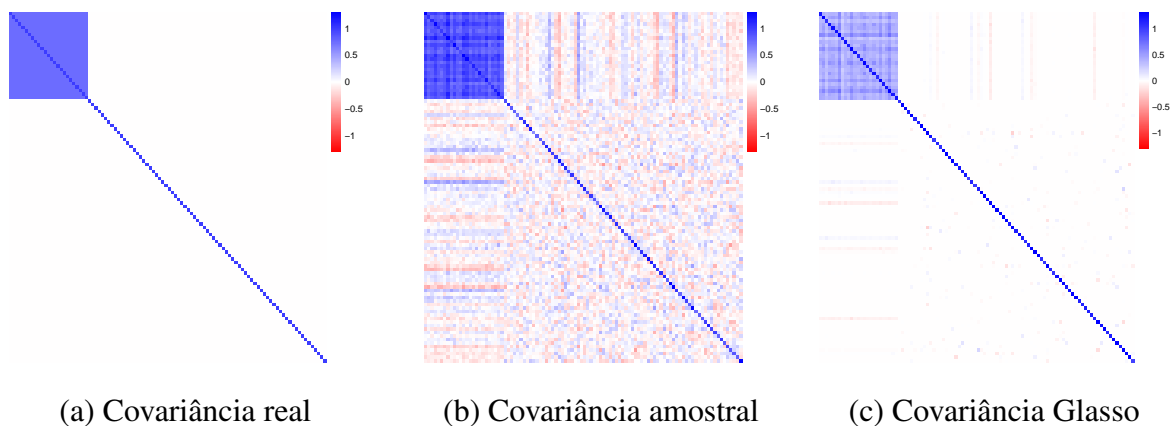


Figura 4.2: Gráficos de calor das matrizes de covariância real (a), estimada por $n^{-1}XX^T$ (b) e por Glasso (c) a partir de dados foram gerados com $p = 100$ e $n = 50$. Fonte: elaboração própria com o uso do pacote *pheatmap* do R.

Observa-se que os elementos não nulos da matriz estimada pelo método Glasso são um pouco menores que os da matriz de covariância amostral (Figura 4.2). Isso deve-se ao efeito de encolhimento da penalidade L_1 .

Tipicamente, os critérios tais como a validação cruzada tendem a escolher redes muito gran-

des para serem úteis, assim sendo, uma melhor escolha de rede é guiada por considerações práticas, como a interpretabilidade e estabilidade da rede (Li et al., 2013; Meinshausen e Bühlmann, 2010). Danaher, Wang e Witten (2012) argumentam que quando a estimativa de rede é realizada para exploração e geração de hipóteses a partir dos dados e quando p é grande, apenas uma estimativa de rede muito esparsa será útil.

4.3.2 Métodos baseados em otimização restrita

Tomando por base o modelo gaussiano, alguns métodos baseados em otimização restrita abordam o problema de estimação da matriz de precisão através de regressões coluna por coluna. Isso faz com que esse métodos sejam computacionalmente mais simples e mais acessíveis em comparação aos métodos de verossimilhança penalizada.

Alguns desses métodos estão implementados no *software* R. Os pacotes *clime* e *fastclime* implementam o método CLIME (do inglês *Constrained L1-minimization for Inverse Matrix Estimation*), proposto por Cai, Liu e Luo (2011). Esse método estima a j -ésima coluna da matriz de precisão Θ resolvendo

$$\hat{\Theta}_j = \operatorname{argmin}_{\Theta_j} \|\Theta_j\|_1, \text{ sujeito a } \|\mathbf{S}\Theta_j - \mathbf{e}_j\|_\infty \leq \rho_j, \text{ para } j = 1, \dots, p,$$

em que \mathbf{S} é a matriz de covariância estimada, \mathbf{e}_j é um vetor com o j -ésimo elemento igual a 1 e o restante sendo 0 e ρ_j é um parâmetro de ajuste.

Relacionado ao estimador CLIME, Liu e Luo (2015) propuseram um novo método chamado SCIO (do inglês *Sparse Column-Wise Inverse Operator*) como um método eficiente para estimar a matriz de precisão. Esse método se encontra implementado no pacote *scio*. Além deste, Liu e Wang (2017) propuseram outro método chamado TIGER (do inglês *Tuning-Insensitive Graph Estimation and Regression*) para resolver o mesmo problema de restrição, o qual está disponível no pacote *camel*.

Esses métodos adotam a ideia de estimar a matriz de precisão coluna por coluna. A diferença

fundamental entre os métodos está em como resolvem o problema de regressão linear esparsa. O estimador CLIME usa o seletor Dantzig (Candes e Tao, 2007), o estimador SCIO usa um algoritmo iterativo de descida de coordenadas (Liu e Luo, 2015) e o estimador TIGER usa o SQRT-Lasso (Belloni, V.Chernozhukov e Wang, 2011).

Para ilustrar o uso desses métodos, foi gerado um conjunto de dados a partir de uma matriz de precisão bloco diagonal representada na Figura 4.3.

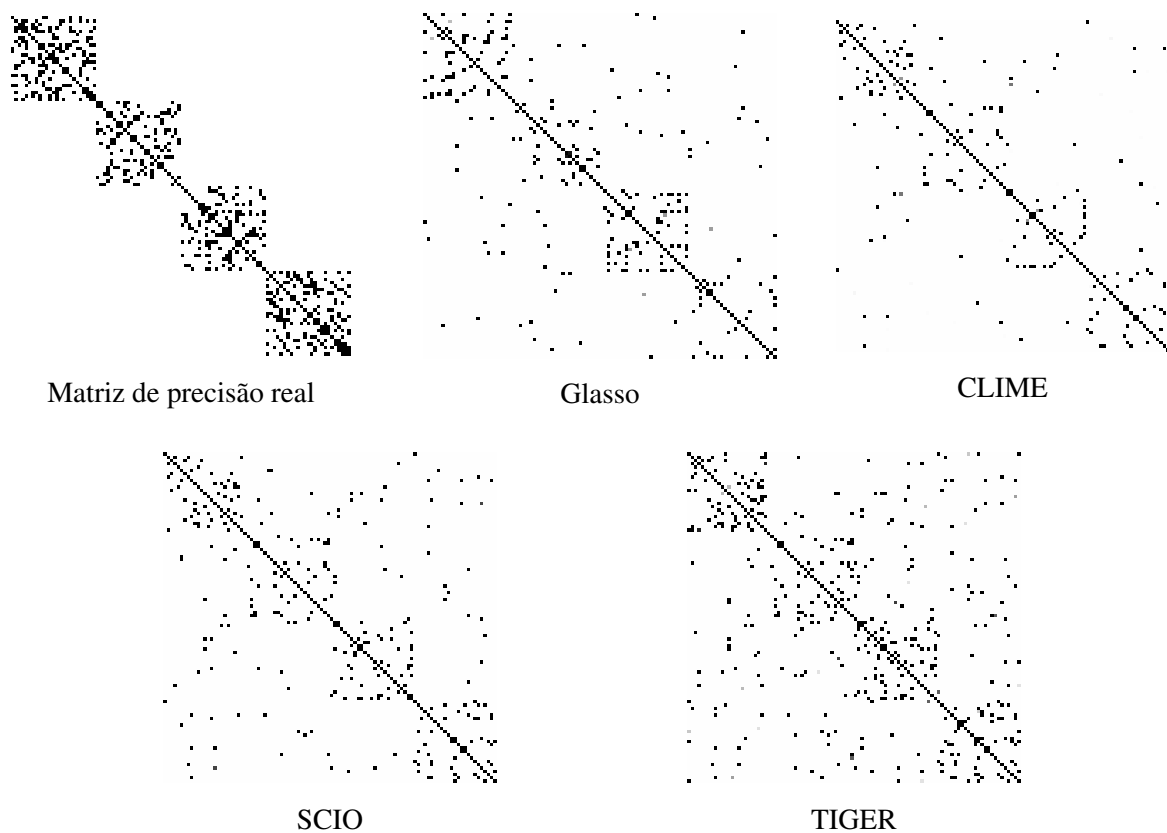


Figura 4.3: Gráficos de calor das matrizes de precisão real e estimadas pelos métodos Glasso, CLIME, SCIO e TIGER a partir de dados gerados com $p = 100$ e $n = 50$. Fonte: elaboração própria com o uso do pacote *pheatmap* do R.

Na Figura 4.3, a matriz de precisão corresponde a uma estrutura gráfica de cluster obtida a partir do pacote *huge*. O primeiro mapa de calor mostra a matriz de precisão verdadeira, os seguintes mostram a matriz estimada por Glasso, CLIME, SCIO e TIGER. Dentre todos os métodos de estimação, observa-se que o resultado do Glasso representa melhor o padrão de

esparcidade da matriz de precisão real.

A Figura 4.4 apresenta as matriz de covariância real, assim como as matrizes estimadas pelos métodos Glasso, CLIME, SCIO e TIGER. Nota-se que todos os métodos capturaram com êxito a esparcidade da matriz real. No entanto, a matriz estimada por Glasso representa melhor os elementos não nulos da matriz real.

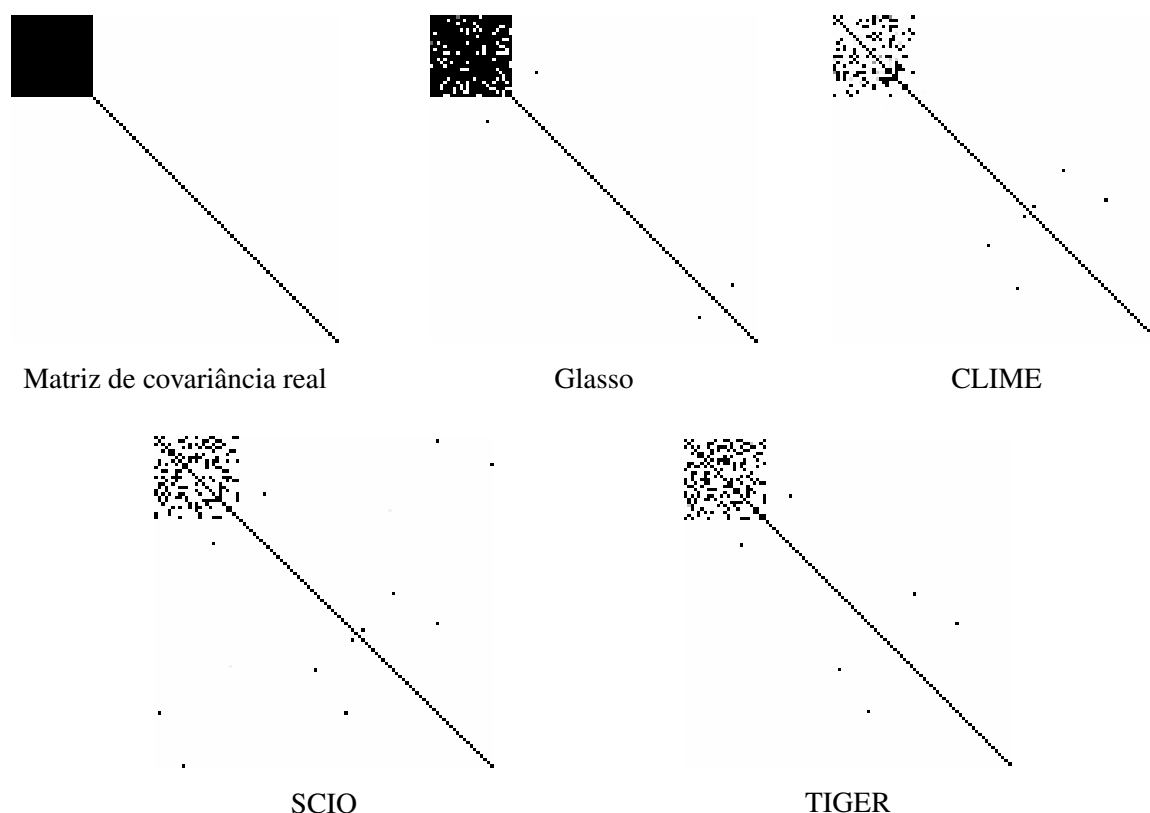


Figura 4.4: Gráficos de calor das matrizes de covariância estimadas por Glasso, CLIME, SCIO e TIGER a partir de dados ($p = 100$ e $n = 50$) gerados pela matriz de covariância real. Fonte: elaboração própria com o uso do pacote *pheatmap* do R.

4.4 Análise de Vias Biológicas (*Reactome Pathway Analysis*)

Uma vez que os transcritos com expressões alteradas foram identificados para cada região cerebral, é pertinente examinar se eles fornecem uma representação significativa de genes que estão envolvidos em uma ou mais vias biológicas (ou sistemas biológicos). Tem-se que um processo

biológico é composto por vias (*pathways*) e processos maiores constituídos pelas atividades de múltiplos produtos gênicos (por exemplo, RNA, proteína etc). Além disso, uma via biológica pode ser interpretada como um conjunto não ordenado de genes relacionados funcionalmente.

Nesse contexto, o pacote ReactomePA (*Reactome Pathway Analysis*) (Yu e He, 2016) incorpora a análise de super-representação (ASR), uma abordagem amplamente utilizada para identificar vias biológicas conhecidas super-representados em uma lista de genes de interesse. Tal pacote utiliza a base de dados colaborativa chamada Reactome de autoria de especialistas e revisada por pares, a qual armazena anotações² sobre uma gama de vias biológicas humanas.

Na ASR, um modelo hipergeométrico (Boyle et al., 2004) é usado para avaliar se o conjunto de genes anotados em uma determinada via biológica estão em um número maior do que seria esperado ao acaso (super-representados) em genes de uma lista transcritos selecionados.

$$\text{p-valor} = 1 - \sum_{i=0}^{k-1} \frac{\binom{M}{i} \binom{N-M}{n-i}}{\binom{N}{n}}.$$

Nesta equação, N é o número total de genes anotados, M é o número de genes anotados em uma determinada via biológica, n é o tamanho da lista de genes de interesse e k denota o número de genes dentro dessa lista que são anotados na via biológica em questão. Nesse sentido, esse método mede a sobreposição entre uma lista dos genes de interesse e o conjunto de genes previamente mapeados em milhares de vias biológicas conhecidas.

A suposição básica por trás da ASR é que uma via biologicamente relevante pode ser identificada pelo excesso de genes previamente mapeados na lista de genes de interesse em comparação com o número esperado ao acaso. Para controlar a taxa de erro do tipo I, utiliza-se geralmente o ajuste dos p-valores para comparações múltiplas pelo método BH.

²uma anotação consiste na descrição e na atribuição de uma informação biológica a um gene

Capítulo 5

Aplicação em Dados de Expressão Gênica da Doença de Alzheimer (DA)

Os valores brutos de expressão gênica foram obtidos diretamente dos arquivos em formato CEL, os quais foram pré-processados usando o método GC-RMA (do inglês *GeneChip Robust Multi-array Average*). Em seguida, os dados de expressão gênica estimados passaram por uma etapa de controle de qualidade. Os passos seguintes envolveram a seleção de transcritos diferencialmente expressos, análise de redes e de vias biológicas em cada região cerebral independentemente. No tocante a análise de redes, estimou-se as redes de co-expressão diferencial e redes de dependência condicional dos transcritos para cada região cerebral. Por fim, utilizou-se à análise de redes para identificar interações transcrito-transcrito com papéis potenciais nas patologias da DA a partir do pacote ReactomePA do R, no qual concentrou-se a análise nas dez vias biológicas mais significativas em cada lista de transcritos selecionada.

5.1 Estimação das Expressões Gênicas e Análise Exploratória

Os dados de expressão gênica da Doença de Alzheimer (DA) utilizados no presente trabalho foram inicialmente estudados por Liang et al. (2007a) e Liang et al. (2008). Os autores reuniram

dados de amostras cerebrais pós-morte de 14 indivíduos saudáveis e 34 indivíduos com a DA estudadas pelos Centros de Doença de Alzheimer da Universidade de Washington (St. Louis, Missouri), de Doença de Alzheimer da Universidade Duke (Durham, North Carolina) e pelo *Sun Health Research Institute* (Sun City, Scottsdale, Arizona).

As expressões gênicas foram medidas separadamente entre casos e controles em 6 regiões do cérebro anatomicamente e funcionalmente distintas, as quais são conhecidas por serem diferencialmente vulneráveis as características da DA. Além disso, as regiões incluídas nos dados procuram fornecer uma cobertura de diferentes zonas funcionais do cérebro a partir das regiões mais representativas: sistema límbico (hipocampo), paralímbico (córtex entorrinal e cíngulo posterior), heteromodal (giro frontal superior), unimodal (giro temporal médio) e zonas sensoriais (córtex visual primário). A partir da Figura 5.1 é possível localizar as regiões do cérebro compreendidas nos dados.

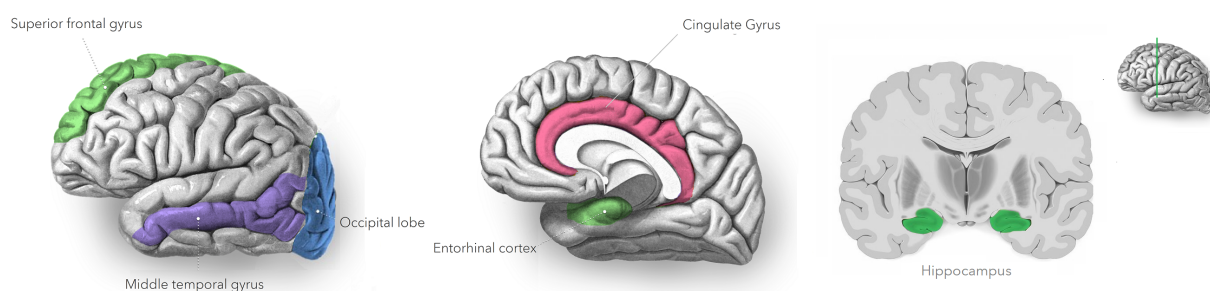


Figura 5.1: Regiões cerebrais coloridas: Cíngulo Posterior (localizado no Giro Cingulado), Córtex Entorrinal, Córtex Visual Primário (localizado no Lobo Occipital), Giro Temporal Médio, Giro Frontal Superior e Hipocampo. Fonte: Adaptado de www.kenhub.com e radiopaedia.org

A base de dados está disponível no site Gene Expression Omnibus (GEO)(número de acesso GSE5281). Ressalta-se que os dados disponibilizados pelos autores passaram por um controle de qualidade prévio, no qual removeram-se observações de baixa qualidade. A Tabela 5.1 apresenta os tamanhos de amostra por região cerebral em ambas as condições (controle e doença

de Alzheimer (DA)). Ao todo, os dados englobam a medição de 54675 transcritos.

Região	Grupo Controle	Grupo DA
Cingulado Posterior (CP)	13	9
Córtex Entorrinal (CE)	13	10
Córtex Visual Primário (CVP)	12	19
Giro Temporal Medial (GTM)	12	16
Giro Frontal Superior (GFS)	11	23
Hipocampo (HIP)	13	10

Tabela 5.1: Tamanho da amostra por região e condição amostral.

5.1.1 Pré-Processamento

A primeira etapa consistiu em realizar o pré-processamento dos dados brutos das seis regiões do cérebro, independentemente. Para tanto, utilizou-se o método GC-RMA via pacote *gcrma* (Wu e James MacDonald Jeff Gentry, 2020) do R.

5.1.2 Controle de qualidade

Após o pré-processamento, espera-se que as distribuições das expressões estimadas tenham formatos e intervalos semelhantes. Caso a distribuição de um chip de microarranjo seja muito diferente das outras, isso pode indicar um problema experimental. Nesse sentido, os chips cujas distribuições são muito diferentes das outras podem não ser adequadas em análises posteriores.

Para uma das regiões do cérebro avaliadas, foram geradas figuras (Figuras 5.2 a 5.7) para avaliação da qualidade dos dados. Cada uma dessas Figuras é composta por quatro gráficos. Os gráficos (a) e (b) apresentam respectivamente *boxplots* das distribuições de expressão gênica por indivíduo na escala original e logarítima com base 2. Os gráficos (c) e (d) apresentam respectivamente as densidades estimadas das expressões gênicas por indivíduos nas escalas original e \log_2 .

Cingulado Posterior (CP)

A Figura 5.2 apresenta os gráficos para controle de qualidade das amostras da região Cingulado Posterior. Visto que as distribuições apresentam-se relativamente homogêneas, conclui-se que todas as amostras estão adequadas para análises posteriores.

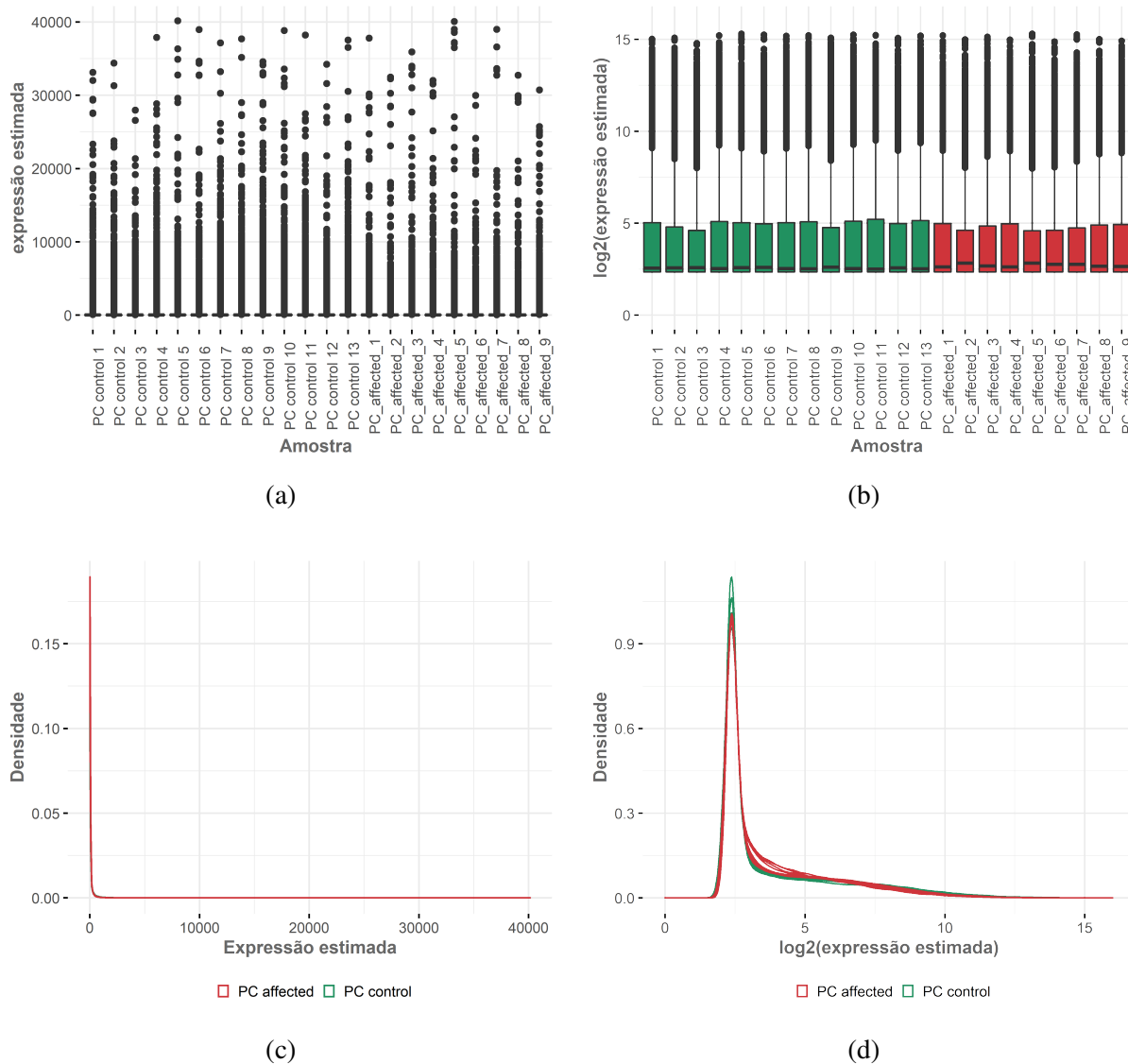


Figura 5.2: Estimativas de expressão na região **Cingulado Posterior** das amostras de controle e de Alzheimer. (a) *Boxplots* das estimativas de expressão na escala original. (b) *Boxplots* dos dados na escala log. (c) Densidade dos dados na escala original. (d) Densidade dos dados na escala log. Fonte: elaboração própria com o uso do pacote *ggplot2* do R.

Córtex Entorrinal (CE)

Observa-se que nenhuma das distribuições dos chips de microarranjo parecem se destacar das demais (Figura 5.3).

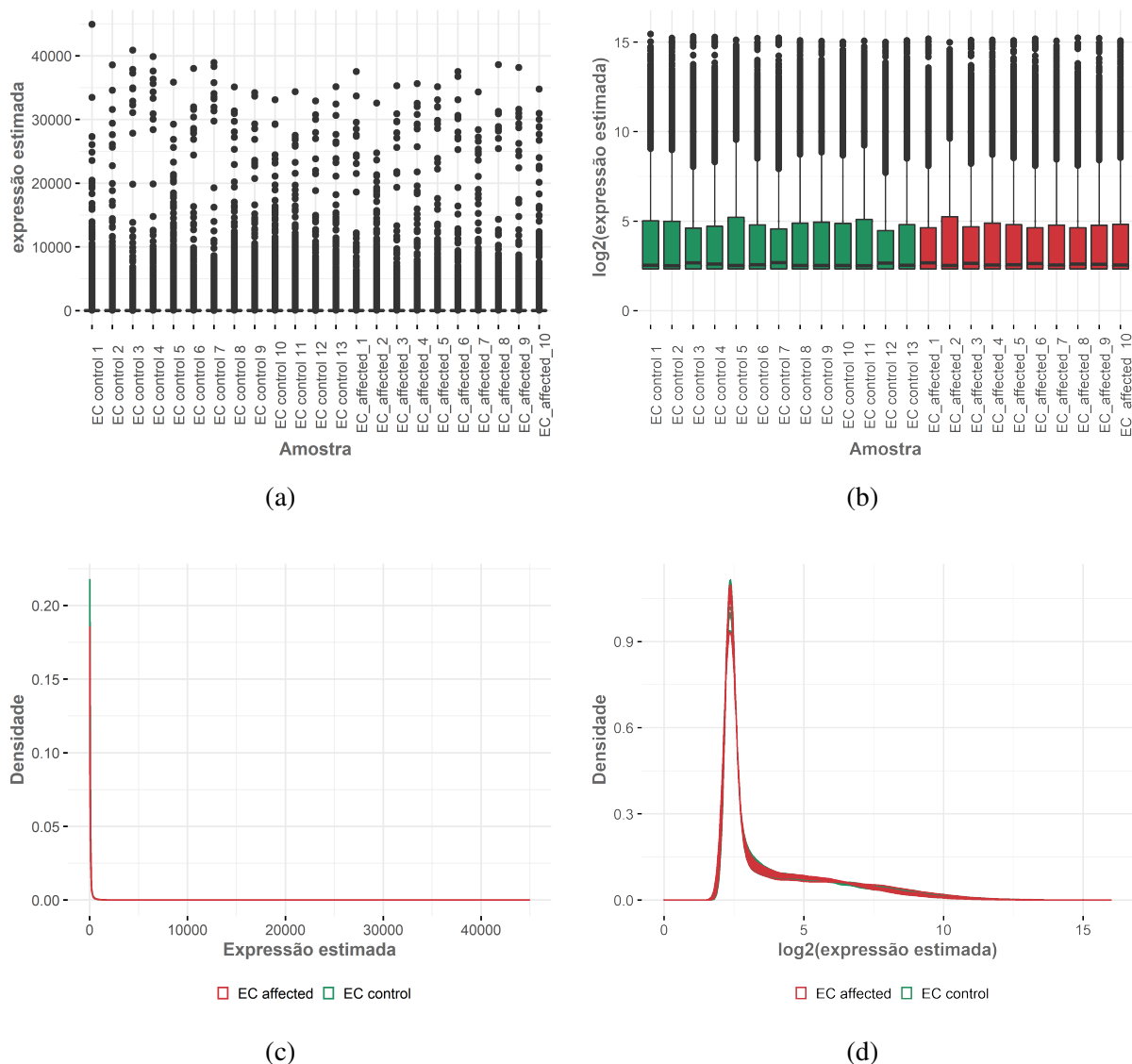


Figura 5.3: Estimativas de expressão na região **Córtex Entorrinal** das amostras de controle e de Alzheimer. (a) *Boxplots* das estimativas de expressão na escala original. (b) *Boxplots* dos dados na escala log. (c) Densidade dos dados na escala original. (d) Densidade dos dados na escala log. Fonte: elaboração própria com o uso do pacote *ggplot2* do R.

Córtex Visual Primário (CVP)

De forma análoga às regiões CE e CP, as amostras para a região CVP mostram-se adequadas para as próximas etapas da análise (Figura 5.4).

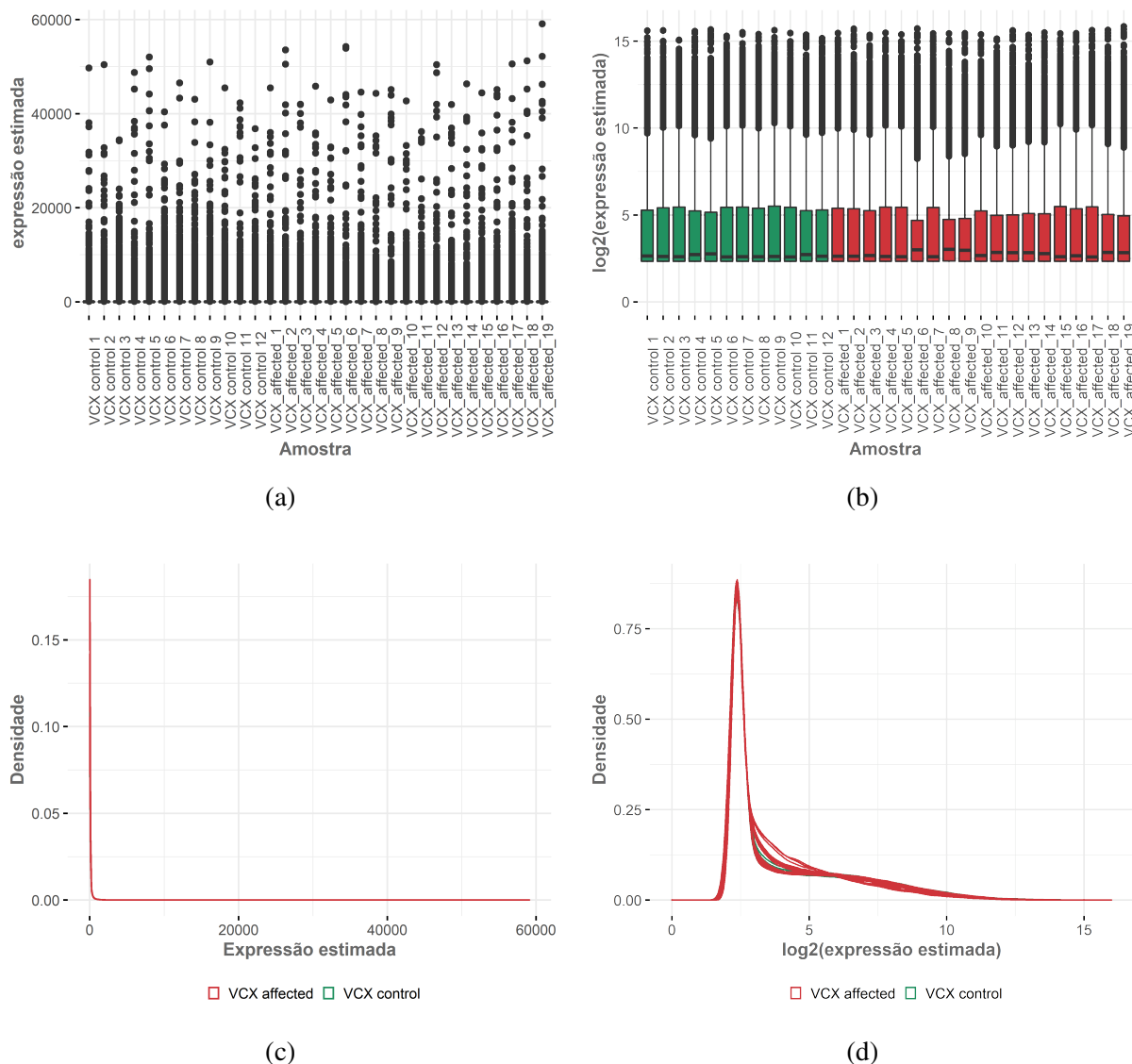


Figura 5.4: Estimativas de expressão na região **Córtex Visual** das amostras de controle e de Alzheimer. (a) *Boxplots* das estimativas de expressão na escala original. (b) *Boxplots* dos dados na escala log. (c) Densidade dos dados na escala original. (d) Densidade dos dados na escala log. Fonte: elaboração própria com o uso do pacote *ggplot2* do R.

Giro Temporal Medial (GTM)

Assim como as regiões previamente analisadas, observa-se que os gráficos para a região GTM não sugerem que as amostras sejam de má qualidade (Figura 5.5).

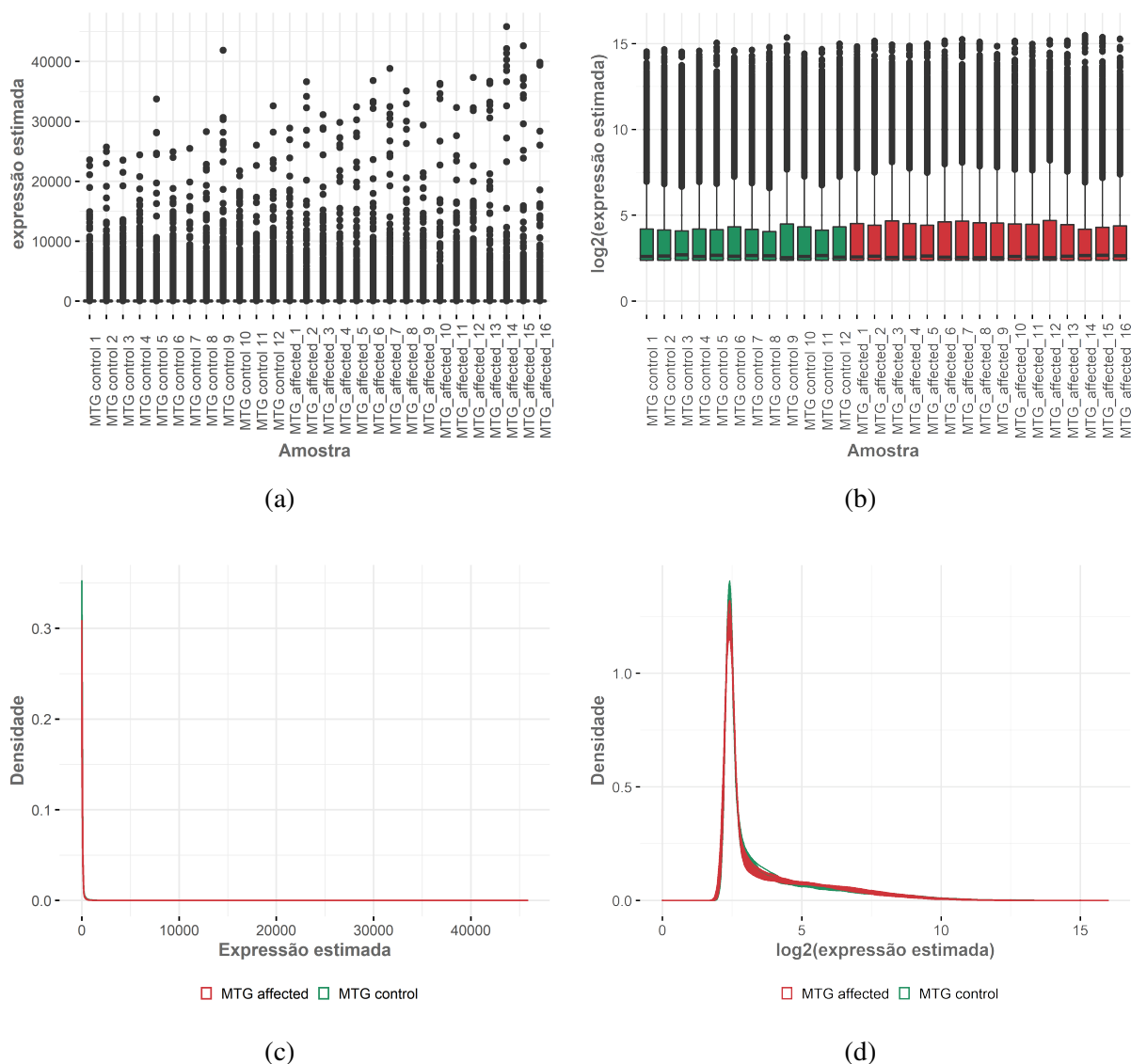


Figura 5.5: Estimativas de expressão na região **Giro Temporal Medial** das amostras de controle e de Alzheimer. (a) *Boxplots* das estimativas de expressão na escala original. (b) *Boxplots* dos dados na escala log. (c) Densidade dos dados na escala original. (d) Densidade dos dados na escala log. Fonte: elaboração própria com o uso do pacote *ggplot2* do R.

Giro Frontal Superior (GFS)

A partir da Figura 5.6, nota-se que a distribuição da expressões estimadas da 11ª amostra de controle diferencia-se nitidamente das demais. Isso sugere uma baixa qualidade dos dados para esse indivíduo (amostra). Portanto, decidiu-se excluí-lo (indivíduo) das análise posteriores.

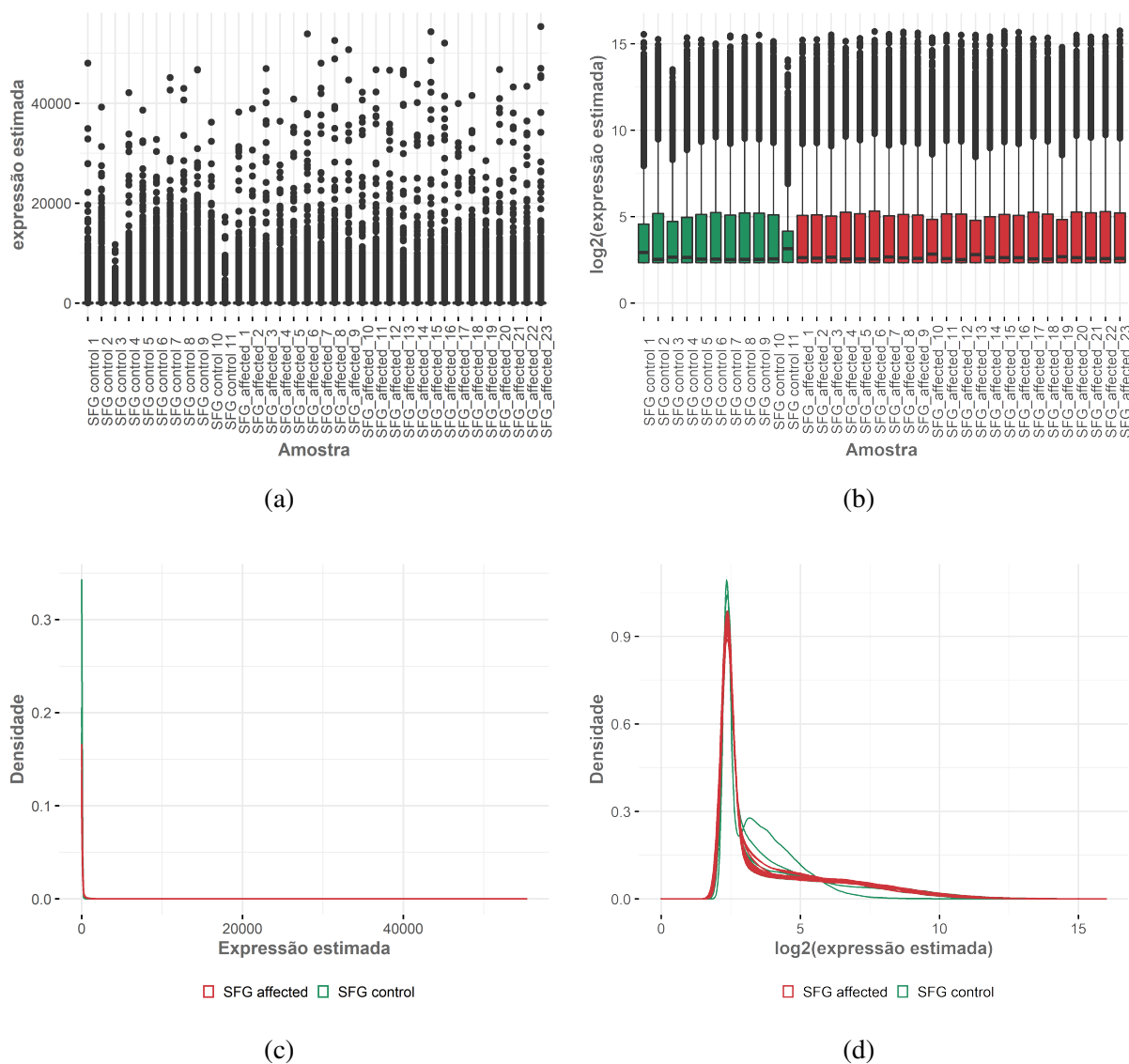


Figura 5.6: Estimativas de expressão na região **Giro Frontal Superior** das amostras de controle e de Alzheimer. (a) *Boxplots* das estimativas de expressão na escala original. (b) *Boxplots* dos dados na escala log. (c) Densidade dos dados na escala original. (d) Densidade dos dados na escala log. Fonte: elaboração própria com o uso do pacote *ggplot2* do R.

Hipocampo (HIP)

Assim como todas as demais regiões, com exceção da região GFS, as amostras do hipocampo apresentam-se devidamente adequadas para as análises posteriores (Figura 5.7).

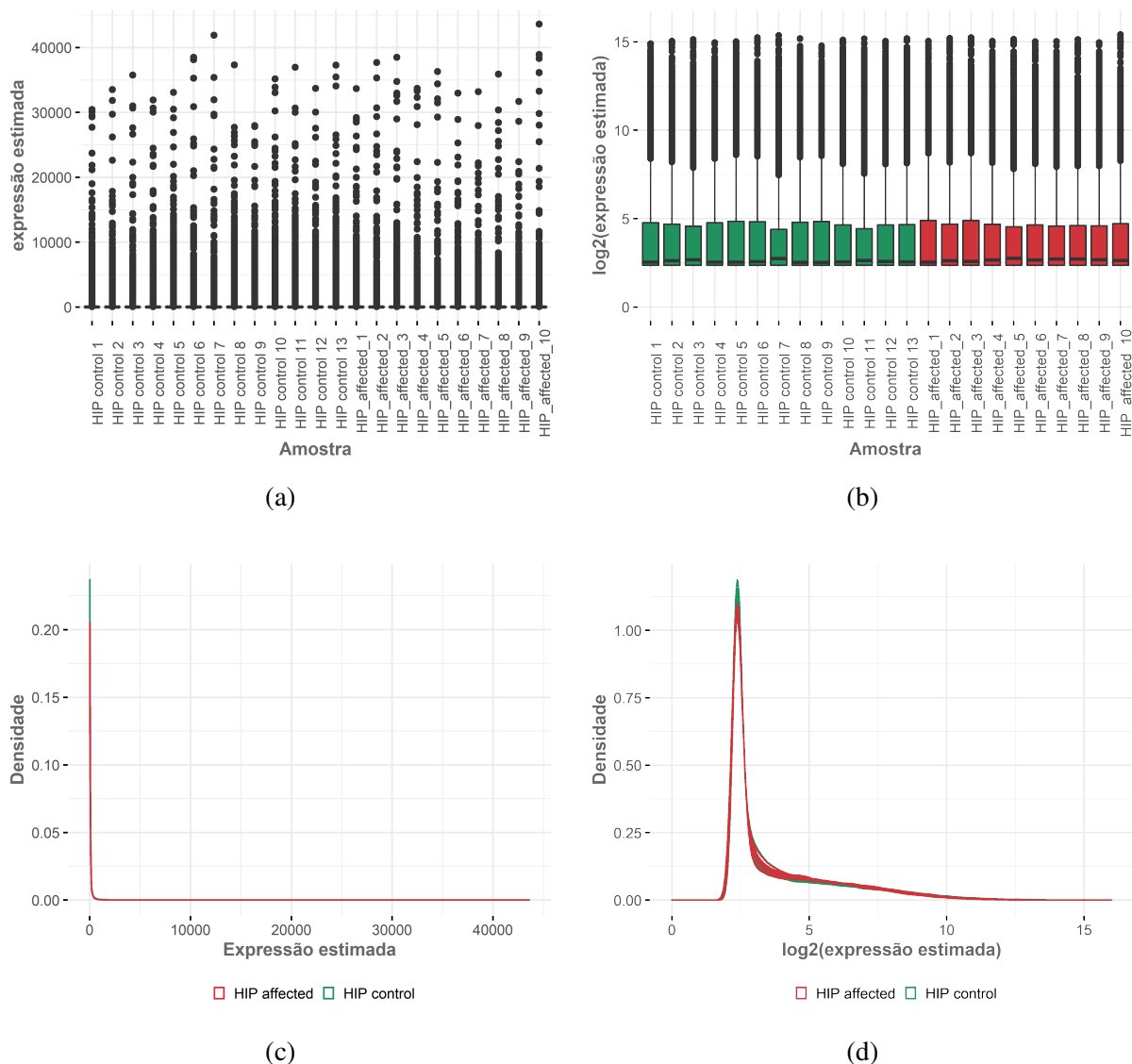


Figura 5.7: Estimativas de expressão na região **Hipocampo** das amostras de controle e de Alzheimer. (a) *Boxplots* das estimativas de expressão na escala original. (b) *Boxplots* dos dados na escala log. (c) Densidade dos dados na escala original. (d) Densidade dos dados na escala log. Fonte: elaboração própria com o uso do pacote *ggplot2* do R.

5.2 Seleção de Transcritos (ou Genes) Diferencialmente Expressos

Após as etapas de pré-processamento e controle de qualidade para cada região do cérebro avaliada, foram selecionados transcritos denominados **diferencialmente expressos** por três diferentes abordagens: uma paramétrica, uma não paramétrica e pelo método PPCLUST. A subseção 5.2.1 descreve os resultados das duas primeiras (paramétrica e não paramétrica) e a subseção 5.2.2, da utilização do método PPCLUST.

5.2.1 Métodos paramétrico e não paramétrico

Para cada abordagem realizada separadamente, os transcritos diferencialmente expressos foram selecionados a partir dos critérios descritos abaixo.

- **Abordagem não-paramétrica:**

- Resultado de teste Mann-Whitney significativo após correção por testes múltiplos pelo método BH (p-valor ajustado $< 0,05$).
- Fold change FC $< 0,25$ ou > 4 , em que FC é a razão entre a mediana de expressão dos casos e a mediana de expressão dos controles para cada transcrito.
- Exclusão dos transcritos cuja mediana das expressões estimadas nas amostras caso e controle estão ambas abaixo do valor 100.

- **Abordagem paramétrica:**

- Resultado de teste t significativo após correção por testes múltiplos pelo método BH (p-valor ajustado $< 0,05$).
- Fold change FC $< 0,25$ ou > 4 , em que FC é a razão entre a média de expressão dos casos e a média de expressão dos controles para cada transcrito.
- Exclusão dos transcritos cuja média das expressões estimadas nas amostras caso e controle estão ambas abaixo do valor 100.

É sabido que os critérios de filtragem e FC são arbitrários, contudo, essas técnicas tem sido popularmente aplicadas no presente contexto devido a sua utilidade. Afinal, tais critérios satisfazem um interesse típico dos pesquisadores ao trabalharem com dados de expressão gênica, que consiste em selecionar apenas aqueles transcritos/genes que tenham significância biológica para o estudo.

Vale ressaltar que os critérios são aplicados de forma independente e sobre o mesmo conjunto inicial de transcritos. Em cada abordagem, a lista resultante de transcritos satisfaz ao mesmo tempo aos três critérios. Nesse sentido, seja T_i o subconjunto de transcritos que satisfaçam o critério i , a lista final de transcritos será composta por $T_1 \cap T_2 \cap T_3$.

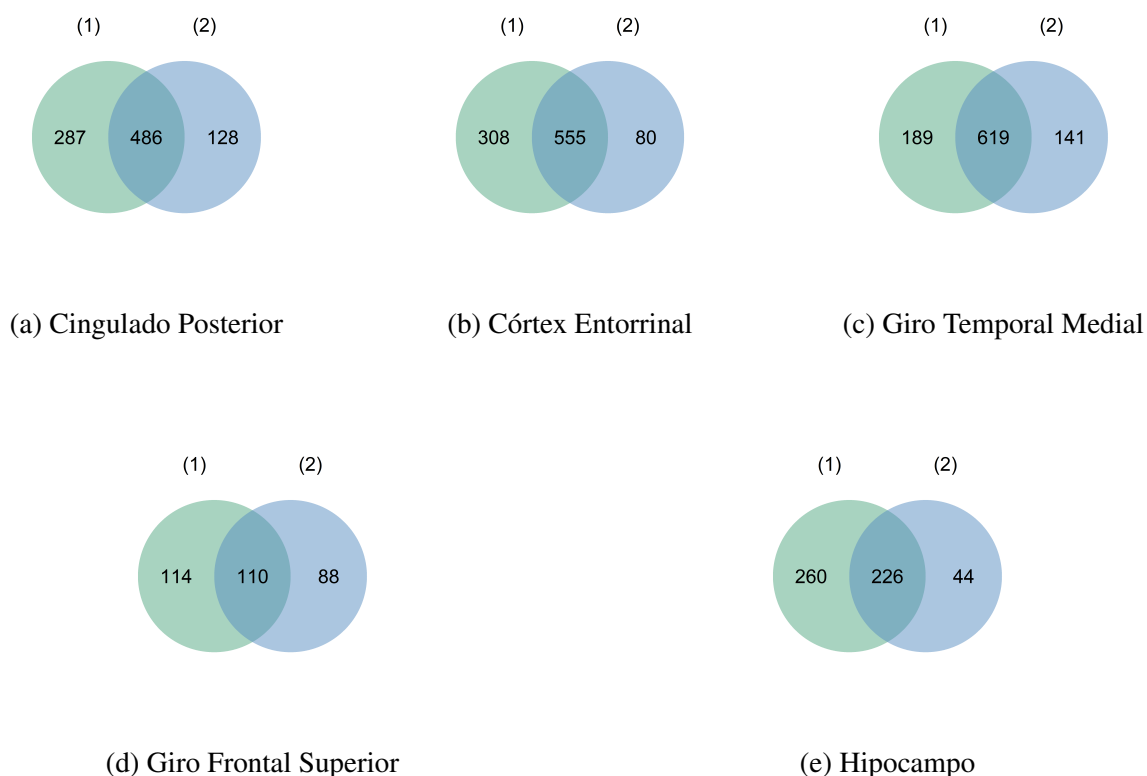


Figura 5.8: Diagramas de Venn, por região cerebral, dos transcritos selecionados pelas duas abordagens. (1) Abordagem não paramétrica; (2) Abordagem paramétrica. Fonte: elaboração própria com o uso do pacote *VennDiagram* do R.

Os diagramas de Venn (Figura 5.8) indicam que boa parte dos transcritos são identificados

como diferencialmente expressos (DE) apesar de um número razoável de transcritos serem detectados por ambas as abordagens (inteseções dos diagramas de Venn), muitos transcritos são identificados apenas pela abordagem não paramétrica. Em parte, esse resultado pode indicar uma desempenho menor do teste t em detectar transcritos DEs, quando comparado ao teste de Mann-Whitney. O contrário seria esperado caso as distribuições fossem aproximadamente normais.

A partir da Figura 5.9 pode-se observar a quantidade de transcritos DEs identificados por cada abordagem e como esses transcritos estão distribuídos nas diferentes regiões do cérebro. Observa-se que alguns transcritos são diferencialmente expressos em mais de uma região, mas a maior parte é identificada como DE em apenas uma região. Tais Figuras ilustram um ponto de partida para a avaliação de diferenças e similaridades de atividade gênica entre regiões do cérebro.

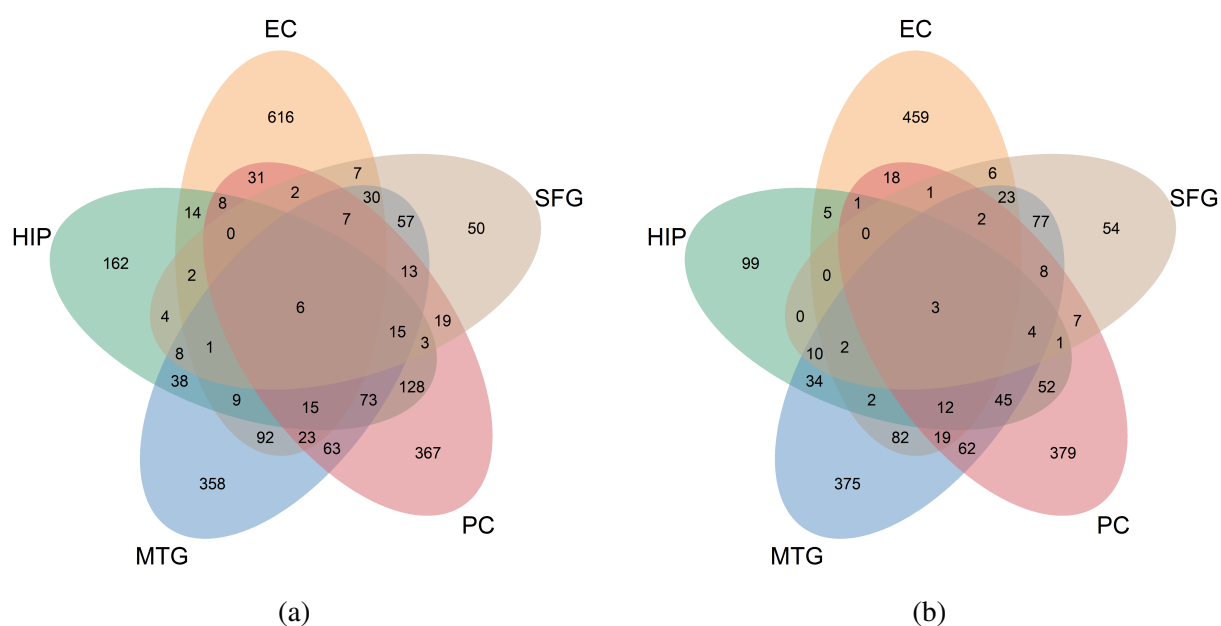
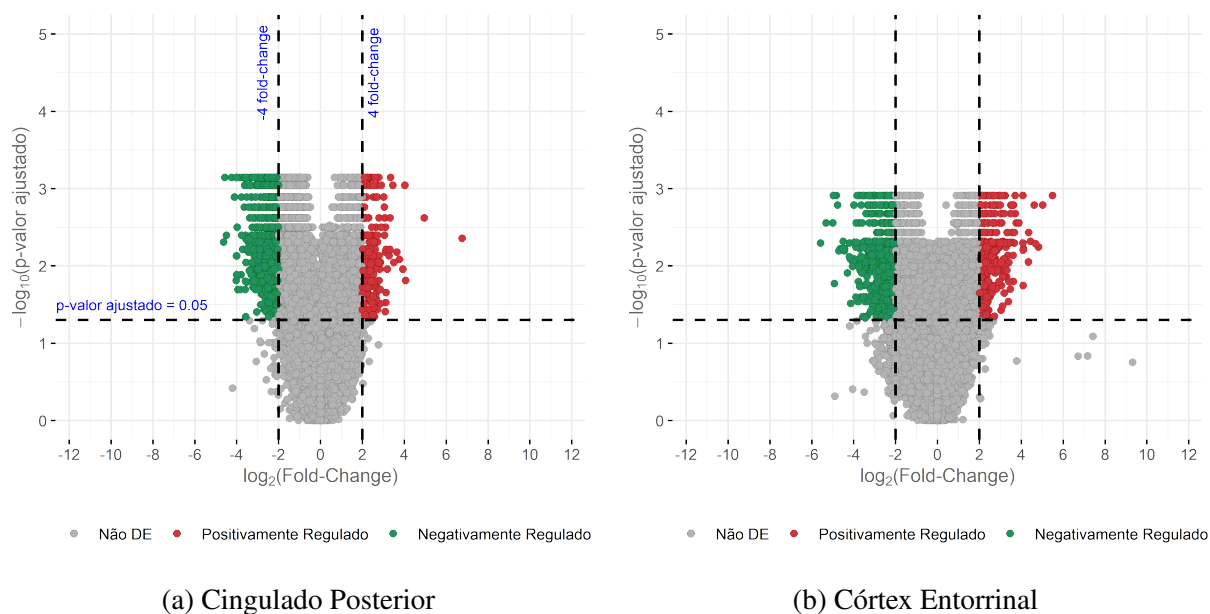


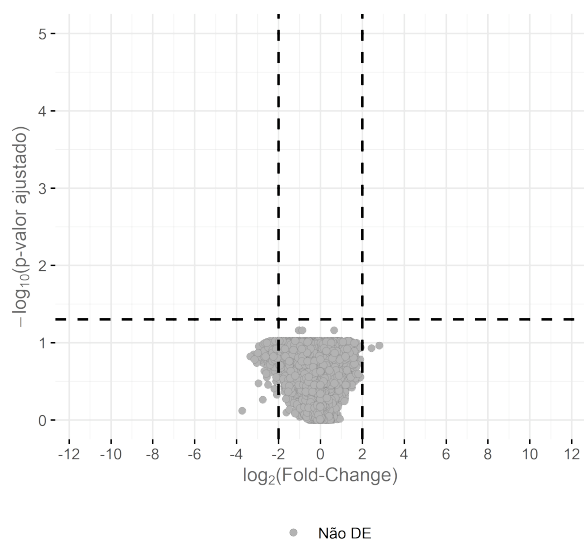
Figura 5.9: **(a)** Diagrama de Venn dos transcritos selecionados utilizando FC na mediana, teste de Mann-Whitney e filtragem na mediana. **(b)** Diagrama de Venn dos transcritos selecionados utilizando FC na média, teste t e filtragem na média. Ressalta-se que a elaboração desta Figura não incorpora nenhuma correção adicional para múltiplos testes de hipótese. Fonte: elaboração própria com o uso do pacote *VennDiagram* do R.

Com base no exposto, utilizou-se as listas de transcritos DE pela abordagem não paramétrica para as análises subsequentes. Dessa forma, o conjunto de elementos foi reduzido à transcritos potencialmente associados com a neuropatologia específica da DA nas diferentes regiões do cérebro.

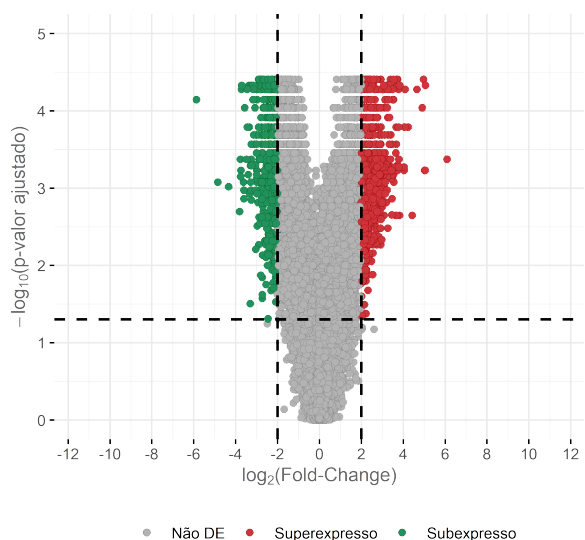
A Figura 5.10 apresentam os gráficos de vulcão para as seis regiões do cérebro. Tais gráficos mostram os cortes que definem a expressão diferencial, com base em p-valores ajustados pelo método BH (na escala $-\log_{10}$) do teste de Mann-Whitney (eixo y) e por *fold change* com base na mediana (eixo x), permitindo a identificação visual dos transcritos que além de estatisticamente significativos, exibem diferenças de grande magnitude (Li, 2012).



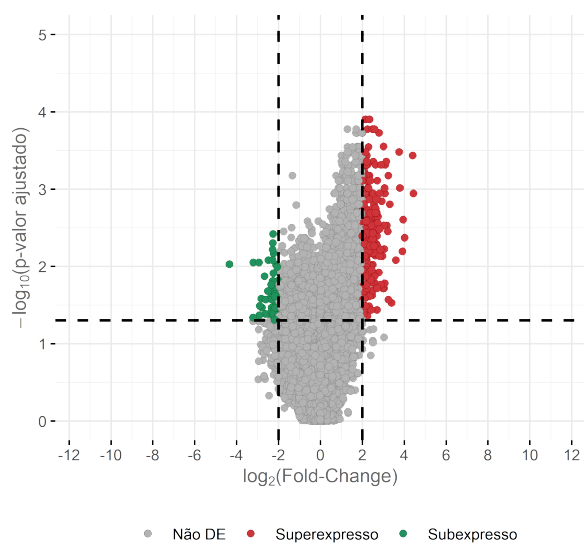
§5.2. Seleção de Transcritos (ou Genes) Diferencialmente Expressos



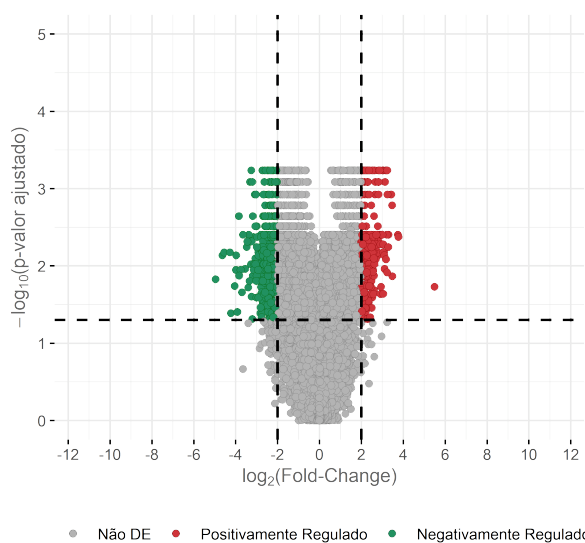
(c) Córtex Visual Primário



(d) Giro Temporal Medial



(e) Giro Frontal Superior



(f) Hipocampo

Figura 5.10: Gráficos de vulcão para as seis regiões cerebrais. Em cada gráfico, destaca-se os transcritos supereexpressos (• p-valor ajustado < 0,05 e $\log_2(FC) > 2$) e transcritos subexpressos (• p-valor ajustado < 0,05 e $\log_2(FC) < -2$). Fonte: elaboração própria com o uso do pacote *ggplot2* do R.

Os gráficos de vulcão indicam que as expressões gênicas diferenciais são relativamente

diferentes entre as seis regiões do cérebro. Em especial, o gráfico de vulcão para o Córtex Visual Primário (CVP) sugere que esta região é distinta durante o desenvolvimento de DA, pois esta foi a única região que não se comporta como as demais. Durante a progressão da DA nas regiões do cérebro, as diferenças de efeito em tal região já foram identificadas em estudos anteriores (Liang et al., 2007b; Liu et al., 2011).

5.2.2 PPCLUST

O método PPCLUST foi utilizado como terceira abordagem utilizada para a seleção de transcritos.

Preparação dos Dados

O passo que antecede aplicação do PPCLUST consiste no cálculo da diferença entre os grupos DA e controle. Para tanto, adotou-se a fórmula abaixo, a qual assemelha-se a medida *fold change*.

$$X_{ij}^* = \log_2(X_{ij}) - \log_2(\bar{Z}_i) = \log_2\left(\frac{X_{ij}}{\bar{Z}_i}\right)$$

em que X_{ij} representa a expressão gênica estimada do transcrito i na amostra j do grupo DA e, \bar{Z}_i é a expressão mediana do transcrito j no grupo controle. Dado que o PPCLUST é invariante a transformações monótonas dos dados, do ponto de vista do algoritmo, essa forma de cálculo torna-se equivalente a $X_{ij}^* = \text{sign}(X_{ij} - \bar{Z}_i)(X_{ij}/\bar{Z}_i)$.

Seleção do Limiar

Uma etapa decisiva na aplicação do PPCLUST e, ainda não explorada pelo autor (von Borries, 2008; von Borries, 2009) e demais trabalhos relacionados consiste na escolha do limiar de agrupamento (von Borries, 2008; von Borries, 2009; Silva, 2012; Souza, 2013). Isto posto, explora-se neste trabalho dois critérios úteis para a escolha do limiar. O primeiro critério consiste na

escolha do limiar a partir do qual o número de grupos resultantes estabiliza-se e, o segundo, em complemento ao primeiro, apoia-se na avaliação do Índice de Rand Ajustado (Subseção 3.2.2) entre os agrupamentos subsequentes.

A Figura 5.11 apresenta o número de grupos obtidos de acordo com diferentes limiares (na escala $-\log_{10}$): 10^{-k} , em que $k = 2, 4, 6, 8, 10, 12, 13, 14, 16, 18$ e 20 . Nota-se que o número de grupos matem-se constante para os limiares apresentados menores ou iguais a 10^{-16} , sugerindo assim, uma estabilização do número de grupos.

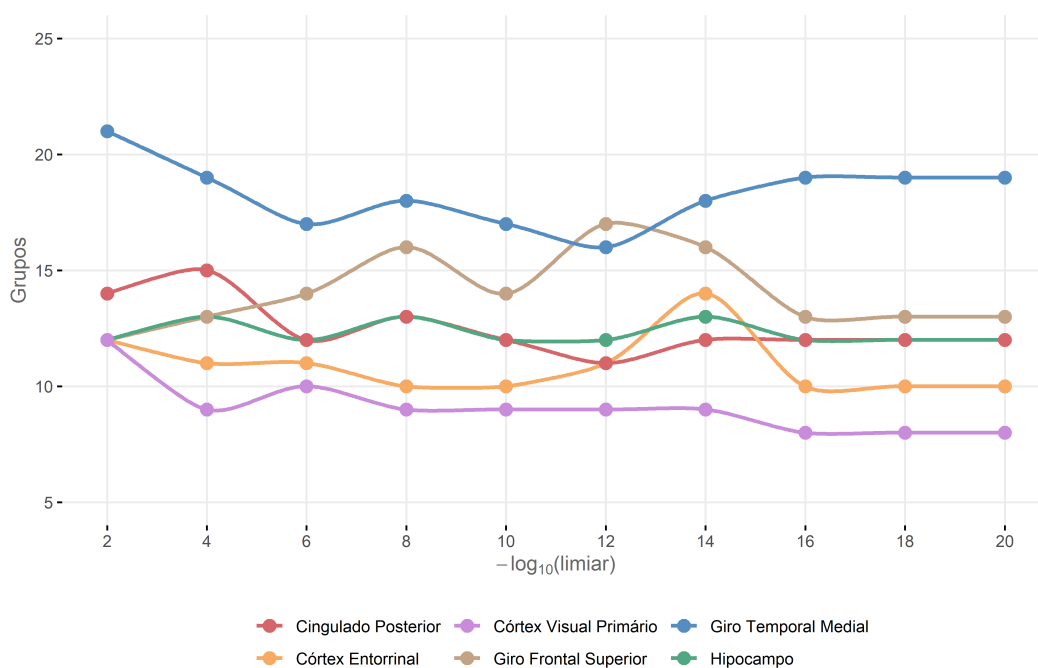


Figura 5.11: Número de grupos obtidos de acordo com os limiares selecionados.

A partir da Figura 5.12 pode-se observar o ARI calculado para cada de agrupamentos subsequentes. Visto que tal índice mede o grau de concordância entre esses agrupamentos, constata-se que o agrupamento resultante é cada vez mais semelhante à medida em que o limiar diminui. Sendo assim, adotou-se o limiar 10^{-16} , haja vista que o agrupamento resultante estabiliza-se para um limiar igual ou inferior a esse.

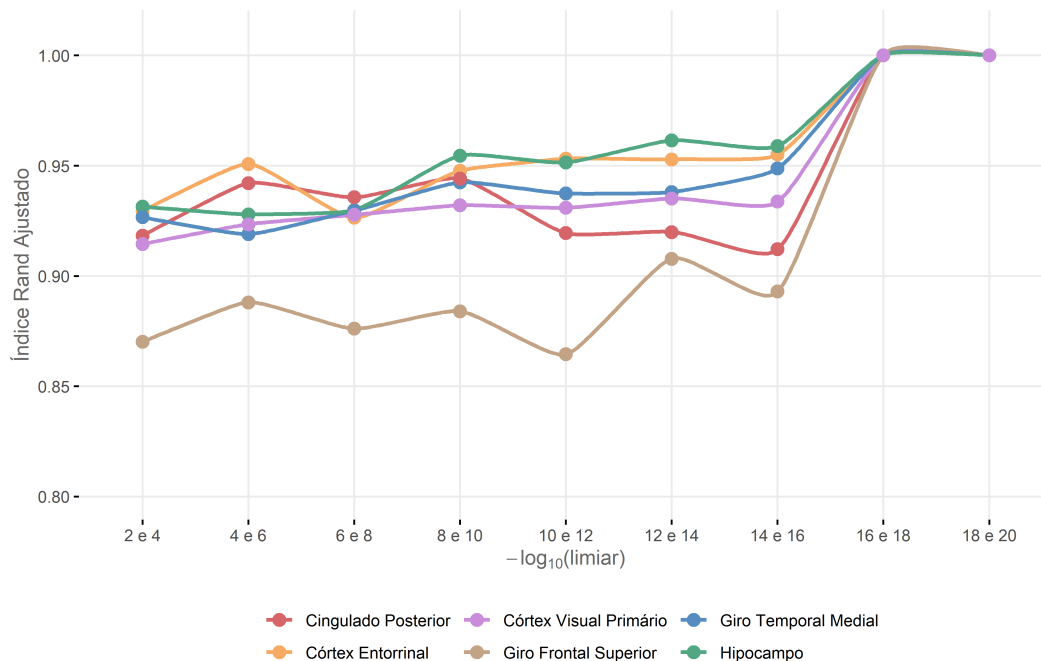


Figura 5.12: Índice Rand Ajustado (ARI) entre os agrupamentos subsequentes.

Para cada região cerebral, dentre os grupos resultantes da aplicação do algoritmo para o limiar 10^{-16} , foram selecionados aqueles com menor e com maior nível de expressão (mais biologicamente significantes). Nesse sentido, tomou-se os grupos com mais baixo nível de expressão cuja soma é maior ou igual a 300 transcritos e, analogamente, os grupos com mais alto nível de expressão cuja soma é maior ou igual 300 transcritos. Sendo assim, tais grupos compõem as listas de transcritos alterados (ver Tabela A.1). Assim sendo, adotou-se a intercessão entre as listas de transcritos diferencialmente expressos obtidos na Subseção 5.2.1 e os das listas obtidas na aplicação do PPCLUST.

5.2.3 Normalidade dos Transcritos Selecionados

A avaliação de normalidade multivariada em dados com muitas variáveis e poucas observações é desafiadora. Os métodos de co-expressão diferencial e de dependência condicional utilizados neste trabalho exigem que os dados tenham uma distribuição aproximadamente normal.

Nesse contexto, o teste Shapiro-Wilk destaca-se como teste de normalidade por ser relativamente mais poderoso em comparação a uma ampla gama de testes alternativos (Thode, 2002). O poder do teste é a probabilidade de se rejeitar a hipótese de normalidade, sendo que de fato, esta hipótese é falsa. Em outras palavras, a probabilidade de detectar transcritos que advém de uma distribuição não normal. Assim, aplicou-se o teste de Shapiro-Wilk para avaliar a normalidade das distribuições dos transcritos selecionados. A Tabela 5.2 apresenta as proporções de transcritos com p-valor $< 0,05$.

Região	Grupo Controle		Grupo Alzheimer	
	Taxa rejeição	<i>n</i>	Taxa rejeição	<i>n</i>
Córtex Entorrinal (CE)	0.216	13	0.208	10
Cingulado Posterior (CP)	0.144	13	0.100	9
Giro Temporal Medial (GTM)	0.165	12	0.081	16
Giro Frontal Superior (GFS)	0.248	10	0.162	23
Hipocampo (HIP)	0.255	13	0.031	10

Tabela 5.2: Taxa de hipóteses de normalidade rejeitadas ao nível de significância de 5% (sem ajuste para múltiplos testes) e tamanhos das amostras por região cerebral e grupo.

O poder do teste é proporcional ao tamanho das amostras, sendo assim, quanto maior o número de amostras, maior será o poder do teste. Contudo, é sabido que o poder dos testes de normalidade é limitado devido ao pequeno tamanho de amostra. Além disso, dados cujas variáveis seguem a distribuição normal univariada não garantem que a distribuição conjunta dessas variáveis seja normal multivariada. Não obstante, os resultados da Tabela 5.2, não revelam sérios desvios de normalidade.

Ao observar a correlação entre os dados e os quantis teóricos da distribuição normal, percebe-se que não há claros indícios de violações preocupantes da suposição de normalidade (Figura 5.13).

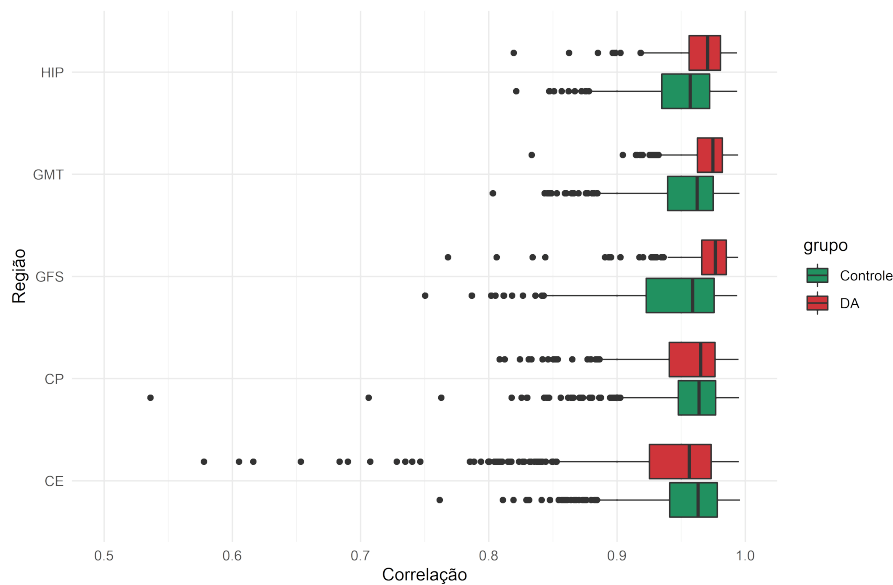


Figura 5.13: Correlação entre os dados e os quantis teóricos da distribuição normal.

5.3 Análise de Redes e de Vias Biológicas

Nesta seção, apresenta-se resultados de estimações das redes de co-expressão diferencial e de dependência condicional para cada uma das regiões do cérebro, exceto a Córtex Visual Primário que não foi analisado por não apresentar transcritos DE. Como também realiza-se a investigação de vias biológicas potencialmente ligadas ao mal de Alzheimer a partir dos transcritos alterados selecionados em conjunto pelos gráficos vulcão (Subseção 5.2.1) e o algoritmo PPCLUST (Subseção 5.2.2) e, com base nas redes estimadas.

5.3.1 Análise de Co-expressão Diferencial e de Vias Biológicas

A análise de rede de expressões gênicas é de especial interesse na análise de dados de expressão gênica, uma vez que entende-se que os genes operam em conjunto. Ademais, a análise de redes podem fornecer uma ferramenta útil para visualizar as relações entre transcritos (ou genes) e para gerar hipóteses biológicas.

A análise de co-expressão diferencial ou análise de correlação diferencial surgiu para obter

informações sobre a diferença nas relações transcrito-transcrito entre as condições de interesse. Em contraste com a análise de expressão diferencial, a co-expressão diferencial opera no nível de pares de transcritos em vez de transcritos individuais. Para tanto, utilizou-se o pacote *DGCA* (do inglês *Differential Gene Correlation Analysis*) (Mckenzie et al., 2016) para identificar as correlações diferenciais entre pares de transcritos nos grupos de casos e controles.

Um ponto importante a destacar sobre os grafos nesta Subseção é que a disposição dos nós é parcialmente aleatória, ou seja, um mesmo grafo pode tornar-se visualmente diferente a cada vez que este é gerado. No entanto, o visual do grafo segue determinados padrões: o grafo tende a ocupar uma área circular, os nós não conectados a nenhum outro localizam-se à margem do grafo e quanto maior a conectividade, maior será a proximidade entre os vértices.

Análise de Vias Biológicas dos Transcritos (ou Genes) Diferencialmente Expressos

Realizou-se análises de vias biológicas com base nas listas de genes diferencialmente expressos para cada região do cérebro, identificados nas seções 5.2.1 e 5.2.2. Inicialmente, vale destacar que os pontos na Figura 5.14 são codificados por cores com base em seus p-valores ajustados correspondentes. Os p-valores indicam quais categorias são mais prováveis de ter relevância biológica. O gradiente de cores varia de vermelho a azul correspondendo a uma ordem crescente de p-valores.

A Figura 5.14 evidencia que a maior parte das vias biológicas identificadas diz respeito a resultados de subexpressão de transcritos entre casos e controles (negativamente regulados), o que indica que tais vias são potencialmente inibidas na presença da doença de Alzheimer. Por exemplo, os genes *CPLX1*, *RAB3A*, *RIMS1* e *STXBP1* estão negativamente regulados na regiões córtex entorrinal, cíngulo posterior, giro temporal medial e hipocampo, os quais estão associados a vias do sistema neuronal e a química cerebral. Por outro lado, as regiões córtex entorrinal e giro frontal superior apresentam os genes *MT1F*, *MT1G*, *MT1X*, *MT2A* positivamente regulados, os quais ligados as vias *Metallothioneins bind metals* e *Response to metal ions*.



Figura 5.14: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos em todas as regiões (ver Tabela A.3). CE: Córtex Entorrinal, CP: Cingulado Posterior, GFS: Giro Frontal Superior, GTM: Giro Temporal Medial, HIP: Hipocampo, NR: Negativamente Regulado, PR: Positivamente Regulado. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

Nas regiões cingulado posterior, giro temporal medial e hipocampo foram encontradas vias subexpressas ligadas a doenças infecciosas em geral e HIV em particular (*Infectious disease* e *HIV infection*). Vale destacar também a presença de vias ligadas ao desgaste dos neurônios como as vias *Cellular response to stress* e *Cellular response to external stimulus*. Isso sugere que tais vias resultem em uma mudança no estado ou atividade dos neurônios como resultado da DA.

Por fim, a região giro frontal superior apresenta um conjunto de vias específicas ativadas: *YAP1- and WWTR1 (TAZ)-stimulated gene expression* (também ativada no giro temporal me-

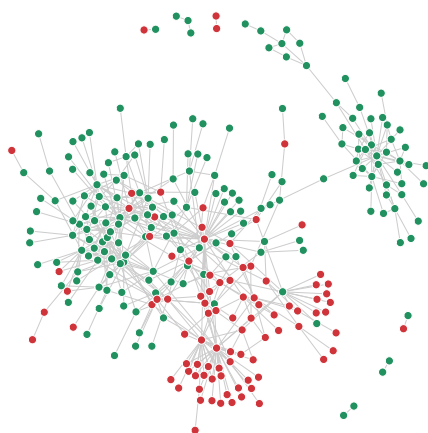
dial), *MET promotes cell motility*, *ECM proteoglycans*, *Signaling by Hippo*, *Initial triggering of complement*, *Syndecan interactions*, *Transport of inorganic cations/anions and amino acids/oligopeptides* e *Post-translational protein phosphorylation*, sendo a última compartilhada com a região hipocampo.

Análise de Vias Biológicas com Base na Co-expressão Diferencial

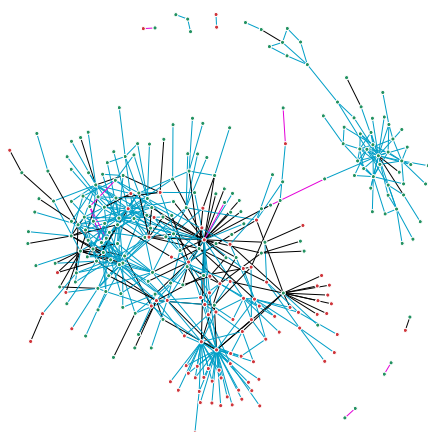
A Figura 5.15 apresenta as redes de co-expressão diferencial para cada região cerebral analisada. Ressalta-se que cada vértice da rede representa um transcrito diferente. As arestas conectam os vértices (transcritos) cuja correlação diferencial é estatisticamente significativa, sendo que apenas os pares de transcritos co-expressos diferencialmente são apresentados.

Com base nas co-expressões diferenciais, dividimos os transcritos (superexpressos e subexpressos) em quatro diferentes listas (correlação com sinais opostos entre os grupos controle e DA; correlação diferente de zero somente no grupo controle; correlação diferente de zero somente no grupo DA; correlação com o mesmo sinal em ambos os grupos) e identificamos as vias biológicas super-representadas (Figura 5.16).

Observa-se que as redes apresentam padrões de conexões diferentes. Na região cíngulo posterior nota-se um padrão de proximidade entre transcritos superexpressos e subexpressos (Figura 5.15a), assim como diferentes transcritos estão conectados apenas em um dos grupos, controle ou DA (Figura 5.15b).



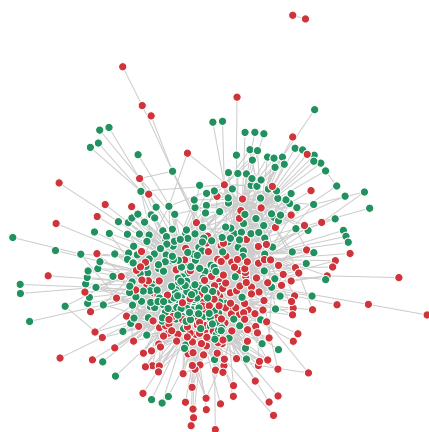
(a) Cingulado Posterior



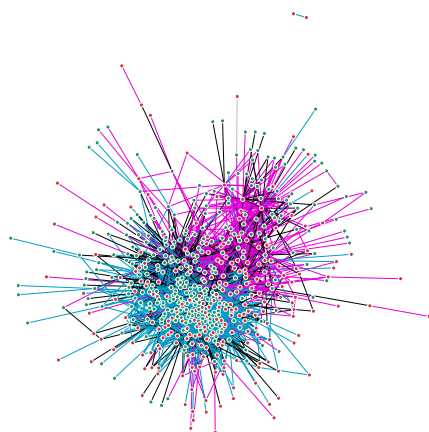
(b) Cingulado Posterior

Figura 5.15: Rede de co-expressão diferencial de transcritos em amostras cerebrais normais (controle) e com DA para a região Cingulado Posterior. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação com sinais opostos entre os grupos controle e DA; — correlação diferente de zero somente no grupo controle; — correlação diferente de zero somente no grupo DA; — correlação com o mesmo sinal em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

Já para a região córtex entorrinal, observa-se uma rede mais complexa e densamente conectada, na qual ocorre uma predominância de duas classificações de transcritos: aqueles correlacionados apenas no grupo controle e outros apenas no grupo DA.



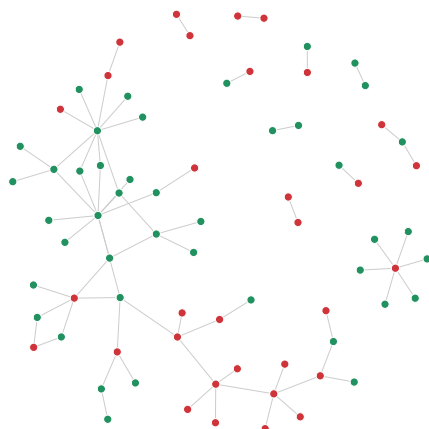
(c) Córtex Entorrinal



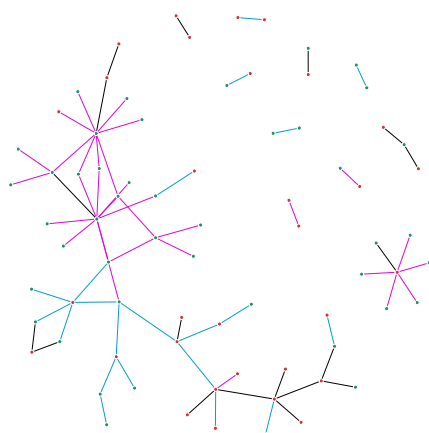
(d) Córtex Entorrinal

Figura 5.15: Rede de co-expressão diferencial de transcritos em amostras cerebrais normais (controle) e com DA para a região Córtex Entorrinal. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação com sinais opostos entre os grupos controle e DA; — correlação diferente de zero somente no grupo controle; — correlação diferente de zero somente no grupo DA; — correlação com o mesmo sinal em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

Por outro lado, a rede para o giro temporal medial é reduzida em comparação às demais (Figuras 5.15e e 5.15e).



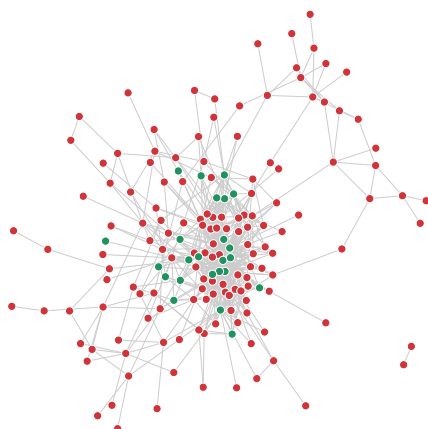
(e) Giro Temporal Medial



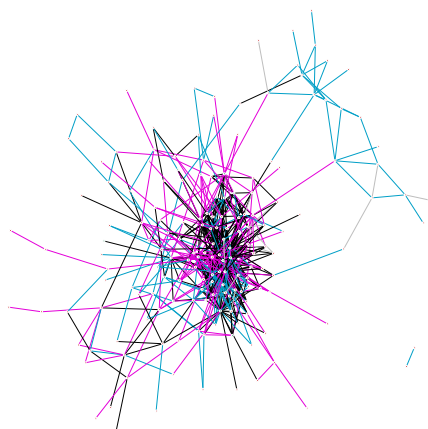
(f) Giro Temporal Medial

Figura 5.15: Rede de co-expressão diferencial de transcritos em amostras cerebrais normais (controle) e com DA para a região Giro Temporal Medial. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação com sinais opostos entre os grupos controle e DA; — correlação diferente de zero somente no grupo controle; — correlação diferente de zero somente no grupo DA; — correlação com o mesmo sinal em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

Ao analisar a rede da região giro frontal superior não se observa um padrão claro entre as co-expressões diferenciais (Figura 5.15h).



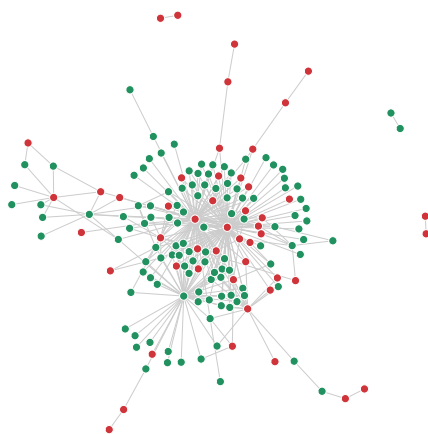
(g) Giro Frontal Superior



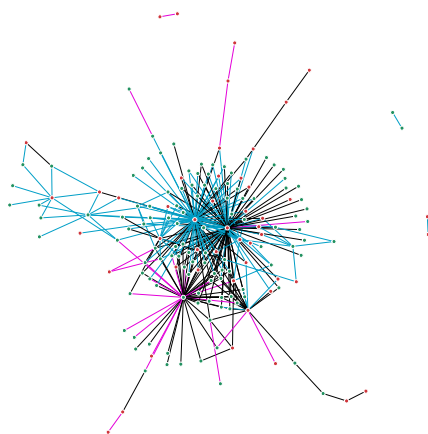
(h) Giro Frontal Superior

Figura 5.15: Rede de co-expressão diferencial de transcritos em amostras cerebrais normais (controle) e com DA para a região Giro Frontal Superior. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação com sinais opostos entre os grupos controle e DA; — correlação diferente de zero somente no grupo controle; — correlação diferente de zero somente no grupo DA; — correlação com o mesmo sinal em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

Por fim, a rede para a região hipocampo exibe ao centro poucos transcritos que são conectados a vários outros (centro da rede).



(i) Hipocampo



(j) Hipocampo

Figura 5.15: Rede de co-expressão diferencial de transcritos em amostras cerebrais normais (controle) e com DA para cada região cerebral. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação com sinais opostos entre os grupos controle e DA; — correlação diferente de zero somente no grupo controle; — correlação diferente de zero somente no grupo DA; — correlação com o mesmo sinal em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

A Figura 5.16 mostra as vias mais significativas encontradas com base na co-expressão diferencial para a região cíngulado posterior. Observa-se a presença de vias ligadas ao processo biológico *electron transport chain*, são elas: *The citric acid (TCA) cycle and respiratory electron transport* e *Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins*. Além disso, para os transcritos cujas correlações são opostas entre o grupo controle e DA (DIF-NR e DIF-PR na Figura 5.16) identificou-se exatamente 8 vias inibidas e 2 vias favorecidas na presença da DA, em sua maioria, relacionados ao funcionamento mitocondrial das células cerebrais. No tocante aos transcritos correlacionados apenas no grupo DA, identificou-se vias propriamente cerebrais, tais como *Neurexins and neuroligins*, *Protein-protein interactions at synapses* e *Neuronal System*.

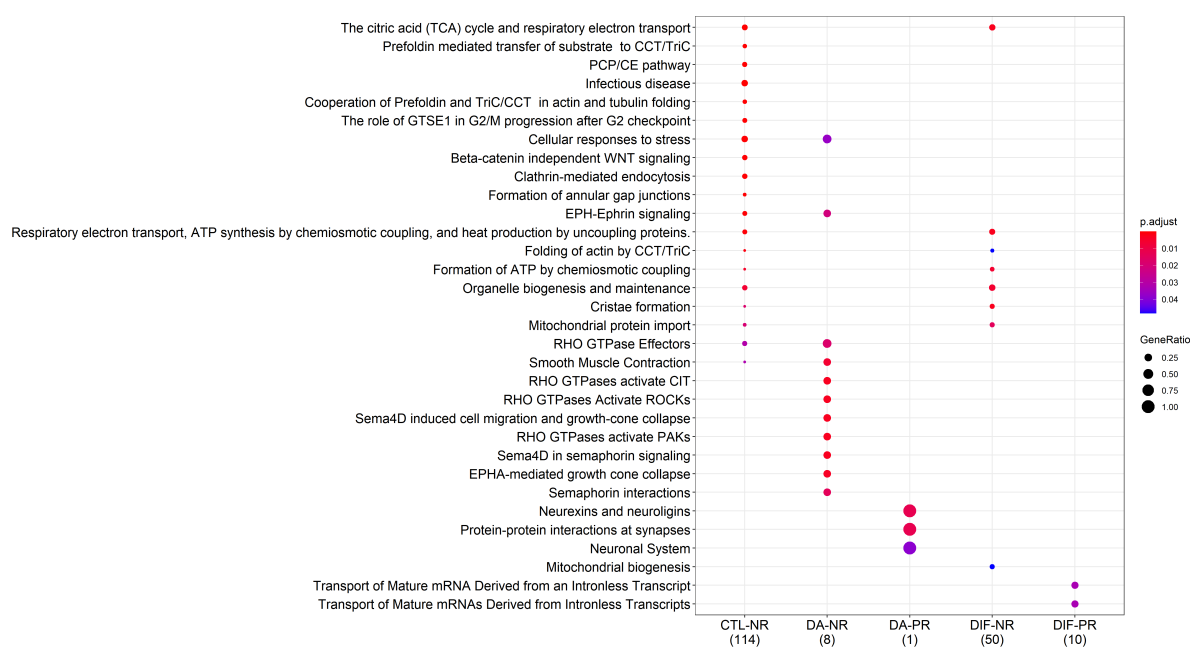


Figura 5.16: As vias biológicas mais significativas identificadas para os transcritos diferencionalmente expressos na região **Cíngulado Posterior** com base na **co-expressão diferencial** (ver Tabela A.6). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *Reac-tomePA* do R.

Os resultados referentes à região córtex entorrinal sugerem uma inibição de vias biológicas

ligadas ao sistema neuronal e aos neurotransmissores, ao passo em que há uma intensificação dos mecanismos resposta, regulação e sinalização (Figura 5.17). Nota-se que algumas vias encontradas para a região córtex entorrinal estão associadas a outras regiões. Essas vias incluem o sistema neuronal de maneira geral e, também a química cerebral (neurotransmissores). Algumas das vias identificadas são específicas desta região, tais como: *Phospholipid metabolism*, *Acetylcholine regulates insulin secretion* e *Free fatty acids regulate insulin secretion*.

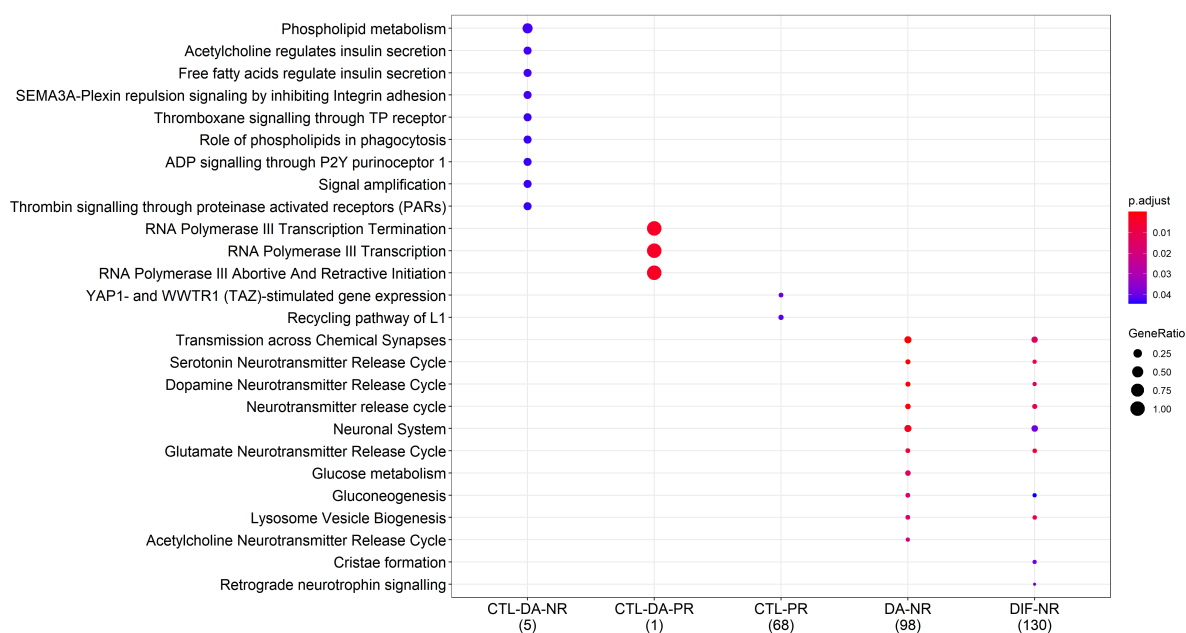


Figura 5.17: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos na região **Córtex Entorrinal** com base na **co-expressão diferencial** (ver Tabela A.4). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

Para a região giro temporal medial, tanto os pares de transcritos co-expressos apenas no grupo controle quanto os pares de transcritos com sinais opostos entre os grupos sugerem um conjunto de vias muito parecidas. Percebe-se também a presença de vias específicas no grupo DA.



Figura 5.18: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos na região **Giro Temporal Medial** com base na **co-expressão diferencial** (ver Tabela A.10). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

Dentre todas as regiões cerebrais, a giro frontal superior foi a que menos retornou vias biológicas super-representadas. Contudo, percebe-se que as vias encontradas também estão super-representadas nas outras regiões.

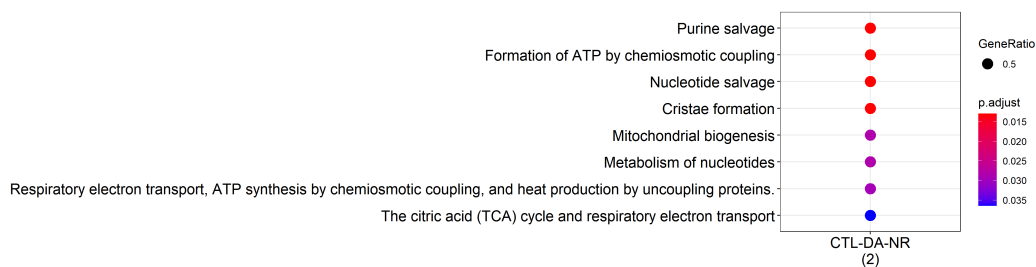


Figura 5.19: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos na região **Giro Frontal Superior** com base na **co-expressão diferencial** (ver Tabela A.12). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

Analisando a Figura 5.20, percebe-se que o conjunto de vias biológicas super-representadas nos transcritos negativamente regulados do grupo controle (CTL-NR) são similares às vias identificadas para os pares de transcritos negativamente regulados cuja co-expressão é diferente entre os grupos (DIF-NR).

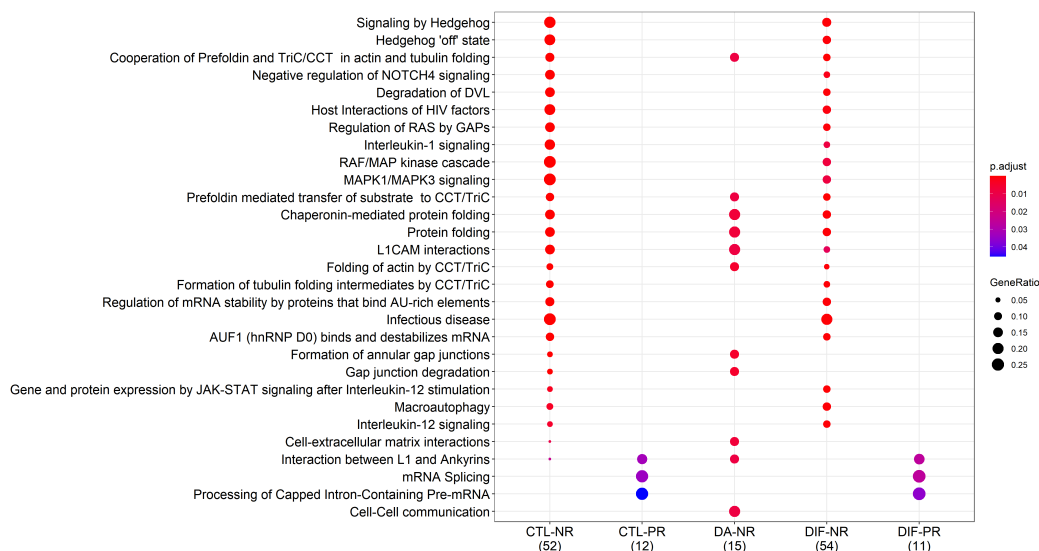


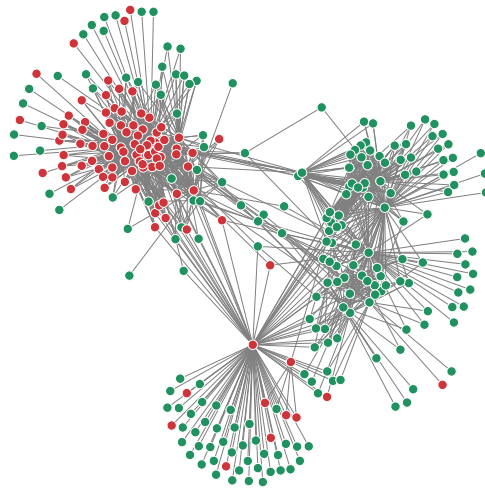
Figura 5.20: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos na região **Hipocampo** com base na **co-expressão diferencial** (ver Tabela A.8). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

5.3.2 **Análise de Dependência Condicional e de Vias Biológicas**

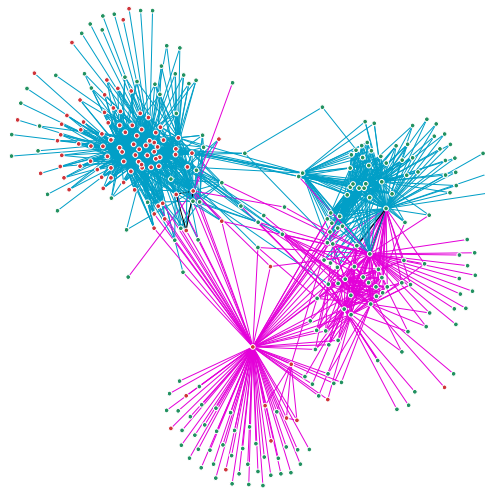
Nesta Subseção, estimam-se as redes de dependência condicional (modelos gráficos) para os grupos controle e doença de Alzheimer (DA), separadamente. A princípio, seria esperado que ambas as redes sejam semelhantes entre si, uma vez que ambas são baseados no mesmo tipo de tecido (neurônios), contudo, há diferenças importantes decorrentes do fato de que as redes de transcritos (ou genes) são frequentemente desreguladas na DA. Isto posto, utilizamos o método glasso para estimar as redes de dependência condicional dos grupos controle e DA (ou seja, duas matrizes de precisão), separadamente e para cada região cerebral.

Uma vez que o objetivo deste trabalho é a visualização e a geração de hipóteses, foi adotado um alto valor para o parâmetro λ (1, 5) de tal forma a obter redes altamente esparsas. Em seguida, para cada região cerebral, comparou-se ambas as matrizes de precisão a fim de identificar os pares de transcritos cujo a correlação parcial é diferente. Então, classificou-se os pares com base nos padrões de mudança entre os grupos controle e DA.

Analisando as redes de dependência condicional, nota-se uma divisão clara entre transcritos conectados em cada um dos grupos (controle ou DA), com exceção da rede para a região giro frontal superior (Figuras 5.21g e 5.21h). Para esta última rede percebe-se um aspecto de emaranhado. Dentre todas as redes, a referente ao hipocampo apresenta a maior esparsidade, revelando um padrão de correlação parcial simplificado em relação às demais redes (Figuras 5.21i e 5.21j).

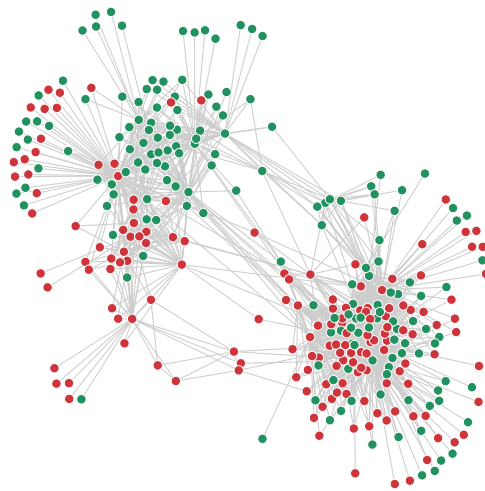


(a) Cingulado Posterior

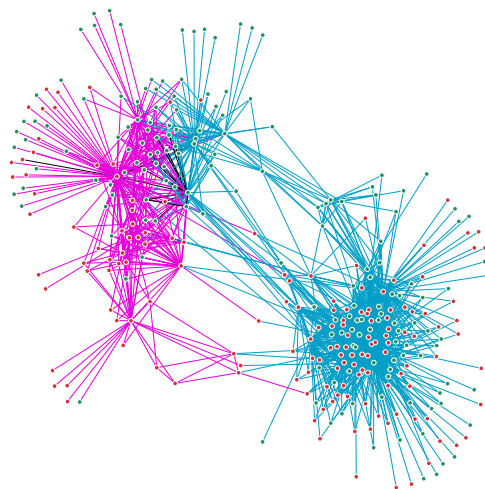


(b) Cingulado Posterior

Figura 5.21: Rede de dependência condicional de transcritos em amostras cerebrais normais (controle) e com DA para a região Cingulado Posterior. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação parcial diferente de zero somente no grupo controle; — correlação parcial diferente de zero somente no grupo DA; — correlação parcial diferente de zero em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

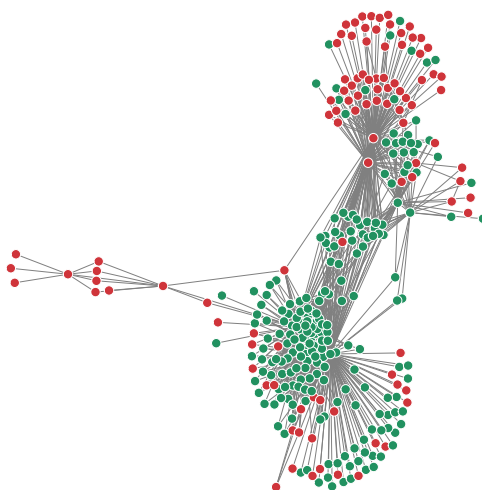


(c) Córtex Entorrinal

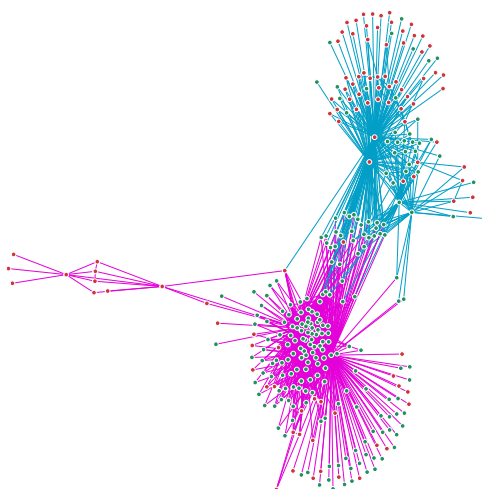


(d) Córtex Entorrinal

Figura 5.21: Rede de dependência condicional de transcritos em amostras cerebrais normais (controle) e com DA para a região Córtex Entorrinal. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação parcial diferente de zero somente no grupo controle; — correlação parcial diferente de zero somente no grupo DA; — correlação parcial diferente de zero em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

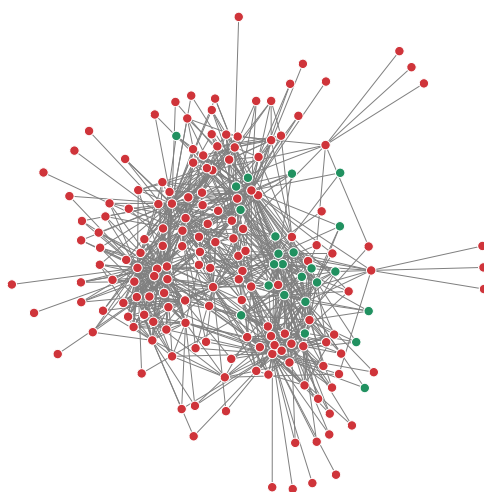


(e) Giro Temporal Medial

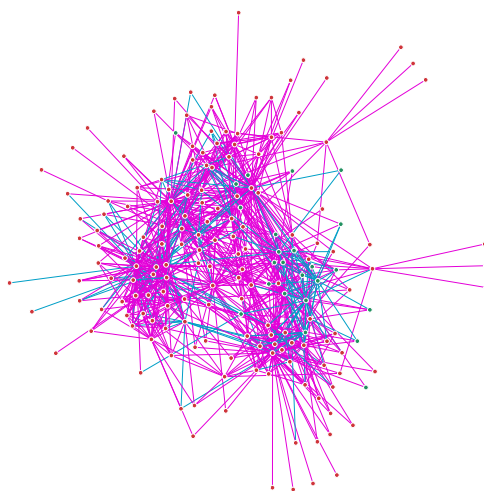


(f) Giro Temporal Medial

Figura 5.21: Rede de dependência condicional de transcritos em amostras cerebrais normais (controle) e com DA para a região Giro Temporal Medial. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação parcial diferente de zero somente no grupo controle; — correlação parcial diferente de zero somente no grupo DA; — correlação parcial diferente de zero em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

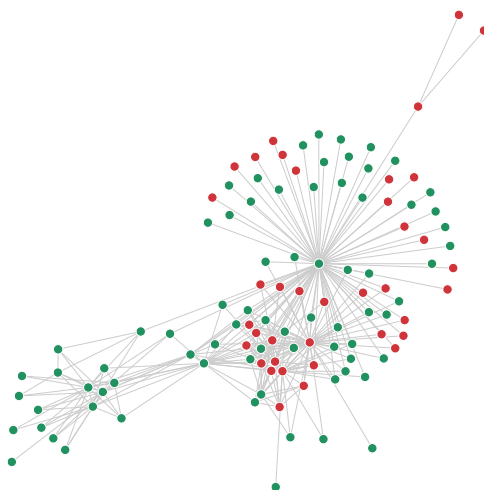


(g) Giro Frontal Superior

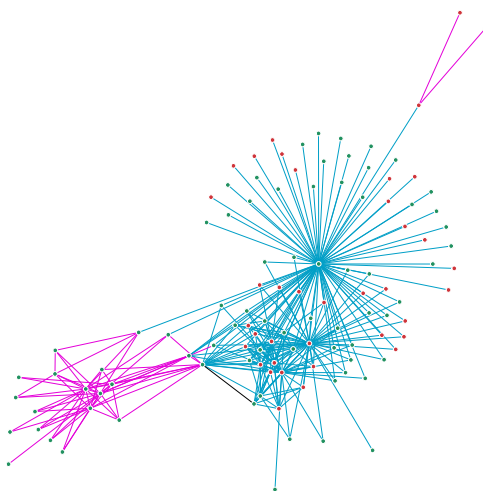


(h) Giro Frontal Superior

Figura 5.21: Rede de dependência condicional de transcritos em amostras cerebrais normais (controle) e com DA para a região Giro Frontal Superior. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação parcial diferente de zero somente no grupo controle; — correlação parcial diferente de zero somente no grupo DA; — correlação parcial diferente de zero em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.



(i) Hipocampo



(j) Hipocampo

Figura 5.21: Rede de dependência condicional de transcritos em amostras cerebrais normais (controle) e com DA para a região Hipocampo. **(a)** ● transcrito subexpresso; ● transcrito superexpresso. **(b)** — correlação parcial diferente de zero somente no grupo controle; — correlação parcial diferente de zero somente no grupo DA; — correlação parcial diferente de zero em ambos os grupos. Fonte: elaboração própria com o uso do pacote *igraph* do R.

A análise de super-representação identificou vias significativas somente para os transcritos negativamente regulados (Figura 5.22). No geral, os transcritos conectados nos diferentes grupos (controle e DA) estão ligados às mesmas vias, com exceção das vias *Neurotransmitter receptors and postsynaptic signal transmission* e *Transmission across Chemical Synapses*, as quais estão também associadas aos transcritos conectados em ambos os grupos (CTL-DA-NR). Ressalta-se também que as vias denominadas *The citric acid (TCA) cycle and respiratory electron transport* (genes ARHGEF7, GABRA1, NCALD, NEFL, NSF, RASGRF2 e SYN2) e *Clathrin-mediated endocytosis* (genes ACTB, ACTG1, AP2S1, ARPC2, CLTB, COPS4 e UBC) apresentam genes distintos das demais. As vias restantes apresentam genes em comum e estão associadas aos grupos controle e DA, independentemente (genes ANAPC5, PSMA1, PSMA2, PSMD1, PSMD6 e UBC).

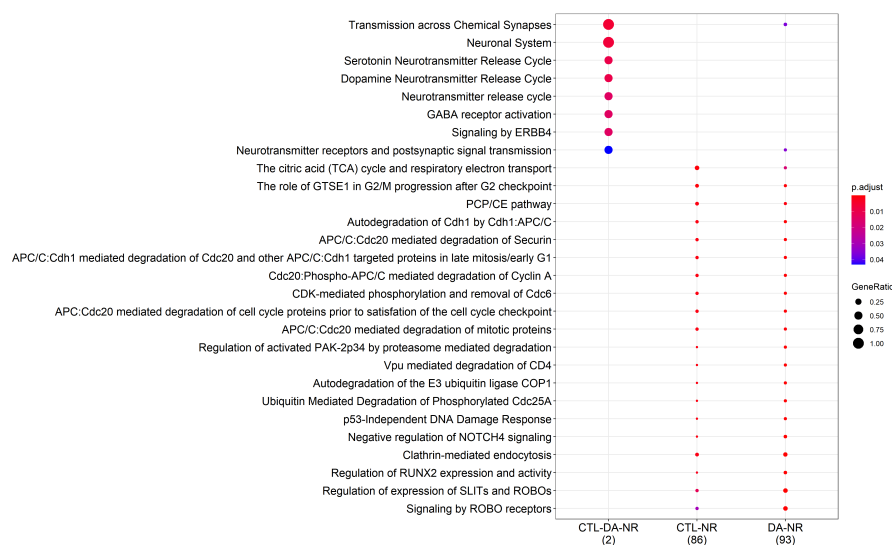


Figura 5.22: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos na região **Cingulado Posterior** com base na **dependência condicional** (ver Tabela A.7). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

Para região córtex entorrinal não foram identificadas vias biológicas significativas para transcritos positivamente regulados no grupo DA (Figura 5.23). Observou-se também que as

vias negativamente reguladas tendem a ser as mesmas, independente do grupo ao qual estão super-representadas.



Figura 5.23: As vias biológicas mais significativas identificadas para os transcritos diferencionalmente expressos na região **Córtex Entorrinal** com base na **dependência condicional** (ver Tabela A.5). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *Reac-tomePA* do R.

A Figura 5.24 mostra as vias mais significativas encontradas com base na dependência condicional para a região giro temporal medial. Somente para esta região encontraram-se vias super-representadas na lista de transcritos positivamente regulados co-expressos somente no grupo DA, são elas: *Initial triggering of complement*, *Regulation of Complement cascade* e *Complement cascade*, as quais estão ligadas aos genes DAC4A, C4B, C4B2.

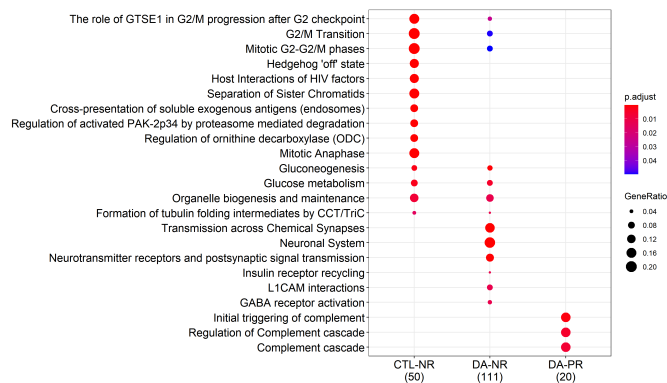


Figura 5.24: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos na região **Giro Temporal Medial** com base na **dependência condicional** (ver Tabela A.11). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

Diferentemente das demais regiões analisadas, destaca-se que as vias identificadas para o giro frontal superior são completamente excludentes de acordo com o padrão de dependência condicional (Figura 5.25).

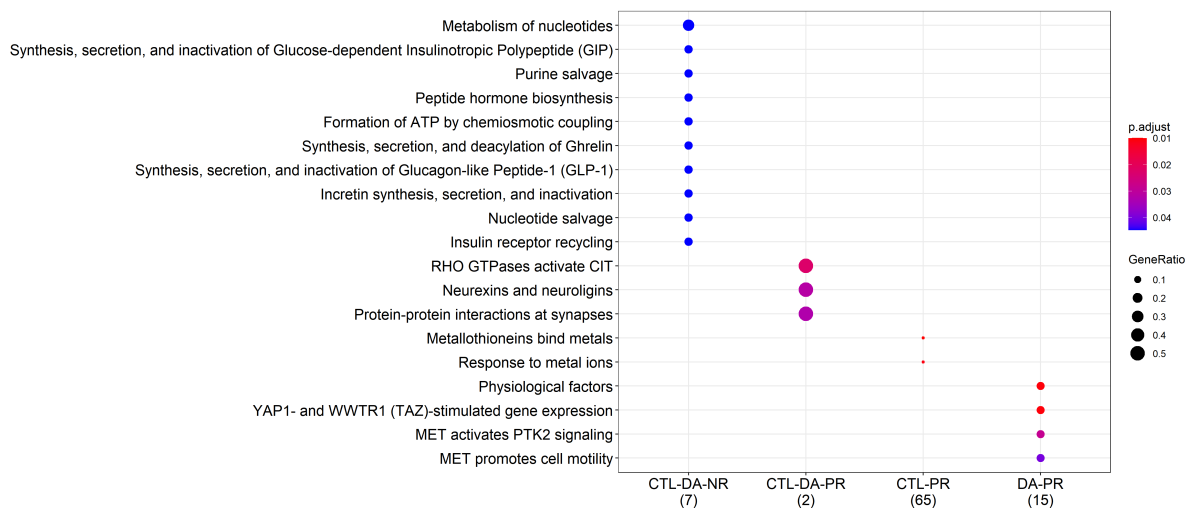


Figura 5.25: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos na região **Giro Frontal Superior** com base na **dependência condicional** (ver Tabela A.13). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

Constata-se que a maior parte das vias identificadas para a região hipocampo são potencialmente inibidas na presença da DA, visto que estas vias estão super-representadas para os transcritos negativamente regulados nessa região (Figura 5.26).

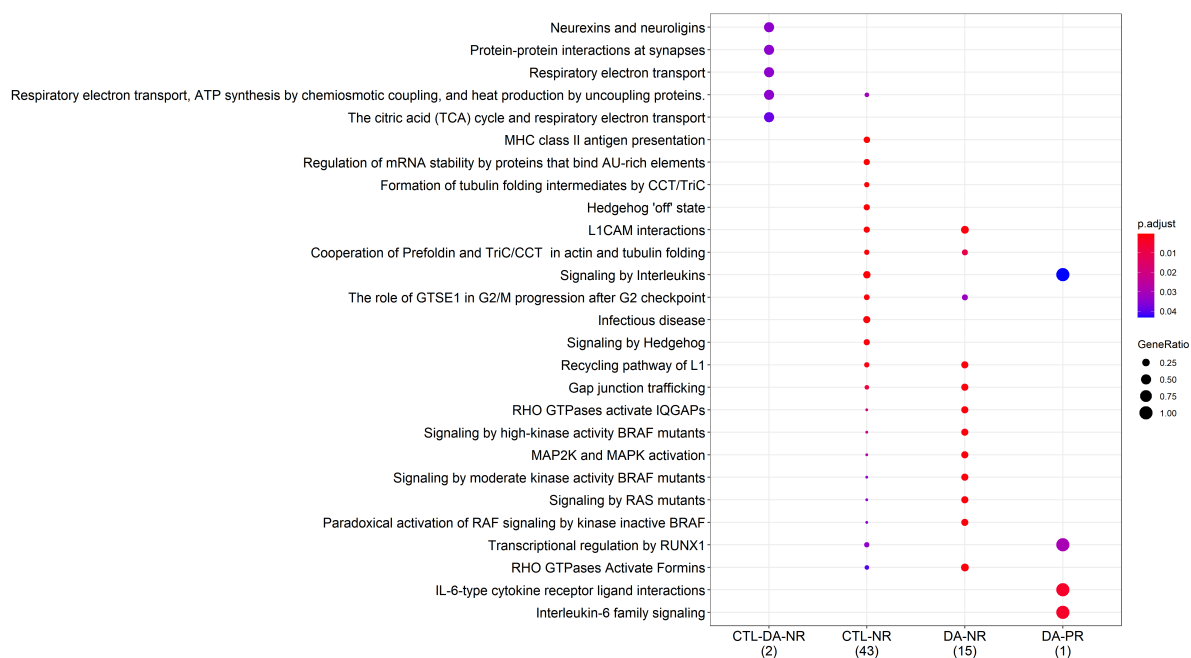


Figura 5.26: As vias biológicas mais significativas identificadas para os transcritos diferencialmente expressos na região **Hipocampo** com base na **dependência condicional** (ver Tabela A.9). CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Fonte: elaboração própria com o uso do pacote *ReactomePA* do R.

Capítulo 6

Discussão e Conclusão

O presente trabalho aplicou técnicas estatísticas para o pré-processamento de dados de microarranjo da doença de Alzheimer (DA) disponíveis publicamente, bem como, identificou mudanças nos perfis de expressão gênica de controles saudáveis e indivíduos com a DA, ao passo que identificou mecanismos biológicos potencialmente associados ao desenvolvimento da doença. Nesse sentido, demonstrou-se um fluxo de trabalho que iniciou-se no preparo dos dados brutos, seleção de transcritos (genes) de interesse e, uma combinação de análise de redes com a análise de super-representação de vias biológicas.

Inicialmente, constatou-se que não houve transcritos diferencialmente expressos para a região Córtex Visual Primário (CVP) de acordo com os critérios adotados, sugerindo que esta região não tem um papel importante na DA e a mesma é poupada na presença da doença, como esperado de acordo com uma literatura prévia. Além disso, embora os genes APP, APOE e PSEN sejam reconhecidos como fatores de risco para a DA, tais genes não foram identificados como diferencialmente expressos em qualquer uma das seis regiões do cérebro em nossas análises ou em Wang e Liu (2019), que utilizaram os mesmos dados.

Como interações transcrito-transcrito podem fornecer um recurso valioso para uma maior exploração dos mecanismos biológicos subjacentes da doença, realizou-se uma investigação acerca da co-expressão diferencial e da dependência condicional em cada uma das regiões ce-

rebrais. As análises de redes de expressão gênica evidenciaram as mudanças significativas que ocorrem na relação entre os transcritos na presença da DA. A partir dos pares de transcritos identificados na análise de redes, procedeu-se com a análise de vias biológicas com o intuito de identificar os potenciais mecanismos biológicos ligados à DA.

Entre todas as vias identificadas com significância estatística, as vias relacionadas ao sistema neuronal, transmissão química entre as sinapses e doenças infecciosas foram reiteradamente observadas. Alguns estudos indicam que algumas doenças infecciosas, bem como a causada pelo vírus herpes tipo I, podem desencadear alguns casos da DA (Itzhaki, 2014). As vias recorrentemente encontradas como a *Infections Disease* e *HIV infection* corroboram com a hipótese de contribuição viral para a DA. Readhead et al. (2018) estuda os mesmos dados analisados neste trabalho com o intuito de esclarecer o papel de doenças infecciosas no desenvolvimento da DA, especialmente o *herpesvirus 6A* e 7. Além disso, estudos revelam danos na cadeia de transporte de elétrons (*electron transport chain*) em pacientes com a DA, sugerindo que essa disfunção mitocondrial desempenha um papel importante na fisiopatologia da doença (Feldhaus et al., 2011).

Com relação ao número pequeno de amostras, constatou-se que a provável perda de poder estatístico não impediu que muitos transcritos diferencialmente expressos fossem encontrados, ao passo que os perfis de expressão gênica alterados fossem identificados mesmo após o ajuste para múltiplos testes. Assim, mesmo com tamanhos de amostra limitados, os dados mostraram-se úteis para investigar a associação de vias biológicas alteradas no mal de Alzheimer.

Esse trabalho estimou as matrizes de precisão dos grupos controle e DA independentemente a partir do estimador *glasso*. Nesse contexto, Danaher, Wang e Witten (2012) apresenta uma extensão do *glasso* para estimar várias matrizes de precisão ao mesmo tempo e, então, comparar as matrizes estimadas. O artigo argumenta que o *joint graphical lasso* leva a estimativas mais acuradas que o *glasso*. Sendo assim, uma possibilidade de trabalho futuro envolveria a comparação entre os resultados obtidos pelos dois métodos.

Por fim, o presente trabalho aplicou uma abordagem baseada em análise de rede pode ser

útil na identificação de mecanismos biologicamente para um estudo mais aprofundado, fornecendo, portanto, uma lista de genes e vias biológicas identificadas para investigações adicionais (Apêndice A).

Apêndice A

Apêndice

Região	Grupo																		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Córtex Entorrinal	3	104	905	47	7283	1421	27620	5308	2106	90									
Cingulado Posterior	5	3	116	882	56	5756	5	4471	32795	6099	1312	100							
Giro Temporal Medial	28	7	208	600	101	2510	151	22	7178	68	45	17	29819	7138	10	3302	974	336	116
Giro Frontal Superior	7	33	717	4120	483	6714	28063	1280	28	5876	4128	1647	134						
Hipocampo	4	16	666	2396	195	7135	17	32401	5153	1937	220	18							

Tabela A.1: Resultado do agrupamento dos transcritos pelo algoritmo PPCLUST para cada região cerebral. Os números de transcritos selecionados pelo algoritmo estão destacados em negrito.

Tabela A.2: **Genes diferencialmente expressos em cada região cerebral.** CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente.

Região-Regulação	<i>N</i>	Genes
CE-PR	210	A2M, ACACB, ACSL1, ADAM10, ADD3, AFF1, AHCYL1, AIF1L, AK3, AMER2, ANKDD1A, ANKRD13A, ANLN, ANP32B, APLNR, ATP1A2, BAG3, BBX, BCAS1, BCL2, BCL6, C21orf91, CDK19, CEBPB, CEP350, CLDN11, CLIC4, CLMN, CPEB4, CREBRF, CTDSP2, CYBRD1, DAAM2, DDIT4, DIP2A, DNAJC1, DTNA, EGFR, ENPP2, EPC1, ERBIN, ERMN, FAM107B, FGF1, FMNL2, FNBP1, FOXN3, FOXO1, FRYL, GABRG1, GALNT15, GATM, GFAP, GNB4, GOLIM4, GPM6B, GPR37, GPRC5B, GRAMD2B, GRB10, HADHA, HCG18, HIPK2, HNRNPDL, HSPA2, ID4, IGF1R, IPW, IRF2BP2, ITPKB, ITPRID2, JPX, KAT2B, KCNJ2, KIF5B, KLF15, KREMEN1, LATS2, LIFR, LINC00844, LOC100507477, LOC728613, LPAR1, LRIG1, LRP4, MACF1, MAFF, MALAT1, MAP4, MBP, MED23, MIR612, MOBP, MSI2, MSN, MT1F, MT1G, MT1X, MT2A, MTURN, MTUS1, MYRF, NASP, ND6, NDRG1, NEAT1, NEK7, NFE2L2, NFIA, NFKBIA, NOTCH2NLA, NPAS3, NTRK2, NUFIP2, PAIP2B, PALLD, PDCD6, PDK4, PECAM1, PHIP, PLSCR4, PMP2, PON2, POU3F3, PPA2, PRDX6, PTAR1, PTCHD1, PWARN, QKI, RAB13, RAB31, RASSF4, RBMX, REEP3, RFX4, RHOBTB3, RHOU, RICTOR, RIN2, RO60, RPS3, RPS3A, RPS6, RPS6KA2, RPS6KA5, RUFY3, S100B, SASH1, SAT1, SCD, SEC14L1, SEC22C, SEL1L2, SEPTIN2, SEPTIN7, SERPINA3, SERPINE2, SGK1, SHTN1, SLAIN1, SLC14A1, SLC44A1, SLC5A3, SLC7A11, SLCO1A2, SMARCC1, SNORD107, SNORD115-13, SNORD115-26, SNORD115-7, SNORD116-22, SNORD116-28, SNORD116-4, SNORD61, SNORD73A, SNRPN, SORBS1, SOX2-OT, SPAG9, SPATA13, SPP1, SPTBN1, STAG2, STON2, SUZ12, SYNPO2, TBL1XR1, TCF12, TEAD1, TGFBR1, TJP2, TNS1, TOB1, TP53INP2, TTBK2, TXNIP, UBXN7, UGT8, USP47, VCAN, VEZF1, WSB1, ZBTB20, ZCCHC24, ZEB2, ZFH3, ZFP36L1, ZNF507, ZNRF3

Tabela A.2 continuação da página anterior

Região-Regulação	<i>N</i>	Genes
CE-NR	369	<p>ABHD14A, ABHD17A, ACAP3, ACOT7, ACTN1, ACTR1A, ACY1, ADGRL1, ADGRL3, AGK, AKAP13, ALDOA, ALDOC, ALKBH6, AMPH, ANKRD9, ANKS1B, AP1G2, AP1M1, AP2A2, AP3S2, AP4S1, ARF3, ARHGAP32, ARHGAP33, ARHGDIS, ARMC2-AS1, ARPIN-AP3S2, ASMTL, ATP13A2, ATP1B1, ATP5F1B, ATP5F1C, ATP5MC1, ATP5MF, ATP5MF-PTCD1, ATP6V0D1, ATP6V1G2, ATP9B, AVPI1, BABAM2, BRSK2, BTF3L4, C12orf10, C1orf50, C4orf48, C9orf16, CA10, CA11, CACNA1C, CACNB3, CACNG3, CALN1, CAMK1G, CAPNS1, CCDC24, CCDC3, CCDC85B, CCDC92, CCKBR, CD200, CHCHD10, COL24A1, COMMD9, CPLX1, CPT1C, CRIMI-DT, CRTAC1, CST3, CTIF, CUEDC2, CUTA, CYB561, CYP46A1, DAB1, DACH2, DCTN1, DDN, DDX41, DGCR5, DIPK1A, DMAC2, DMD, DNAAF1, DNAJC4, DNAJC8, DNMI, DPM3, DPYSL4, DTNB, EDF1, EEF1A2, EFN3, EIF3C, EIF3CL, EIF3G, ELOB, EPHA5, ERICH3, F8A1, F8A2, F8A3, FABP3, FAHD2CP, FAM174B, FAM89B, FBXO44, FIG4, FKBP1B, FLYWCH1, FRMPD2, FRMPD2B, FSD1, FXYD6, FXYD7, G6PD, GABBR2, GABRA1, GABRA2, GABRA4, GABRG2, GBAP1, GGCX, GLS2, GNA11, GNB5, GNG3, GOT2, GP1BB, GPR88, GPS2, GPX4, GSS, GUCY1A1, HCFC1R1, HECW1, HERC6, HGS, HSPA9, HSPBP1, HTR2A, HTT, IDH3G, IDS, INA, INAFM1, KCNH3, KCNIP4, KCNK12, KCTD13, KCTD4, KIAA1211L, KIFC2, KNDC1, KPNA1, KYAT3, LAGE3, LDHA, LDOC1, LGALS8, LIN37, LINC00205, LINC00632, LINC01961, LMF1, LMF2, LNX1, LOC100233156, LOC101929800, LOC440896, LRPAP1, LSP1P4, LSP1P5, LUZP1, LYNX1, MACROH2A1, MAFIP, MANBAL, MAP1S, MAP6, MAPRE2, MAST2, MAST3, MCHR1, MEA1, MED16, MGRN1, MIEN1, MIR4691, MIR7113, MKRN1, MLLT11, MROH1, MRPL36, MRPL41, MRPL9, MRPS15, MRTFA, MZF1, NAA38, NAGPA, NAPA, NDUFA3, NDUFB4, NDUFB7, NDUFS7, NDUFS8, NECAB2, NECAB3, NEFL, NME7, NMRAL1, NPDC1, NPLOC4, NPRL3, NRSN2, NRXN2, NSDHL, NUP88, OLA1, OLFM3, OPTN, ORC5, P2RX5-TAX1BP3, PABPC1L2A, PABPC1L2B, PAM, PCNX2, PCSK1N, PCSK2, PCYT2, PEX11B, PFKM, PGLS, PHF24, PHPT1, PIN1, PINK1, PIP5K1C, PKNOX2, PLD3, PLPBP, PNMA8B, POLR2I, POLR2J, POLR2L, PDPDF, PPIB, PPME1, PPP1R11, PPP1R7, PRDX2, PRKCG, PRR36, PRSS3, PRXL2B, PSMB2, PSMB3, PSMF1, PTCD1, PTOV1, PTP4A3, PTPRR, PWWP2B, R3HCC1, RAB3A, RAB40C, RABEP1, RABEP2, RASL10A, RBFOX1, RBFOX3, REEP2, RETREG2, RGS7BP, RHBDD2, RIMS1, RITA1, RNF123, RNF145, RPH3A, RTL8C, RTN4R, RUVBL2, RYR2, SAMM50, SCN3B, SDF4, SELENOF, SELENOM, SEPT5-GP1BB, SEPTIN5, SEZ6L2, SF3A2, SH3GL2, SH3GLB2, SHFL, SLC17A7, SLC1A4, SLC25A11, SLC25A16, SLC2A6, SLC4A3, SLC4A5, SLIRP, SNED1, SNRNP27, SPATA2L, SPHKAP, SPTAN1, SQLE, SRPRA, SRRM1, STMN2, STOX2, STUM, STX8, STXBP1, SUB1, SV2B, SYN1, SYN2, SYN3, SYNGR1, SYNGR3, SYP, SYT13, TAF1A, TAX1BP3, TCTA, TEKT4P2, TENM3, TESK1, THAP7, TMEM205, TMEM222, TMEM269, TMEM97, TMX4, TPGS2, TPI1, TPRG1L, TRAP1, TRAPPC5, TRIR, TTC9B, TUBA4A, TUBB, TUBB3, TUBB4B, TUBG1, USP11, USP25, USP42, UTP4, VKORC1, VPS16, VPS51, WARS1, WBP2, WDR45, YARS1, YPEL3, YWHAB, YWHAZ, ZDHHC16, ZGPAT, ZNF8, ZNF827</p>

Tabela A.2 continuação da página anterior

Região-Regulação	N	Genes
CP-NR	362	ACER3, ACO2, ACP1, ACTB, ACTG1, ACYP2, AFTPH, ALDOA, AMPH, ANAPC5, ANKMY2, ANXA6, AP1S2, AP2S1, APLP2, APOO, ARFGEF3, ARHGAP26, ARHGEF7, ARHGEF9, ARPC2, ARPC3, ARPP19, ATMIN, ATP5F1A, ATP5F1B, ATP5F1C, ATP5MC2, ATP5MC3, ATP5PB, ATP6AP1, ATP6V1B2, ATP6V1E1, ATP6V1H, ATRX, ATXN10, BAALC, BCAS2, BCL2L2, BECN1, BEX4, BRWD1, BSCL2, BTF3, C17orf49, C18orf32, C2orf74, C8orf88, CA10, CANX, CASD1, CCDC34, CCK, CCNDBP1, CCNH, CCT4, CCT7, CD99L2, CDC123, CHCHD2, CHMP2A, CHP1, CHRM1, CLDND1, CLTB, CLTC, CLU, CMAS, CMSS1, CNST, COA3, COPS2, COPS4, COPS5, COPS7A, COX4I1, COX5A, COX5B, COX6C, CSDE1, CSE1L, CTSB, CTXN3, CUL3, DAD1, DCAF6, DCTN5, DCTN6, DDX1, DDX10, DHCR24, DHRS7, DLD, DMXL2, DNAJB6, DNER, DYNC1H1, DYNC111, DYNLL1, ECPAS, EDF1, EEF1A1, EEF1A2, EEF1E1, EFCAB7, EFR3B, EIF3C, EIF3CL, EIF3H, EIF3K, EIF5, EMC3, EML6, ENDOD1, ENPP4, ERGIC2, ESD, ETFB, EXOC1, FAF1, FAIM2, FAM149A, FAM162A, FAR2, FHOD3, FIBP, FIS1, FTH1, FXYD6, GABARAPL2, GABRA1, GAPDH, GATB, GGH, GHITM, GLO1, GLS2, GMPR2, GNAS, GOT2, GSPT2, GSTO1, H2AZ1, HACD3, HDDC2, HDLBP, HMOX2, HNRNPUL2-BSCL2, HOPX, HSP90AB1, HSP90B1, HSPA8, IAH1, IDH3A, IDH3B, IFT22, IMMT, ITGB1BP1, ITPR1, KARS1, KIAA1841, KIF1B, KIF9, KIZ, KLC1, KPNA1, LDHB, LGALS1, LIPA, MAGED1, MAL2, MAP2K4, MCTS1, MDH1, MDH2, MDN1, MED21, MEF2C, METAP2, METTL5, MIF, MIR3652, MKKS, MR-FAP1L1, MRPS9, MSRB2, MTCH2, MYH10, MYL12A, MYL12B, NAA20, NACA, NAMPT, NAP1L1, NAP1L5, NCALD, NDFIP2, NDUFA10, NDUFA13, NDUFA5, NDUFB10, NDUFB5, NDUFC2, NDUFC2-KCTD14, NDUFS3, NDUFV2, NEBL, NECAP1, NEFL, NFU1, NIPAL3, NRN1, NSF, NSFL1C, NUDT3, NUP88, OAT, OBI1, ORMDL1, OXCT1, PDCD6, PEBP1, PEG3, PFDN5, PGK1, PGM2L1, PGRMC1, PHLPP2, PIK3R4, PITPNA, PKIA, PLEKHB2, PLPBP, POLR2K, PPA1, PPARGC1A, PPIA, PPP6C, PRDX3, PRDX5, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, PSMG1, PTGDS, PTGR1, PTS, RAB11A, RAB2A, RABEP1, RAP1GDS1, RAPH1, RASGRF2, RCHY1, RCOR3, REPS2, RGS4, RNASEK, RNASEK-C17orf49, RPL10A, RPL15, RPL17, RPL17-C18orf32, RPL3, RPL4, RPL8, RPRD1A, RTN1, SAMM50, SAP18, SCG2, SCOC, SERINC3, SGIP1, SLC16A6, SLC25A4, SLC7A14, SLI-TRK4, SMARCA2, SMARCA4, SMIM14, SMYD3, SNAPC5, SNCA, SNORA29, SNORD139, SNORD14C, SNORD14D, SNORD16, SNORD18A, SNORD18B, SNORD18C, SNORD83B, SOD1, SPATS2L, SPCS2, SQLE, SRSF7, SSBP1, STMN2, STS, SUB1, SV2B, SYN2, SYNJ2, TAF7, TCEA2, TCP1, THYN1, TIMM23, TIMM23B, TMEM126A, TMEM135, TMEM30A, TMEM65, TMX4, TOMM20, TPD52, TPGS2, TPRG1L, TPT1, TRO, TSPAN5, TTL, TUB, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B, TXN, UBA1, UBC, UBE2L3, UBE2V2, UQCRC1, UQCRC1, USO1, USP11, USP31, VAMP1, VAMP7, VDAC1, VPS29, VPS50, VSNL1, VTI1B, WDR61, WDR7, XRCC6, YWHAZ, ZC2HC1A, ZC3H15, ZNF184, ZNF365

Tabela A.2 continuação da página anterior

Região-Regulação	<i>N</i>	Genes
CP-PR	80	ACACB, ALMS1, ANAPC16, ANKRD13D, ARHGEF40, C9orf64, CACNB2, CBR4, CCDC152, CDK13, CEBPB, CLIP1, COL27A1, CTTN, DLGAP1, DLGAP4, EPG5, ERC2-IT1, FAM161B, FBXO32, FBXW12, FKSG49, FRMD4A, FTX, GRAMD2B, KDM4B, LOC101927166, LOC101930112, LOC202181, LOC286437, LOC729966, MALAT1, MAPKBP1, MIR4691, MIR612, MIR7113, NAV1, NDUFS8, NEAT1, NFASC, NLN, NOTCH2NLA, PCSK5, PDE4C, PIGT, PLGLB1, PLGLB2, POLR1B, PRDX2, PRR11, PTPN2, RAB11FIP3, RAB18, RANBP2, RGPD1, RGPD2, RGPD3, RGPD4, RGPD5, RGPD6, RGPD8, RHBDD3, RIOK3, SBF1, SERPINB6, SLC22A3, SLC25A16, SLC35E1, SPG7, SPTAN1, SYMPK, THRA, TMED10, TMEFF2, TRA2A, UBE2D3, UBE3A, USP34, USP36, ZC3H7B
GFS-NR	25	ATP5F1C, ATP6V1G2, ATRNL1, BEX5, BRWD1, CADPS, EEF1E1, EIF5A2, FAR2, FGF12, FRMPD2, FRMPD2B, GHITM, HPRT1, KIF9, KIFAP3, MKKS, NAP1L5, NME1, PCSK1, PRKAG2-AS1, RGS4, SST, TBC1D30, TUBB
GFS-PR	130	ACACB, ACKR3, AHNAK, AJAP1, AKAP10, ANKRD36B, ARHGEF40, ARRB1, ASCL1, ATP1A2, ATP1B2, BAG3, BDP1, C4A, C4B, C4B_2, CADM1, CCDC88A, CIT, COL27A1, COL5A3, CTTN, CUX2, DDIT4, DLGAP1, EMX2, EP400, EPAS1, EZR, FAM107A, FAM107B, FBXO32, FGFR3, FN1, FOXC1, FOXO1, GFAP, GJA1, GOLIM4, GPRC5B, HEPH, HIP1R, HIPK2, ID4, IGFBP7, IL17RB, IQCA1, ITGB5, ITPKB, KCNJ10, KLC1, LOC202181, LOC286437, LRP4, MALAT1, MAP4K4, METTL7A, MIR1178, MIR612, MT1H, MT1M, MT2A, MYO10, NACC2, NEAT1, NFASC, NHSL1, NIPBL, NOTCH2, NOTCH2NLA, NPIPA5, NPIPB11, NPIPB13, NPIPB3, NPIPB4, NPIPB5, NPL, NR2E1, NR2F2, NTRK2, PALLD, PBXIP1, PCSK5, PDCD6, PDK4, PLSCR4, PMP2, PON2, PTN, PXDC1, RAB11FIP3, RAB13, RAB31, RASEF, RBM25, REL1, RHOJ, RIN2, RNPC3, SASH1, SLC12A7, SLC1A3, SLC25A18, SLC39A12, SLC7A11, SLC7A2, SOX9, SRGN, SRRM2, STOM, STON2, SYMPK, TGFBR3, TIMP3, TJP2, TNFRSF10B, TNS1, TNS3, TRIL, UBE3A, VCAN, VIM, WWTR1, XAF1, YAP1, ZBTB20, ZC3H7B, ZFP36L1, ZFP36L2, ZIC2

Tabela A.2 continuação da página anterior

Região-Regulação	<i>N</i>	Genes
GTM-NR	254	AAK1, ACOT7, ACP1, ADAM22, ADAM23, AHNAK2, AK5, AK6, ALS2, ANKMY2, ANXA6, AP3M2, ARHGEF7, ARPP19, ASPHD2, ATP1A1, ATP2A2, ATP5F1B, ATP5F1C, ATP5MC3, ATP6V1B2, ATP6V1E1, ATP6V1G2, ATP6V1H, ATRNL1, AZIN1, BEX5, BRWD1, BSCL2, BTF3L4, CA10, CACNA1D, CACNA2D3, CADPS, CALB1, CALM1, CALM2, CALM3, CCT4, CDC40, CDH18, CERS6, CHGB, CHURC1, CKMT1A, CKMT1B, CLTB, CLTC, CNRIP1, CNST, COPG2, COPG2IT1, COPS3, COPS4, COX5B, COX6B1, CRYM, CSE1L, CUL1, CUL3, DDX1, DDX10, DHCR24, DNAJA4, DOP1A, DPP10-AS1, EFNA5, EGR1, EIF2S2, EMC3, ENO2, EPB41L3, EPDR1, EPRS1, ERICH3, ERLEC1, FAM162A, FAM49B, FIG4, FOCAD, FRMPD4, FXYD6, GABBR2, GABRA1, GABRB3, GABRG2, GATB, GLO1, GLS, GLS2, GNG3, GOT1, GOT2, GPI, GPR158, GSTO1, GUCY1B1, HINT1, HNRNPUL2-BSCL2, HOPX, HPRT1, HSP90AB1, IARS1, IMPAD1, INA, JAZF1, KAT14, KCNIP4, KIAA1211L, KITLG, KLC1, LAMP5, LARGE1, LDHA, LGALS8, LYRM9, MAP1B, MAP2K1, MAP4, MAST3, MDH1, MEF2C, MFSD6, MICU1, MKKS, MLLT11, MRPL15, MRPS22, MTX2, NAA20, NAP1L5, NAV3, NCALD, NDFIP2, NDUFA8, NDUFAB1, NDUFV2, NEFL, NMNAT2, NRN1, NRXN1, NSF, NSG1, NSG2, NUP50, OCIAD2, OPN3, PAFAH1B1, PAK1, PARP2, PART1, PCNX2, PDCD6, PDK3, PET117, PGK1, PGM2L1, PHB, PHF24, PLCL2, PPA1, PPIA, PPP1R7, PPP3CA, PRDM2, PREPL, PRMT8, PRORP, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, PTPRS, PTPRT, R3HDM1, RAB11A, RAB2A, RAB3C, RAB6B, RAN, RAPIGDS1, RBFOX1, REEP1, RETREG1, RGS4, RNF175, RPH3A, RPL15, RTN1, RTN3, RUFY3, SAP18, SARS1, SCN2B, SELENOI, SERINC3, SERPINI1, SGIP1, SLC25A3, SLC25A4, SLC39A10, SLC7A14, SLC9A6, SLIRP, SLIT2, SLITRK5, SMYD2, SNCA, SNRPN, SNURF, SNX10, SOD1, STMN2, SULT4A1, SV2B, SYN2, SYNE1, SYT1, SYT13, TAF9, TAF9A, TAGLN3, TARBP1, THYN1, TIMM23, TIMM23B, TMEM70, TMOD1, TOMM20, TPBG, TUBA1B, TUBB, TUBB3, TUBB4B, UBE2QL1, UCHL1, UGP2, UQCRRS1, UQCRH, UQCRHL, VDAC3, VPS35, VSNL1, WIPF3, YWHAH, YWHAZ, ZMAT4, ZNF204P, ZNF365, ZNHIT3

Tabela A.2 continuação da página anterior

Região-Regulação	<i>N</i>	Genes
GTM-PR	213	ACACB, AJAPI, ALDH6A1, AMER2, ANAPC16, ANKRD13D, ANLN, ANO6, ANP32B, AQP4, ARH-GAP5, ARHGEF26-AS1, ARRB1, ASCL1, BBX, BCL6, BMPR1B, C4A, C4B, C4B_2, C9orf64, CAMTA1, CDC42EP4, CFLAR, CHD9, CHERP, COL27A1, COX3, CSPG5, DAAM1, DDR1, DDX17, DNAJC1, DTNA, EMP1, EP400, EPC1, EPS8, ERBIN, FAM120A, FAM168A, FBXL17, FERMT2, FLT1, FNBP1, FOXO1, FXR1, GASK1B, GNA13, GPAM, GRAMD2B, HERPUD2, HIPK2, ID4, IGFBP7, IL6ST, IPW, IQCA1, IRF2BP2, ITGA7, ITPKB, KANK1, KAT2B, KCNJ10, KLF9, KLHL5, KMT2E, KTN1, LAMP2, LIFR, LINC00537, LLGL1, LOC101930112, LOC202181, LOC286437, LOC729732, LPP, LRP4, MAFF, MALAT1, MAP4K4, MARK3, MCC, MEGF10, MFHAS1, MIR1304, MIR4640, MIR612, MSRB3, N4BP2L2, NACC2, NADK2, NASP, NEAT1, NFIA, NFKBIA, NLN, NPAS3, NPL, NRXN2, NSUN6, PALLD, PCDHGA1, PCDHGA10, PCDHGA11, PCDHGA12, PCDHGA2, PCDHGA3, PCDHGA4, PCDHGA5, PCDHGA6, PCDHGA7, PCDHGA8, PCDHGA9, PCDHGB1, PCDHGB2, PCDHGB3, PCDHGB4, PCDHGB5, PCDHGB6, PCDHGB7, PCDHGC3, PCDHGC4, PCDHGC5, PCSK5, PDK4, PDLIM5, PGM2, PHAX, PLGLB1, PLGLB2, PLOD2, PM20D2, PMP2, PNISR, PPA2, PPARD, PPFIBP1, PRDM16, PRDX2, PSAT1, PTRPZ1, PWARSN, QKI, RAD51L3-RFFL, RBM25, RBMS3, RFFL, RFX4, RHOBTB3, RHOQ, RIOK3, SCAF11, SCAMP1, SELENOP, SEMA3F, SFT2D2, SLC18B1, SLC1A2, SLC39A12, SLC7A2, SMAD1, SMC3, SMOX, SNORA1, SNORA18, SNORA32, SNORA40, SNORA8, SNORD107, SNORD115-13, SNORD115-26, SNORD115-7, SNORD116-22, SNORD116-28, SNORD116-4, SNORD5, SOX2, SOX2-OT, SPAG9, SPG7, SRCIN1, SRGAP1, SRRM2, STAG2, SYMPK, TAF1D, TAF3, TBL1X, TCF4, TEAD1, TGFBR3, THRA, TJP2, TMEM47, TNRC6B, TNS1, TOB2, TRA2A, TSPAN31, USP42, VPS13C, WNK1, XAF1, YAP1, ZBTB20, ZCCHC24, ZEB2, ZFP36L1, ZHX2, ZNF37A, ZNF462, ZNRF3
HIP-NR	157	ACO2, ACPI, ACTB, ACTG1, ACYP2, ADAM23, ADARB1, ADD1, AHNK2, ANK2, AP2A2, AP2M1, APLP2, ARHGAP5, ATP1A1, ATP5PB, ATRN, BOD1L1, BRWD1, CANX, CARMIL1, CCND2, CCT2, CCT4, CCT7, CDC123, CHMP2A, CIRBP, COPS6, COX4I1, CPNE3, CSNK2B, CTBP1, CUL3, DMAP1, DYNC1H1, DYNLL1, EDF1, EEF1G, EFR3B, EIF1, EIF3C, EIF3CL, EMC3, EMC7, ENDOD1, EPB41L3, FABP5, FAM162A, FAM50A, FDFT1, GABARAPL2, GALNT18, GAPDH, GATC, GLO1, GNAS, GOT1, GPR89A, GPR89B, HDHC2, HSPA8, IK, ITGB1BP1, KARS1, KDM1A, KIFAP3, KIZ, KPNA2, LAMP1, LARGE1, LEPROTL1, LETMD1, LRRC49, MAP2K1, MAP4, MDH1, MIR3654, MKKS, MRPL15, MRPL24, MRPS23, MTX2, NACA, NDUFA10, NDUFA11, NDUFAB1, NECAP1, NPEPPS, OPA1, PD6D6, PDHA1, PEBP1, PFDN2, PFDN5, PHLPP2, PITPNA, PITRM1, PLPBP, POP4, PPA1, PPIA, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RABAC1, RBX1, RCOR3, RPL29, RPLP0, RSPH10B, RSPH10B2, SARNP, SDHA, SET, SETSIP, SETX, SIRPA, SLC25A3, SLC7A14, SLF2, SNCA, SNORD14C, SNORD14D, SOD1, SPOCK1, SRSF9, SULF2, SUMO1, SUPV3L1, SV2B, SYN2, TALDO1, THYN1, TMC01, TOMM34, TRAPPC12, TSG101, TTC19, TUBA1B, TUBB3, U2AF1, UBC, UBE2L3, UBE2N, UBE2QL1, UQCRFS1, USP47, VPS29, XRCC5, XRCC6, YWHAZ, ZFAND2A, ZNF532

Tabela A.2 continuação da página anterior

Região-Regulação	<i>N</i>	Genes
HIP-PR	90	ADGRB3, AJAP1, ALMS1, ANKRD13D, ANKRD36B, ARHGAP21, BBX, C5orf24, CCDC88A, CELF2, CFLAR, COL27A1, CPEB4, CSRNP3, DDX17, DUSP8, ELAVL3, EP300, EP400, ESF1, FAM120A, FMNL2, FTX, FXR1, GGA3, GOLGA2, GPR155, GRK3, HRK, HSP90B1, KDM4B, KTN1, LIFR, LMO4, LOC101927166, LOC202181, MAGI2, MALAT1, MAP4K3-DT, MIAT, MIB1, MINK1, MIR3652, MIR612, NAV1, NEAT1, NFASC, NFAT5, NIPBL, NLN, NOVA2, NSUN6, PLGLB1, PLGLB2, PPA2, PRDX2, RBM25, RBM39, RIOK3, RNF165, RNPC3, SCAF4, SCAMP1, SEC22C, SECISBP2L, SEPTIN7P2, SILC1, SLC35E2A, SLC35E2B, SLC8A1, SMC3, SPEN, SPTAN1, SRCIN1, SRRM1, SRRM2, SYMPK, TBL1XR1, THRA, TOP1, TRA2A, TTBK2, UBE2D3, UBE2W, UBE3A, VPS13B, WASL, WNK1, YLPM1, ZBTB7A

Tabela A.3: **As vias biológicas mais significativas para os transcritos diferencialmente expressos em cada região cerebral.** CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CE-NR (200)	Transmission across Chemical Synapses	AP2A2, CACNB3, CACNG3, CPLX1, GABBR2, GABRA1, GABRA2, GABRA4, GABRG2, GLS2, GNB5, GNG3, NEFL, PRKCG, RAB3A, RIMS1, SLC17A7, STXBPI, SYN1, SYN2, SYN3, TUBA4A, TUBB3, TUBB4B
CE-NR (200)	Serotonin Neurotransmitter Release Cycle	CPLX1, RAB3A, RIMS1, STXBPI, SYN1, SYN2, SYN3
CE-NR (200)	Neuronal System	AP2A2, CACNB3, CACNG3, CPLX1, GABBR2, GABRA1, GABRA2, GABRA4, GABRG2, GLS2, GNB5, GNG3, KCNH3, NEFL, NRXN2, PRKCG, RAB3A, RIMS1, SLC17A7, STXBPI, SYN1, SYN2, SYN3, TUBA4A, TUBB3, TUBB4B
CE-NR (200)	Dopamine Neurotransmitter Release Cycle	CPLX1, RAB3A, RIMS1, STXBPI, SYN1, SYN2, SYN3
CE-NR (200)	Neurotransmitter release cycle	CPLX1, GLS2, RAB3A, RIMS1, SLC17A7, STXBPI, SYN1, SYN2, SYN3
CE-NR (200)	Glutamate Neurotransmitter Release Cycle	CPLX1, GLS2, RAB3A, RIMS1, SLC17A7, STXBPI

Tabela A.3 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CE-NR (200)	G2/M Transition	ACTR1A, DCTN1, LIN37, NME7, OPTN, PPME1, PSMB2, PSMB3, PSMF1, TUBA4A, TUBB, TUBB3, TUBB4B, TUBG1
CE-NR (200)	Cristae formation	ATP5F1B, ATP5F1C, ATP5MC1, ATP5MF, HSPA9, SAMM50
CE-NR (200)	Mitotic G2-G2/M phases	ACTR1A, DCTN1, LIN37, NME7, OPTN, PPME1, PSMB2, PSMB3, PSMF1, TUBA4A, TUBB, TUBB3, TUBB4B, TUBG1
CE-NR (200)	Neurotransmitter receptors and postsynaptic signal transmission	AP2A2, CACNG3, GABBR2, GABRA1, GABRA2, GABRA4, GABRG2, GNB5, GNG3, NEFL, PRKCG, TUBA4A, TUBB3, TUBB4B
CE-NR (200)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	ATP5F1B, ATP5F1C, ATP5MC1, ATP5MF, NDUFA3, NDUFB4, NDUFB7, NDUFS7, NDUFS8, TRAP1
CE-NR (200)	The citric acid (TCA) cycle and respiratory electron transport	ATP5F1B, ATP5F1C, ATP5MC1, ATP5MF, IDH3G, LDHA, NDUFA3, NDUFB4, NDUFB7, NDUFS7, NDUFS8, TRAP1
CE-PR (121)	Metallothioneins bind metals	MT1F, MT1G, MT1X, MT2A
CE-PR (121)	Response to metal ions	MT1F, MT1G, MT1X, MT2A
CE-PR (121)	Regulation of FOXO transcriptional activity by acetylation	FOXO1, KAT2B, TXNIP
CE-PR (121)	Signaling by Receptor Tyrosine Kinases	ADAM10, AHCYL1, EGFR, ERBIN, FGF1, GFAP, GRB10, IGF1R, LRIG1, NTRK2, RICTOR, RPS6KA2, RPS6KA5, S100B, SPP1
CP-NR (245)	The citric acid (TCA) cycle and respiratory electron transport	ACO2, ATP5F1A, ATP5F1B, ATP5F1C, ATP5MC2, ATP5MC3, ATP5PB, COX4I1, COX5A, COX5B, COX6C, DLD, ETFB, GLO1, IDH3A, IDH3B, LDHB, MDH2, NDUFA10, NDUFA13, NDUFA5, NDUFB10, NDUFB5, NDUFC2, NDUFS3, NDUFV2, UQCRC1, UQCRCF1, VDAC1
CP-NR (245)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	ATP5F1A, ATP5F1B, ATP5F1C, ATP5MC2, ATP5MC3, ATP5PB, COX4I1, COX5A, COX5B, COX6C, ETFB, NDUFA10, NDUFA13, NDUFA5, NDUFB10, NDUFB5, NDUFC2, NDUFS3, NDUFV2, UQCRC1, UQCRCF1

Tabela A.3 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CP-NR (245)	The role of GTSE1 in G2/M progression after G2 checkpoint	HSP90AB1, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B, UBC
CP-NR (245)	Infectious disease	AP1S2, AP2S1, ATP6V1H, CANX, CCNH, CHMP2A, CLTC, HSP90AB1, KPNA1, MAP2K4, NUP88, POLR2K, PPIA, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, RPL10A, RPL15, RPL17, RPL3, RPL4, RPL8, SLC25A4, SV2B, TAF7, UBC, VAMP1, XRCC6
CP-NR (245)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	ACTB, CCT4, CCT7, PFDN5, TCP1, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B
CP-NR (245)	HSP90 chaperone cycle for steroid hormone receptors (SHR)	DCTN5, DCTN6, DYNC1H1, DYNC1H1, DYNLL1, HSP90AB1, HSPA8, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B
CP-NR (245)	Cellular responses to stress	ANAPC5, ATP6V1B2, ATP6V1E1, ATP6V1H, DCTN5, DCTN6, DNAJB6, DYNC1H1, DYNC1H1, DYNLL1, EEF1A1, H2AZ1, HSP90AB1, HSPA8, MAP2K4, NUP88, PRDX3, PRDX5, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, SOD1, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B, TXN, UBC
CP-NR (245)	Prefoldin mediated transfer of substrate to CCT/-TriC	ACTB, CCT4, CCT7, PFDN5, TCP1, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B
CP-NR (245)	Cellular responses to external stimuli	ANAPC5, ATP6V1B2, ATP6V1E1, ATP6V1H, DCTN5, DCTN6, DNAJB6, DYNC1H1, DYNC1H1, DYNLL1, EEF1A1, H2AZ1, HSP90AB1, HSPA8, MAP2K4, NUP88, PRDX3, PRDX5, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, SOD1, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B, TXN, UBC
CP-NR (245)	Hedgehog 'off' state	GNAS, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B, UBC

Tabela A.3 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CP-NR (245)	Signaling by Hedgehog	CUL3, EFCAB7, GNAS, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B, UBC
CP-NR (245)	HIV Infection	AP1S2, AP2S1, ATP6V1H, CCNH, CHMP2A, KPNA1, NUP88, POLR2K, PPIA, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, SLC25A4, TAF7, UBC, XRCC6
CP-NR (245)	Cristae formation	APOO, ATP5F1A, ATP5F1B, ATP5F1C, ATP5MC2, ATP5MC3, ATP5PB, IMMT, SAMM50
CP-NR (245)	G2/M Transition	CCNH, DYNC1H1, DYNLL1, HSP90AB1, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B, UBC
CP-NR (245)	Mitotic G2-G2/M phases	CCNH, DYNC1H1, DYNLL1, HSP90AB1, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B, UBC
CP-NR (245)	Negative regulation of NOTCH4 signaling	PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC, YWHAZ
CP-NR (245)	Degradation of DVL	CUL3, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC
CP-NR (245)	Regulation of RAS by GAPs	CUL3, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC
CP-NR (245)	Regulation of mRNA stability by proteins that bind AU-rich elements	HSPA8, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC, YWHAZ
CP-NR (245)	Downstream signaling events of B Cell Receptor (BCR)	PPIA, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC
CP-NR (245)	Beta-catenin independent WNT signaling	AP2S1, CLTB, CLTC, ITPR1, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC
CP-NR (245)	Signaling by the B Cell Receptor (BCR)	ITPR1, PPIA, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC

Tabela A.3 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CP-NR (245)	CLEC7A (Dectin-1) signaling	ITPR1, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC
CP-NR (245)	Fc epsilon receptor (FCER1) signaling	ITPR1, MAP2K4, PSMA1, PSMA2, PSMA5, PSMB1, PSMB3, PSMB6, PSMD1, PSMD4, PSMD6, UBC
CP-NR (245)	Transmission across Chemical Synapses	AP2S1, ARHGEF7, ARHGEF9, GABRA1, GLS2, HSPA8, NCALD, NEFL, NSF, RASGRF2, SYN2, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B
CP-NR (245)	Neurotransmitter receptors and postsynaptic signal transmission	AP2S1, ARHGEF7, ARHGEF9, GABRA1, NCALD, NEFL, NSF, RASGRF2, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B
CP-NR (245)	Neuronal System	AP2S1, ARHGEF7, ARHGEF9, GABRA1, GLS2, HSPA8, NCALD, NEFL, NSF, RASGRF2, SLITRK4, SYN2, TUBA1B, TUBA1C, TUBA3C, TUBA3D, TUBB3, TUBB4B
GFS-PR (83)	Metallothioneins bind metals	MT1H, MT1M, MT2A
GFS-PR (83)	Response to metal ions	MT1H, MT1M, MT2A
GFS-PR (83)	YAP1- and WWTR1 (TAZ)-stimulated gene expression	HIPK2, WWTR1, YAP1
GFS-PR (83)	MET promotes cell motility	COL27A1, COL5A3, FN1, TNS3
GFS-PR (83)	ECM proteoglycans	COL5A3, FN1, ITGB5, LRP4, VCAN
GFS-PR (83)	Signaling by Hippo	TJP2, WWTR1, YAP1
GFS-PR (83)	Initial triggering of complement	C4A, C4B, C4B_2
GFS-PR (83)	Syndecan interactions	COL5A3, FN1, ITGB5
GFS-PR (83)	Transport of inorganic cations/anions and amino acids/oligopeptides	SLC12A7, SLC1A3, SLC25A18, SLC7A11, SLC7A2
GFS-PR (83)	Post-translational protein phosphorylation	C4A, C4B, FN1, IGFBP7, VCAN
GTM-NR (167)	Downstream signaling events of B Cell Receptor (BCR)	CALM1, CALM2, CALM3, CUL1, PPIA, PPP3CA, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1
GTM-NR (167)	Signaling by the B Cell Receptor (BCR)	CALM1, CALM2, CALM3, CUL1, PPIA, PPP3CA, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1
GTM-NR (167)	The role of GTSE1 in G2/M progression after G2 checkpoint	HSP90AB1, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, TUBA1B, TUBB3, TUBB4B

Tabela A.3 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
GTM-NR (167)	CLEC7A (Dectin-1) signaling	CALM1, CALM2, CALM3, CUL1, PPP3CA, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1
GTM-NR (167)	Neuronal System	ARHGEF7, CACNA2D3, CALM1, CALM2, CALM3, EPB41L3, GABBR2, GABRA1, GABRB3, GABRG2, GLS, GLS2, GNG3, NCALD, NEFL, NRXN1, NSF, PTPRS, RTN3, SLITRK5, SYN2, SYT1, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	Beta-catenin independent WNT signaling	CALM1, CALM2, CALM3, CLTB, CLTC, GNG3, PPP3CA, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1
GTM-NR (167)	Transmission across Chemical Synapses	ARHGEF7, CACNA2D3, CALM1, CALM2, CALM3, GABBR2, GABRA1, GABRB3, GABRG2, GLS, GLS2, GNG3, NCALD, NEFL, NSF, SYN2, SYT1, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	Negative regulation of NOTCH4 signaling	CUL1, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, YWHAZ
GTM-NR (167)	Fc epsilon receptor (FCERI) signaling	CALM1, CALM2, CALM3, CUL1, PAK1, PPP3CA, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1
GTM-NR (167)	Infectious disease	ATP6V1H, CALM1, CALM2, CALM3, CLTC, HSP90AB1, MAP2K1, NUP50, PPIA, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, RAN, RPL15, SLC25A4, SV2B, SYT1, TAF9
GTM-NR (167)	Signaling by Hedgehog	CUL1, CUL3, ERLEC1, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	Hedgehog 'off' state	CUL1, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	Degradation of DVL	CUL3, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1
GTM-NR (167)	G2/M Transition	CUL1, HSP90AB1, PAFAH1B1, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, TUBA1B, TUBB, TUBB3, TUBB4B

Tabela A.3 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
GTM-NR (167)	Mitotic G2-G2/M phases	CUL1, HSP90AB1, PAFAH1B1, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, TUBA1B, TUBB, TUBB3, TUBB4B
GTM-NR (167)	Neurotransmitter receptors and postsynaptic signal transmission	ARHGEF7, CALM1, CALM2, CALM3, GABBR2, GABRA1, GABRB3, GABRG2, GNG3, NCALD, NEFL, NSF, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	Regulation of RAS by GAPs	CUL3, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1
GTM-NR (167)	The citric acid (TCA) cycle and respiratory electron transport	ATP5F1B, ATP5F1C, ATP5MC3, COX5B, COX6B1, GLO1, LDHA, NDUFA8, NDUFAB1, NDUFV2, PDK3, UQCRFS1, UQCRH
GTM-NR (167)	Regulation of mRNA stability by proteins that bind AU-rich elements	PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, YWHAZ
GTM-NR (167)	HIV Infection	ATP6V1H, NUP50, PPIA, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, RAN, SLC25A4, TAF9
GTM-NR (167)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	ATP5F1B, ATP5F1C, ATP5MC3, COX5B, COX6B1, NDUFA8, NDUFAB1, NDUFV2, UQCRFS1, UQCRH
GTM-NR (167)	Cellular responses to stress	ATP6V1B2, ATP6V1E1, ATP6V1G2, ATP6V1H, DNAJA4, HSP90AB1, NUP50, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, SOD1, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	Cellular responses to external stimuli	ATP6V1B2, ATP6V1E1, ATP6V1G2, ATP6V1H, DNAJA4, HSP90AB1, NUP50, PSMA1, PSMA5, PSMA6, PSMB2, PSMB3, PSMB4, PSMB7, PSMD1, SOD1, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	Signaling by Receptor Tyrosine Kinases	ARHGEF7, ATP6V1B2, ATP6V1E1, ATP6V1G2, ATP6V1H, CALM1, CALM2, CALM3, CLTC, GABRA1, GABRB3, GABRG2, KITLG, MAP2K1, MEF2C, PAK1, PTPRS
GTM-NR (167)	Cristae formation	ATP5F1B, ATP5F1C, ATP5MC3, MTX2
GTM-NR (167)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	CCT4, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	HSP90 chaperone cycle for steroid hormone receptors (SHR)	DNAJA4, HSP90AB1, TUBA1B, TUBB3, TUBB4B
GTM-NR (167)	Glutamate Neurotransmitter Release Cycle	GLS, GLS2, SYT1
GTM-NR (167)	Neurotransmitter release cycle	GLS, GLS2, SYN2, SYT1

Tabela A.3 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
GTM-NR (167)	Prefoldin mediated transfer of substrate to CCT/- TriC	CCT4, TUBB3, TUBB4B
GTM-PR (103)	YAP1- and WWTR1 (TAZ)-stimulated gene ex- pression	HIPK2, KAT2B, TEAD1, YAP1
HIP-NR (107)	Infectious disease	AP2A2, AP2M1, CANX, CHMP2A, KPNA2, MAP2K1, PPIA, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, RPL29, RPLP0, SV2B, TSG101, UBC, XRCC5, XRCC6
HIP-NR (107)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	ACTB, CCT2, CCT4, CCT7, PFDN2, PFDN5, TUBA1B, TUBB3
HIP-NR (107)	Negative regulation of NOTCH4 signaling	PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, UBC, YWHAZ
HIP-NR (107)	Degradation of DVL	CUL3, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, UBC
HIP-NR (107)	Prefoldin mediated transfer of substrate to CCT/- TriC	ACTB, CCT2, CCT4, CCT7, PFDN2, PFDN5, TUBB3
HIP-NR (107)	HIV Infection	AP2A2, AP2M1, CHMP2A, PPIA, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, TSG101, UBC, XRCC5, XRCC6
HIP-NR (107)	Hedgehog 'off' state	GNAS, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, TUBA1B, TUBB3, UBC
HIP-NR (107)	Regulation of mRNA stability by proteins that bind AU-rich elements	HSPA8, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, SET, UBC, YWHAZ
HIP-NR (107)	Regulation of RAS by GAPs	CUL3, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, UBC
HIP-NR (107)	Signaling by Hedgehog	CUL3, GNAS, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, TUBA1B, TUBB3, UBC
HIP-NR (107)	The role of GTSE1 in G2/M progression after G2 checkpoint	PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, TUBA1B, TUBB3, UBC
HIP-NR (107)	G2/M Transition	DYNC1H1, DYNLL1, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, TUBA1B, TUBB3, UBC
HIP-NR (107)	Mitotic G2-G2/M phases	DYNC1H1, DYNLL1, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, TUBA1B, TUBB3, UBC
HIP-NR (107)	Downstream signaling events of B Cell Receptor (BCR)	PPIA, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, UBC

Tabela A.3 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
HIP-NR (107)	CLEC7A (Dectin-1) signaling	PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, UBC, UBE2N
HIP-NR (107)	The citric acid (TCA) cycle and respiratory electron transport	ACO2, ATP5PB, COX4I1, GLO1, NDUFA10, NDUFA11, NDUFAB1, PDHA1, SDHA, UQCRFS1
HIP-NR (107)	Beta-catenin independent WNT signaling	AP2A2, AP2M1, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, UBC
HIP-NR (107)	Signaling by the B Cell Receptor (BCR)	PPIA, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, UBC
HIP-NR (107)	Fc epsilon receptor (FCER1) signaling	PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, UBC, UBE2N
HIP-NR (107)	HSP90 chaperone cycle for steroid hormone receptors (SHR)	DYNC1H1, DYNLL1, HSPA8, TUBA1B, TUBB3
HIP-NR (107)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	ATP5PB, COX4I1, NDUFA10, NDUFA11, NDUFAB1, SDHA, UQCRFS1
HIP-NR (107)	Cellular responses to stress	DYNC1H1, DYNLL1, HSPA8, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, SOD1, TUBA1B, TUBB3, UBC
HIP-NR (107)	Cellular responses to external stimuli	DYNC1H1, DYNLL1, HSPA8, PSMA1, PSMB2, PSMB6, PSMC2, PSMD4, PSMD6, RBX1, SOD1, TUBA1B, TUBB3, UBC
HIP-PR (45)	Nephrin family interactions	MAGI2, SPTAN1, WASL

Tabela A.4: As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Córtex Entorrinal com base na co-expressão diferencial. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
DA-NR (98)	Transmission across Chemical Synapses	AP2A2, CACNG3, CPLX1, GABBR2, GABRA1, GABRA2, GABRG2, GLS2, GNB5, NEFL, RAB3A, STXBP1, SYN2, SYN3
DA-NR (98)	Serotonin Neurotransmitter Release Cycle	CPLX1, RAB3A, STXBP1, SYN2, SYN3

Tabela A.4 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
DA-NR (98)	Dopamine Neurotransmitter Release Cycle	CPLX1, RAB3A, STXBP1, SYN2, SYN3
DA-NR (98)	Neurotransmitter release cycle	CPLX1, GLS2, RAB3A, STXBP1, SYN2, SYN3
DA-NR (98)	Neuronal System	AP2A2, CACNG3, CPLX1, GABBR2, GABRA1, GABRA2, GABRG2, GLS2, GNB5, NEFL, RAB3A, STXBP1, SYN2, SYN3
CTL-DA-PR (1)	RNA Polymerase III Transcription Termination	NFIA
CTL-DA-PR (1)	RNA Polymerase III Transcription	NFIA
CTL-DA-PR (1)	RNA Polymerase III Abortive And Retractive Initiation	NFIA
DA-NR (98)	Glutamate Neurotransmitter Release Cycle	CPLX1, GLS2, RAB3A, STXBP1
DIF-NR (130)	Glutamate Neurotransmitter Release Cycle	CPLX1, GLS2, RAB3A, SLC17A7, STXBP1
DIF-NR (130)	Neurotransmitter release cycle	CPLX1, GLS2, RAB3A, SLC17A7, STXBP1, SYN2
DIF-NR (130)	Serotonin Neurotransmitter Release Cycle	CPLX1, RAB3A, STXBP1, SYN2
DIF-NR (130)	Lysosome Vesicle Biogenesis	APIG2, AP1M1, AP4S1, HGS, SH3GL2
DA-NR (98)	Glucose metabolism	ALDOA, ALDOC, GOT2, NUP88, PFKM, SLC25A11
DIF-NR (130)	Transmission across Chemical Synapses	AP2A2, CPLX1, GABBR2, GABRA1, GABRA4, GLS2, GNB5, PRKCG, RAB3A, SLC17A7, STXBP1, SYN2
DA-NR (98)	Gluconeogenesis	ALDOA, ALDOC, GOT2, SLC25A11
DA-NR (98)	Lysosome Vesicle Biogenesis	APIG2, AP1M1, AP4S1, SH3GL2
DIF-NR (130)	Dopamine Neurotransmitter Release Cycle	CPLX1, RAB3A, STXBP1, SYN2
DA-NR (98)	Acetylcholine Neurotransmitter Release Cycle	CPLX1, RAB3A, STXBP1
DIF-NR (130)	Neuronal System	AP2A2, CPLX1, GABBR2, GABRA1, GABRA4, GLS2, GNB5, KCNH3, NRXN2, PRKCG, RAB3A, SLC17A7, STXBP1, SYN2
DIF-NR (130)	Cristae formation	ATP5F1C, ATP5MC1, HSPA9, SAMM50
DIF-NR (130)	Retrograde neurotrophin signalling	AP2A2, DNM1, SH3GL2
CTL-PR (68)	YAP1- and WWTR1 (TAZ)-stimulated gene expression	HIPK2, KAT2B, TEAD1
CTL-PR (68)	Recycling pathway of L1	MSN, RPS6KA2, RPS6KA5, SHTN1
CTL-DA-NR (5)	Phospholipid metabolism	PIP5K1C, PLD3
CTL-DA-NR (5)	Acetylcholine regulates insulin secretion	GNA11
CTL-DA-NR (5)	Free fatty acids regulate insulin secretion	GNA11
CTL-DA-NR (5)	SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion	PIP5K1C
CTL-DA-NR (5)	Thromboxane signalling through TP receptor	GNA11

Tabela A.4 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-DA-NR (5)	Role of phospholipids in phagocytosis	PLD3
CTL-DA-NR (5)	ADP signalling through P2Y purinoceptor 1	GNA11
CTL-DA-NR (5)	Signal amplification	GNA11
CTL-DA-NR (5)	Thrombin signalling through proteinase activated receptors (PARs)	GNA11
DIF-NR (130)	Gluconeogenesis	ALDOA, ALDOC, GOT2, SLC25A11

Tabela A.5: As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Córtex Entorrinal com base na dependência condicional. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (71)	Neurotransmitter receptors and postsynaptic signal transmission	GABBR2, GABRA1, GABRA4, GABRG2, GNB5, GNG3, PRKCG, TUBA4A, TUBB3, TUBB4B
CTL-NR (71)	Transmission across Chemical Synapses	GABBR2, GABRA1, GABRA4, GABRG2, GNB5, GNG3, PRKCG, RAB3A, TUBA4A, TUBB3, TUBB4B
CTL-NR (71)	Neuronal System	GABBR2, GABRA1, GABRA4, GABRG2, GNB5, GNG3, KCNH3, NRXN2, PRKCG, RAB3A, TUBA4A, TUBB3, TUBB4B
DA-NR (46)	Transmission across Chemical Synapses	GABRA1, GABRA2, GABRG2, GLS2, NEFL, SLC17A7, TUBA4A, TUBB3
DA-NR (46)	G2/M Transition	DCTN1, NME7, OPTN, PSMB3, TUBA4A, TUBB, TUBB3
DA-NR (46)	Mitotic G2-G2/M phases	DCTN1, NME7, OPTN, PSMB3, TUBA4A, TUBB, TUBB3
CTL-NR (71)	GABA receptor activation	GABBR2, GABRA1, GABRA4, GABRG2, GNG3
DA-NR (46)	Recruitment of NuMA to mitotic centrosomes	DCTN1, NME7, TUBA4A, TUBB, TUBB3
CTL-NR (71)	COPI-mediated anterograde transport	ARF3, DCTN1, NAPA, TUBA4A, TUBB3, TUBB4B
CTL-DA-PR (1)	GRB2 events in EGFR signaling	EGFR
CTL-DA-PR (1)	ERBB2 Activates PTK6 Signaling	EGFR

Tabela A.5 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-DA-PR (1)	SHC1 events in EGFR signaling	EGFR
CTL-DA-PR (1)	Constitutive Signaling by EGFRvIII	EGFR
CTL-DA-PR (1)	Signaling by EGFRvIII in Cancer	EGFR
CTL-DA-PR (1)	ERBB2 Regulates Cell Motility	EGFR
CTL-DA-PR (1)	TFAP2 (AP-2) family regulates transcription of growth factors and their receptors	EGFR
CTL-DA-PR (1)	GRB2 events in ERBB2 signaling	EGFR
CTL-DA-PR (1)	PI3K events in ERBB2 signaling	EGFR
CTL-DA-PR (1)	GAB1 signalosome	EGFR
CTL-NR (71)	Translocation of SLC2A4 (GLUT4) to the plasma membrane	TUBA4A, TUBB3, TUBB4B, YWHAB, YWHAZ
CTL-DA-PR (1)	Signaling by ERBB4	EGFR
CTL-NR (71)	The role of GTSE1 in G2/M progression after G2 checkpoint	PSMB3, PSMF1, TUBA4A, TUBB3, TUBB4B
CTL-NR (71)	MHC class II antigen presentation	AP1M1, DCTN1, SH3GL2, TUBA4A, TUBB3, TUBB4B
CTL-NR (71)	Microtubule-dependent trafficking of connexons from Golgi to the plasma membrane	TUBA4A, TUBB3, TUBB4B
CTL-NR (71)	Golgi-to-ER retrograde transport	ARF3, DCTN1, NAPA, TUBA4A, TUBB3, TUBB4B
CTL-NR (71)	G2/M Transition	DCTN1, PSMB3, PSMF1, TUBA4A, TUBB3, TUBB4B
CTL-NR (71)	Mitotic G2-G2/M phases	DCTN1, PSMB3, PSMF1, TUBA4A, TUBB3, TUBB4B
CTL-NR (71)	Recruitment of NuMA to mitotic centrosomes	DCTN1, TUBA4A, TUBB, TUBB3, TUBB4B
DA-NR (46)	Neurotransmitter receptors and postsynaptic signal transmission	GABRA1, GABRA2, GABRG2, NEFL, TUBA4A, TUBB3
DA-NR (46)	Organelle biogenesis and maintenance	ATP5F1B, ATP5F1C, DCTN1, MCHR1, TUBA4A, TUBB, TUBB3
DA-NR (46)	Translocation of SLC2A4 (GLUT4) to the plasma membrane	TUBA4A, TUBB3, YWHAB, YWHAZ
DA-NR (46)	Neuronal System	GABRA1, GABRA2, GABRG2, GLS2, NEFL, SLC17A7, TUBA4A, TUBB3
DA-NR (46)	Recruitment of mitotic centrosome proteins and complexes	DCTN1, NME7, TUBA4A, TUBB
DA-NR (46)	Centrosome maturation	DCTN1, NME7, TUBA4A, TUBB
CTL-PR (51)	YAP1- and WWTR1 (TAZ)-stimulated gene expression	HIPK2, KAT2B, TEAD1

Tabela A.5 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (71)	Recruitment of mitotic centrosome proteins and complexes	DCTN1, TUBA4A, TUBB, TUBB4B
CTL-NR (71)	Centrosome maturation	DCTN1, TUBA4A, TUBB, TUBB4B
DA-NR (46)	GABA receptor activation	GABRA1, GABRA2, GABRG2
CTL-NR (71)	Organelle biogenesis and maintenance	ATP5F1B, ATP5F1C, DCTN1, TUBA4A, TUBB, TUBB3, TUBB4B
CTL-DA-NR (9)	Neurotransmitter receptors and postsynaptic signal transmission	GABRA1, GABRG2, TUBB3
CTL-DA-NR (9)	GABA receptor activation	GABRA1, GABRG2
CTL-DA-NR (9)	Signaling by ERBB4	GABRA1, GABRG2
CTL-DA-NR (9)	Transmission across Chemical Synapses	GABRA1, GABRG2, TUBB3
CTL-DA-NR (9)	Translocation of SLC2A4 (GLUT4) to the plasma membrane	TUBB3, YWHAZ
DA-NR (46)	Microtubule-dependent trafficking of connexons from Golgi to the plasma membrane	TUBA4A, TUBB3
DA-NR (46)	The role of GTSE1 in G2/M progression after G2 checkpoint	PSMB3, TUBA4A, TUBB3

Tabela A.6: As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Cingulado Posterior com base na co-expressão diferencial. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (114)	The citric acid (TCA) cycle and respiratory electron transport	ATP5F1A, ATP5F1C, ATP5MC2, COX5A, COX6C, GLO1, IDH3A, IDH3B, LDHB, MDH2, NDUFB10, NDUFB5, UQCRCF1
CTL-NR (114)	Prefoldin mediated transfer of substrate to CCT/-TriC	ACTB, CCT4, PFDN5, TCP1, TUBA3C, TUBA3D
CTL-NR (114)	PCP/CE pathway	AP2S1, CLTB, CLTC, PSMA1, PSMA2, PSMB1, PSMB3, PSMB6, UBC

Tabela A.6 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (114)	Infectious disease	AP2S1, CCNH, CHMP2A, CLTC, MAP2K4, NUP88, POLR2K, PSMA1, PSMA2, PSMB1, PSMB3, PSMB6, RPL10A, RPL8, SLC25A4, SV2B, UBC
CTL-NR (114)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	ACTB, CCT4, PFDN5, TCP1, TUBA3C, TUBA3D
CTL-NR (114)	The role of GTSE1 in G2/M progression after G2 checkpoint	PSMA1, PSMA2, PSMB1, PSMB3, PSMB6, TUBA3C, TUBA3D, UBC
CTL-NR (114)	Cellular responses to stress	ATP6V1B2, DCTN6, DYNC1H1, DYNLL1, MAP2K4, NUP88, PRDX3, PRDX5, PSMA1, PSMA2, PSMB1, PSMB3, PSMB6, SOD1, TUBA3C, TUBA3D, TXN, UBC
CTL-NR (114)	Beta-catenin independent WNT signaling	AP2S1, CLTB, CLTC, ITPR1, PSMA1, PSMA2, PSMB1, PSMB3, PSMB6, UBC
CTL-NR (114)	Clathrin-mediated endocytosis	ACTB, ACTG1, AP2S1, ARPC2, CLTB, CLTC, COPS4, NECAP1, REPS2, UBC
CTL-NR (114)	Formation of annular gap junctions	ACTB, ACTG1, CLTB, CLTC
CTL-NR (114)	EPH-Ephrin signaling	ACTB, ACTG1, AP2S1, ARPC2, CLTB, CLTC, MYL12A, MYL12B
CTL-NR (114)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	ATP5F1A, ATP5F1C, ATP5MC2, COX5A, COX6C, NDUFB10, NDUFB5, UQCRCF1
CTL-NR (114)	Folding of actin by CCT/TriC	ACTB, CCT4, TCP1
DIF-NR (50)	Cristae formation	APOO, ATP5F1A, ATP5F1B, ATP5MC2
DIF-NR (50)	The citric acid (TCA) cycle and respiratory electron transport	ATP5F1A, ATP5F1B, ATP5MC2, COX5A, LDHB, NDUFB10, NDUFB5
DIF-NR (50)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	ATP5F1A, ATP5F1B, ATP5MC2, COX5A, NDUFB10, NDUFB5
DA-NR (8)	RHO GTPases activate CIT	MYL12A, MYL12B
DA-NR (8)	RHO GTPases Activate ROCKs	MYL12A, MYL12B
DA-NR (8)	Sema4D induced cell migration and growth-cone collapse	MYL12A, MYL12B
DA-NR (8)	RHO GTPases activate PAKs	MYL12A, MYL12B
DA-NR (8)	Sema4D in semaphorin signaling	MYL12A, MYL12B
DA-NR (8)	EPHA-mediated growth cone collapse	MYL12A, MYL12B
CTL-NR (114)	Formation of ATP by chemiosmotic coupling	ATP5F1A, ATP5F1C, ATP5MC2

Tabela A.6 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
DIF-NR (50)	Organelle biogenesis and maintenance	APOO, ATP5F1A, ATP5F1B, ATP5MC2, CCT4, DYNLL1, RAB11A, TCP1
DIF-NR (50)	Formation of ATP by chemiosmotic coupling	ATP5F1A, ATP5F1B, ATP5MC2
DA-NR (8)	Smooth Muscle Contraction	MYL12A, MYL12B
CTL-NR (114)	Organelle biogenesis and maintenance	ATP5F1A, ATP5F1C, ATP5MC2, CCT4, DYNLL1, RAB11A, SSBP1, TCP1, TUBA3C, TUBA3D
DA-PR (1)	Neurexins and neuroligins	DLGAP1
DA-PR (1)	Protein-protein interactions at synapses	DLGAP1
DA-NR (8)	Semaphorin interactions	MYL12A, MYL12B
DIF-NR (50)	Mitochondrial protein import	ATP5F1A, ATP5F1B, SLC25A4, TOMM20
CTL-NR (114)	Cristae formation	ATP5F1A, ATP5F1C, ATP5MC2
DA-NR (8)	RHO GTPase Effectors	MYL12A, MYL12B, TUBA1C
CTL-NR (114)	Mitochondrial protein import	ATP5F1A, CHCHD2, SLC25A4, TOMM20
DA-NR (8)	EPH-Ephrin signaling	MYL12A, MYL12B
CTL-NR (114)	RHO GTPase Effectors	ACTB, ACTG1, ARPC2, DYNC111, DYNLL1, MYL12A, MYL12B, TUBA3C, TUBA3D
CTL-NR (114)	Smooth Muscle Contraction	ANXA6, MYL12A, MYL12B
DIF-PR (10)	Transport of Mature mRNA Derived from an Intronless Transcript	RANBP2, SYMPK
DIF-PR (10)	Transport of Mature mRNAs Derived from Intronless Transcripts	RANBP2, SYMPK
DA-NR (8)	Cellular responses to stress	ANAPC5, EEF1A1, TUBA1C
DA-PR (1)	Neuronal System	DLGAP1
DIF-NR (50)	Folding of actin by CCT/TriC	CCT4, TCP1
DIF-NR (50)	Mitochondrial biogenesis	APOO, ATP5F1A, ATP5F1B, ATP5MC2

Tabela A.7: As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Cingulado Posterior com base na dependência condicional. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
DA-NR (93)	Regulation of expression of SLITs and ROBOs	GSPT2, PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, RPL10A, RPL3, RPL4, RPL8, UBC
DA-NR (93)	Negative regulation of NOTCH4 signaling	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC, YWHAZ
DA-NR (93)	Clathrin-mediated endocytosis	ACTB, ACTG1, AMPH, ARPC3, COPS4, NE-CAP1, REPS2, SGIP1, SYNJ2, UBC
DA-NR (93)	Signaling by ROBO receptors	GSPT2, PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, RPL10A, RPL3, RPL4, RPL8, UBC
DA-NR (93)	Regulation of RUNX2 expression and activity	PPARGC1A, PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	Regulation of activated PAK-2p34 by proteasome mediated degradation	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	Vpu mediated degradation of CD4	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	Autodegradation of the E3 ubiquitin ligase COP1	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	Ubiquitin Mediated Degradation of Phosphorylated Cdc25A	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	p53-Independent DNA Damage Response	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	Autodegradation of Cdh1 by Cdh1:APC/C	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	APC/C:Cdc20 mediated degradation of Securin	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	APC/C:Cdh1 mediated degradation of Cdc20 and other APC/C:Cdh1 targeted proteins in late mitosis/early G1	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	Cdc20:Phospho-APC/C mediated degradation of Cyclin A	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	CDK-mediated phosphorylation and removal of Cdc6	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	APC:Cdc20 mediated degradation of cell cycle proteins prior to satisfaction of the cell cycle checkpoint	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	APC/C:Cdc20 mediated degradation of mitotic proteins	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
DA-NR (93)	The role of GTSE1 in G2/M progression after G2 checkpoint	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC

Tabela A.7 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (86)	The citric acid (TCA) cycle and respiratory electron transport	ACO2, ATP5F1C, ATP5MC2, ATP5MC3, COX5A, IDH3A, IDH3B, NDUFS3, UQCRC1, UQCRC1
CTL-NR (86)	The role of GTSE1 in G2/M progression after G2 checkpoint	HSP90AB1, PSMA1, PSMA2, PSMD1, PSMD6, TUBA1C, UBC
DA-NR (93)	PCP/CE pathway	PSMA1, PSMB1, PSMB3, PSMD4, PSMD6, UBC
CTL-NR (86)	PCP/CE pathway	AP2S1, CLTB, PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	Autodegradation of Cdh1 by Cdh1:APC/C	ANAPC5, PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	APC/C:Cdc20 mediated degradation of Securin	ANAPC5, PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	APC/C:Cdh1 mediated degradation of Cdc20 and other APC/C:Cdh1 targeted proteins in late mitosis/early G1	ANAPC5, PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	Cdc20:Phospho-APC/C mediated degradation of Cyclin A	ANAPC5, PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	CDK-mediated phosphorylation and removal of Cdc6	ANAPC5, PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	APC:Cdc20 mediated degradation of cell cycle proteins prior to satisfaction of the cell cycle checkpoint	ANAPC5, PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	APC/C:Cdc20 mediated degradation of mitotic proteins	ANAPC5, PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	Regulation of activated PAK-2p34 by proteasome mediated degradation	PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	Vpu mediated degradation of CD4	PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	Autodegradation of the E3 ubiquitin ligase COP1	PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	Ubiquitin Mediated Degradation of Phosphorylated Cdc25A	PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	p53-Independent DNA Damage Response	PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	Negative regulation of NOTCH4 signaling	PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-NR (86)	Clathrin-mediated endocytosis	ACTB, ACTG1, AP2S1, ARPC2, CLTB, COPS4, UBC
CTL-NR (86)	Regulation of RUNX2 expression and activity	PSMA1, PSMA2, PSMD1, PSMD6, UBC
CTL-DA-NR (2)	Transmission across Chemical Synapses	GABRA1, SYN2
CTL-DA-NR (2)	Neuronal System	GABRA1, SYN2
CTL-DA-NR (2)	Serotonin Neurotransmitter Release Cycle	SYN2
CTL-DA-NR (2)	Dopamine Neurotransmitter Release Cycle	SYN2

Tabela A.7 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (86)	Regulation of expression of SLITs and ROBOs	PSMA1, PSMA2, PSMD1, PSMD6, RPL8, UBC
CTL-DA-NR (2)	Neurotransmitter release cycle	SYN2
CTL-DA-NR (2)	GABA receptor activation	GABRA1
CTL-DA-NR (2)	Signaling by ERBB4	GABRA1
DA-NR (93)	The citric acid (TCA) cycle and respiratory electron transport	ATP5F1C, ATP5PB, COX4I1, IDH3B, NDUFA13, VDAC1
CTL-NR (86)	Signaling by ROBO receptors	PSMA1, PSMA2, PSMD1, PSMD6, RPL8, UBC
DA-NR (93)	Neurotransmitter receptors and postsynaptic signal transmission	ARHGEF7, GABRA1, NCALD, NEFL, NSF, RASGRF2
DA-NR (93)	Transmission across Chemical Synapses	ARHGEF7, GABRA1, NCALD, NEFL, NSF, RASGRF2, SYN2
CTL-DA-NR (2)	Neurotransmitter receptors and postsynaptic signal transmission	GABRA1

Tabela A.8: As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Hipocampo com base na co-expressão diferencial. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (52)	Signaling by Hedgehog	CUL3, GNAS, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, TUBA1B, TUBB3, UBC
CTL-NR (52)	Hedgehog 'off' state	GNAS, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, TUBA1B, TUBB3, UBC
CTL-NR (52)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	ACTB, CCT2, CCT4, CCT7, PFDN5, TUBA1B, TUBB3
CTL-NR (52)	Negative regulation of NOTCH4 signaling	PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, UBC, YWHAZ
CTL-NR (52)	Degradation of DVL	CUL3, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, UBC
CTL-NR (52)	Host Interactions of HIV factors	AP2A2, AP2M1, PPIA, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, UBC

Tabela A.8 continuação da página anterior

Expressão	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (52)	Regulation of RAS by GAPs	CUL3, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, UBC
CTL-NR (52)	Interleukin-1 signaling	MAP2K1, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, UBC, UBE2N
CTL-NR (52)	RAF/MAP kinase cascade	ACTB, ACTG1, CUL3, MAP2K1, PEBP1, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, UBC
CTL-NR (52)	MAPK1/MAPK3 signaling	ACTB, ACTG1, CUL3, MAP2K1, PEBP1, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, UBC
CTL-NR (52)	Prefoldin mediated transfer of substrate to CCT/-TriC	ACTB, CCT2, CCT4, CCT7, PFDN5, TUBB3
CTL-NR (52)	Chaperonin-mediated protein folding	ACTB, CCT2, CCT4, CCT7, CSNK2B, PFDN5, TUBA1B, TUBB3
CTL-NR (52)	Protein folding	ACTB, CCT2, CCT4, CCT7, CSNK2B, PFDN5, TUBA1B, TUBB3
CTL-NR (52)	L1CAM interactions	ACTB, ACTG1, AP2A2, AP2M1, CSNK2B, MAP2K1, TUBA1B, TUBB3
CTL-NR (52)	Folding of actin by CCT/TriC	ACTB, CCT2, CCT4, CCT7
CTL-NR (52)	Formation of tubulin folding intermediates by CCT/TriC	CCT2, CCT4, CCT7, TUBA1B, TUBB3
CTL-NR (52)	Regulation of mRNA stability by proteins that bind AU-rich elements	PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, UBC, YWHAZ
CTL-NR (52)	Infectious disease	AP2A2, AP2M1, MAP2K1, PPIA, PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, RBX1, UBC, XRCC6
CTL-NR (52)	AUF1 (hnRNP D0) binds and destabilizes mRNA	PSMA1, PSMB2, PSMB6, PSMD4, PSMD6, UBC
CTL-NR (52)	Formation of annular gap junctions	ACTB, ACTG1, AP2M1
CTL-NR (52)	Gap junction degradation	ACTB, ACTG1, AP2M1
DIF-NR (54)	Prefoldin mediated transfer of substrate to CCT/-TriC	CCT2, CCT4, CCT7, PFDN2, TUBB3
DIF-NR (54)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	CCT2, CCT4, CCT7, PFDN2, TUBB3
DIF-NR (54)	Gene and protein expression by JAK-STAT signaling after Interleukin-12 stimulation	PITPNA, PPIA, RPLP0, SOD1, TALDO1
DIF-NR (54)	Infectious disease	AP2M1, CHMP2A, KPNA2, PPIA, PSMB6, PSMD4, PSMD6, RPL29, RPLP0, UBC, XRCC6
DIF-NR (54)	Interleukin-12 signaling	PITPNA, PPIA, RPLP0, SOD1, TALDO1
DIF-NR (54)	Regulation of mRNA stability by proteins that bind AU-rich elements	HSPA8, PSMB6, PSMD4, PSMD6, SET, UBC

Tabela A.8 continuação da página anterior

Expressão	Via Biológica (<i>Pathway</i>)	Genes
DIF-NR (54)	Formation of tubulin folding intermediates by CCT/TriC	CCT2, CCT4, CCT7, TUBB3
DIF-NR (54)	Macroautophagy	CHMP2A, CSNK2B, DYNLL1, GABARAPL2, HSPA8, UBC
DIF-NR (54)	Chaperonin-mediated protein folding	CCT2, CCT4, CCT7, CSNK2B, PFDN2, TUBB3
DIF-NR (54)	AUF1 (hnRNP D0) binds and destabilizes mRNA	HSPA8, PSMB6, PSMD4, PSMD6, UBC
DIF-NR (54)	Signaling by Hedgehog	CUL3, GNAS, PSMB6, PSMD4, PSMD6, TUBB3, UBC
DIF-NR (54)	Degradation of DVL	CUL3, PSMB6, PSMD4, PSMD6, UBC
DIF-NR (54)	Protein folding	CCT2, CCT4, CCT7, CSNK2B, PFDN2, TUBB3
DIF-NR (54)	Folding of actin by CCT/TriC	CCT2, CCT4, CCT7
DIF-NR (54)	Hedgehog 'off' state	GNAS, PSMB6, PSMD4, PSMD6, TUBB3, UBC
DIF-NR (54)	Regulation of RAS by GAPs	CUL3, PSMB6, PSMD4, PSMD6, UBC
DIF-NR (54)	Host Interactions of HIV factors	AP2M1, PPIA, PSMB6, PSMD4, PSMD6, UBC
DIF-NR (54)	Negative regulation of NOTCH4 signaling	PSMB6, PSMD4, PSMD6, UBC
CTL-NR (52)	Gene and protein expression by JAK-STAT signaling after Interleukin-12 stimulation	PPIA, SOD1, TALDO1
CTL-NR (52)	Macroautophagy	CSNK2B, DYNLL1, GABARAPL2, UBC
CTL-NR (52)	Interleukin-12 signaling	PPIA, SOD1, TALDO1
DA-NR (15)	Folding of actin by CCT/TriC	ACTB, CCT7
DA-NR (15)	Formation of annular gap junctions	ACTB, ACTG1
DA-NR (15)	Gap junction degradation	ACTB, ACTG1
DA-NR (15)	Cell-extracellular matrix interactions	ACTB, ACTG1
DA-NR (15)	Chaperonin-mediated protein folding	ACTB, CCT7, CSNK2B
DA-NR (15)	Protein folding	ACTB, CCT7, CSNK2B
DIF-NR (54)	RAF/MAP kinase cascade	CUL3, PEBP1, PSMB6, PSMD4, PSMD6, UBC
DIF-NR (54)	MAPK1/MAPK3 signaling	CUL3, PEBP1, PSMB6, PSMD4, PSMD6, UBC
DIF-NR (54)	Interleukin-1 signaling	PSMB6, PSMD4, PSMD6, UBC
DA-NR (15)	L1CAM interactions	ACTB, ACTG1, CSNK2B
DA-NR (15)	Prefoldin mediated transfer of substrate to CCT/TriC	ACTB, CCT7
DA-NR (15)	Cell-Cell communication	ACTB, ACTG1, SIRPA
DA-NR (15)	Interaction between L1 and Ankyrins	ACTB, ACTG1
DA-NR (15)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	ACTB, CCT7
CTL-NR (52)	Cell-extracellular matrix interactions	ACTB, ACTG1
DIF-NR (54)	L1CAM interactions	AP2M1, CSNK2B, HSPA8, TUBB3
CTL-NR (52)	Interaction between L1 and Ankyrins	ACTB, ACTG1

Tabela A.8 continuação da página anterior

Expressão	Via Biológica (<i>Pathway</i>)	Genes
DIF-PR (11)	Interaction between L1 and Ankyrins	NFASC, SPTAN1
DIF-PR (11)	mRNA Splicing	RNPC3, SRRM2, SYMPK
CTL-PR (12)	Interaction between L1 and Ankyrins	NFASC, SPTAN1
CTL-PR (12)	mRNA Splicing	RNPC3, SRRM2, SYMPK
DIF-PR (11)	Processing of Capped Intron-Containing Pre-mRNA	RNPC3, SRRM2, SYMPK
CTL-PR (12)	Processing of Capped Intron-Containing Pre-mRNA	RNPC3, SRRM2, SYMPK

Tabela A.9: As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Hipocampo com base na dependência condicional. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (43)	MHC class II antigen presentation	AP2A2, AP2M1, CANX, DYNLL1, KIFAP3, TUBA1B, TUBB3
CTL-NR (43)	Regulation of mRNA stability by proteins that bind AU-rich elements	HSPA8, PSMB2, PSMB6, PSMD6, SET, YWHAZ
CTL-NR (43)	Formation of tubulin folding intermediates by CCT/TriC	CCT2, CCT7, TUBA1B, TUBB3
CTL-NR (43)	Hedgehog 'off' state	GNAS, PSMB2, PSMB6, PSMD6, TUBA1B, TUBB3
CTL-NR (43)	L1CAM interactions	AP2A2, AP2M1, HSPA8, MAP2K1, TUBA1B, TUBB3
CTL-NR (43)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	CCT2, CCT7, TUBA1B, TUBB3
CTL-NR (43)	Signaling by Interleukins	BRWD1, CANX, HSPA8, MAP2K1, PSMB2, PSMB6, PSMD6, SOD1, UBE2N, YWHAZ
CTL-NR (43)	The role of GTSE1 in G2/M progression after G2 checkpoint	PSMB2, PSMB6, PSMD6, TUBA1B, TUBB3
CTL-NR (43)	Infectious disease	AP2A2, AP2M1, CANX, KPNA2, MAP2K1, PSMB2, PSMB6, PSMD6, XRCC5

Tabela A.9 continuação da página anterior

Expressão	Via Biológica (<i>Pathway</i>)	Genes
DA-NR (15)	RHO GTPases activate IQGAPs	ACTB, ACTG1, TUBB3
DA-NR (15)	Signaling by high-kinase activity BRAF mutants	ACTB, ACTG1, PEBP1
DA-NR (15)	L1CAM interactions	ACTB, ACTG1, CSNK2B, TUBB3
DA-NR (15)	MAP2K and MAPK activation	ACTB, ACTG1, PEBP1
DA-NR (15)	RHO GTPases Activate Formins	ACTB, ACTG1, DYNC1H1, TUBB3
DA-NR (15)	Gap junction trafficking	ACTB, ACTG1, TUBB3
DA-NR (15)	Recycling pathway of L1	ACTB, ACTG1, TUBB3
DA-NR (15)	Signaling by moderate kinase activity BRAF mutants	ACTB, ACTG1, PEBP1
DA-NR (15)	Signaling by RAS mutants	ACTB, ACTG1, PEBP1
DA-NR (15)	Paradoxical activation of RAF signaling by kinase inactive BRAF	ACTB, ACTG1, PEBP1
CTL-NR (43)	Signaling by Hedgehog	GNAS, PSMB2, PSMB6, PSMD6, TUBA1B, TUBB3
CTL-NR (43)	Recycling pathway of L1	AP2A2, AP2M1, TUBA1B, TUBB3
DA-PR (1)	IL-6-type cytokine receptor ligand interactions	LIFR
DA-PR (1)	Interleukin-6 family signaling	LIFR
CTL-NR (43)	Gap junction trafficking	AP2M1, TUBA1B, TUBB3
DA-NR (15)	Cooperation of Prefoldin and TriC/CCT in actin and tubulin folding	ACTB, TUBB3
CTL-NR (43)	RHO GTPases activate IQGAPs	TUBA1B, TUBB3
CTL-NR (43)	Signaling by high-kinase activity BRAF mutants	MAP2K1, PEBP1
CTL-NR (43)	MAP2K and MAPK activation	MAP2K1, PEBP1
DA-PR (1)	Transcriptional regulation by RUNX1	LIFR
CTL-NR (43)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	ATP5PB, NDUFA11, UQCRCF1
DA-NR (15)	The role of GTSE1 in G2/M progression after G2 checkpoint	PSMD4, TUBB3
CTL-NR (43)	Signaling by moderate kinase activity BRAF mutants	MAP2K1, PEBP1
CTL-NR (43)	Signaling by RAS mutants	MAP2K1, PEBP1
CTL-NR (43)	Paradoxical activation of RAF signaling by kinase inactive BRAF	MAP2K1, PEBP1
CTL-NR (43)	Transcriptional regulation by RUNX1	CCND2, PSMB2, PSMB6, PSMD6
CTL-DA-NR (2)	Neurexins and neuroligins	EPB41L3
CTL-DA-NR (2)	Protein-protein interactions at synapses	EPB41L3
CTL-DA-NR (2)	Respiratory electron transport	UQCRCF1

Tabela A.9 continuação da página anterior

Expressão	Via Biológica (<i>Pathway</i>)	Genes
CTL-DA-NR (2)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	UQCRFS1
CTL-DA-NR (2)	The citric acid (TCA) cycle and respiratory electron transport	UQCRFS1
CTL-NR (43)	RHO GTPases Activate Formins	DYNLL1, TUBA1B, TUBB3
DA-PR (1)	Signaling by Interleukins	LIFR

Tabela A.10: As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Giro Temporal Medial com base na co-expressão diferencial. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão	Via Biológica (<i>Pathway</i>)	Genes
DIF-NR (8)	Downstream signaling events of B Cell Receptor (BCR)	PPP3CA, PSMB4, PSMD1
DIF-NR (8)	Creatine metabolism	CKMT1A, CKMT1B
DIF-NR (8)	CLEC7A (Dectin-1) signaling	PPP3CA, PSMB4, PSMD1
DIF-NR (8)	Interleukin-1 signaling	MAP2K1, PSMB4, PSMD1
DIF-NR (8)	Signaling by the B Cell Receptor (BCR)	PPP3CA, PSMB4, PSMD1
DIF-NR (8)	Metabolism of amino acids and derivatives	CKMT1A, CKMT1B, PSMB4, PSMD1
DIF-NR (8)	Fc epsilon receptor (FCER1) signaling	PPP3CA, PSMB4, PSMD1
DIF-NR (8)	Interleukin-1 family signaling	MAP2K1, PSMB4, PSMD1
DIF-NR (8)	C-type lectin receptors (CLRs)	PPP3CA, PSMB4, PSMD1
DIF-NR (8)	Beta-catenin independent WNT signaling	PPP3CA, PSMB4, PSMD1
DIF-NR (8)	Cross-presentation of soluble exogenous antigens (endosomes)	PSMB4, PSMD1
DIF-NR (8)	Regulation of activated PAK-2p34 by proteasome mediated degradation	PSMB4, PSMD1
DIF-NR (8)	Regulation of ornithine decarboxylase (ODC)	PSMB4, PSMD1
DIF-NR (8)	Vpu mediated degradation of CD4	PSMB4, PSMD1
DIF-NR (8)	Autodegradation of the E3 ubiquitin ligase COP1	PSMB4, PSMD1

Tabela A.10 continuação da página anterior

Expressão	Via Biológica (<i>Pathway</i>)	Genes
DA-PR (4)	Activation of gene expression by SREBF (SREBP)	ACACB, TBL1X
DA-PR (4)	Regulation of cholesterol biosynthesis by SREBP (SREBF)	ACACB, TBL1X
DIF-NR (8)	Infectious disease	MAP2K1, PSMB4, PSMD1
DA-NR (15)	Neurotoxicity of clostridium toxins	SV2B, SYT1
DA-NR (15)	Creatine metabolism	CKMT1A, CKMT1B
DIF-NR (8)	Host Interactions of HIV factors	PSMB4, PSMD1
CTL-NR (11)	Downstream signaling events of B Cell Receptor (BCR)	PPIA, PSMA1, PSMB4
CTL-NR (11)	Signaling by the B Cell Receptor (BCR)	PPIA, PSMA1, PSMB4
CTL-NR (11)	Host Interactions of HIV factors	PPIA, PSMA1, PSMB4
CTL-NR (11)	Infectious disease	PPIA, PSMA1, PSMB4, SV2B
CTL-NR (11)	Disorders of transmembrane transporters	PSMA1, PSMB4, SLC9A6
CTL-NR (11)	Cross-presentation of soluble exogenous antigens (endosomes)	PSMA1, PSMB4
CTL-NR (11)	Regulation of activated PAK-2p34 by proteasome mediated degradation	PSMA1, PSMB4
CTL-NR (11)	Regulation of ornithine decarboxylase (ODC)	PSMA1, PSMB4
CTL-NR (11)	Vpu mediated degradation of CD4	PSMA1, PSMB4
CTL-NR (11)	Autodegradation of the E3 ubiquitin ligase COP1	PSMA1, PSMB4
CTL-NR (11)	CLEC7A (Dectin-1) signaling	PSMA1, PSMB4
CTL-NR (11)	Interleukin-1 signaling	PSMA1, PSMB4
DIF-NR (8)	Disorders of transmembrane transporters	PSMB4, PSMD1
CTL-NR (11)	Fc epsilon receptor (FCERI) signaling	PSMA1, PSMB4
CTL-NR (11)	Interleukin-1 family signaling	PSMA1, PSMB4
CTL-NR (11)	C-type lectin receptors (CLRs)	PSMA1, PSMB4
CTL-NR (11)	Beta-catenin independent WNT signaling	PSMA1, PSMB4
CTL-NR (11)	Neurotoxicity of clostridium toxins	SV2B
DA-PR (4)	Metabolism of steroids	ACACB, TBL1X
DA-NR (15)	Uptake and actions of bacterial toxins	SV2B, SYT1
DA-NR (15)	Gluconeogenesis	ENO2, GPI
DIF-NR (8)	Uptake and actions of bacterial toxins	MAP2K1
DA-PR (4)	Biotin transport and metabolism	ACACB
DA-PR (4)	RHO GTPases activate KTN1	KLC1
DA-PR (4)	Import of palmitoyl-CoA into the mitochondrial matrix	ACACB
DA-PR (4)	Loss of function of MECP2 in Rett syndrome	TBL1X

Tabela A.10 continuação da página anterior

Expressão	Via Biológica (<i>Pathway</i>)	Genes
DA-PR (4)	Pervasive developmental disorders	TBL1X
DA-PR (4)	RORA activates gene expression	TBL1X
CTL-NR (11)	Uptake and actions of bacterial toxins	SV2B
DA-PR (4)	Other interleukin signaling	PTPRZ1

Tabela A.11: **As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Giro Temporal Medial com base na dependência condicional.** CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (50)	The role of GTSE1 in G2/M progression after G2 checkpoint	HSP90AB1, PSMA1, PSMB3, PSMB4, PSMB7, PSMD1, TUBB3, TUBB4B
DA-NR (111)	Transmission across Chemical Synapses	ARHGEF7, CACNA2D3, GABBR2, GABRA1, GABRB3, GABRG2, GLS, GLS2, GNG3, NCALD, NEFL, NSF, SYN2, SYT1, TUBA1B, TUBB3, TUBB4B
DA-NR (111)	Neuronal System	ARHGEF7, CACNA2D3, GABBR2, GABRA1, GABRB3, GABRG2, GLS, GLS2, GNG3, NCALD, NEFL, NRXN1, NSF, RTN3, SLITRK5, SYN2, SYT1, TUBA1B, TUBB3, TUBB4B
CTL-NR (50)	G2/M Transition	HSP90AB1, PAFAH1B1, PSMA1, PSMB3, PSMB4, PSMB7, PSMD1, TUBB, TUBB3, TUBB4B
CTL-NR (50)	Mitotic G2-G2/M phases	HSP90AB1, PAFAH1B1, PSMA1, PSMB3, PSMB4, PSMB7, PSMD1, TUBB, TUBB3, TUBB4B
CTL-NR (50)	Hedgehog 'off' state	PSMA1, PSMB3, PSMB4, PSMB7, PSMD1, TUBB3, TUBB4B
CTL-NR (50)	Host Interactions of HIV factors	PSMA1, PSMB3, PSMB4, PSMB7, PSMD1, RAN, SLC25A4
CTL-NR (50)	Separation of Sister Chromatids	PAFAH1B1, PSMA1, PSMB3, PSMB4, PSMB7, PSMD1, TUBB3, TUBB4B

Tabela A.11 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-NR (50)	Cross-presentation of soluble exogenous antigens (endosomes)	PSMA1, PSMB3, PSMB4, PSMB7, PSMD1
CTL-NR (50)	Regulation of activated PAK-2p34 by proteasome mediated degradation	PSMA1, PSMB3, PSMB4, PSMB7, PSMD1
CTL-NR (50)	Regulation of ornithine decarboxylase (ODC)	PSMA1, PSMB3, PSMB4, PSMB7, PSMD1
CTL-NR (50)	Mitotic Anaphase	PAFAH1B1, PSMA1, PSMB3, PSMB4, PSMB7, PSMD1, TUBB3, TUBB4B
DA-NR (111)	Gluconeogenesis	ENO2, GOT1, GOT2, GPI, MDH1, PGK1
DA-NR (111)	Neurotransmitter receptors and postsynaptic signal transmission	ARHGEF7, GABBR2, GABRA1, GABRB3, GABRG2, GNG3, NCALD, NEFL, NSF, TUBA1B, TUBB3, TUBB4B
DA-PR (20)	Initial triggering of complement	C4A, C4B, C4B_2
CTL-NR (50)	Gluconeogenesis	GOT1, GPI, PGK1
CTL-NR (50)	Glucose metabolism	GOT1, GPI, PGK1, PGM2L1
DA-NR (111)	Glucose metabolism	ENO2, GOT1, GOT2, GPI, MDH1, PGK1, PGM2L1
DA-PR (20)	Regulation of Complement cascade	C4A, C4B, C4B_2
DA-PR (20)	Complement cascade	C4A, C4B, C4B_2
CTL-NR (50)	Organelle biogenesis and maintenance	ATP5F1B, ATP5F1C, PAFAH1B1, TUBB, TUBB3, TUBB4B
DA-NR (111)	Formation of tubulin folding intermediates by CCT/TriC	CCT4, TUBA1B, TUBB3, TUBB4B
DA-NR (111)	Insulin receptor recycling	ATP6V1B2, ATP6V1E1, ATP6V1G2, ATP6V1H
DA-NR (111)	L1CAM interactions	CLTC, MAP2K1, PAK1, SCN2B, TUBA1B, TUBB3, TUBB4B
DA-NR (111)	Organelle biogenesis and maintenance	ATP5F1B, ATP5F1C, ATP5MC3, CCT4, MEF2C, MKKS, MTX2, TUBA1B, TUBB, TUBB3, TUBB4B
DA-NR (111)	GABA receptor activation	GABBR2, GABRA1, GABRB3, GABRG2, GNG3
CTL-NR (50)	Formation of tubulin folding intermediates by CCT/TriC	TUBB3, TUBB4B
DA-NR (111)	The role of GTSE1 in G2/M progression after G2 checkpoint	HSP90AB1, PSMA1, TUBA1B, TUBB3, TUBB4B
DA-NR (111)	G2/M Transition	CUL1, HSP90AB1, PSMA1, TUBA1B, TUBB, TUBB3, TUBB4B
DA-NR (111)	Mitotic G2-G2/M phases	CUL1, HSP90AB1, PSMA1, TUBA1B, TUBB, TUBB3, TUBB4B

Tabela A.12: As vias biológicas mais significativa para os transcritos diferencialmente expressos na região Giro Frontal Superior com base na co-expressão diferencial. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-DA-NR (2)	Purine salvage	HPRT1
CTL-DA-NR (2)	Formation of ATP by chemiosmotic coupling	ATP5F1C
CTL-DA-NR (2)	Nucleotide salvage	HPRT1
CTL-DA-NR (2)	Cristae formation	ATP5F1C
CTL-DA-NR (2)	Mitochondrial biogenesis	ATP5F1C
CTL-DA-NR (2)	Metabolism of nucleotides	HPRT1
CTL-DA-NR (2)	Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	ATP5F1C
CTL-DA-NR (2)	The citric acid (TCA) cycle and respiratory electron transport	ATP5F1C

Tabela A.13: As vias biológicas mais significativas para os transcritos diferencialmente expressos na região Giro Frontal Superior com base na dependência condicional. CTL: Controle, DA: Doença de Alzheimer, NR: Negativamente Regulado, PR: Positivamente Regulado, DIF: Diferente. Os números apresentados entre parênteses indicam a quantidade de genes anotados nas respectivas vias biológicas.

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
CTL-PR (65)	Metallothioneins bind metals	MT1H, MT1M, MT2A
DA-PR (15)	Physiological factors	HIPK2, WWTR1
DA-PR (15)	YAP1- and WWTR1 (TAZ)-stimulated gene expression	HIPK2, WWTR1
CTL-PR (65)	Response to metal ions	MT1H, MT1M, MT2A
CTL-DA-PR (2)	RHO GTPases activate CIT	CIT

Tabela A.13 continuação da página anterior

Expressão-Regulação	Via Biológica (<i>Pathway</i>)	Genes
DA-PR (15)	MET activates PTK2 signaling	COL27A1, FN1
CTL-DA-PR (2)	Neurexins and neuroligins	DLGAP1
CTL-DA-PR (2)	Protein-protein interactions at synapses	DLGAP1
DA-PR (15)	MET promotes cell motility	COL27A1, FN1
CTL-DA-NR (7)	Metabolism of nucleotides	HPRT1, NME1
CTL-DA-NR (7)	Synthesis, secretion, and inactivation of Glucose-dependent Insulinotropic Polypeptide (GIP)	PCSK1
CTL-DA-NR (7)	Purine salvage	HPRT1
CTL-DA-NR (7)	Peptide hormone biosynthesis	PCSK1
CTL-DA-NR (7)	Formation of ATP by chemiosmotic coupling	ATP5F1C
CTL-DA-NR (7)	Synthesis, secretion, and deacylation of Ghrelin	PCSK1
CTL-DA-NR (7)	Synthesis, secretion, and inactivation of Glucagon-like Peptide-1 (GLP-1)	PCSK1
CTL-DA-NR (7)	Incretin synthesis, secretion, and inactivation	PCSK1
CTL-DA-NR (7)	Nucleotide salvage	HPRT1
CTL-DA-NR (7)	Insulin receptor recycling	ATP6V1G2

Referências Bibliográficas

- Association, Alzheimer's (2019). *Alzheimer's disease facts and figures*. <https://alz.org/media/Documents/alzheimers-facts-and-figures-2019-r.pdf>.
- Belloni, A., V.Chernozhukov e Wang, L. (2011). "Square-root lasso: pivotal recovery of sparse signals via conic programming". *Biometrika* 98.4, pp. 791–806. ISSN: 00063444.
- Benjamini, Yoav e Hochberg, Yosef (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". *Journal of the Royal Statistical Society. Series B (Methodological)* 57.1, pp. 289–300. ISSN: 00359246.
- Bolstad, B. M. et al. (fev. de 2003). "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias". English (US). *Bioinformatics* 19.2, pp. 185–193. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/19.2.185.
- Boyle, Elizabeth I. et al. (ago. de 2004). "GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes". *Bioinformatics* 20.18, pp. 3710–3715. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bth456. eprint: <https://academic.oup.com/bioinformatics/article-pdf/20/18/3710/522506/bth456.pdf>. URL: <https://doi.org/10.1093/bioinformatics/bth456>.
- Brown, Patrick e Botstein, David (fev. de 1999). "Exploring the New World of the Genome With DNA Microarrays". *Nature genetics* 21, pp. 33–7. DOI: 10.1038/4462.
- Buhlmann, Peter et al., ed. (2016). *Handbook of Big Data*. Chapman e Hall/CRC. ISBN: 978-1-4822-4907-1.

- Cai, Tony, Liu, Weidong e Luo, Xi (2011). “A Constrained L1 Minimization Approach to Sparse Precision Matrix Estimation”. *Journal of the American Statistical Association* 106.494, pp. 594–607. DOI: 10.1198/jasa.2011.tm10155.
- Candes, Emmanuel e Tao, Terence (2007). “The Dantzig selector: Statistical estimation when p is much larger than n ”. *The Annals of Statistics* 35.6, 2313â2351. ISSN: 0090-5364. DOI: 10.1214/009053606000001523.
- Chaudhuri, Joydeep (mar. de 2005). “Genes arrayed out for you: The amazing world of microarrays”. *Medical science monitor : international medical journal of experimental and clinical research* 11, RA52–62.
- Cox, David Roxbee e Wermuth, Nanny (1996). “Multivariate Dependencies: Models, Analysis and Interpretation”.
- Cummings, Jeffrey et al. (2020). “Alzheimer’s disease drug development pipeline: 2020”. *Alzheimer’s and Dementia: Translational Research and Clinical Interventions* 6.1, e12050. DOI: <https://doi.org/10.1002/trc2.12050>.
- Danaher, Patrick, Wang, Pei e Witten, Daniela M. (2012). *The joint graphical lasso for inverse covariance estimation across multiple classes*. arXiv: 1111.0324 [stat.ME].
- Feldhaus, P. et al. (2011). “Evaluation of respiratory chain activity in lymphocytes of patients with Alzheimer disease”. *Metabolic brain disease* 26.3, 229â236. DOI: <https://doi.org/10.1007/s11011-011-9253-y>.
- Fieller, E. C., Hartley, H. O. e Pearson, E. S. (1957). “Tests for Rank Correlation Coefficients. I”. *Biometrika* 44.3/4, pp. 470–481. ISSN: 00063444.
- Friedman, Jerome, Hastie, Trevor e Tibshirani, Robert (ago. de 2008). “Sparse inverse covariance estimation with the graphical LASSO”. *Biostatistics (Oxford, England)* 9, pp. 432–41. DOI: 10.1093/biostatistics/kxm045.
- G., Yap (2002). “Affymetrix, Inc”. *Pharmacogenomics* 3(5), pp. 709–711. DOI: 10.1517/14622416.3.5.709.

- Goedert, Michel e Spillantini, Maria Grazia (2006). “A Century of Alzheimer’s Disease”. *Science* 314.5800, pp. 777–781. ISSN: 0036-8075. DOI: 10.1126/science.1132814.
- Hardin, Johanna (2007). “A Review of: âDNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experimentsâ, by D. B. Allison, G. P. Page, T. M. Beasley, and J. W. Edwards (Eds.)” *Journal of Biopharmaceutical Statistics* 17.1, pp. 187–190. DOI: 10.1080/10543400601001550.
- Hubert, Phipps e Arabie, Lawrence (1985). “Comparing partitions”. *Journal of Classification* 2.1, pp. 193–218.
- Illumina*. <https://www.illumina.com/techniques/microarrays.html>. Accessed: 2021-10-10.
- Irizarry, Rafael A et al. (abr. de 2003). “Exploration, normalization, and summaries of high density oligonucleotide array probe level data”. *Biostatistics* 4.2, pp. 249–264.
- Itzhaki, Ruth F. (2014). “Herpes simplex virus type 1 and Alzheimer’s disease: increasing evidence for a major role of the virus”. *Frontiers in Aging Neuroscience* 6, p. 202. ISSN: 1663-4365. DOI: 10.3389/fnagi.2014.00202.
- Jaluria, Pratik et al. (fev. de 2007). “A perspective on microarrays: Current applications, pitfalls, and potential uses”. *Microbial cell factories* 6, p. 4. DOI: 10.1186/1475-2859-6-4.
- Koller, D. e Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Adaptive computation and machine learning. MIT Press. ISBN: 9780262013192.
- Lauritzen, S. L. (1996). *Graphical Models*. Oxford, UK: Oxford University Press.
- Li, Shuang et al. (2013). “Bootstrap inference for network construction with an application to a breast cancer microarray study”. *The Annals of Applied Statistics* 7.1. ISSN: 1932-6157. DOI: 10.1214/12-aoas589.
- Li, Wentian (dez. de 2012). “Volcano plots in analyzing differential expressions with mRNA microarrays”. *Journal of bioinformatics and computational biology* 10, p. 1231003. DOI: 10.1142/S0219720012310038.

- Liang, Winnie et al. (mar. de 2007a). “Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain”. *Physiological genomics* 28, pp. 311–22. DOI: 10.1152/physiolgenomics.00208.2006.
- Liang, Winnie S. et al. (2007b). “Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain”. *Physiological Genomics* 28.3, pp. 311–322. ISSN: 1094-8341. DOI: 10.1152/physiolgenomics.00208.2006.
- Liang, Winnie S. et al. (2008). “Alzheimers disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons”. *Proceedings of the National Academy of Sciences* 105.11, pp. 4441–4446. ISSN: 0027-8424. DOI: 10.1073/pnas.0709259105.
- Lipshutz, Robert et al. (fev. de 1999). “High Density Synthetic Oligonucleotide Arrays”. *Nature genetics* 21, pp. 20–4. DOI: 10.1038/4447.
- Liu, Han e Wang, Lie (2017). “TIGER: A tuning-insensitive approach for optimally estimating Gaussian graphical models”. *Electron. J. Statist.* 11.1, pp. 241–294. DOI: 10.1214/16-EJS1195.
- Liu, Weidong e Luo, Xi (2015). “Fast and adaptive sparse precision matrix estimation in high dimensions”. *Journal of Multivariate Analysis* 135, pp. 153 –162. ISSN: 0047-259X. DOI: 10.1016/j.jmva.2014.11.005.
- Liu, Zhi-Ping et al. (2011). “Detecting and analyzing differentially activated pathways in brain regions of Alzheimer’s disease patients”. *Molecular bioSystems* 7.5, 1441â1452. ISSN: 1742-206X. DOI: 10.1039/c0mb00325e.
- Maathuis, Marloes et al. (2018). *Handbook of Graphical Models*. 1st. USA: CRC Press, Inc. ISBN: 1498788629.
- Mckenzie, Andrew et al. (dez. de 2016). “DGCA: A comprehensive R package for Differential Gene Correlation Analysis”. *BMC Systems Biology* 10, p. 106. DOI: 10.1186/s12918-016-0349-1.

- Meinshausen, Nicolai e Buhlmann, Peter (2010). “Stability selection”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.4, pp. 417–473. DOI: 10.1111/j.1467-9868.2010.00740.x.
- Organization, World Health (2019). *Dementia*. <https://www.who.int/news-room/fact-sheets/detail/dementia>.
- Pearl, Judea (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. ISBN: 0934613737.
- R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <http://www.R-project.org/>.
- Rand, William M. (1971). “Objective Criteria for the Evaluation of Clustering Methods”. *Journal of the American Statistical Association* 66.336, pp. 846–850. DOI: 10.1080/01621459.1971.10482356.
- Readhead, B. et al. (2018). “Multiscale Analysis of Independent Alzheimer’s Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus”. *Neuron* 99, pp. 64–82. ISSN: 1662-5099. DOI: 10.1016/j.neuron.2018.05.023.
- Silva, Alex Pena Tosta (2012). *Implementação, análise e aplicação de algoritmos de agrupamento de dados superdimensionados, longitudinais e com amostras pequenas*. Dissertação (Mestrado em Estatística), Universidade de Brasília.
- Souza, Thaysa Guimarães (2013). *Tópicos em análise de experimentos longitudinais para aplicações em estudos de sinais biopotenciais*. Dissertação (Mestrado em Estatística), Universidade de Brasília.
- Sulaimanov, Nurgazy e Koepl, Heinz (2016). “Graph reconstruction using covariance-based methods”. *EURASIP journal on bioinformatics and systems biology* 2016.1, p. 19. ISSN: 1687-4145. DOI: 10.1186/s13637-016-0052-y.
- Thode, Henry C. (2002). *Testing For Normality*. 1st. Boca Raton: CRC Press. ISBN: 9780429213250. DOI: 10.1201/9780203910894.

- von Borries, George (2008). *Partition clustering of High Dimensional Low Sample Size data based on P-Values*. Tese (Doutorado em Estatística).
- (2009). *A SAS/JMP Integration for Implementation of a Clustering Algorithm for High Dimensional Low Sample Size Data*. SAS Global Forum. Paper 283.
- Wang, Haiyan e Akritas, Michael G. (2004). “Rank tests for anova with large number of factor levels”. *Journal of Nonparametric Statistics* 16.3-4, pp. 563–589. DOI: 10.1080/10485250310001624774.
- Wang, Lihua e Liu, Zhi-Ping (2019). “Detecting Diagnostic Biomarkers of Alzheimer’s Disease by Integrating Gene Expression Data in Six Brain Regions”. *Frontiers in Genetics* 10, p. 157. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00157.
- Weirauch, Matthew (abr. de 2011). *Gene Coexpression Networks for the Analysis of DNA Microarray Data*. Vol. 1, pp. 215–250. DOI: 10.1002/9783527638079.ch11.
- Whittaker, Joe (2009). *Graphical Models in Applied Multivariate Statistics*. Wiley Publishing. ISBN: 0470743662.
- Wu, Jean e James MacDonald Jeff Gentry, Rafael Irizarry with contributions from (2020). *gcrma: Background Adjustment Using Sequence Information*. R package version 2.60.0.
- Wu, Zhijin et al. (2004). “A Model-Based Background Adjustment for Oligonucleotide Expression Arrays”. *Journal of the American Statistical Association* 99.468, pp. 909–917. DOI: 10.1198/016214504000000683.
- Yu, Guangchuang e He, Qing-Yu (2016). “ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization”. *Mol. BioSyst.* 12 (2), pp. 477–479. DOI: 10.1039/C5MB00663E.