Université Paris Cité

Inserm
La science pour la santé
From science to health

Université Paris Cité

Ecole doctorale Bio Sorbonne Paris Cité BioSPC ED562

Laboratoire des maladies neurodéveloppementales et neurovasculaires NeuroDiderot INSERM U1141

Multi-omic molecular signatures from peripheral blood samples of patients with mood disorders

-Par MOKHTARI Amazigh-

-Thèse de doctorat en Génétique-

Dirigée par Pu-PH Andrée Delahaye-Duriez
Et par Dr Pierre-Eric Lutz

Présentée et soutenue publiquement le 23 Novembre 2022

Devant un jury composé de :

Dr. FRIES Gabriel Assistant Professor, UTHealth Houston.	**Rapporteur**
Dr. VAN DONGEN Jenny Assistant Professor, Vrije Universiteit Amsterdam.	**Rapporteur**
Dr. BAUDOT Anaïs DR CNRS, Université Aix Marseille.	**Examinatrice**
Pu BINDER Elisabeth PU, Max Planck Institute of Psychiatry.	**Examinatrice**
Dr. RAMOZ Nicolas CR-HDR, INSERM, Université Paris-Cité.	**Examinateur**
Pr TENENHAUS Arthur PU, Université Paris-Saclay.	**Examinateur**
Pr DELAHAYE-DURIEZ Andrée Professeur des Universités-Praticien Hospitalier, Université Sorbonne Paris Nord.	**Directrice de thèse**
Dr LUTZ Pierre-Eric CR-HDR CNRS, Université De Strasbourg.	**Co-Directeur de thèse**

Acknowledgement

First of all, I would like to express my deepest gratitude to my two supervisors Pr Andrée Delahaye-Duriez and Dr Pierre-Eric Lutz, for having placed their trust in me through this opportunity to collaborate with them during the four years that preceded this thesis, but also for their patience and valuable feedback that allowed me to progress and move forward intellectually, professionally and personally. I will be forever indebted to you.

This thesis and the presented work result from multiple collaborations. I would like to especially thank Dr. Arnaud Gloaguen for his valuable contributions and feedback on this work, the enriching scientific exchanges that we had, which helped me understand multi-block integration methods, and his precious help in the formalization of this thesis. I would also like to thank Dr. David Peter Cohen and Margot Derouin for their valuable contributions in carrying out part of the analyses presented below. Finally, I would also like to thank all the members of the CT-ScarBip consortium: Pr. Bruno Etain, Dr. Cynthia Marie Claire, Pr. Raoul Belzeaux and Dr. El Chérif Ibrahim for the various and enriching scientific exchanges.

My thanks also go to the colleagues I had the opportunity to meet during these years spent in the NeuroDiderot lab unit, and who became friends whose support was precious and appreciated: Cindy, Alexandre, Julie, and Amélia, and in particular, Alexandra and Eléonore for all the billiard games of a poor sporting level but always punctuated with laughter and good humor. Thanks also to Baptiste and Christophe for the unlikely but always humorous conversations and the liters of shared coffee.

All my gratitude also goes to my family, my sister Thinhinane, my brother Amayas, and above all to my parents, without whom all my adventure from my native village in Algeria to the achievement of this work would not have been possible.

And finally, I would like to thank my wife Sana, who has been a looter with daily and unwavering support for almost eight years. Thank you for guiding me and making me a better man. I am Thankful our paths have crossed.

To my grand mother Yema Fatma

Contents

Résumé substantiel vi

Summary vi

Liste of figures xi

Liste of tables xii

Liste of supplemantary figures and tables xiii

Abreviations xv

I State of the art 10

1 Chapter 1 : Understanding mood disorders - a shifted paradigm 12

1.1 Mood disorders in DSM and ICD: phenomenology, symptoms, and causes . 13

1.1.1 Mood episodes 13

1.1.2 Major Depressive Disorder 15

1.1.3 Bipolar Disorder 16

1.1.4 Mood Disorders comorbidities 17

1.2 The etiology, susceptibility and risk factors of mood disorders 17

1.2.1 Genetics 17

1.2.2 Sex differences 18

1.2.3 The impact of childhood Trauma and stressful life events 18

1.3 The pathophysiology of mood disorders 19

1.3.1 The hypothalamic pituitary-adrenal axis hypothesis 19

1.3.2 The immune system responses 21

1.3.3 The need for a new paradigm 22

2 Chapter 2: Translational research for psychiatric disorders : From single to multi-omics approaches 23

2.1 Omics-based translational psychiatry research 24

2.1.1 Genomics 24

2.1.2 GWAS 25

2.1.3 CNVs 27

2.1.4 Transcriptomics 27

2.1.5 Transcriptomics of psychiatric disorders 29

2.1.6 Epigenetics . . . 33
2.1.7 Covalent modification of DNA . . . 34
2.1.8 Histone's post-translational modification . . . 38
2.1.9 Non coding RNAs . . . 39
2.2 Multi-omics : The whole is greater than the sum of its parts . . . 42
2.2.1 Untangle mood disorders through multi-Omics approaches . . . 42
2.2.2 Review Paper . . . 43

3 Chapter 3: Multi-Omics data integration approaches 62
3.1 Multi-omics nomenclature and classification . . . 64
3.1.1 Step-wise Integration . . . 64
3.1.2 Horizontal and Vertical Integration . . . 64
3.1.3 Simultaneous Data Integration . . . 67
3.1.4 Unsupervised and Supervised Integration . . . 67
3.2 Integration strategies . . . 68
3.2.1 Network based . . . 68
3.2.2 Machine learning based . . . 78
3.3 Challenges in multi-omics integration . . . 90
3.3.1 Data heterogenity . . . 90
3.3.2 Multi-collinearity . . . 90
3.3.3 Missing data . . . 91
3.3.4 Cross-Validation in unsupervised analysis . . . 91
3.3.5 Validation . . . 91

II Contribution 92

4 Chapter 4: Sex-specific and multi-omic integration to enhance accuracy of peripheral blood biomarkers of major depressive disorder 94
4.1 Introduction . . . 96
4.2 Material and methods . . . 98
4.2.1 Human samples . . . 98
4.2.2 DNA methylation arrays . . . 100
4.2.3 RNA isolation . . . 100
4.2.4 miRNA-sequencing . . . 101
4.2.5 RNA-sequencing . . . 101
4.2.6 Raw data processing . . . 101
4.2.7 Analyses of covariates . . . 102

4.2.8 Differential analyses . . . 102
4.2.9 Step-wise multi-omic integration . . . 103
4.2.10 Advanced integration . . . 103
4.2.11 Cross-validation . . . 104
4.2.12 Bootstrapping . . . 105
4.3 Results . . . 105
4.3.1 Single-omic analyses reveal mRNA, micro-RNA and DNA methylation changes in males and females with MDD . . . 105
4.3.2 Comparison of genome-wide changes in males and females with MDD reveals shared and sex-specific single-omic signatures . . . 107
4.3.3 Step-wise integration identified male and female MDD-associated gene co-expression modules enriched for MDD-related transcriptional and epigenetic dysregulations and mood-related SNPs . . . 109
4.3.4 Advanced integration with supervised feature selection and non-supervised clustering prioritize mRNAs and DNA methylation CpGs as more efficient biomarkers of MDD . . . 114
4.3.5 Different omics analyses and integration strategies shows common features serving as potential biomarkers and molecular signature for sex-specific MDD . . . 117
4.4 Discussion . . . 117
4.4.1 Sex specific molecular signatures of MDD . . . 117
4.4.2 Added value and limitations of our multi-omic data integration framework . . . 120
4.4.3 Sample size limitations . . . 122
4.4.4 Is blood a relevant tissue for mood disorder research ? . . . 123
4.4.5 Confounding factors . . . 123
4.5 Perspectives . . . 124
4.6 Conclusion . . . 125
4.7 Supplementary material . . . 125
4.7.1 Adjusting data for batch effects with ComBat while preserving group differences may lead to an over estimation of the significance of the differential analyses. . . . 125
4.7.2 Label Propagation . . . 127
4.8 Supplementary Figures and Tables . . . 128

5 Chapter 5: A multi-Omics based childhood trauma signature in bipolar disorder 135
5.1 Cohort Description . . . 136

5.1.1 Human Samples . . . 137
5.1.2 Data Analysis . . . 137
5.1.3 The Childhood Trauma Questionnaire Scores . . . 138
5.2 Preliminary Results . . . 139
5.2.1 The distribution of the CTQ scores in the BD cohort . . . 139
5.2.2 Correlations of mRNA, miRNA and DNAm with CTQ scores . . . 140
5.2.3 Step-Wise Integration . . . 144
5.2.4 sex-specific effect of Childhood trauma . . . 146

III Conclusion 151

6 Chapter 6: General Discussion and Conclusion 153
6.1 General Discussion . . . 153
6.2 Highlighted contributions: . . . 157
6.3 Perspectives and conclusion: . . . 157

References 196

Appendix 196

List of Figures

1.1 **Mood disorder spectrum:** BD type I is characterized by at least one manic episode. Type II BD is defined by at least one depressed and one hypomanic episode during a lifetime, but with no manic episodes. Episodes of hypomania or mania do not occur in MDD. 14

1.2 **Hypothalamic Pituitary-Adrenal Axis (HPA).** Adapted from Hiller-Sturmhöfel and Bartke (1998) 20

2.1 **Representation of the polygenetic characteristics of non-mendelian diseases** such as psychiatric disorders. 26

2.2 **Statistical power calculations for current and future psychiatric disorders and heights GWASs.** Figure extracted from O'Connell and Coombes (2021) 27

3.1 **Matrix factorization**: example of a data matrix with n samples and p features decomposed into a set of k components and k corresponding loading vectors. 80

3.2 **Different design matrices in RGCCA:** multiomics data integration. . . . 82

3.3 **Conceptual fusion configuration for four dataset types:** $\varepsilon_1 \ldots \varepsilon_4$, equivalently represented by the graph of relations between dataset types. Adapted from Žitnik and Zupan (2015a) 87

4.1 **Study summary. a** Descriptive statistics of the major depressive cohort. **b.** Graphical abstract. 99

4.2 **Single-omics differential analyses results and enrichements: a.** Male and female volcano plots of differentially expressed genes (threshold: $p_{adj} < 0.05$). **b.** Male and female volcano plots of differentially expressed micro-RNAs (threshold: $p_{adj} < 0.05$). **c.** Male and female volcano plots of differentially methylated probes (threshold: $p_{adj} < 0.1$). **d.** GSEA results of male and female DEG comparison with Wittenberg's quantitative review. **e.** GSEA results of male and female DEmiRs comparison with van den Berg's results. **f.** GSEA results of male and female DMPs comparison with Huls, Roberson and Walker's results. Rank-rank hypergeometric overlap results: **g.** miRNA male vs female comparaison; **h.** mRNA male vs female comparaison; **i.** interpretation of the RRHO results; **d.** DNAm male vs female comparaison; **k.** Overlap of male/female discordant gene with Seney's sex specific depression signature. **l.** Graphical representation of the KEGG enrichement pathways for male and female respective DEGs. 109

4.3 **WGCNA module analysis: a.** Distribution of WGCNA's Z-summary statistics, the module's preservation statistics, in an external whole blood and PBMCs datasets from GTEX database and Krebs's et al. controls (Krebs et al., 2020), respectively. Male and female DEG enriched module Z-summary values are represented respectively in blue and red. Modules are preserved if Z-summary > 2. **b.** Jaccard similarity matrix of the DEG enriched modules, the p-values of the module's enrichment for SNPs associated with BMI and psychiatric phenotypes are represented at the bottom and left side of the matrix. **c and d** Distribution of the p-values of the enrichment for BMI and psychiatric phenotypes GWASs in all the generated modules, the values of the DEG enriched modules are represented with vertical lines. 110

4.4 **Visualization of the modules of interest a.** Circos plot: Concentric circles representing the different levels of the modules enriched for MDD-related transcriptional dysregulation, each ring represents a different level of enrichment and each triangle represents a module. Male modules are highlighted in blue and female ones in red. A color gradient is applied for each enrichment according to its direction or p-value if the module's enrichement met statisitcal significance criteria (p-value < 0.05). The first circle from the outside to the inside represents the results of the GSEA with the DEGs, the second the overlaps of a module's genes with the biologically validated genes target of mirs, the third for the GSEAs with the DMPs, the 4th and 6th are enrichment for SNPs associated with the risk of developing MDD and childhood trauma, respectively, and finally the 5th and 7th are for the respective correlations of module's eigengenes with depression and childhood trauma score phenotypes. **b and c** Graphical representations of the M:ME48 and F:ME129 networks: The vertices in the form of circles represent the co-expressed genes in each module and the triangles are miRs from the enrichement analysis, the edges colored in grey are the weighted co-expressions calculated by WGCNA between the module's genes and those in red represent the epigenetic regulations of miR resulting from crossing with the mirBase database. The size of the vertices is proportional to their degree of connection 113

4.5 **Graphical visualization of the multi-omic integration and classification framework**: For each of the 25 splits of the 5-fold cross validation, all the combinations of the omics data and the clinical severity matrix of depression (represented by the score of the 17 questions from Hamilton) are given as inputs to 6 different jDR methods, the resulting factor matrices are then correlated with the MDD vs Controls phenotype and the top 10% of the omics factors contributing to the best correlation are extracted to reconstruct a new set of matrices on the train dataset. A similarity matrix is generated from the new train matrices using SNF, and a label propagation procedure is applied to predict the group of each sample. 115

4.6 **Comparison of prediction accuracy performances of the selection feature methods and omics combination:** In the boxplots, the mean accuracy of the 5 splits of each iteration of the cross validation is represented with a single point. The boxes are colored according to the 6 jDR and differential analysis methods. The results are divided according to the investigated omics combinations. Boxplots **a and b** respectively represent the results of the framework applied to the male and female datasets and **c** those of the pooled cohort. 117

5.1 **Graphical summary:** single and step-wise omic association with childhood trauma questionnaire. 139

5.2 **Box plots of the Childhood Trauma Questionnaire Scores A.** The distribution of the CTQ score according to its 5 items. **B.** The distribution of the cumulative scores of the abuse, neglect and total items 140

5.3 **Linear associations between CTQ scores and gene expression in the BD cohort. A.** Density distribution of the absolute values of the Pearson's correlation coefficients (ρ) between CTQ scores and adjusted mRNA expression. Top correlations ($|\text{coefficient}| > 0.2$) are depicted in red. **B.** Top results for Gene Set Enrichment Analyses of most significant correlations ($|\text{coefficient}| > 0.2$, p-value < 0.01) 141

5.4 **Heatmaps with hierarchical clustering dendrogram**: most significant Pearson's correlations (Top 2% of absolute ρ values) between omic features and CTQ scores in the BD cohort. 143

5.5 **Network-based Step-Wise multi-omic integration: A.** Tile plot of the enrichment of the 65 BD WGCNA modules with CT scores, genes enriched for GWAS's SNPs associated with psychiatric phenotype, methylated CpG probes, mRNAs and microRNAs correlated with the CT score. Significant enrichments (p-value < 0.05) are represented in colored tiles. **B.** Gene ontology biological processes and Reactome pathways enrichment results for the module of interest MEred. 145

5.6 **Network visualization of the protein-protein interactions among the genes of the CT associated MEred module:** Genes and microRNA are respectfully represented in circular and triangular shapes, genes of which only their: i) expressions are significantly correlated with CTQ are cored in red, ii) associated CpG methylation probe are correlated with the CTQ in yellow. Genes which exhibited significant CTQ correlations of both their expression and associated CpG methylation probe are colored in orange. The size of the module are proportional to their number of connection (degree) in the network. 148

5.7 **Linear associations of sex specific molecular variations with the CTQ scores in the bipolar disorder cohort: A.** Density distribution of the ρ correlation coeffetients of the mRNAs with the CT score in male and females. **B.** Descriptive table of the differences in the top n mRNA correlations with the abuse CTs in males and females (Trauma: CTQ score item, Sex: Male and Female, mean: mean of the top correlation's coefficients ,n: number of significantly correlated genes, SE: standard mean error) . **C.** Pathway enrichment analysis of the significant mRNA correlation ($|\rho| > 0.2$, p-value < 0.05 in males and females. **D.** Sex based ranked geometric overlaps between mRNA, miRNA and DNAm correlations with total CT scores 150

List of Tables

1 **Mood disorders associated genome wide transcriptome studies.** . . . 32
2 **Recent and main genome-wide DNA methylation studies on mood affective disorders.** . . . 36
3 **Recent and main genome-wide miRNA studies on mood affective disorders.** . . . 41

4 **Descriptive statistics of the Bipolar Disorder cohort.** . . . 137
5 **Descriptive statistics of the top correlation coefficients,($|\rho| > 0.2$, p-value < 0.01), with each CTQ sub-score. Mean of the coefficients, Number of genes, and Their Standard Error Mean .** . . . 140

List of Supplementary Figures And Tables

4.1 **The impact of preserving group differences in ComBat batch effect correction: A.** Density distribution of the p-values from the differential analysis with and without the preservation option. **B.** Scatter plot of the 850K t-values from the differential analyzes with and without the preservation option. 126

4.2 **Variance partition analyses results:** The violin plots **a,c** indicate the explained variances in overall gene expression in females and males, respectively. The correlograms **b,d** show the correlation between the blood cell composition measurements in females and males, respectively. The violin plots **e** indicate the explained variances in overall microRNA expression in the pooled cohort. The heatmap **e** indicates the explained variances in the overall DNAm methylation of the pooled cohort following a singular value decomposition analysis. 129

4.3 **Dot plot visualization:** GO, Kegg, and Reactome pathways enrichement analysis results for the female F:ME129 and male M:ME48 modules. The size of the dots indicates the $log2(oddratios)$ and the y-axis represent the $-log10(p-value)$. 130

4.4 **Violin Plots:** Distribution of the pairwise Jaccard similarity indices for each male and female module combination. MDD-related modules are depicted in brown. 131

4.5 **Dot plot visualization:** Crossing overlaps between the stable mRNA features of the predictive model and the DEGs found in the modules M:ME48 and F:ME129. 132

4.6 **Count tables:** Number of significantly (adjusted p-value < 0.05 and ranking $> 80\%$) stable features in the male, female and pooled cohorts. 133

4.7 **Size of the Genome Wide Association Studies used for module enrichment** 134

6.1 Mood disorders associated genome wide association studies. 198

6.2 Common dysregulated miRNAs across the miRNA studeis 199

6.3 Short form of the Childhood trauma questionnaire (CTQ). 200

6.4 Hamilton Depression Rating Scale (HDRS) Form 201

Abreviations

ADHD Attention Deficit Hyperactivity Disorder.

BD Bipolar Disorder.

ChIP Chromatin immunoprecipitation.

CNV Copy Number Variations.

CT Childhood Trauma.

CTQ Childhood Trauma Questionnaire.

CV Cross-Validation.

DEG Differentially Expressed Genes.

DEmiR Differentially Expressed miRNA.

DMP Differentially Methylated Probe.

DMR Differentially Methylated Region.

DSM Diagnostic and Statistical Manual of Mental Disorders.

ELA Early Life Adversity.

EWAS Epigenome-Wide Association Study.

GSEA Gene Set Enrichment Analysis.

GWAS Genome-wide Association Studies.

HDRS Hamilton Rating Scale for Depression.

HPA Hypothalamic Pituitary-Adrenal Axis.

ICD International Classification of Diseases.

lncRNA long non-coding RNA.

MDD Major Depressive Disorder.

MDRS Montgomery Asberg Depression Rating Scale.

MDs Mood Disorders.

miRNA microRNA.

ncRNA Non-coding RNA.

ORA Over Representation Analysis.

PD Personality Disorders.

PGC Psychiatric Genomics Consortium.

PPI Protein-Protein Interaction.

RNA-Seq RNA-sequencing.

scRNA-seq Single-cell RNA-sequencing.

SCZ Schizophrenia.

siRNA small interfering RNA.

SNF Similarity Network Fusion.

SNP Single Nucleotide Polymorphism.

SUD Substance Use Disorders.

WES Whole Exome Sequencing.

WGCNA Weighted Gene Co-expression Network Analysis.

WGS Whole Genome Sequencing.

YMRS Young Mania Rating Scale.

Résumé substantiel en français

Signatures moléculaires multi-omiques issues de prélevements sanguins de patients atteint de troubles de l'humeur

L'humeur se définit comme un "état émotionnel caractérisé par ses composantes hédoniques (plaisir-déplaisir) et d'activation-déactivation". Elle peut se présenter avec une valence (positive ou négative) et une intensité (le degré auquel elle est ressentie) variables. Les troubles de l'humeur ou TH sont des troubles psychiatriques communs et souvent débilitants, qui représentent des fardeaux sociaux et économiques importants. Ils comprennent : i) le trouble dépressif caractérisé TDC et il) les troubles bipolaires. La dépression est reconnue comme l'une des principales causes d'invalidité dans l'étude Global Burden of Disease 2019. Bien que moins fréquents, les TB sont néanmoins associés à des troubles fonctionnels et à des comorbidités considérablement plus graves. En outre, les TH s'accompagnent fréquemment de dysfonctionnements cognitifs, y compris un ralentissement des processus psychologiques, ainsi qu'un déficit d'attention, des troubles du raisonnement déductif et de la mémoire. Des troubles somatoformes sont également présents et comprennent une perte de poids importante ou une insomnie. Des épisodes dépressifs accompagnés d'hyperphagie et d'hypersomnie sont également possibles. Enfin, les idées et comportements suicidaires représentent le risque associé aux épisodes dépressifs le plus sévère et alarmant.

Les résultats d'études épidémiologiques ont montré précédemment que l'émergence de symptômes des TH se produit plus fréquemment chez les femmes que chez les hommes. Si le sexe est clairement déterminant dans les TH unipolaires, avec un sex-ratio avoisinant les 2:1, cela reste plus nuancé pour les TH bipolaires, où seules des études récentes ont rapporté une prévalence plus élevée chez les femmes.

Outre l'impact du sexe biologique sur le déclenchement des TH, diverses études ont montré que les expériences traumatiques durant l'enfance (ou Early Life Adversity, ELA) contribuent à accroître la vulnérabilité individuelle aux troubles mentaux et sont associées à leur apparition plus précoce. Malgré cette association claire, la relation causale entre les ELA et les TH est probablement bidirectionnelle, car des circonstances de vie stressantes peuvent induire des TH tandis qu'à l'inverse, les TH peuvent contribuer à déclencher des événements stressants. De plus, bien que la plupart de ces données soient rétrospectives, quelques études prospectives ont suggéré des liens entre les ELAs et le développement a posteriori de TDC. De plus, les expériences de vie sont, par essence, individuelles et subjectives, et leur perception, ou le potentiel de résilience, peuvent varier considérablement

d'un patient à l'autre, entraînant des trajectoires de vie et de maladie différentes.

L'étude des différences sexuelles dans le TDC (ainsi que dans l'impact des expériences traumatiques durant l'enfance) est un domaine très actif de la recherche moderne en psychiatrie moléculaire, dans lequel des études transcriptomiques à grande échelle commencent à fournir de nouvelles informations, indiquant des contributions biologiques de plus en plus documentées du sexe et des traumatismes de l'enfance à la dépression.

En effet, au cours des dernières décennies, l'avènement des techniques de séquençage à haut débit de nouvelle génération a conduit à des progrès rapides en psychiatrie moléculaire, avec des études génomiques, transcriptomiques et épigénomiques entraînant une expansion significative de notre compréhension des phénomènes moléculaires et bio comportementaux sous-jacents aux TH. En conséquence, il est possible de supposer que dans les années à venir, un nouveau paradigme émergera, dans lequel les signatures moléculaires et les biomarqueurs de la maladie mentale pourraient devenir des contributeurs cliniquement fiables au diagnostic et à l'évaluation du pronostic chez les personnes touchées, et ainsi complémenter les catégorisations nosologiques sur la base des observations phénotypiques, comportementales et neuropsychologiques, qui sont actuellement à l'œuvre dans la définition des troubles psychiatriques et qui peinent à être généralisées. Avec cette disponibilité toujours croissante de données fiables, la médecine de précision cherche désormais à intégrer les variations génétiques interindividuelles et les variables environnementales. Cependant, jusqu'à récemment, le domaine de la psychiatrie moléculaire ne semble pas avoir pleinement exploité le potentiel de telles approches. Après des années de recherche fondamentale en neurobiologie et en neurogénomique, la psychiatrie de précision pourrait enfin entrer dans une nouvelle ère, propulsée par les données omiques. En conséquence, parallèlement à l'introduction des techniques de séquençage de nouvelle génération, le nombre d'études de biomarqueurs en psychiatrie a rapidement augmenté. Cependant, la psychiatrie de précision s'est rapidement heurtée à des difficultés de reproductibilité, reflétant probablement, entre autres facteurs, la taille limitée des échantillons et l'hétérogénéité des phénotypes psychiatriques étudiés.

Visant à démêler ces interactions biologiques complexes sous-jacentes aux TH, l'intégration de plus d'une modalité (autrement nommée intégration multi-omique) apparaît comme un cadre de recherche approprié. Les méthodes multi-omiques se sont montrées particulièrement utiles pour la stratification des patients, en permettant une décomposition de la variabilité d'une population en sous-groupes cohérents distincts, caractérisés par des signatures moléculaires spécifiques (par exemple une signature de la réponse des patients aux traitements). De plus, les techniques omiques ne se limitent pas à la stratification des patients, mais peuvent également être utilisées pour le développement de nouveaux médicaments,

en identifiant des changements dans l'expression, ou la régulation épigénétique, de gènes ou de voies de biologiques spécifiques, qui pourraient être utilisées comme nouvelles cibles.

Toutefois, la découverte de biomarqueurs diagnostiques pour le TDC s'est avérée difficile. Cela s'explique probablement par le fait que de multiples facteurs, allant de la prédisposition génétique aux influences environnementales, contribuent à la maladie et entravent la compréhension de sa physiopathologie moléculaire.

Cependant, des études post mortem disponibles sur le cerveau et le sang périphérique ont maintenant établi un aperçu solide des voies les plus fréquemment dérégulées dans le TDC, incluant : le système immunitaire, la réponse au stress, les facteurs neurotrophiques cérébraux et les neurotransmetteurs (Mariani et al., 2021). Il est important de noter que ces études fournissent également des preuves émergentes que les dysrégulations fonctionnelles associées au TDC peuvent varier de manière significative entre les hommes et les femmes (Seney et al., 2022), incitant à examiner directement les différences entre les sexes, plutôt que d'ajuster les données dans les comparaisons cas/témoins de TDC (Parel & Peña, 2022). Un autre facteur crucial à considérer est celui des nombreuses modalités moléculaires et systèmes de régulation qui modulent l'expression des gènes. Il est peu probable que la caractérisation des différences entre les patients atteints de TDC et les témoins au niveau de données omiques individuelles puisse capturer la chaîne complète des événements moléculaires qui conduisent au TDC, en partie en raison de la variabilité interindividuelle des mécanismes de la maladie, ainsi que des limitations techniques inhérentes à chaque méthodologie.

Afin de palier à ces limitations, l'intégration de plusieurs modalités moléculaires est une stratégie qui s'est avérée efficace dans une variété d'autres pathologies, au-delà du champ de la psychiatrie (Li et al., 2018 ; Rappoport and Shamir, 2018). Cependant, en ce qui concerne les TDCs, les études disponibles ont principalement utilisé l'intégration "séquentielle", dans laquelle des différences sont identifiées au niveau de chaque modalité omique considérée individuellement, puis agrégées ou recoupées à travers les différentes couches omiques analysées. Cette approche n'exploite pas complètement le potentiel des données disponibles (voir Bhak et al., 2019 et Mokhtari et al., 2022). En revanche, une intégration multi-omique à l'échelle de l'ensemble du génome devrait générer des progrès significatifs pour l'identification de biomarqueurs diagnostiques et pour la stratification des patients.

Sur la base de ces considérations, dans les travaux présentés dans cette thèse, nous avons dans un premier temps caractérisé et intégré des données transcriptomiques avec 2 couches de régulation épigénomique (méthylation de l'ADN et microARN), générées à partir d'échantillons de sang périphérique de patients atteints de TDC (n = 80) et de témoins

sains (n = 89). Toutes les analyses ont été effectuées séparément pour chaque sexe.

Tout d'abord, chaque omique a été analysé individuellement afin d'extraire des signatures pangénomiques du TDC, qui ont ensuite été validées par comparaison avec des études externes ou des méta-analyses. Deuxièmement, nous avons défini une organisation en réseau de co-expression des gènes, permettant d'identifier des modules (sous-groupe de gènes fortement co-exprimés) associés au TDC et montrant un enrichissement en régulateurs épigénomiques différentiellement exprimés. Enfin, une intégration plus exhaustive, à l'échelle de l'ensemble du génome et multi-omique, a été développée, combinant les approches Momix (Multi-Omics Mix) (Cantini et al., 2021) et Similarity Network Fusion (SNF), afin d'extraire des variables moléculaires sanguines prédictives du TDC.

Le modèle prédictif repose sur l'algorithme de SNF qui calcule des matrices de similarité au niveau de l'échantillon, pour chaque modalité omique, avant de les fusionner. Bien que la méthode ait été conçue pour utiliser toutes les variables disponibles, seul un sous-ensemble est susceptible d'être pertinent dans le contexte du MDD. Par conséquent, ici, nous avons d'abord implémenté une étape de sélection de variables représentant le mieux la variance de chaque omique, en tenant compte de leur variance partagée avec la sévérité du TDC.

En complément d'autres méthodes d'intégration multi-omique disponibles dans la littérature (dont Momix), et pour éviter tout sur-apprentissage, les méthodes de réduction de dimension adaptée de Momix ont été encapsulées dans une procédure de validation croisée (Cross-Validation, CV). Enfin, la stabilité des caractéristiques sélectionnées a été évaluée par une procédure d'échantillonnage avec remise (bootstrap).

Cette approche a permis l'identification de sous-ensembles de variables (c'est-à-dire des ARNm individuels, des micro-RNAs ou des sites de méthylation de l'ADN) qui distinguaient le plus efficacement les cas de TDC des témoins saints. Collectivement, nos résultats soutiennent les hypothèses selon lesquelles les biomarqueurs du TDC sont en partie différents selon le sexe et que l'intégration multi-omique présente des taux de prédiction du phénotype dépressif meilleurs que les méthodologies précédemment utilisées (mono-omique, ou étape par étape). Ces résultats ont donc des implications directes pour le développement de biomarqueurs cliniques multimodaux.

Dans un second temps, je rapporte également les résultats préliminaires d'études d'intégration simple et par étapes de l'expression des gènes et des variations des régulateurs épigénétiques associées aux traumatismes de l'enfance (CT) chez les patients atteints de trouble bipolaire. Nous avons caractérisé à la fois le transcriptome (expression génique) et la régulation épigénétique (méthylation de l'ADN et microARN) à partir d'échantillons de sang total de patients BD (N = 156), en utilisant le même pipeline précédemment décrit pour

l'analyse des données TDC.

Afin d'associer ces données omiques issus de tissus périphériques aux traumatismes de l'enfance, nous avons d'abord extrait les variables omiques uniques (ARNm, microARN et Sondes de méthylation) qui étaient significativement corrélées aux scores CTQ (pour childhood trauma questionnaire), qui représente le type et l'intensité des traumas dans l'enfance rapportés rétrospectivement à l'âge adulte. De manière similaire à l'étude précédente, nous avons construit des réseaux de gènes co-exprimés et identifié ceux associés aux scores CTQ et enrichis en variables mono-omiques préalablement identifiées comme corrélées au CT. De plus, ces modules de gènes ont été testés pour l'enrichissement en polymorphismes nucléotidiques (SNP) liés aux troubles psychiatriques (issues de données de GWAS externes). Enfin, sur la base de toutes ces analyses, nous avons priorisé un module qui pourrait potentiellement servir de signature moléculaire pour les traumatismes de l'enfance chez les patients BD.

Dans l'ensemble, nos résultats supportent la notion émergente d'importantes différences liées au sexe dans les conséquences moléculaires et les mécanismes des traumas dans le BD. Cependant, ces analyses sont préliminaires et nécessiteront des analyses plus approfondies. Nous envisageons notamment d'inclure dans notre approche une procédure de bootstrap dans l'analyse des associations linéaires entre les scores phénotypiques des traumatismes de l'enfance et les variations transcriptomes et épigénétiques. Cela nous permettra d'obtenir une meilleure approximation de l'amplitude et de la variabilité des corrélations entre ces variabes, et de fournir une meilleure estimation de leur stabilité. En matière d'intégration étape par étape, une annotation plus détaillée est à explorer, et les co-expressions des modules sont à croiser avec les données MDD traitées dans la précédente analyse.

En conclusion, lors de ce travail de thèse, nous avons montré que l'intégration de données multi-omiques et l'analyse distincte des données obtenues chez les hommes et les femmes améliorent la probabilité de distinguer les groupes de patients atteints de troubles de l'humeur des témoins, à l'aide de mesures moléculaires réalisées sur des échantillons de sang périphérique. Sur la base de ces résultats, il est possible de supposer que l'application de telles méthodes pourrait aider, dans le futur, à identifier des sous-groupes de patients atteints de TH susceptibles de mieux répondre à certains traitements, dans une perspective de médecine personnalisée.

Summary

Multi-omic molecular signatures from peripheral blood samples of patients with mood disorders

Mood disorders (MDs) are common, complex and chronic mental disorders with significant social and economic impact. These disorders include several variations of two main conditions: a) major depressive disorder (MDD) and b) bipolar disorder (BD), both recognized as leading causes of disability. Results from epidemiological studies on mood disorders consistently indicate that the sex ratio is roughly 2:1 for major depressive disorder (MDD), occurring more frequently in women than in men, suggesting potentially distinct pathology across sexes. Aside from sex, several variables, including genetics and environmental factors (childhood trauma and stressful life events), contribute to the disease's variability, as well as to its onset and severity. As a result, identifying biomarkers and providing suitable treatment has proven difficult.

Over the last decade, researchers have increasingly relied on high-throughput technologies (omics) to bridge the gap between a lack of molecular understanding of mood disorders and the identification of clinically useful biomarkers. However, the interaction between these multiple omic levels has been shown to be non-linear and to entail complicated mechanisms such as transcriptomic adaptations under epigenetic control. In this context, multi-omic data integration is presented as a collection of tools that aim to identify mechanisms that contribute to these interactions, and that may provide new avenues for the understanding of their molecular etiologies.

As a result, in the present PhD work we hypothesized that combining multiple molecular modalities (multi-omic integration) could be an appropriate framework for untangling complicated biological interactions in mood disorders, and may ultimately contribute to improve patients follow-up and treatment, by: i) identifying specific biological signals for mood disorders and its associated aggravating factors, such as childhood maltreatment and ii) reducing population variability by estimating a relevant patients stratification based on subgroup specific underlying mechanisms. To do so, we implemented a multi-omic integrative framework, in order to extract peripheral blood molecular signatures of MDD, and evaluate the added value of combining several types of omic data for the classification of patients and healthy controls. We analyzed transcriptomic (RNA-Sequencing) and 2 types of epigenomic (microRNA-Sequencing and DNA methylation arrays) processes in a well-characterized cohort of individuals with MDD (n=80), and healthy controls (n=89), and

combined several strategies. First, we applied gene co-expression network and annotation enrichment analyses, and prioritized gene modules that showed significant enrichments for i) differentially expressed genes, ii) differentially methylated DNA sites, and iii) regulation by differentially expressed micro-RNAs, and also iv) presented significant associations with SNPs implicated in the risk for MDD(from genome-wide association studies). Second, we compared several statistical methods for joint dimensionality reduction, to select features (individual protein-coding genes, micro-RNAs, and DNA methylation probes) that best discriminate between patients and controls. This allowed us to identify a subset of features that maximized the prediction accuracy of these 2 groups of subjects. Importantly, in both the pooled and male only cohort, clustering based on the combination of mRNA and DNAm outperformed single omic- based classification. In females, mRNA feature slightly outperformed the mRNA / DNAm combination.

In conclusion, this work provides new evidence suggesting that sex is a critical biological factor that must be considered for the discovery of discriminating MDD biomarkers. It also supports the hypothesis that, compared to single omics, advanced integration methods fully harnessing the power of multi-omics data may significantly improve patient clustering, with implication for biomarker development.

Keywords:

Mood disorders, Multi-omic, Transcriptomics, Epigenetics, Major Depressive Disorder, Bipolar Disorder.

Résumé

Signatures moléculaires multi-omiques issues de prélevements sanguins de patients atteint de troubles de l'humeur

Les troubles de l'humeur (TH) sont des troubles psychiatriques courants et chroniques incluant : i) les troubles dépressif majeur (TDM), et ii) les troubles bipolaires (TB), tous deux reconnus comme des causes majeures d'invalidité. Les résultats d'études épidémiologiques sur les THs montrent que le ratio femmes/hommes est d'environ 2 pour 1, les femmes étant atteintes deux fois plus fréquemment que les hommes, suggérant ainsi une dérégulation et/ou des mécanismes différents selon le sexe. Outre le sexe, plusieurs facteurs, notamment le patrimoine génétique et des facteurs environnementaux (traumatismes dans l'enfance et à l'âge adulte), influencent la variabilité de la maladie, ainsi que l'apparition et la gravité des symptômes. Face à ces multiples facteurs, identifier des biomarqueurs et proposer des options thérapeutiques appropriées représente un défi considérable. Dernièrement, les recherches se sont appuyées sur les technologies de séquençage à haut débit (omiques) afin de combler l'écart entre le manque de compréhension, à l'échelle moléculaire, des troubles de l'humeur et leurs graves conséquences sociétales. Or, les interactions entre ces différents processus s'avèrent être non linéaires, et impliquent des mécanismes complexes, tels que des adaptations transcriptomiques sous le contrôle de mécanismes de régulations épigénétiques. Dans ce contexte, l'intégration de données multiomiques pourrait représenter un ensemble d'outils permettant de tester des hypothèses sur les étiologies moléculaires de ces troubles. Par conséquent, tout au long du présent travail de thèse, nous avons émis l'hypothèse que la combinaison de multiples modalités, et leur intégration multiomique, serait approprié pour démêler les interactions biologiques complexes des THs. Cette approche pourrait, sur le long terme, améliorer le suivi et le traitement des patients, en permettant d'identifier des signaux biologiques associés aux TH, et en réduisant la variabilité inter-individuelle grâce à une stratification pertinente des patients. Pour ce faire, nous avons implémenté un cadre d'analyse intégratif et multiomique, permettant d'extraire les signatures moléculaires du TDM à partir d'échantillons de sang périphérique, et pour évaluer la valeur ajoutée par la combinaison de plusieurs types d'omiques lors de la classification de patients et des témoins. Nous avons analysé le transcriptome (ARNm) et 2 types de processus épigénomiques (séquençage de microARNs et puces de méthylation de l'ADN -ADNm) dans une cohorte d'individus atteints de TDM (n=80) et de témoins sains (n=89). Puis, nous avons combiné plusieurs stratégies. Premièrement, nous avons appliqué des analyses de réseau de coexpression de gènes et d'enrichissements fonctionnels, et prior-

isé les modules de gènes qui montraient des enrichissements significatifs pour les gènes et micro- ARNs différentiellement exprimés, les sites d'ADN différentiellement méthylés, et présentant des associations significatives avec les SNP impliqués dans le risque de TDM. Deuxièmement, nous avons comparé plusieurs méthodes statistiques de réduction de dimension, afin de sélectionner les variables (gènes, micro-ARNs et sondes d'ADNm) qui discriminent le mieux les patients des témoins. Enfin, nous avons identifié des sous-ensembles de variables qui maximisent la précision de la classification des patients atteints de TDM et des témoins sains. En conclusion, nos résultats montrent que le sexe est un signal biologique majeur qui doit être pris en compte lors de l'identification de biomarqueurs du TDM. Ils étayent également l'hypothèse selon laquelle, par rapport à l'analyse d'un seul type de mécanisme moléculaire, les méthodes d'intégration multiomique avancées pourraient améliorer de manière significative la classification des patients et le développement de biomarqueurs des troubles de l'humeur.

Mots-clés:

Troubles de l'humeur, Multi-omique, Transcriptomique, Epigénétique, Trouble dépressif majeur , Trouble bipolaire.

Part I

State of the art

Chapter 1

Chapter 1 : Understanding mood disorders - a shifted paradigm

1.1 Mood disorders in DSM and ICD: phenomenology, symptoms, and causes . . . 13
1.1.1 Mood episodes . . . 13
1.1.2 Major Depressive Disorder . . . 15
1.1.3 Bipolar Disorder . . . 16
1.1.4 Mood Disorders comorbidities . . . 17
1.2 The etiology, susceptibility and risk factors of mood disorders . . . 17
1.2.1 Genetics . . . 17
1.2.2 Sex differences . . . 18
1.2.3 The impact of childhood Trauma and stressful life events 18
1.3 The pathophysiology of mood disorders . . . 19
1.3.1 The hypothalamic pituitary-adrenal axis hypothesis . . . 19
1.3.2 The immune system responses . . . 21
1.3.3 The need for a new paradigm . . . 22

In this chapter, we present an overview of the phenomenology of Mood Disorders (MDs), focusing on the symptoms and etiology of Bipolar Disorder (BD) and Major Depressive Disorder (MDD), the 2 conditions on which we conducted molecular biomarker studies presented in Chapter 3 and 4, respectively. In particular, this overview seeks to emphasize the spectrum aspect of these 2 categories, by discussing factors that characterize differences between them.

1.1 Mood disorders in DSM and ICD: phenomenology, symptoms, and causes

Mood is defined as a "prototypical affective state characterized by its hedonic tone (pleasure-displeasure) and activation-arousal (energy-tension) components". It can be characterized in terms of its valence (positive or negative) and intensity (the degree to which it is experienced). Stress has been found to be associated with both negative and positive mood states, although the relationship is more robust for negative mood states. MD are prevalent and often debilitating psychiatric disorders with significant social and economic burdens. They include several forms of two main conditions: i) MDD and ii) BD. Depression is recognized as one of the primary causes of disability in the Global Burden of Disease 2019 Study (Mental & Collaborators, 2022). While less common, BD is nonetheless associated with considerably more severe functional impairments and comorbidities.

The current categorization of MD is based on the Diagnostic and Statistical Manual of Mental Disorders (DSM), published by the American Psychiatric Association. This manual references the criteria necessary for the diagnosis of all mental disorders, including MD. The last fifth version of the DSM, referred to as DSM-5, was released in 2013 (Association, 2013). In addition to the DSM, the International Classification of Diseases (ICD) is another nosography that includes clinical definitions of mental diseases, with its latest edition being the 11th (World Health Organization(WHO), 1993). Some definitions may slightly or significantly differ between the DSM-5 and ICD10.

1.1.1 Mood episodes

A clear definition of mood episodes must be provided to fully grasp the concept of MD. The DSM-5 describes 4 types of changes in mood states or episodes: manic, hypomanic, depressed, and mixed (Note that mixed episodes were removed from the 5th edition of the DSM). Short descriptions of these episodes are provided in the following subsections:

1.1.1.1 Major Depressive Episodes - Depressed mood

According to the DSM, a depressive episode is described as a period during which a low or sad mood is present almost all day, practically every day. Ruminations, a sense of helplessness, uselessness, remorse, and incurability accompany the sad mood. The individual loses interest in the outer world and falls into anhedonia, i.e. the inability to find pleasure in his regular activities. On the other hand, World Health Organization(WHO) (1993) characterizes a MDE by at least 2 of the following symptoms: depressed mood, loss of interest, and fatigue.

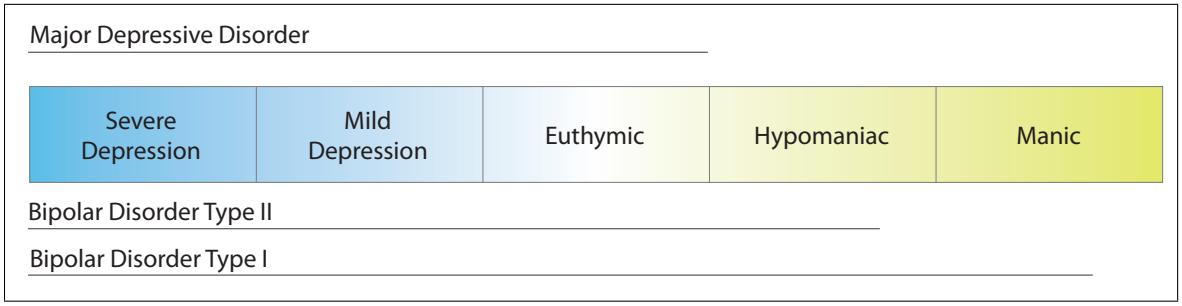


Figure 1.1: Mood disorder spectrum: BD type I is characterized by at least one manic episode. Type II BD is defined by at least one depressed and one hypomanic episode during a lifetime, but with no manic episodes. Episodes of hypomania or mania do not occur in MDD.

Furthermore, depressive episodes are frequently accompanied by cognitive dysfunction, including the slowing of psychological processes, as well as deficiencies in attention, reasoning, and memory. Somatic alterations are also present, and include significant weight loss, or sleep disturbances, with sleeplessness and insomnia. In contrast, depression episodes accompanied by hyperphagia and hypersomnia are also possible. Finally, suicidal ideation and behavior represent the most severe risk during a depressive episode.

1.1.1.2 Manic Episodes

A manic episode is defined in the DSM-5 as "a period during which the mood is abnormally elevated and persistent, lasting at least one week". The symptoms of manic episodes are perceived as mirroring that of depressive episodes, as the patient presents a hyperactive emotional activity, with an enhanced affect perception. Moreover, emotional states are labile, leading to shifts between euphoric and anxious states.

A cognitive process impairment is also observed in manic patients, mainly characterized by i) tachypsychia (racing thoughts), which is a psychopathological symptom characterized by an abnormal acceleration of the rhythm of thought, creating a state of over-excitement, and ii) logorrhea, which is a language symptom characterized by an increased and incessant verbal flow, accompanied by the use of neologisms, indicating an attention deficit.

1.1.1.3 Hypomanic Episodes

Hypomania is a milder type of mania in which the symptoms are not severe enough to need hospitalization. Physical and cerebral hyperactivity, as well as sleep difficulties, are among the most common symptoms.

1.1.1.4 Mixed Episode

The definition of mixed state by the DSM requires the presence of symptoms characterizing both depressed and manic episodes. However, this category has been removed in the latest edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM).

1.1.2 Major Depressive Disorder

Major Depressive Disorder (MDD) is a chronic psychiatric condition that manifests episodically (MDE, see section 1.1.1.1). It ranks amongst the most prevalent disorders worldwide. Indeed, despite significant geographical disparities, according to Gutiérrez-Rojas et al. (2020), its lifetime prevalence can rise to 21% in European countries. Furthermore, the lifetime probability of suicide attempts in patients with MDD is estimated to be around 27–34% (Dong et al., 2019), with a 2-5% risk of completing suicide during a major depressive episode (Cai et al., 2021). Besides, depressive disorders were recognized as one of the most significant sources of burden in the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2010. In 2019, MDD was ranked as the 13th leading cause of Disability-adjusted life-years (DALYs, an estimation of the global burden taking into account years lived with disability, and years of life lost due to premature mortality). Considering the ranking based only on the years lived with disability, MDD was the 2nd leading cause worldwide in both 1990 and 2019 (Mental & Collaborators, 2022). MDD also contributes to the burden of ischemic heart disease (Ferrari et al., 2013). These findings highlight the need to establish depressive disorders as a public health priority, and to adopt efficient treatment strategies to lower their prevalence. Moreover, Santomauro et al. (2021) showed that the Covid-19 pandemic had raised the pressure for most countries to urgently upgrade their mental health systems.

A classification of MDD severity based on its impact on occupational and social functioning is proposed in the DSM-5 (Association, 2013), including mild, moderate, and severe episodes. Additionally, severe episodes may also present with psychosis, delusions, and hallucinations. Moreover, a notion of quantification of the number of symptoms experienced by the patient is added to the severity classification in the World Health Organization(WHO) (1993)'s ICD. Indeed, a mild state has two or three symptoms that cause minor impairment in everyday activities. A moderate state has four or more symptoms and affects the patient without hindering him from proceeding with daily life tasks. Finally, a severe condition is characterized by several signs that are bothersome to everyday activities, with the frequent presence of suicidal ideation and psychosis.

1.1.2.1 Depression Scale

Several scales have been developed to harmonize diagnosis across countries and clinicians. Among these, the Hamilton Rating Scale for Depression (HDRS) (Annex 6.4) is frequently used to assess the frequency and severity of depressive symptoms, for instance, before, during and after treatment. It takes 15 to 20 minutes to complete and score. Ratings are given on a 5- or 3-point scale, with a maximum score of 61 HAMILTON (1967). Scores between 7-17 represent mild depression, 18-24 moderate depression, and scores above 25 represent severe depression. While its latest version comprises 21 items, the original 17 items described in the first version are more reliable. The 4 additional items are intended for subtyping depression but are sometimes, wrongly, used to rate severity. The HDRS considers symptoms that occurred during the preceding week. Some items are defined as a sequence of intensity increments, while others are defined by a set of requirements of equal value. Among the 17 items, the following are characterized: depressed mood; guilt; suicide; early night, midnight, and morning insomnia; work and activities; lateness; agitation; psychic anxiety; somatic anxiety; gastrointestinal symptoms; general somatic symptoms; genital symptoms; hypochondria; weight loss; and insight.

1.1.3 Bipolar Disorder

Bipolar Disorder (BD) is a mood disorder defined by manic and hypomanic episodes (see sections 1.1.1.2 and 1.1.1.3 above). In the DSM, BD is divided into type I, type II, unspecified, and cyclothymic disorder, as briefly detailed below.

1.1.3.1 BD type I

Manic episodes that continue for at least seven days, or manic symptoms that are so severe that the person requires emergency medical treatment, are the critical characteristics of BD type I. Depressive episodes generally last at least two weeks and are also common. Mania or depression symptoms must differ significantly from the patient's euthymic state.

1.1.3.2 BD type II

BD type II is characterized by a series of depressive episodes that alternate with hypomanic episodes, but with no manic or mixed periods.

1.1.3.3 Unspecified BD

Unspecified BD is diagnosed whenever a patient's symptoms do not match the clinical definition of type I or II. The symptoms may not continue long enough, or the patient may have too few symptoms, for type I or II to be diagnosed. The symptoms, however, are different from the patient's typical behavior.

1.1.3.4 Cyclothymia

Cyclothymia is characterized by episodes of hypomania that alternate with mild depression for at least two years.

1.1.3.5 Prevalence of BD

For BD I, the lifetime prevalence in the general population is roughly 1% (Pini et al., 2005). However, when taking into account the full spectrum of BD subtypes, their overall prevalence rises to 3% (Akiskal et al., 2000). Furthermore, the risk of suicide among BD patients is around 10 to 13 times greater than in the general population. Available data indicate that up to 20% of BD patients commit suicide, and 20-60% attempt suicide at least once in their lifetime (Dome et al., 2019). Similarly to MDD, BD also represents a heavy societal and economic burden (Alonso et al., 2011; He et al., 2020).

1.1.4 Mood Disorders comorbidities

Depression is common among people with preexisting medical conditions, and such comorbidity is related to poorer treatment outcomes for both conditions (Hirschfeld, 2001). Furthermore, comorbidities contribute to the misdiagnosis of MD, interfering with treatment response and worsening its evolution (McIntyre et al., 2019). Not surprisingly, anxiety disorders and Substance Use Disorders (SUD) are the most frequent comorbidities amongst those suffering from MD (Mineka et al., 1998). In addition, other psychiatric comorbidities may also be associated with MD, including: Personality Disorders (PD) (Friborg et al., 2014), Attention Deficit Hyperactivity Disorder (ADHD) (Schaffer et al., 2012), and eating disorders (Álvarez Ruiz & Gutiérrez-Rojas, 2015).

1.2 The etiology, susceptibility and risk factors of mood disorders

1.2.1 Genetics

There is extensive evidence that genetic variation plays a significant role in the risk for MD, with a substantial rate of heritability. Heritability is an estimation of the proportion of risk attributable to genetic variation. For instance, twin and family studies estimate the heritability of BD to be as high as 70% (Berrettini, 2004). MD genetic factors do not appear specific to BD and MDD, but rather overlap with genetic variants associated with other psychiatric disorders (PGC-Consortium, 2013). Further details on the genetics of MD are provided later in chapter 2, section 2.1.1.

1.2.2 Sex differences

Results from epidemiological studies on MD demonstrated that they occur more frequently in women than in men. If sex is clearly important in unipolar MD, with a sex ratio roughly equal to 2:1 (Diflorio & Jones, 2010; Kennedy et al., 2005; Seedat et al., 2009; Seney & Sibille, 2014), only recent studies have reported a higher prevalence of BD in female patients (Dell'Osso et al., 2021; Gordovez & McMahon, 2020). These differences in MD prevalence as a function of sex are also modulated by age. While there is no clear female predominance of depression in children (Hankin & Abramson, 1999), this sex difference appears during adolescence, and then decreases with age (Stordal et al., 2001). Sex differences in help-seeking behavior and symptom reporting have also been reported (Mackenzie et al., 2011; Salk et al., 2017), suggesting that such biases could be part of the explanation for the sex ratios observed. However, in other epidemiological studies, sex differences have been described not only for the presented and reported MDD symptoms, but also for some of its co-morbid disorders, risk factors, and antidepressant treatment response. Cross-national meta-analyses indicate that sex differences for MDD are found regardless of culture, community, or nation (Salk et al., 2017). In addition, there are well-documented biological contributions of sex to depression (considering its genetic, hormonal, and gender aspects). Deconstructing the explanations of sex differences in MDD (as well as in the impact of childhood adversity) is a highly active area in modern molecular psychiatry research, in which large-scale transcriptomics studies are starting to provide new insights (Labonté et al., 2017a; Parel & Peña, 2022; Seney et al., 2022). Our own work (see Chapter 4) was conducted in part with the goal of differentiating sex-specific molecular peripheral biomarkers of MD.

1.2.3 The impact of childhood Trauma and stressful life events

Various studies have shown that Early Life Adversity (ELA) contributes to increase individual vulnerability to mental disorders, and is associated with earlier onset of these mental disorders. Accordingly, multiple studies have documented mainly links between MD and parental apathy, neglect, and physical and sexual abuse (Aas et al., 2016; Berthelot et al., 2022; Green et al., 2010; Quidé et al., 2020). ELA also determines the clinical course of MD. Among other findings, Etain et al. (2017) showed that a history of ELA associates with a more severe course of BD, with a higher number of manic episodes. Furthermore, ELA has been linked to earlier onset of MD, more severe symptoms, increased suicidality, higher rates of psychosis leading to hospitalization, and higher rates of psychiatric comorbidity, particularly substance use and post-traumatic stress disorders (Danese, 2020; Hett et al., 2022; Lucero et al., 2022; McKay et al., 2021; Moraes et al., 2018).

Despite this clear association, the causal relationship between ELA and MD is likely

bidirectional, as stressful life circumstances may induce MDs while, conversely, MDs can contribute to trigger stressful life events. Moreover, although most of these data are retrospective, a few prospective studies have found links between ELA and the development of MDD later in life. In addition, life experiences are, by essence, individual and subjective, and their perception, or potential for resilience, may vary significantly from one patient to another, resulting in different life and disease trajectories.

1.2.3.1 Evaluation of childhood trauma

The Childhood Trauma Questionnaire (CTQ), presented in Annex 6.3, is a retrospective evaluation of ELA. It evaluates two main subtypes of childhood trauma: abuse and neglect. The CTQ is subdivided into 5 subscales: emotional abuse, physical abuse, sexual abuse, emotional neglect, and physical neglect. The CTQ scale was initially proposed by Bernstein et al. (2003), and has since been validated across various cohorts (Karos et al., 2014; Thombs et al., 2009). A detailed description of this scale is provided in this manuscript (see section 5.1.3).

1.3 The pathophysiology of mood disorders

The following paragraphs will briefly summarize core biological pathways that are tought to underlie MD and their therapies. MDs have long been connected with various alterations in biochemical pathways and neurocircuitry. Here, we will mainly focus on most frequently investigated pathophysiological mechanisms: imbalance in neurotransmitters and neurotrophic factors, abnormal immune system responses, neuroinflammation, and dysfunction of the Hypothalamic Pituitary-Adrenal Axis (HPA) stress axis.

1.3.1 The hypothalamic pituitary-adrenal axis hypothesis

The HPA axis has a predominant role in regulating several biological processes, including the immune system and metabolism, and plays a crucial role in neurobiological processes, including learning and memory. As shown in the figure below 1.2, different hormones regulate the activity of the HPA axis: corticotropin-releasing hormone (CRH) and vasopressin (AVP) from the hypothalamus, adrenocorticotrophic hormone (ACTH) from the pituitary, as well as glucocorticoids (cortisol in humans) from the adrenal cortex.

Rapid recruitment of the HPA axis during acute stress is adaptive, and allows the mobilization of physiological resources to face ongoing challenge. In contrast, high exposure to stress for an extended period of time might result in chronic alteration of the HPA axis response (Chrousos, 2009). The hypothalamus initiates HPA axis activation and may be

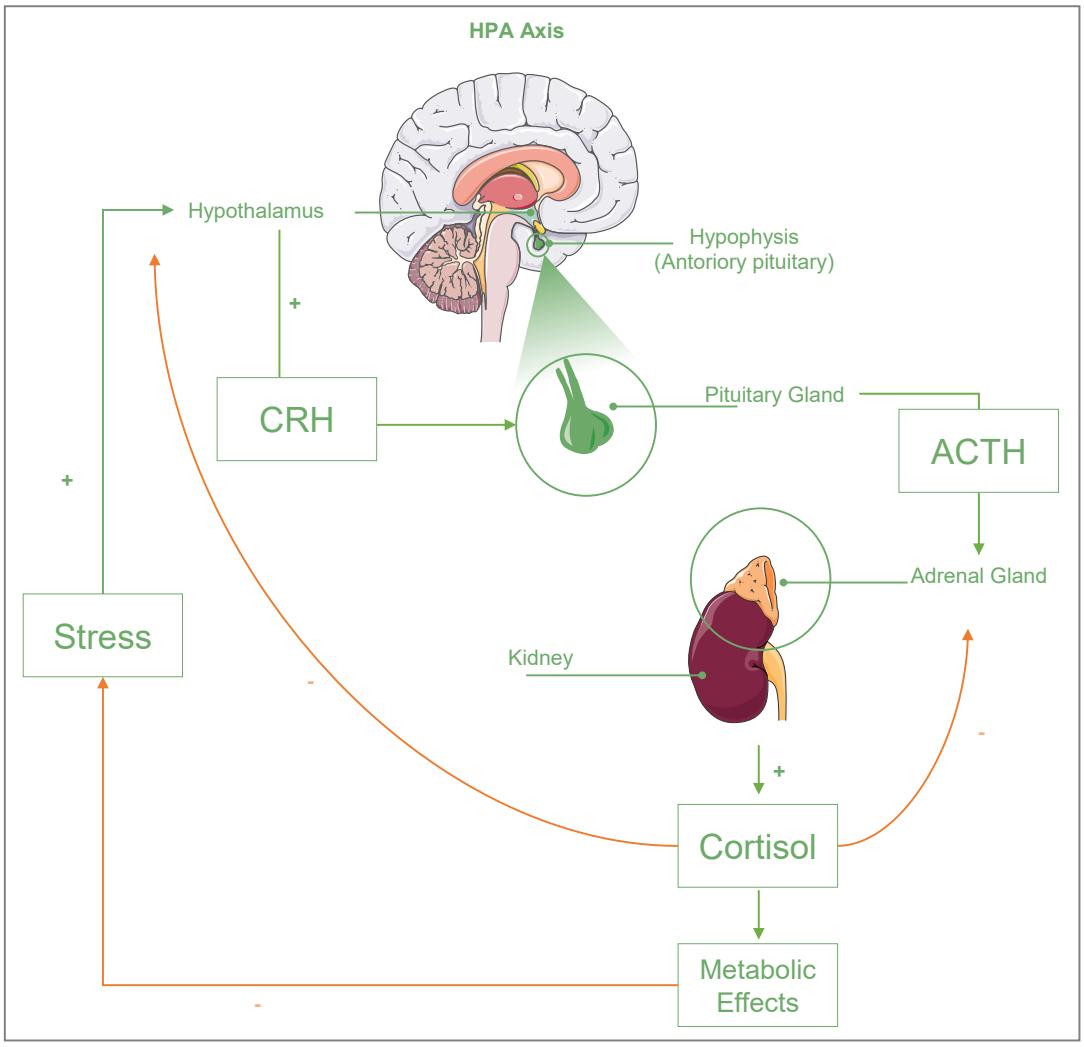


Figure 1.2: Hypothalamic Pituitary-Adrenal Axis (HPA). Adapted from Hiller-Sturmhöfel and Bartke (1998)

stimulated by external stressors, or internal physiological processes. When activated, parvocellular neurons located in the paraventricular nucleus of the hypothalamus release CRH (Herman et al., 2016), which in turn triggers the release of ACTH by the pituitary gland into the bloodstream. This initiates the release of the stress hormone cortisol from adrenal glands.

Under physiological conditions, cortisol exerts a series of negative feedbacks onto the HPA axis, through corticoid receptors, in order to maintain homeostatic control over its activity (Chrousos, 2009). However, when stress is experienced too frequently, or for too long, excessive corticoid production is thought to trigger alterations in the structural and functional integrity of the HPA axis, as well as systemic adaptations, thereby contributing to the development of MD.

This phenomenon is further supported by the high occurrence of depressive episodes in individuals with Cushing's syndrome, which is characterized by persistent hypercortisolemia (Parker et al., 2003; Sonino & Fava, 2001). Further studies showed that increased cortisol levels are associated with an increased risk of depression in both the short- and long-term (Campbell & Ehlert, 2012; McEwen, 2008). Moreover, high levels of glucocorticoids associated with glucocorticoid resistance, defined as an inefficient glucocorticoid-mediated negative feedback on the HPA axis, are observed in many depressed patients (Pace et al., 2007). This dysfunction in glucocorticoid negative feedback loop has also been linked with a higher risk of suicide (O'Connor et al., 2016; Pandey, 2013).

1.3.2 The immune system responses

Several studies indicate that MDs are associated with increased inflammatory markers in the blood, which may be caused by stress, infection, or autoimmune responses. However, it is still unclear if depression could be categorized as a neuroinflammatory disorder and, if so, whether the inflammation is a cause or a consequence of the disease (Mechawar & Savitz, 2016).

Initial activation of the innate immune system is called the acute-phase response. This response is mediated by immune cells' release of cytokines, such as interleukin-1 (*IL-1*) and interleukin-6 (*IL-6*), by immune cells. It has been well documented that inflammation and associated chronic immune responses trigger the development of MD (Dantzer et al., 2008; Slavich & Irwin, 2014). Distinct types of inflammation, acute or chronic, elicit different responses from the body in case of physical injury, wound, or infection. Acute inflammation is characterized by the rise of specific cytokines, chemokines, and other substances that draw cells of the immune system to an injury site. This results in tissue repair and healing. In contrast, chronic inflammation is defined by the presence of activated cells of the immune system and the production of cytokines and other substances for prolonged periods. Chronic inflammation can lead to the development of diseases such as arthritis, autoimmune diseases, and cancer. The inflammatory response system is mitigated by the presence of pro-inflammatory cytokines such as IL-1β, IL-6, and TNF-α. The corresponding genes have been associated with MD in multiple studies. Furthermore, it has been shown that patients who suffer from significant depression show elevated levels of pro-inflammatory cytokines in the plasma and cerebrospinal fluid. These cytokines have been linked to more severe depressive symptoms. Cytokines may also trigger microglia to secrete neurotrophic factors such as brain-derived neurotrophic factor (BDNF), which is necessary for the survival and proliferation of neurons and can influence synaptic plasticity.

1.3.3 The need for a new paradigm

As shown in this introductory chapter, multiple nosological categorization (e.g., ICD; DSM) are used to defined psychiatric disorders, based on clusters of phenotypical behavioral and/or neuropsychological observations. Meanwhile, the development of next-generation high-throughput technologies for genomics, transcriptomics, and epigenomics investigations during the past decade have led to rapid progress in molecular psychiatry, resulting in a significant expansion of our understanding of the molecular and biobehavioral phenomena underlying MD. As a result, it is possible to speculate that in the coming years, a new paradigm will emerge, whereby molecular signatures and biomarkers in mental disease may become clinically reliable contributors to diagnosis and evaluation of prognosis in affected individuals.

Chapter 2

Chapter 2: Translational research for psychiatric disorders : From single to multi-omics approaches

2.1 Omics-based translational psychiatry research 24
2.1.1 Genomics 24
2.1.2 GWAS 25
2.1.3 CNVs 27
2.1.4 Transcriptomics 27
2.1.5 Transcriptomics of psychiatric disorders 29
2.1.6 Epigenetics 33
2.1.7 Covalent modification of DNA 34
2.1.8 Histone's post-translational modification 38
2.1.9 Non coding RNAs 39
2.2 Multi-omics : The whole is greater than the sum of its parts . . . 42
2.2.1 Untangle mood disorders through multi-Omics approaches 42
2.2.2 Review Paper 43

The drug development paradigm is currently prospective, implying that many pharmacological targets are screened with high throughput technologies. Drug development aims to balance safety with effectiveness, in order to maximize molecular efficacy in the broadest possible range of clinical populations. As a result, patients' biological diversity is overlooked for efficiency and, mostly, economic considerations. For instance, subgroups of patients presenting distinct symptoms and etiologies, as in MD, are often prescribed non-personalized treatments. However, the pharmaceutical industry is quickly changing and embracing new technologies to drive personalized clinical decisions.

Over the last decade, technological advancement has enabled omic investigations to be deployed at a large scale while increasing their genomic or cellular resolution, and cutting costs. However, despite the availability of large volumes of data, the intricacies of the biological systems under study appear subtle, and cannot be fully described using single-omic analyses. Hence, multi-omics have recently gained attraction as a primary focus for personalized treatment. Indeed, several omic-based personalized medicine studies, mainly in oncology, have emerged recently, and seem to be reaching a tipping point for diagnosis and prognosis applications (Chen & Snyder, 2012).

With this ever-growing availability of reliable data, precision medicine now seeks to incorporate inter-individual genetic variation and environmental variables. However, until recently, the field of molecular psychiatry does not appear to have fully exploited the potential of such approaches. Following years of fundamental research in neurobiology and neurogenomics, precision psychiatry may finally be entering a new era, propelled by omics data. Accordingly, concomitantly with the introduction of next-generation sequencing technologies for genomic, epigenetic, and transcriptomic research, the number of biomarker studies in psychiatry quickly rose (Fernandes et al., 2017; Sullivan & Geschwind, 2019). However, precision psychiatry quickly stumbled on reproducibility issues, likely reflecting, among other factors, limited sample size and disease heterogeneity.

Aiming at untangling the complex biological interaction at play in MD, integrating more than one modality (i.e. multi-omic integration) appears as a suitable framework. Multi-omic methods aim at patient stratification, by breaking down population variability, with their specific molecular signatures (e.g. for drug response). Moreover, omic technologies are not limited to patient stratification and can also be used for drug development, by identifying changes in the expression, or epigenetic regulation, of specific genes or pathways, which could be used as new targets.

In this context, the present chapter provides a description of omic-based medicine, with a particular emphasis on psychiatry-aimed application, first by describing single-omic biomarkers and molecular signature discovery. Then, a systematic review of multi-omic precision psychiatry in MD is provided in Chapter 3 (Mokhtari et al., 2022).

2.1 Omics-based translational psychiatry research

2.1.1 Genomics

Historically, with the discovery of DNA in 1953, a research field has emerged to dissect the molecular mechanisms that may regulate human physiology. The following decades

witnessed profound advances in the knowledge of the building elements of human life. However, it took more than half a century to integrate DNA technology and genomics into translational medicine, and it is only recently that it has started to be integrated into translational psychiatry.

For several years, the search for genomic biomarkers of psychiatric diseases was hindered by the lack of genome-wide approaches. Studies were focused on candidate genes, which were selected based on limited insights into the neurobiology of psychiatric disorder. Typically, single candidate genes and their genetic variants, such as for example the Serotonin Transporter Promoter Variant *5-HTTLPR* (Karg et al., 2011), were examined for large effect-size in small cohorts, with limited success due to frequent lack of reproducibility. The first psychiatric genetic studies were based on family, twin, adoption studies, and linkage analyses. Lately, following subsequent technological progress with Genome-wide Association Studies (GWAS) and genomics sequencing (Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS)), investigators were able to identify common and rare variants associated with psychiatric phenotypes. (Fabbri, 2022; Gandal et al., 2016).

2.1.2 GWAS

Since the first Genome-wide Association Studies (GWAS) published in 2005 (Klein et al., 2005) on age-related macular degeneration, high-resolution genome-wide screens became ever more popular (Loos, 2020; Uffelmann et al., 2021). Their underlying assumption is that the phenotypes under study, including psychiatric disorders, are polygenic and due to multiple common genomic variants. Hence, genomic markers represented by single nucleotide variation, also called Single Nucleotide Polymorphism (SNP), are associated with such phenotypes. Most SNPs individually have a marginal or no effect on biological systems. However, a reduced number of them could have functional consequences in particular contexts, tissues, or cell types, for instance altering transcription factor binding affinity and mRNA stability, or could have a direct functional impact as observed with the *BDNF* val66met polymorphism, which was associated with hippocampal volume and memory in Human (Egan et al., 2003).

This shift to genome-wide investigations allowed an assumption-free exploration of causal variants, and reinforced the notion that psychiatric and mental health-related phenotypes involve a continuum of a large number of small effects (Zhang et al., 2018), as framed in figure 2.1. Furthermore, this shift was accompanied by a need for large cohorts to yield detection and reliable estimates of these small effects. Such large data cohort can only be collected through organized international collaborations. Hence, the Psychiatric Genomics Consortium (PGC) was developed to collect and analyze case-control GWAS related to psy-

chiatric disorders. The PGC initiative has proved very effective, with the publication in 2011 of a first GWAS study on schizophrenia (SCZ) (Ripke et al., 2011), followed shortly by a larger-scale study (36,989 cases and 113,075 controls) that identified 108 genes with a SNP showing statistically significant association with this disorder (Ripke et al., 2014). This study, among others, confirmed the neurotransmitter alteration hypothesis in SCZ, with significant results for SNP associated with the D2 dopamine receptor *DRD2* gene, which is known as the pharmacological target of some antipsychotics. Regarding MD, the main results from GWASs are summarized in the "Genomics" section of our systematic review (see section 2.2.2) and in table 6.1.

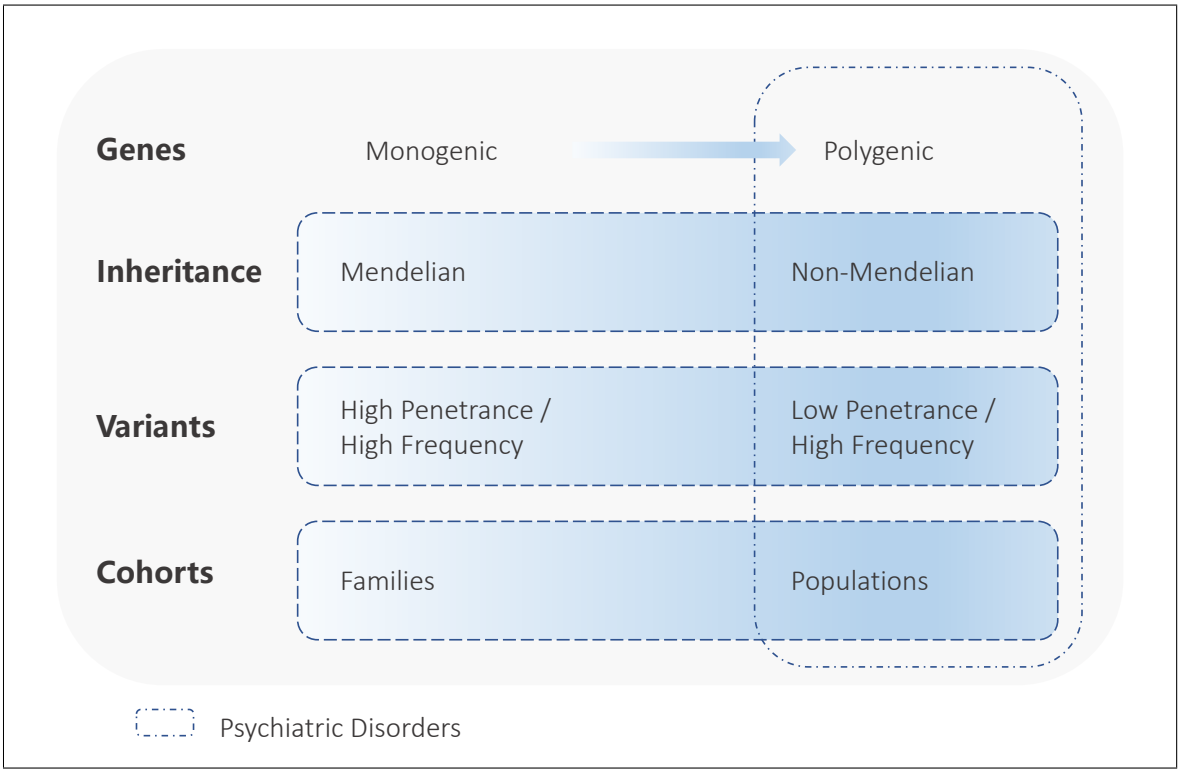


Figure 2.1: Representation of the polygenetic characteristics of non-mendelian diseases such as psychiatric disorders.

However, despite a growing effort into identifying common variants associated with MD and other psychiatric traits, GWASs yielded mitigated success for BD and evermore for MDD. Indeed, as shown in Figure 2.2, reproduced from (O'Connell & Coombes, 2021), the estimated GWAS sample sizes required to capture 50% of the genetic variance of MDD is estimated to be 6 time higher than the sample size required for BD and 10 time higher than for SCZ. These results suggest that beyond sample size limitation, further phenotypical and pathophysiological heterogeneity may play a substantial role in the genetic variance of MDD Ormel et al., 2019.

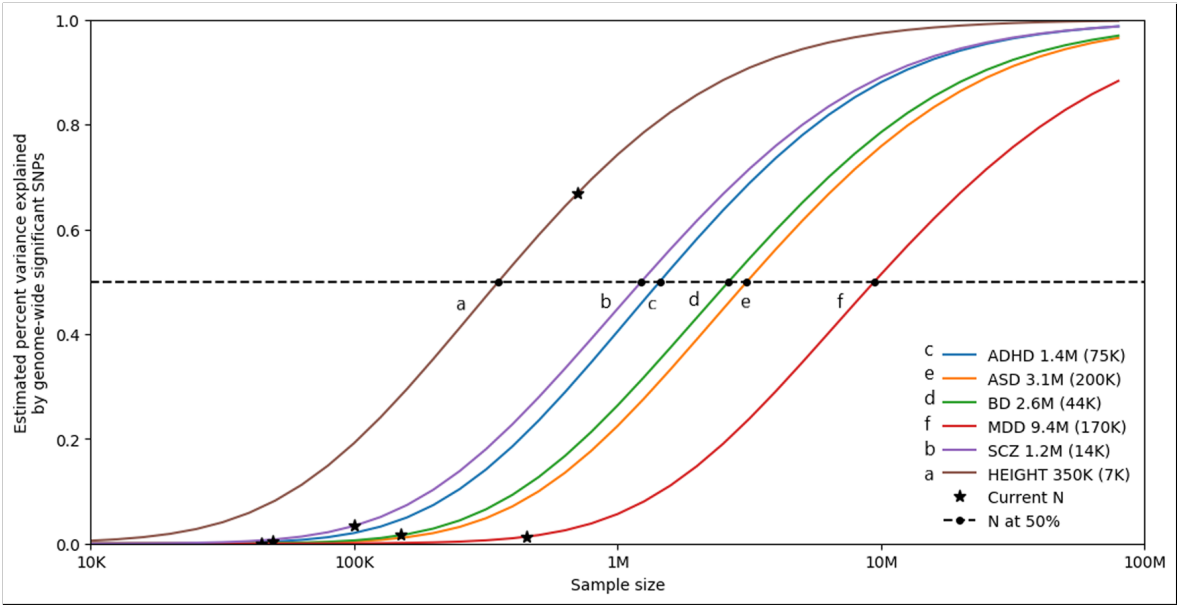


Figure 2.2: Statistical power calculations for current and future psychiatric disorders and heights GWASs. Figure extracted from O'Connell and Coombes (2021)

2.1.3 CNVs

Beyond SNP, understanding genomic variation associated with psychiatric disorders has been enriched with the analysis of Copy Number Variations (CNV). CNVs are small chromosomal segments of at least 1000 bases (1 kb) that are deleted or repeated across individuals (Kehrer-Sawatzki, 2007). They are known to play a predominant role in complex multifactorial diseases such as cancer (Pös et al., 2021; Yeh et al., 2010). Due to reasons yet to be understood, CNVs are highly associated with neurodevelopmental and cognitive diseases (Kendall et al., 2017; Kirov, 2015; Männik et al., 2015; Thygesen et al., 2021), which include the well-established association between the 22q11.2 deletion and SCZ. Regarding MD, while little association with CNVs has been reported (Gordovez & McMahon, 2020; O'Connell & Coombes, 2021), a noteworthy example corresponds to BD and the 16p11.2 duplication, despite its rarity ($< 1\%$ of the BD population, see Green et al., 2016).

2.1.4 Transcriptomics

Although genomic-based studies have revealed variants associated with psychiatric disorders, the mechanistic understanding necessary for the development of treatment may be more a matter of transcriptomics and functional genomic analyses. Indeed, measuring mRNA levels can provide a snapshot of molecular biological activity, which is a good proxy for functional processes at protein level. Furthermore, determining a gene expression profile during neurodevelopment, or the progression phases of psychiatric syndromes, may

allow a more precise inference of the mechanisms involved in their onset and evolution. However, this is challenging for research on psychiatric disease, since access to brain tissue, which is at the origin of the pathology, is only possible postmortem, while longitudinal investigations can only be conducted on peripheral tissues, which do not enable direct inference of causality.

Technologies that simultaneously sequence the expression of thousands of different RNA molecules have advanced rapidly, from microarrays that detect pre-specified RNA molecules, to RNA-sequencing that analyzes the entire population of RNA molecules, and now single-cell RNA-sequencing, which performs such analyses at the level of individual cells. The following section presents most popular technologies.

2.1.4.1 Transcriptomic Methods

Microarrays: These methods appeared in the 1990s. The basic concept of microarrays is to label the cDNA strands with a fluorescent dye. Simultaneously, a DNA chip containing DNA segments (probes) of the desired genes to be analyzed is attached to a plate's surface (BeadArray). Then, the cDNA strands hybridize with DNA probes, based on nucleotide complementarity. Finally, the level of hybridization is detected by fluorescence and compared across different experimental conditions. Several microarray platforms have been developed, with Affymetrix and Illumina the most commonly used.

Next generation Sequencing (Bulk RNA seq): Next-generation RNA-sequencing (RNA-Seq) provides an unbiased measurement of transcriptomic activity, improving over microarrays. Indeed, while short reads from microarrays suffer from repeated sequences, longer RNA-seq reads (generally up to 100-150bp) allow better mapping, broader detection range (covering lowly expressed genes) and higher reproducibility (Zhao et al., 2014).

Next generation Sequencing (Single cell RNA seq): In conventional RNA sequencing, transcript abundance is measured in bulk tissue, with unknown contributions from the various cell types in the tissue of interest. Recently, Single-cell RNA-sequencing (scRNA-seq) has allowed the unmasking of this cellular heterogeneity, by measuring gene expression in individual cells. Tang et al. (2009) presented the first method to sequence the whole transcriptome (mRNA) of a single cell and detected the expression of 5270 genes in a single mouse blastomere. Since then, multiple improved technologies have become available. However, their use in psychiatric disorders remains at its debut (Antunes & Martins-de Souza, 2022; Nagy et al., 2020).

2.1.5 Transcriptomics of psychiatric disorders

Ciobanu et al. (2016) and Hess et al. (2019) previously reviewed transcriptomic studies of MDs. Below, we briefly detail more recent advancements over the last couple of years 1, with 22 studies that were conducted on human biological tissues using genome-wide methods. This bibliographical snapshot of the PubMed database was captured on May 27, 2021, and included only recent human-based genome-wide transcriptomic studies of MD. Main findings, cohort size, biological tissue, and technologies used by these studies are summarized in Table 1. They typically identify Differentially Expressed Genes (DEG) between patients and controls, followed by functional enrichment annotation (carried out mainly using Gene Ontology, KEGG and Reactome databases).

In the past years, some of these transcriptomic studies have incorporated co-expression gene network analysis, mainly via Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder & Horvath, 2008), thus taking into account the modularity of the organization of gene expression to identify subgroups of genes ("modules") associated with BD or MDD. A detailed explanation of WGCNA is provided in section 3.2.1.2. Out of the 22 studies (4 meta-analyses aside): 7 performed differential expression only, 1 used WGCNA only, while 6 combined both approaches; 1 combined co-expression networks with aberrant gene expression analysis (which relies on Mahalanobis distance to assess case/control expression discrepancies); the 3 remaining studies used either a likelihood ratio test (LRT, which compares a gene-by-gene fitted full model including all desired factors to its reduced version), a gene-based association test, or machine learning.

Regarding the type of biological sample investigated, 9 studies used brain tissue exclusively, while 10 used blood samples solely, with additional important diversity within each tissue. Indeed, researchers used RNA from either whole blood samples collected using PAXgene tubes (Ciobanu et al., 2018; Ciobanu et al., 2020; Hori et al., 2018; Jansen et al., 2016; Krebs et al., 2020; Leday et al., 2018; Mostafavi et al., 2014), or from peripheral blood mononuclear cells (PBMC) (Le et al., 2018; Mellon et al., 2016). For brain tissue, various structures with distinct cellular composition and physiological function were investigated: the dorsal prefrontal cortex (Akula et al., 2014; Arion et al., 2017), cingulate cortex (Guan et al., 2019), hippocampus (Kim et al., 2016; Mahajan et al., 2018), and dorsal striatum (Pacifico & Davis, 2017). Besides, Labonté et al. (2017b) investigated 6 brain regions (ventromedial prefrontal cortex; Brodmann area 25, orbitofrontal cortex, dorsolateral prefrontal cortex, anterior insula, nucleus accumbens, and ventral subiculum), while (Seney et al., 2018) covered 3 brain regions in their meta-analysis (dorsolateral prefrontal cortex, subgenual anterior cingulate cortex, and basolateral amygdala). Of note, Gamazon et al. (2019) and Ciobanu et al. (2016) have conducted cross-tissue studies covering 10 brain

structures (dorsal prefrontal cortex, hippocampus, or cerebellum, for example) along with peripheral tissues (e.g. adrenal gland, colon transverse, and whole blood), while Hess et al. (2019) and Wittenberg et al. (2020) conducted a meta-analysis based on datasets obtained from leukocytes, monocytes or whole blood. In the following paragraph, we will first describe studies on brain and blood separately, before concluding on possible convergence between the 2.

Some blood transcriptomic analyses failed to identify genome-wide significant DEGs, even in the large cohort of MDD patients (n=463) and controls (n=459) investigated by Mostafavi et al., 2014. A commonly used method to circumvent this issue is to analyze pathway enrichment and/or biological function of genes showing nominal significance (Gene ontology biological processes, KEGG Reactome pathways; Gene set enrichment analysis, GSEA). This can be done using an arbitrary significance threshold, to restrict the DEG to be considered (classical over-representation analysis, ORA), or a more holistic view, to test for functional enrichment in a threshold-free manner (e.g. using GSEA). Several biological functions and pathways were identified using these approaches: ribosome-related pathways were found significantly enriched in a study of stress vulnerability and MDD in a general population cohort, and confirmed in independent MDD case-control datasets, with significant upregulations of *RPL17* and *RPL34* (Hori et al., 2018). More importantly, pathways and/or biological functions related to immunity were reported in 7 studies out of the 10 that used peripheral tissues (Ciobanu et al., 2020; Hess et al., 2019; Jansen et al., 2016; Le et al., 2018; Leday et al., 2018; Mellon et al., 2016). Alterations of the immune system's biological function and oxidative stress and apoptosis pathways were also reported as significantly enriched in both DEG and WGCNA co-expression modules in a meta-analysis of 7 studies on both BD and schizophrenia (Hess et al., 2019). These results are of interest to us, as our own work on MDD also pointed towards immune related GO terms (see Chapter 4). More recently, a novel machine learning method using an algorithm combining WGCNA and random-forest was applied to a large cohort (n=521), and identified transferrin receptor, a protein involved in iron transport and immune system function, as a predictor of patient status (with an accuracy of 63%) (Ciobanu et al., 2020). Of note, in a transcriptomic study of whole blood of BD patients (n=240) and controls (n=204), taking lithium treatment into account, a significant enrichment for the immune response system was found among lithium-associated genes, suggesting a treatment bias in BD DEG analyses results (Krebs et al., 2020). This underscores the importance of considering drug effects as covariates in such studies, which, surprisingly, was the case in only 2 of the 22 studies that we identified.

Regarding brain tissue, despite the diversity of brain structures investigated across studies, findings tended to converge. Genome-wide significant DEGs were found in 5 (Gama-

zon et al., 2019; Guan et al., 2019; Kim et al., 2016; Pacifico & Davis, 2017; Ramaker et al., 2017) of the 7 studies: Gamazon et al. (2019) and Guan et al. (2019) and Pacifico and Davis (2017) identified DEG associated either specifically with BD, or more generally with BD, schizophrenia, and autism; Kim et al. (2016) highlighted DEG correlated to both MDD and BD; Mahajan et al. (2018) studied the hippocampal dentate gyrus in MDD, and conducted multiple comparisons: single vs. multiple episodes, suicidal vs non-suicidal groups and male/female patients compared to controls. They reported a significant overlap of DEG between results from those comparisons, and common immune system-related pathways enrichment. Among the biological pathways identified in these brain studies, the immune system and inflammation-related gene ontology terms stood out, with consistent findings reported across several brain regions: the hippocampus Kim et al. (2016), cingulate cortex (Guan et al., 2019), caudate nucleus and putamen (Pacifico & Davis, 2017) from BD patients. Finally, regarding sex specificity, Labonté et al. (2017b) studied gene expression and co-expression modules in 6 different brain regions in MDD, and found limited DEG overlap across genders, suggesting a strong gender-specific effect at play in MDD. Again, our own results (Chapter 4) concur with the later notion.

Table 1: Mood disorders associated genome wide transcriptome studies.

Study	Cohort	Biological sample	Technologies	Main analysis method	Covariate correction	Relevent results
Akula et al. 2014	11 BD, 8 HC	Brain (BA46, DLPFC)	Illumina GA-IIx system	DEG analysis	Matching samples	Transmembrane receptor protein phosphatase activity enrichment of BD associated DEGs and DETs.
Arion et al. 2017	19 BD, 19 BD, 19 SCZ, 19 HC	Brain (DLPFC)	Affymetrix GeneChip U219	DEG analysis	Age, gender, PMI, brain pH, RIN, death by suicide, medication and diagnosis	Only SCZ significant transcriptome alteration identified.
Ciobanu et al. 2016	(MDD)	Brain and blood	Micoarrays and RNA-seq	Meta-analysis	Not mentioned	Brain and blood replicated genes indicating shared comorbidity with cardiovascular diseases.
Ciobanu et al. 2018	27 recurrent MDD, 494 non recurrent MDD	Whole blood	Illumina HumanHT-12 v4	DEG and WGCNA	Matching samples	MDD associated modules enriched in innate and adaptive immune response signaling.
Ciobanu et al. 2020	27 recurrent MDD, 494 non recurrent MDD	Whole blood	Illumina HumanHT-12 v4	Machine learning	Not mentioned	Transferrin receptor as a predictor of patient status in a model with a 63% accuracy.
Gamazon et al. 2019	SCZ, BD, ADHD, major and broad depression	Brain, whole blood, colon transverse and adrenal gland	Micoarrays and RNA-seq	Gene based association test	Age, gender, brain pH and PMI	Cross tissue ADHD, BD, SCZ and MDD associated genes.
Guan et al. 2019	47 AUT, 31 SCZ, 25 BD and 83 HC	Brain	High throughput sequencing	Aberrant gene expression analysis and WGCNA	Age, gender and brain region	A module eigengene associated with BD and SCZ enriched in negative regulation of biosynthetic and nitrogen compound metabolic process.
Hess et al. 2019	95 BD & 111 HC and 258 SCZ & 241 HC	Whole blood	Micoarrays and RNA-seq	Meta-analysis and WGCNA	Age, gender, study site and significant surrogate variables	DEG and co-expression modules from WGCNA described alterations of the immune system, oxidative stress and apoptosis in BD patients.
Hori et al. 2018	MDD: 20 resilient, 20 vulnerable and 20 resistant subjects	Whole blood	Agilent SurePrint G3 v3 human microarrays	DEG analysis	Age, gender, study site and significant surrogate variables	Upregulated DEGs associated with ribosomal function.
Jansen et al. 2016	414 MDD (base line & 2 year follow-up)	Whole blood	Affymetrix U219 arrays	DEG analysis	Age, gender, BMI, smoking status, blood cell composition and technical covariates	Genes associated with MDD enriched for interleukin-6 and NK cells involved in the immune response system.
Kim et al. 2016	15 BD, 1 5MDD, 15 SCZ and 15 HC	Hippocampus	High throughput sequencing	DEG and WGCNA	Matching samples	No concerved immune/inflammation-related co-expression modules in BD and MDD
Krebs et al. 2019	23 BD, 23 SCZ, 26 HC	Whole blood	Illumina HiSeq 2500	DEG and WGCNA	Age, gender, ethnicity and PMI	6 genes significantly associated to BD independently from lithium treatment: *c6orf163,COG4,PVT1,DOCK3,BBS9,ENSG00000267702*.
Labonté et al 2017	26 MDD and 22 HC	Brain (6 regions)	Illumina HiSeq 2500	DEG and WGCNA	Age, PMI, tissue pH and RIN	Similar gene modules in MDD but appearing in different brain regions across genders.
Le et al. 2018	78 MDD and 79 HC	PBMC	Illumina Hiseq 3000	WGCNA	Age, gender, BMI and batch	Modules associated with Montgomery-Åsberg Depression Scale enriched for apoptosis and downstream signaling by B cell receptor.
Leday et al. 2018	207 MDD and 157 HC	Whole blood	Affymetrix Human Genome U133	DEG and WGCNA	Age, gender, batch and anxiety diagnosis	165 DEGs related to the immune response system.
Mahajan et al. 2018	23 MDD and 23 HC	Hippocampus	Illumina NextSeq 500	DEG analysis	Age, PMI, tissue pH, BMI and RNA quality index	30 DEGs with down-regulated genes involved in the inflammatory response and up-regulated ones in cytokine function
Mellon et al. 2016	20 MDD and 20 HC	PBMC	Affymetrix Human Genome U133	DEG analysis	Matching samples	DEGs involved in transcriptional activity of cAMP, oxidative stress and immune responses.
Mostafavi et al. 2014	463 MDD and 459 HC	Whole blood	Illumina HiSeq 2000	Likelihood ratio test	Age, gender, BMI, smoking and medication	No significantly genome-wide MDD associated gene identified.
Pacifico and Davis 2017	18 BD and 17 HC	Brain (dorsal striatum)	Illumina HiSeq 2000	DEG and WGCNA	Age, gender, PMI and RIN	DEGs and modules enriched in an immune and inflammatory response.
Ramaker et al. 2017	24 BD, 24 MDD, 24 SCZ and 24 HC	Brain	Illumina HiSeq 2000	DEG analysis	Ahe, gender, PMI and percentage of reads uniquely aligned	No significantly genome-wide BD, SCZ or MDD associated gene identified.
Seney et al. 2018	26 MDD men and 24 MDD women	Brain (3 regions)	Illumina HiSeq 2000	Meta-analysis	Age, gender, PMI, brain pH, RIN, death by suicide, medication and alcohol dependence	Opposite transcriptional alteration in man and woman with major depressive disorder.
Wittenberg et al. 2020	(MDD)	Peripheral tissue	Micoarrays and RNA-seq	Meta-analysis	Age, gender and batch	Common genes associated with the immune response system and especially neutrophil activation.

Legends : BA, Brodmann area; BD, bipolar disorder; DEG, differentially expressed gene; DLPFC, dorsolateral prefrontal cortex; HC, healthy control; MDD, major depressive disorder; PBMC, peripheral blood mononuclear cells; PMI, post mortem interval; RIN, RNA integrity number; SCZ, schizophrenia; WGCNA, weighted correlation network analysis.

2.1.6 Epigenetics

The heritability, or h index, estimates the proportion of variance attributed to genetic factors, assessed through studies of monozygotic twins or, more broadly, population studies. The heritability of MD was greatly overestimated by the first measurements carried out in monozygotic twins, reaching more than 85% for BD (McGuffin et al., 2003) and more than 44% for MDD (Bierut et al., 2015). In later population-based studies, more nuanced results found heritabilities around 60% and 40% for BD and MDD, respectively (Fernandez-Pujals et al., 2015; Johansson et al., 2019). However, these values are still believed to be overestimated (Owen & Williams, 2021). Furthermore, despite promising results from large-scale GWAS on tens or even hundreds of thousands of samples, Lee et al. (2013) estimated that, on average, the proportion of variance in psychiatric disorders explained by common genomic variants may not exceed 30% (MDD having the lowest SNP-based heritability with 17%) and Mistry et al. (2018) attributed less than 2% of both MDD and BD phenotypic variance to their polygenic risk scores. This sparked debate about the so-called "missing heritability" of psychiatric disorders.

Although statistical limitations of effect size can partly explain this missing heritability (which is not specific to psychiatric phenotypes but is more broadly a general observation in multifactorial diseases), epigenetic mechanisms of interactions between genes and environmental factors have been suggested as another promising contributing factor (Trerotola et al., 2015). This hypothesis stems partly from the recognition, driven by epidemiological studies, that environmental exposures are significant risk factors for psychiatric traits. Accordingly, available evidence demonstrates a strong association between stressful life events and a higher risk for MD (Hillegers et al., 2004; Kendler et al., 1999; Low et al., 2011; Rowland & Marwaha, 2018). While the causality of this interaction is still debated, interest in the incorporation of environmental factors as a source of variance in MD has grown rapidly. This is now frequently used in gene x environment models, with epigenetics conceived as molecular substrates of these interactions.

Epigenetic mechanisms correspond to reversible alterations to DNA or DNA-associated proteins that influence gene expression without modifying the DNA sequence. They explain differences in cell paths and gene expression programs, although they share and rely on the same genetic code (Allis & Jenuwein, 2016; Waterland & Michels, 2007). They include 3 main categories: i) covalent modifications of the DNA, ii) post-translational modifications of histone proteins, and iii) non-coding RNAs. The mechanisms involved in these categories are further detailed in the next sections, followed by a review of epigenome-wide analyses of MD.

2.1.7 Covalent modification of DNA

DNA methylation is carried out by enzymes called DNMT DNA methyltransferases, which add a methyl group at the 5th position of a cytosine base (5mC), followed by a guanine G (CpG dinucleotide). Depending on these modifications, genomic localization and chromatin context, gene expression can be modulated. Although DNA methylation generally acts as a repressor on gene expression, the presence of methylated CpG groups in gene bodies has been associated with an activating effect in some cell lines and tissue types (Greenberg & Bourc'his, 2019; Jones, 2012).

A wide range of technologies have been implemented over the years to measure DNA methylation levels, including: immunoprecipitation of methylated DNA, or Methyl-CpG binding domain-based capture, coupled to next-generation sequencing (MeDIP-seq or MBDCap-seq); the use of restriction enzymes that are either sensitive (MRE-seq) or insensitive (MspI, in Reduced Representation Bisulfite sequencing, RRBS) to DNA methylation presence, or microarrays (e.g. 450K, EPIC arrays) (Xing et al., 2018). Among these approaches, the most comprehensive, but also the most expensive, is whole genome sequencing of converted DNA, with the classical gold standard of chemical conversion (bisulfite conversion, WGBS) now rapidly replaced by enzymatic conversion (Enzymatic Methylation sequencing, EM-seq; see Vaisvila et al. (2021)). These conversions correspond to the fact that non-methylated cytosines become uracils, while methylated cytosines are unaffected, such that the methylation status of individual cytosines is revealed by changes in the nucleotide sequence of sequencing reads. However, up-to-date, sequencing-based approaches for DNA methylation analyses have remained costly. Therefore, most studies have relied on microarrays for the analysis of large cohorts, or on Methyl-CpG binding domain-based capture and sequencing.

For DNA methylation arrays, inspired by SNP genomic phenotyping chips, Bibikova et al. (2009) designed hybridization chips of methylated or non-methylated sequences, following bisulfite treatment, to their pre-identified complementary probes. The first implementation, HumanMethylation27 chips, incorporated 27,000 CG probes located in the promoter region of nearly 15,000 genes. Later on, Bibikova et al. (2011) extended this technology by adding type I and type II probes to increase genomic coverage and expand the number of investigated probes to 450,000 (450k arrays). Lately, Moran et al. (2016) doubled this number, with the last generation of EPIC arrays now covering 850,000 methylation sites in the human genome, including a set of enhancer regions.

Until recently, the assessment of the role of DNA methylation in MD and other psychiatric traits was limited to a handful of candidate genes, mainly: Brain-derived neu-

rotrophic factor *BDNF*, Glucocorticoid receptor *NR3C1*, and Serotonin transporter *SLC6A4*. Now, genome-wide scaled DNA methylation studies are conducted on MDD and BD, and have tried to describe e.g. part of the relationship between childhood trauma and MDD (Peng et al., 2018). Based on studies that leveraged such genome-wide technologies, accumulating evidence now suggests an implication of DNA methylation in gene x environment interactions in psychiatric phenotypes (Alladi, 2018; Bagot & Meaney, 2010; Binder, 2017; Kubota et al., 2014; Van Dongen et al., 2016). Recent studies have also tried to differentiate methylation changes influenced by different environmental factors that are associated with MD from MDD genetic risk, using models including polygenic scores, methylation scores, and Mendelian randomization (Barbu et al., 2022; Shen et al., 2022).

Building on previous reviews that examined the role of DNA methylation in MDD and BD(Li et al., 2019; Teroganova et al., 2016), we identified (n=20) publications investigating DNAm in MD, which are detailed in Table 2. This results from a systematic bibliographic search (PubMed database) conducted on May 27, 2021.

Table 2: Recent and main genome-wide DNA methylation studies on mood affective disorders.

Study	Cohort	Biological sample	Technologies	Main analysis Method	Covariate correction	Relevent results
Aberg et al. 2020	Brain derived : 31 MDD and 30 HC Blood derived: 812 MDD and 320 HC	Brain and blood	Micoarrays (Illumina Human-Methylation 450K BedChip) and MBD-seq (Illumina NextSeq 500)	MWAS	Age, gender and technical variables	Significant overlap between blood and brain MWAS with genes associated to neurogenesis.
Barbu et al. 2020	1223 MDD and 1824 HC	Whole blood	Illumina Human-MethylationEPIC BeadChip	LASSO penalised regression model	Age, gender and and genetic PCs	DNAme risk scores (MRS) independently associated with current and future MDD status .
Bustamante et al. 2018	29 MDD and 29 HC	Brain (Prefrontal cortex)	Illumina Human-Methylation 450K BedChip	DMA and WGCNA	Age, gender, batch, suicide status and PMI	WDR11 probe genome-wide significant; probes and WGCNA modules enriched for synapse function, DNA binding, and mitochondrion.
Chan et al. 2019	Brain derived : 113 MDD and 93 HC Blood derived: 812 MDD and 320 HC	Brain and whole blood	Micoarrays (Illumina Human-Methylation 450K BedChip) and MBD-seq (Illumina NextSeq 500)	MWAS	Age, gender and technical variables	T-cell and monocyte methylome-wide associations in MDD and genes involved in innate immune toll-like receptor signaling.
Clark et al. 2019	581 MDD (6-year follow-up)	Whole blood	MeDIP-seq (Illumina NextSeq 500)	MWAS	Gender, smoking status, MDD severity, technical artefacts and blood cells composition	Association of DNAme to stress and signaling mechanisms linked to immune cell migration and inflamation.
Comes et al. 2020	96 BD	Whole blood	Illumina Human-MethylationEPIC BeadChip	Linear mixed-efects model	Agen gender, batch, smoking and blood cells composition	*POU6F2* suggestively associated to stressful life events.
Dempster et al. 2014	SCZ, BD, ADHD and major & broad depression	Buccal cells	Illumina Human-Methylation 450K BedChip	DMA	Matching samples	DMPs in saliva associated with depression (*DUSP4,ACP1,SSTR5,NR3C1,GARBRA3*) and enriched for genes involved in nervous system development.
Fries et al. 2017	22 BD, 16 BD siblings and 20 HC	Whole blood	Illumina Human-Methylation 450K BedChip	DMA an,d epigenetic clock	Age, gender and blood cells composition	BD patients exibited an accelerated epigenetic aging.
Gaine et al. 2019	50 BD and 31 HC	Brain (BA46)	Methyl-Seq (Illumina HiSeq2000/2500)	DMA	Age, gender, PMI, brain pH and hidden factors	Axonal guidance signaling pathway associated with BD and opioid signaling involved in methyated genes in suicidal BD patients.
Han et al. 2018	811 MDD and 319 HC	Whole blood	MBD-seq (SOLiD5500 wildfire)	Epigenetic clock	Age, gender, study site and significant surrogate variables	MDD patients exibited an accelerated epigenetic aging.
Ho et al. 2019	20 BD, 10 MDD and 10 I[a]	Brain (DLPFC & BA38)	Illumina Human-Methylation 850K BedChip	DMA	Age and gender	Intra and inter tissue and condition comparaison revealed pathways involving melatonin signaling and cyrcadian rythm associated with BD; axonal guidance and opiod signaling associated with MDD.
Hüls et al. 2020	608 MDD	Brain (DLPFC)	Illumina Human-Methylation 450K BedChip	EWAS	Age , gender, PMI, proportion of neurons and the first three genetic PCs	Altered DNAme levels in the *YOD1/ PFKFB2* loci associated with MDD.
Jeremian et al. 2017	18 highly suicidal and 22 low suicidal BD	White blood cells	Illumina Human-Methylation 450K BedChip	DMA and epigenetic clock	Age, batch effect and cells composition	Hypomethylation of *MPP4* and *TBC1D16* and hypermethylation of *NUP113* in BD patient with suicidal history accompagned with altered molecular aging.
Lutz et al. 2017	52 MDD and 26 HC	Brain (Anterior Cingulate Cortex)	RRBS (Illumina HiSeq 2000)	DMA	Age, gender and RIN	*LINGO3, POU3F1* and *ITGB1* differentially methylated genes involved in the myelin and oligodendrocytes in the anterior cingulate cortex associated to childhood abuse in MDD patients.
Malki et al. 2016	97 MZ twins pairs discordant for MDD	White blood cells	Illumina 8.1-K CpG island microarray	Machine learning and network analysis	Batch and unwanted variance (PCs)	Altered methylation levels of genes and networks involved in the inflamatory response associated with MDD.
Marie-Claire et al. 2020	15 BD lithium responders and 11 BD non responders	Whole blood	Roche NimbleGen SeqCap Epi CpGiant	DMA and sPLS	Age, gender, phenotype and mood stabilizer medication status	7 DMRs with a good discriminatory power for response to Li (AUC=0.806) including : *EIF2B5, VWA5B2* and *RALGAPA1*.
Okazaki et al. 2020	30 BD and 30 HC	Whole blood	Illumina Human-Methylation 450K BedChip	Epigenetic clock	Matching samples	Decreased epigenetic aging in patients with BD associated with a potential repression effect of mood stabilizers.
Ruzicka et al. 2018	8 BD, 8 SCZ and 8 HC	Hippocampus	Illumina Human-Methylation 450K BedChip	DMA	Matching samples	11 significant DMRs involved in pathways including *WNT* signaling factors and zinc finger of the cerebellum genes familly members.
Shimada et al. 2018	20 MDD and 27 HC	White blood cells	Illumina Human-Methylation 450K BedChip	EWAS	Age, gender, BMI and cell proportion	DMGs involved in regulation of G-protein coupled receptor protein signaling pathway.
Walker et al. 2016	19 MDD, 3 BD and 23 I[a]	Whole blood	Illumina Human-Methylation 450K BedChip	DMA	Age, gender and hidden factors	No significantly genome-wide DMR and top gene enrichement included: nervous system development and function.

Legends : BA, Brodmann area; BD, bipolar disorder; DLPFC, dorsolateral prefrontal cortex; DMG, differentially methylated gene; DMP, differentially methylated probe; DMR, differentially methylated region; DNAme, DNA methylation; EWAS, epigenome-wide association study; HC, healthy control; MDD, major depressive disorder; PMI, post mortem interval; RIN, RNA integrity number; SCZ, schizophrenia; WGCNA, weighted correlation network analysis.

As observed in transcriptomics studies, the tissues characterized in these DNAm studies vary widely: 6 (Bustamante et al., 2018; Gaine et al., 2019; Ho et al., 2019; Hüls et al., 2020; Lutz et al., 2017; Ruzicka et al., 2018) out of 20 evaluated brain tissue, while 12 focused on peripheral tissues. Among the latter, Barbu et al. (2020), Clark et al. (2019), Comes et al. (2020), Fries et al. (2017), Han et al. (2018), Marie-Claire et al. (2020), Okazaki et al. (2020), and Walker et al. (2015) focused on whole blood samples, while Jeremian et al. (2017), Malki et al. (2016), and Shimada et al. (2018) analyzed white blood cells, and Dempster et al. (2014) used buccal cells. Distinctly, Aberg et al. (2020) and Chan et al. (2019) investigated DNAm across both brain and blood. To some extent, results obtained using DNAm concur with transcriptomics, and involve genes related to immune toll-like receptors (Chan et al., 2019) and neurogenesis (Aberg et al., 2020) across brain and blood. Regarding peripheral tissues in MDD, immune cell migration pathways were found enriched in a whole blood-based follow-up study (Clark et al., 2019), as well as pathways implicated in the inflammatory response (Malki et al., 2016) in white blood cells. These pathways were also found in the whole blood of unrelated individuals, at high familial risk, who later developed BD or MDD (Walker et al., 2015). Indeed, despite this study's lack of genome-wide significance, the Differentially Methylated Region (DMR) identified were linked to *IL1RAPL1*, a gene that encodes the X-linked interleukin-1 receptor accessory protein-like 1. Surprisingly, Dempster et al. (2014) found differentially methylated genomic sites associated with nervous system development in buccal cells of MDD patients, while Shimada et al. (2018) found differentially methylated genes involved in the regulation of G protein-coupled receptor signaling pathways in white blood cells of an MDD cohort. Recently, Comes et al. (2020) identified that DNAm levels at the *POU6F2* locus were associated with stressful events in the whole blood of patients with BD. In contrast, when considering brain analyses, a study comparing individuals with BD who died by suicide or from another cause, and non-psychiatric controls, found that genes associated with DMR identified in BD patients compared to controls, regardless of suicide, were enriched in the UVB-induced *MAPK* signaling, while suicidal BD associated with the opioid signaling pathway (Gaine et al., 2019). Furthermore, altered DNAm levels in the *YOD1/PFKFB2* locus were associated with MDD in the DLPFC region (Hüls et al., 2020). In the same brain region, Ho et al. (2019) compared BD patients and controls and identified differentially methylated genes associated with Melatonin signaling. DMRs were associated with circadian rhythm and Ephrin receptor signaling pathways in the same BD patients' temporal cortex. In addition, in the temporal cortex, this study highlighted the involvement of Dopamine *DARPP32* Feedback in *cAMP* signaling in MDD patients, which was also found enriched in whole blood DMRs of an MD high-risk family with a predominant representation of MDD patients (Walker et al., 2015). Additionally, the *Wnt* signaling pathway was also found to be differentially methylated in the hippocampus of BD patients (Ruzicka et

al., 2018), and in the whole blood of Walker et al. (2015)'s discovery cohort. Moreover, Lutz et al. (2017) found that MDD and childhood trauma associated with cell-type specific DMRs affecting glial cells, in particular oligodendrocytes, which were accompanied by alterations in myelin ensheathment. Finally, regarding biomarkers of treatment efficacy, Marie-Claire et al. (2020) proposed a methylomic signature of lithium response in BD patients. Altogether, these findings suggest that MD implicate the dysregulation of gene expression in part through alteration of the DNAm landscape.

Of note, over the last decade, the development of the concept of the epigenetic clock has complemented the differentially methylated gene approach (Horvath, 2015), and allowed the emergence of a hypothesis linking MD and stress with the dysregulation of epigenetic aging (Fries et al., 2017; Jeremian et al., 2017) and (Fries et al., 2020) for review). Accordingly, Han et al. (2018) showed that this aging mechanism might be accelerated in patients with MDD who suffered from Childhood Trauma (CT), while Tang et al. (2020) similarly found that emotional and physical abuse were associated with age acceleration in girls (using the "Horvath" DNA methylation clock), and Okazaki et al. (2020) provided evidence suggesting the beneficial effects of mood stabilizers may correlate with slower epigenetic aging in BD patients. Of note, a growing interest has recently focused on non-CG methylation. This form of DNA methylation is particularly enriched in brain tissue, non-randomly distributed (He & Ecker, 2015), and is thought to be associated with child abuse and MDD (Lutz et al., 2021).

2.1.8 Histone's post-translational modification

The condensation state of chromatin, which is made of nucleosomes consisting of 8 histone proteins (H2A, H2B, H3, and H4) around which the DNA is wrapped, determines transcriptional activity. Histones' post-translational modifications can alternate the regional chromatin state between a euchromatic, transcriptionally open and active state, and a heterochromatic, closed and transcriptionally silenced state (Bannister & Kouzarides, 2011). Histone chemical modifications can occur through over 100 known mechanisms (Zhao & Garcia, 2015), regrouped in categories that include acetylation, methylation and phosphorylation, which are the most studied ones. Histone acetylation occurs with the addition of an acetyl group by histone acetyltransferases or HATs on the lysine amino acids situated at the histone tail. Histone acetylation has an activation effect on gene expression. Histone methylation done by histone methyltransferases or HMTs has a repressing effect. Conversely, histones demethylases (HDMs) reversibly remove the added methyl group, promoting gene expression.

Genome-wide histone analyses Chromatin immunoprecipitation (ChIP) has evolved into a powerful and widely used technique for studying DNA-Protein interactions. ChIP-Seq is accomplished through the following steps: initially, cells are treated with an agent that covalently bonds DNA-interacting proteins (histones) to the DNA. The genomic DNA is broken down into 300bp-long fragments. An antibody recognizes the protein of interest, and immunoprecipitation is used to isolate protein-DNA complexes. These DNA fragments are then sequenced.

Histone modification and psychiatric disorders Chromatin remodeling plays a significant role in the pathophysiology of several psychiatric phenotypes (Nestler et al., 2016; Sun et al., 2013; Uchida et al., 2018). Moreover, Li et al. (2020) reviewed and reported epigenetic effects of early life stress events in animal-based models, including histone modifications in the hippocampus. However, various limitations, ranging from tissue accessibility to many confounding factors, hamper the association of a change at a single modification to a clinical readout. The technologies to investigate chromatin architecture and states are multiple, and not all linked to post-translational modification of histone proteins (HiC and ATACseq). After the first ChIP-seq H3K27ac assays across different brain regions, the PsychENCODE consortium now investigates not only bulk-tissue brain samples, but also at the single cell scale (Hu et al., 2021; Wang et al., 2018). Thus, further investigations are ongoing.

2.1.9 Non coding RNAs

The involvement of Non-coding RNA (ncRNA) in several epigenetic processes has recently emerged, and has been included under a broad definition of epigenetics. ncRNAs vary in size and function and contain microRNA (miRNA), small interfering RNA (siRNA), and long non-coding RNA (lncRNA).

miRNAs: miRNAs are small RNA molecules that are about 22-nucleotide long and can post-transcriptionally regulate the expression of their target genes. There are currently over 2000 known miRNAs in Human. MiRNAs bind by sequence complementarity to their targets, completely or partially. Complete binding can induce target degradation, whereas incomplete binding, typical in 3 UTR regions, can induce translational repression of targeted genes. Each mRNA may be regulated by multiple miRNAs, while each miRNA can target multiple mRNAs. MiRNAs are transcribed into primary miRNAs by RNA polymerase II, and then processed in the nucleus into precursor miRNAs by the RNase III Drosha and *DGCR8*. They are then exported into the cytoplasm by exportin 5, and cleaved into final functional mature miRNAs by Dicer.

miRNAs and psychiatric disorders: ncRNAs, especially miRNAs, play critical roles in several biological brain functions, from neurogenesis to neural development and synaptic plasticity (Brennan & Henshall, 2020; Hollins & Cairns, 2016; Prodromidou & Matsas, 2019). Therefore, there is a growing interest in whether disruption of miRNA regulation might contribute to psychiatric disorders. These aspects in relation to BP and MDD were recently reviewed by Fries et al. (2018) and Żurawek and Turecki (2021a).

On May 27, 2021, we conducted a comprehensive literature search on human-based investigations of microRNA in mood disorders. To meet genome-wide inclusion criteria, we only considered articles screening a number of miRNAs above the arbitrary threshold of 700. We identified 13 publications investigating MD. These studies, listed in table 3, mostly investigated cohorts of modest size and, as observed for other omic mechanisms reviewed above, tissues that were investigated varied across studies, and included the brain (Amoah et al., 2019; Yoshino et al., 2020a; 2020b) and peripheral tissues (whole blood, plasma, PBMCs, and lymphoblastic cell lines, LCLs) (Gururajan et al., 2016; Kim et al., 2019; Lim et al., 2016; Maffioletti et al., 2016; Mendes-Silva et al., 2019; min Fan et al., 2014; Squassina et al., 2020; Tabano et al., 2019; Yrondi et al., 2020; Zhao et al., 2019). Various technologies were used (Nanostring, microarrays, or high-throughput sequencing). Combined with different statistical analysis approaches and cutoffs, this results in a few significant differentially expressed miRNAs common to several studies 6.2. Some of these results also support the hypothesis that circulating miRNAs could mediate brain-blood cross-tissue relationships Roy et al. (2020). For example, whole-blood expression of a new miRNA identified in chromosome 11, pmiR-chr11, correlated with hippocampal volume in individuals with MDD Zhao et al. (2019). In addition, 5 differential miRNA analysis studies, using either brain or peripheral tissues, reported enrichment of their targeted genes for the neural development process (Maffioletti et al., 2016; min Fan et al., 2014; Tabano et al., 2019; Yoshino et al., 2020a; 2020b). Pathway enrichment analyses of the miRNA-targeted genes also concur with those uncovered at the level of other molecular layers previously described. Accordingly, differentially expressed miRNAs in BD and MDD target genes were enriched for the immune-inflammatory response Kim et al. (2019) and Mendes-Silva et al. (2019). Thus, miRNAs could play a role in this pathophysiological mechanism of MD. Indeed, under environmentally stressful conditions, miRNAs are thought to act as a response vector of the immune-inflammatory system (Mehta & Baltimore, 2016). Furthermore, miRNAs were also identified as potential biomarkers of BD, schizophrenia (Amoah et al., 2019), MDD (Gururajan et al., 2016), suicidal behavior in BD patients (Squassina et al., 2020), and nausea intensity in MDD patients treated with mood stabilizers (Yrondi et al., 2020). Altogether, these results suggest that miRNAs are promising diagnostic and predictive biomarkers for patient stratification in psychiatric diseases (Clausen et al., 2022; Gibbons et al., 2020).

Table 3: Recent and main genome-wide miRNA studies on mood affective disorders.

Study	Cohort	Biological sample	Technologies	Analysis method	Covariate correction	Differentially expressed miRNA (direction of change)	Relevent results
Amoah et al. 2019	26 BD, 29 SCZ and 25 HC	Brain (OFC)	NanoString nCounter miRNA	Standard error to the mean	Matching samples	hsa-miR-132-3p(↑), hsa-miR-28-5p(↓), hsa-miR-223-3p(↓), hsa-miR-548ar-3p(↑), hsa-miR-1268a(↓),hsa-miR-28-3p(↓), hsa-miR-874-5p(↓), hsa-miR-15b-5p(↓), hsa-miR-193a-5p+hsa-miR-193b-5p(↓), hsa-miR-548b-3p(↑), hsa-miR-326(↑), hsa-miR-1291(↓), hsa-miR-374b-5p(↓), hsa-miR-198(↑), hsa-miR-493-3p(↓), hsa-miR-193b-3p(↓), hsa-miR-522-3p(↓), hsa-miR-374c-5p(↓), hsa-miR-361 5p(↓), hsa-miR-454-3p(↓), hsa-miR-936(↑), hsa-miR-152-3p(↓), hsa-miR-1908-5p(↑), hsa-miR-365b-5p(↑), hsa-miR-4755-5p(↑), hsa-miR-155-5p(↓), hsa-miR-1245b-3p(↓), hsa-miR-924(↓), hsa-miR-3136-5p(↓), hsa-miR-23a-3p(↓), hsa-miR-320a(↓), hsa-miR-1290(↓), hsa-miR-204-5p(↓), hsa-miR-892a(↑), hsa-miR-181b-2-3p(↑), hsa-miR-4454+hsa-miR-7975(↓),hsa-miR-1185-1-3p(↑), hsa-let-7c-5p(↓), hsa-miR-203a-3p(↑),hsa-miR-3161(↓),hsa-let-7b-5p(↓), hsa-miR-92d(↑), hsa-miR-411-5p(↑), hsa-miR-630(↓), hsa-miR-211-5p(↑),hsa-let-7d-5p(↓), hsa-miR-154-5p(↑), hsa-miR-1205(↑), hsa-miR-378b(↓), hsa-miR-487a-3p(↑), hsa-miR-580-3p(↓), hsa-miR-532-5p(↓), hsa-miR-3613-5p(↑), hsa-miR-4461(↑), hsa-miR-874-3p(↓),hsa-miR-3605-3p(↑),hsa-miR-125a-5p(↑), hsa-miR-543(↑),hsa-miR-4521(↓),hsa-miR-10a-5p(↓), hsa-miR-513c-5p(↑), hsa-miR-143-3p(↓),hsa-miR-624-3p(↑), hsa-miR-498(↓), hsa-miR-199a-3p+hsa-miR-199b-3p(↓), hsa-miR-16-5p(↓), hsa-miR-671-3p(↓), hsa-miR-361-3p(↓),hsa-miR-199a-5p(↓), hsa-miR-608(↑), hsa-miR-4284(↓), hsa-miR-196a-5p(↓), hsa-miR-106a-5p+hsa-miR-17-5p(↓), hsa-miR-561-5p(↓), hsa-miR-483-5p(↓), hsa-miR-627-5p(↓).	miR-193b-3p(↓)and miR-28a-3p(↓) in both SCZ and BD.
Fan et al. 2014	3 MDD and 3 HC	PBMC	Affymetrix miRNA 3.0 array	Mann Whitney U-test	Matching samples	miRNA-1244(↑), miRNA-1972(↑), miRNA-874(↑),miRNA-29b(↑), miRNA-4448(↑), miRNA-933(↑), miRNA-1291(↑), miRNA-3147(↑), miRNA-4743(↑), miRNA-4498(↑), miRNA-4669(↑),miRNA-26b(↑), miRNA-4485(↑), miRNA-146b-5p(↑), miRNA-4776-5p(↑), miRNA-3679-5p(↑), miRNA-4298(↑), miRNA-928(↑), miRNA-4640-5p(↑),miRNA-4521(↑),miRNA-4327(↑), miRNA-338-5p(↓), miRNA-210(↓),miRNA-4443(↓),miRNA-143(↓).	miRNA targets pathways involved in the nervous system and brain functions.
Gururajan et al. 2016	40 MDD and 20 HC	Whole blood	Exiqon Human miRCURY LNA microRNA Array	Differential expression analysis	None	hsa-let-7b-5p (↓), hsa-let-7c-5p (↓)	hsa-let-7b(↓) and hsa-let-7c(↓) with targeted genes involved in PI3k-Akt-mTOR signaling pathway were significantly differentially expressed in MDD patients
Kim et al . 2019	450 MDD (8 weeks follow-↑) , 361 MDD (6 weeks follow-↑) and 475 MDD (8 weeks follow-↑)	Plasma	Ion Proton sequencer	Machine learning	None	hsa-miR-21-5p, hsa-miR-144-3p, hsa-let-7c-5p	hsa-miR-21-5p , hsa-miR-144-3p and hsa-let-7c-5p ranked for duloxetine treatment in the top 10th percentile in at least two out of the three investigated studies with targeted genes involved in the cytokine-mediated signaling pathway.
Lim et al. 2016	10 BD (12 weaks follow-↑)	Whole blood	Affymetrix GeneChip miRNA 4.0 array	Mann Whitney U-test	Matching samples	Asenapine Gro↑: hsa-miR-18a-5p(↑), hsa-miR-19b-3p(↑), hsa-miR-145-5p(↑), hsa-miR-27a-3p(↑), hsa-miR-148b-3p(↑), hsa-miR-210-3p(↑), hsa-miR-17-3p(↑), hsa-miR-30b-5p(↑), hsa-miR-106b-5p(↑),hsa-miR-339-5p(↑), hsa-miR-106a-5p(↑), hsa-miR-20a-5p(↑), hsa-miR-17-5p(↑), hsa-miR-15a-5p(↑), hsa-miR-92b-5p(↓), hsa-miR-1343-5p(↓). Risperidone Gro↑: hsa-miR-664b-5p(↓), hsa-miR-6778-5p(↓), hsa-miR-1460-5p(↓)	Target genes enriched in pathways associated with immune system response and apoptosis.
Maffioletti et al. 2016	20 BD, 20 MDD and 20 HC	Whole blood	Affymetrix miRNA 3.0 array	Standard deviation	Matching samples	miRNA-1244(↑), miRNA-1972(↑), miRNA-874(↑),miRNA-29b(↑), miRNA-4448(↑), miRNA-933(↑), miRNA-1291(↑), miRNA-3147(↑), miRNA-4743(↑), miRNA-4498(↑), miRNA-4669(↑),miRNA-26b(↑), miRNA-4485(↑), miRNA-146b-5p(↑), miRNA-4776-5p(↑), miRNA-3679-5p(↑), miRNA-4298(↑), miRNA-928(↑), miRNA-4640-5p(↑),miRNA-4521(↑),miRNA-4327(↑), miRNA-338-5p(↓), miRNA-210(↓),miRNA-4443(↓),miRNA-143(↓).	miR-345-5p(↑),miR-330-3p(↑) and miR-29c-5p(↑) significantly differentially expressed across MDs with enrichments of target genes involved in neural function including Wnt, mTOR, ErbB, and insulin signaling pathways.
Mendes-Silva et al. 2019	19 MDD and 19 HC	Plasma	Ion Proton sequencer	Differential expression analysis	Matching samples	hsa-miR-184(↑), hsa-miR-1-3p(↑).	miR-184 ↑regulated in late-life depressive disorder with targeted genes involved in the immune-inflammatory response, neuroplasticity and apoptosis.
Squassina et al. 2020	12 BD and 13 HC	LCL (LiCl treated/untreated)	NanoString nCounter miRNA	Student t-test	Matching samples	Untreated LCLs : hsa-miR-428(↑), hsa-miR−337−3p(↑). LiCl treated LCLs : hsa-miR-428(↑), hsa-miR−186−5p(↑), hsa-miR−423−5p(↑).	miR-428(↑) as a potential BD biomarker of suicide with insulin resistance enriched targeted genes.
Tabano et al. 2019	15 BD and 9 HC	Plasma	NanoString nCounter miRNA	Mann Whitney U-test	Matching samples	hsa-miR-21-5p(↑), hsa-miR-379-5p(↓), hsa-miR-644a(↓), hsa-miR-520f-3p(↓), hsa-miR-199a-3p+hsa-miR-199b-3p(↑), hsa-let-7g-5p(↑), hsa-miR-520h(↓), hsa-miR-548q(↓), hsa-miR-548ah-5p(↓), hsa-let-7f-5p(↑),hsa-miR-498(↓), hsa-let-7a-5p(↑), hsa-miR-1290(↓), hsa-miR-212-3p(↓), hsa-miR-128-1-5p(↓), hsa-miR-196a-5p(↓),hsa-miR-587(↓), hsa-miR-219b-3p(↓), hsa-miR-1297(↓), hsa-miR-553(↓), hsa-miR-16-5p(↑), hsa-miR-23a-3p(↑), hsa-miR-376a-3p(↓), hsa-miR-107(↑), hsa-miR-15b-5p(↑),hsa-let-7d-5p(↑), hsa-miR-549a(↓), hsa-miR-4536-5p(↓), hsa-miR-146a-5p(↑),hsa-miR-98-5p(↑), hsa-miR-1253(↓), hsa-miR-640(↓), hsa-miR-26b-5p(↑).	Targets of BD differentially expressed miRNA involved in neurodevelopment and neurogenesis.
Yoshino et al. a 2020	22 MDD and 25 HC	Brain (anterior cingulate cortex)	Illumina NextSeq 500	Differential expression analysis	Age, gender, brain pH, PMI, alcohol abuse, antidepressant and cause of death.	hsa-miR-6077 (↑), hsa-miR-4632−3p(↑), hsa-miR-6789−3p(↑), hsa-miR-648(↑), hsa-miR-4498(↑), hsa-miR-6084(↑), hsa-miR-4433-a-3p(↑), hsa-miR-4638−5p(↑), hsa-miR-4258(↑), hsa-miR-6850−5p(↑), hsa-miR-4510(↑), hsa-miR-3651(↑), hsa-miR-4497(↑), hsa-miR-668−5p(↑), hsa-miR-6800−3p(↑), hsa-miR-7108−3p(↑), hsa-miR-4761−3p(↑),hsa-miR-4426(↑),hsa-miR-4720−3p(↑), hsa-miR-4281(↑),hsa-miR-1273h-3p(↑), hsa-miR-4736(↑), hsa-miR-3155a(↑),hsa-miR-1227−5p(↑), hsa-miR-6802−5p(↑), hsa-miR-572(↑), hsa-miR-1470(↑), hsa-miR-520f-3p(↑), hsa-miR-92b-3p(↑), hsa-miR-5572(↑), hsa-miR-4746−3p(↑), hsa-miR-550b-2−5p(↑), hsa-miR-518e-5p(↑), hsa-miR-6781−5p(↑), hsa-miR-6890−3p(↑), hsa-miR-760(↑), hsa-miR-516a-3p(↑),hsa-miR-4695−5p(↑), hsa-miR-501−3p(↑), hsa-miR-1180−5p(↑), hsa-miR-4728−3p(↑), hsa-miR-523−5p(↑),hsa-miR-1193(↑), hsa-miR-4318(↑), hsa-miR-5580−5p(↑), hsa-miR-7112−5p(↑), hsa-miR-6805−3p(↑), hsa-miR-5587−5p(↑), hsa-miR-6851−5p(↑), hsa-miR-6814−3p(↑), hsa-miR-4663(↑), hsa-miR-6075(↑), hsa-miR-1343−3p(↑), hsa-miR-6737−5p(↑), hsa-miR-6879−3p(↑),hsa-miR-4701−3p(↑), hsa-miR-92b-5p(↑), hsa-miR-6753−5p(↑), hsa-miR-4485−5p(↑), hsa-miR-6769b-5p(↑), hsa-miR-519b-3p(↑), hsa-miR-4430(↑), hsa-miR-6511b-5p(↑), hsa-miR-595(↑), hsa-miR-3591−5p(↑), hsa-miR-6892−5p(↑), hsa-miR-6769a-5p(↑), hsa-miR-6815−3p(↑), hsa-miR-4662(↑), hsa-miR-4758−5p(↑),hsa-miR-622(↑), hsa-miR-6851−3p(↑), hsa-miR-521(↑), hsa-miR-521−1(↑), hsa-miR-521−2(↑), hsa-miR-636(↑), hsa-let-7c-3p(↑),hsa-miR-6810−3p(↑), hsa-miR-1247−5p(↑), hsa-miR-3648(↑), hsa-mir-3648−1(↑), hsa-mir-3648−2(↑), hsa-miR-484(↑), hsa-miR-3621(↑), hsa-let-7e-5p(↑), hsa-miR-519d-3p(↑), hsa-miR-486−3p(↑), hsa-miR-3189−3p(↑), hsa-miR-5189−5p(↑), hsa-miR-452 2(↑), hsa-miR-302c-5p(↑), hsa-miR-328−3p(↑), hsa-miR-1−5p(↑), hsa-miR-1234−3p(↑), hsa-miR-6826−3p(↑), hsa-miR-378d(↑), hsa-miR-30d-5p(↑), hsa-miR-7156−3p(↑), hsa-miR-519e-5p(↑), hsa-miR-6869−5p(↑), hsa-miR-3689f(↑), hsa-miR-4512(↑), hsa-miR-6857−5p(↑), hsa-miR-6839−3p(↑), hsa-miR-6131(↑), hsa-miR-6510−5p(↑), hsa-miR-4505(↑), hsa-miR-4685−3p(↑), hsa-miR-4530(↑),hsa-miR-6794−5p(↑), hsa-miR-125b-2−3p(↑), hsa-miR-4538(↑), hsa-miR-663b(↑), hsa-miR-3609(↑), hsa-miR-2861(↑), hsa-miR-4783−5p(↑), hsa-miR-671−3p(↑), hsa-miR-3928−5p(↑), hsa-miR-6821−5p(↑),hsa-miR-6853−3p(↑), hsa-miR-116−5p(↓), hsa-miR-553b-3p(↓), hsa-miR-367−3p(↓), hsa-miR-5590−5p(↓), hsa-miR-337−5p(↓), hsa-miR-1251−5p(↓), hsa-miR-543(↓), hsa-miR-138−5p(↓), hsa-miR-20b-3p(↓), hsa-miR-495−5p(↓), hsa-miR-6808-3p(↓),hsa-miR-4705(↓), hsa-miR-376a-3p(↓), hsa-miR-4796−3p(↓), hsa-miR-495−3p(↓), hsa-miR-154−5p(↓), hsa-miR-496(↓), hsa-miR-548ar-3p(↓), hsa-miR-670−3p(↓), hsa-miR-488−3p(↓), hsa-miR-29b-2−5p(↓), hsa-miR-497−5p(↓), hsa-miR-335−5p(↓),hsa-miR-487a-3p(↓), hsa-miR-544a(↓), hsa-miR-342−5p(↓),hsa-miR-1197(↓),hsa-miR-153−5p(↓), hsa-miR-684-6(↓), hsa-miR-758−3p(↓), hsa-miR-22−5p(↓), hsa-miR-29b-3p(↓), hsa-miR-377−3p(↓), hsa-miR-153−3p(↓),hsa-miR-2277−5p(↓), hsa-miR-9−3p(↓), hsa-miR-1307-5p(↓), hsa-miR-548a-3p(↓),hsa-miR-494−3p(↓), hsa-miR-22−3p(↓), hsa-miR-3689a-5p(↓), hsa-miR-146a-5p(↓), hsa-miR-4477b(↓), hsa-miR-582−3p(↓), hsa-miR-7−1−3p(↓), hsa-miR-376c-3p(↓), hsa-miR-135b-5p(↓), hsa-miR-373a-5p(↓), hsa-miR-128−2−5p(↓),hsa-miR-376c-5p(↓), hsa-miR-5582−3p(↓), hsa-miR-548ac(↓), hsa-miR-3182(↓), hsa-miR-5 48a-5p(↓).	miRNA targets pathways involved in morphometric changes and neuronal atrophy.
Yoshino et al. b 2020	15 MDD and 15 HC	Brain (DLPFC)	Illumina NextSeq 500	Differential expression analysis	Age, gender, brain pH, PMI, alcohol abuse, antidepressant and cause of death.	hsa-miR-217-5p(↓), hsa-miR-372-3p(↓), hsa-miR-33a-5p(↓),hsa-miR-32-5p(↓), hsa-miR-431-5p(↓), hsa-miR-30e-5p(↓), hsa-miR-205-5p(↓), hsa-miR-193a-5p(↑), hsa-miR-223-3p(↑), hsa-miR-376a-3p(↓), hsa-miR-374a-3p(↓), hsa-miR-455-3p(↑), hsa-miR-487a-3p(↓), hsa-miR-136-3p(↓), hsa-miR-181c-5p(↓), hsa-miR-376c-3p(↓), hsa-miR-629-5p(↑), hsa-miR-324-5p(↓)	Enrichment of miRNA target genes in DLPFC region of MDD patients with glutamatergic, ERK/MAPK, neuregulin, estrogen receptor, PI3K and telomeres pathways.
Yrondi et al. 2020	160 MDD (2weak follow-↑)	Whole blood	Illumina HiSeq 4000	ANOVA	Age and gender	miR20b-5p(↑), miR106a-5p(↑), miR20a-5p(↑), miR17-5p(↑), miR185-5p(↑), miR103a-3p(↑), miR103b(↑), miR1301-3p(↓), miR182-5p(↑), miR15b-5p(↑), miR93-5p(↑), miR660-5p(↑)	miR185-5p (↑) negatively correlated with the intensity of nausea in MDD patients
Zhao et al. 2019	10 MDD and 10 HC	Whole blood	Illumina HiSeq 2500	Student t-test	Matching samples	hsa-miR-3611(↑), pmiR-chr11(↑), hsa-miR-3158-3p(↑), hsa-miR-433(↑), hsa-miR-3944-5p(↑), hsa-miR-1275(↑), hsa-miR-4532(↑), hsa-miR-4516(↑), hsa-miR-30e-3p(↓), hsa-miR-148b-3p(↓)	pmiR-chr11 (↑) negatively correlated with the hippocampal volume of patients with MDD

Legends : ANOVA, analysis of variance; BD, bipolar disorder; DLPFC, dorsolateral prefrontal cortex; HC, healthy control; MDD, major depressive disorder; miR, micro RNA; LCL, lymphoblastoid cell lines; LiCL, lithium chloride; OFC, orbitofrontal cortex; PBMC, peripheral blood mononuclear cells; PMI, post mortem interval; SCZ, schizophrenia

2.2 Multi-omics : The whole is greater than the sum of its parts

As demonstrated in the preceding section, the biological variation with small effect sizes that have been uncovered by genome-wide and large-scale single-omic studies on MD and other mental disorders, are widely distributed across various biological modalities. Therefore, there is a natural incentive to simultaneously analyze these data to identify molecular signatures with stronger statistical associations. Beyond assessing the relevance of single-omic results, increasing availability of omic data from the PGC and PsychEncode projects (Akbarian et al., 2015; Sullivan, 2018) provides various advantages and opportunities. They enable the exploration of functional and pathophysiological markers that are associated with psychiatric disorders.

This abundance of data is often referenced as big data. Big data became a broad term to impart the accumulation of large volumes of complex measures, and the velocity at which data is generated. Furthermore, the combination of now-established machine learning methods with big data is commonly used in life sciences, and has already allowed predictive models for complex pathologies such as cancer. Its application to psychiatry, while still scarce, is gaining growing interest (Ressler & Williams, 2021) and may lead to better treatments and stratification of patients.

Multidimensional data may be integrated in various ways, with some fundamental differences. As a result, two primary types of paradigms could be considered: bottom-up and top-down. Bottom-up strategies are frequently hypothesis-driven. The data modalities are first examined independently, then curated accordingly after a clustering step. In contrast, the top-down paradigm is fundamentally holistic, in which case data are "viewed" simultaneously and evaluated using advanced statistical techniques (Menyhárt & Győrffy, 2021).

2.2.1 Untangle mood disorders through multi-Omics approaches

In order to anchor our research work on multi-omic data integration using blood samples from patients with MD (detailed in Chapter 04) in the existing literature, we initiated a systematic review of the literature on the use of multimodal integrative approaches on MDs. To do so, we followed the PRISMA (Page et al., 2021) guidelines for writing and reading systematic reviews and meta-analyses. This bibliographical snapshot of the PubMed database was captured on May 27, 2021, and reflects the literature at that time. We have restricted our search to studies using human biological samples only, and have excluded articles focusing on sets of target variables. Thus, we retained only those studies with genome-wide dimensions in their designs. As multi-omic approaches were the main focus

of this review, we were only interested in studies that investigated at least 2 datasets of the same omic type across different cohorts (horizontal integration), or at least 2 types of omic data within a single cohort (Vertical Integration).

2.2.2 Review Paper

In the following review, I provide a brief introduction to analytical notions that are further detailed in section 3.1.4, and elaborate on the application of multi-omic to MD.

Progress in Neuropsychopharmacology & Biological Psychiatry 116 (2022) 110520

Contents lists available at ScienceDirect

Progress in Neuropsychopharmacology & Biological Psychiatry

journal homepage: www.elsevier.com/locate/pnp

The molecular pathophysiology of mood disorders: From the analysis of single molecular layers to multi-omic integration

Amazigh Mokhtari [a], Baptiste Porte [a], Raoul Belzeaux [b,c,d], Bruno Etain [e,f], El Cherif Ibrahim [b], Cynthia Marie-Claire [f], Pierre-Eric Lutz [g,h,1,**], Andrée Delahaye-Duriez [a,i,j,*,1]

[a] NeuroDiderot, Inserm U1141, Université de Paris, F-75019 Paris, France
[b] Aix Marseille Université CNRS, Institut de Neurosciences de la Timone, F-13005 Marseille, France
[c] Fondation FondaMental, F-94000 Créteil, France
[d] Assistance Publique Hôpitaux de Marseille, Pôle de psychiatrie, pédopsychiatrie et addictologie, F-13005 Marseille, France
[e] Assistance Publique des Hôpitaux de Paris, GHU Lariboisière-Saint Louis-Fernand Widal, DMU Neurosciences, Département de psychiatrie et de Médecine Addictologique, F-75010 Paris, France
[f] Université de Paris, INSERM UMR-S 1144, Optimisation thérapeutique en neuropsychopharmacologie, OTeN, F-75006 Paris, France
[g] Centre National de la Recherche Scientifique, Université de Strasbourg, Fédération de Médecine Translationnelle de Strasbourg, Institut des Neurosciences Cellulaires et Intégratives UPR3212, F-67000 Strasbourg, France
[h] Douglas Mental Health University Institute, McGill University, QC H4H 1R3 Montréal, Canada
[i] Assistance Publique des Hôpitaux de Paris, Unité de médecine génomique, Département BioPhaReS, Hôpital Jean Verdier, Hôpitaux Universitaires de Paris Seine Saint Denis, F-93140 Bondy, France
[j] Université Sorbonne Paris Nord, F-93000 Bobigny, France

ARTICLE INFO

Keywords:
Mood disorders
Bipolar disorder
Major depressive disorder
Multi-omics
Genomics
Transcriptomics
Epigenetics

ABSTRACT

Next-generation sequencing now enables the rapid and affordable production of reliable biological data at multiple molecular levels, collectively referred to as "omics". To maximize the potential for discovery, computational biologists have created and adapted integrative multi-omic analytical methods. When applied to diseases with traceable pathophysiology such as cancer, these new algorithms and statistical approaches have enabled the discovery of clinically relevant molecular mechanisms and biomarkers. In contrast, these methods have been much less applied to the field of molecular psychiatry, although diagnostic and prognostic biomarkers are similarly needed. In the present review, we first briefly summarize main findings from two decades of studies that investigated single molecular processes in relation to mood disorders. Then, we conduct a systematic review of multi-omic strategies that have been proposed and used more recently. We also list databases and types of data available to researchers for future work. Finally, we present the newest methodologies that have been employed for multi-omics integration in other medical fields, and discuss their potential for molecular psychiatry studies.

1. Introduction

Mood disorders are defined in the 5th Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) as a group of psychiatric disorders characterized by the recurrence of clinically significant changes in mood state, energy, cognitive processes, sleep, or appetite. Among this group, Bipolar Disorder (BD) and Major Depressive Disorder (MDD) are the most prevailing syndromes in the general population. Indeed, BD affects around 1% of the global population (Ferrari, 2015), whereas MDD reaches a 12-months prevalence of 10.4% and a lifetime prevalence of up to 20.6% (Hasin et al., 2018). Mood disorders impose a substantial burden on patients and caregivers due to their recurrence and the frequently associated psychiatric comorbidities, including substance use, anxiety disorders, and suicidal behaviors (Isometsä, 2014). Consequently, they represent an economic and public health issue that needs to be addressed. Unfortunately, their heterogeneous etiology, ranging from genetic predisposition to environmental factors, complicates understanding underlying biological mechanisms. Furthermore,

* Correspondence to: A. Delahaye-Duriez, NeuroDiderot, Inserm U1141, Hôpital Robert Debré, 48 boulevard Sérurier, 75019 Paris, France..
** Correspondence to: P. –E. Lutz, INCI UPR3212, 8 allée du général Rouvillois, 67000 Strasbourg, France,.
E-mail addresses: pierreeric.lutz@gmail.com (P.-E. Lutz), andree.delahaye@inserm.fr (A. Delahaye-Duriez).
[1] These authors share senior authorship

https://doi.org/10.1016/j.pnpbp.2022.110520
Received 7 October 2021; Received in revised form 22 January 2022; Accepted 22 January 2022
Available online 30 January 2022
0278-5846/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the development of prognostic and diagnostic biomarkers in psychiatry is still at its debut, and the need for personalized approaches to improve therapeutic outcomes is rising (Nemeroff, 2020).

Over the last decade, researchers have turned to high-throughput technologies to try and bridge the gap between the lack of biological insights into mood disorders and their worldwide impact. Indeed, the emergence of Next Generation Sequencing (NGS) tools and their financial and technological accessibility has enabled the production of petabytes of biological data at different molecular levels, thereby accelerating the shift from a monogenic paradigm to a holistic view, where numerous molecular regulatory mechanisms interact to drive phenotypes. The relationship between these different omic layers turns out, in most cases, to be non-linear and to involve intricated mechanisms, such as transcriptomic adaptations under epigenetic regulation. In this context, multi-omic data integration is proposed as a set of tools to identify the main processes guiding these interactions and test hypotheses on molecular etiologies. This approach, which takes full advantage of multidimensional databases, creates new research avenues to explore and new challenges to overcome. Applied to psychiatry, it is expected to help refine diagnosis and patient stratification and improve therapeutic decisions (Fig. 01).

The present systematic review aims to describe the contribution of omics data integration to psychiatry, focusing on mood disorders. Omics data integration can be classified into two main categories, as illustrated in Fig. 02:

i) **Horizontal integration (indirect):** This "meta-analysis" strategy combines datasets related to the same molecular layer, generated on distinct clinical cohorts but with a given phenotype as a common denominator.
ii) **Vertical integration (direct):** This approach associates data on multiple layers generated from the same cohort of individuals, allowing a more comprehensive view of the different layers' interactions and identifying omics-phenotype causal relationships with higher statistical power.

In the first part of our systematic literature review, we first briefly summarize the most significant contributions and recommend dedicated reviews on mood disorders studies based on unique or horizontally (indirect) integrated single omic data. Then, we present our systematic literature search focused on identifying studies that combined at least 2 omic approaches (as defined in the Methods section below; see Fig. 03), to assess which strategies have been prioritized for data integration in the field, and what results were obtained when compared to previous single-layer approaches. Finally, we discuss how similar endeavors are ongoing in other medical areas and may guide future developments in molecular psychiatry.

2. Materials and methods

2.1. Search strategy

A systematic literature search was performed using the MEDLINE, Crossref, and Scopus databases to identify research and review articles published between January 1st, 2000 and May 15, 2021. Selected articles had to include one of the following terms: "Genome-wide association study (GWAS)", "RNA-Seq","Transcriptomics", "Transcriptome", "Epigenetics", "miRNA", "DNA Methylation", "Histone methylation", "Chromatin Conformation," and "Proteomics." These terms were associated with either "Bipolar Disorder" or "Major Depressive Disorder". Furthermore, the term "Multiomics" was also associated with each of the following: "Bipolar Disorder", "Major Depressive Disorder", and "Psychiatry". Papers resulting from these queries were investigated according to the Preferred Reporting Items for Systematic Reviews (PRISMA) protocol (Moher et al., 2009), as shown in Fig. 03.

2.2. Inclusion and exclusion criteria

Papers investigating BP or MDD patients and integrating more than two datasets either horizontally (meta-analyses of a single type of omics data – single-omics) or vertically (multi-omics) were included. Publications that did not include patients with either BD or MDD in their discovery cohort were excluded. No other specific study design was required for inclusion in our corpus, but a description of samples (biological origin, size, and diagnosis) was necessary. Research on non-human species was considered out-of-scope and excluded. Regarding multi-omics studies, we considered relevant all genome-wide studies that conducted an integration of more than one "omic" level.

3. Results

Below, we present the meta-analyses for patients with mood disorders identified during the systematic search summarizing the main findings from each single omic approach.

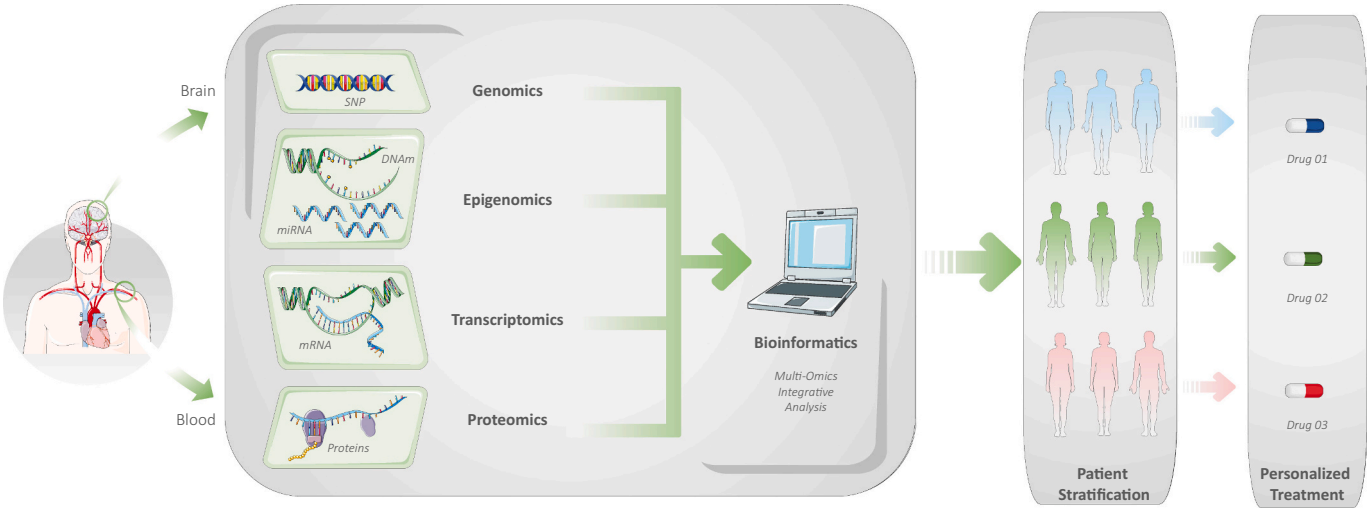


Fig. 01. Illustration of the concept of multi-omics data integration enhancing patient stratification and precision medicine. Different omics data-levels are collected from the brain and peripheral tissues which are then integrated using computational biology tools, providing complementary information on the studied diseases resulting in better patient stratification and adapted treatments.

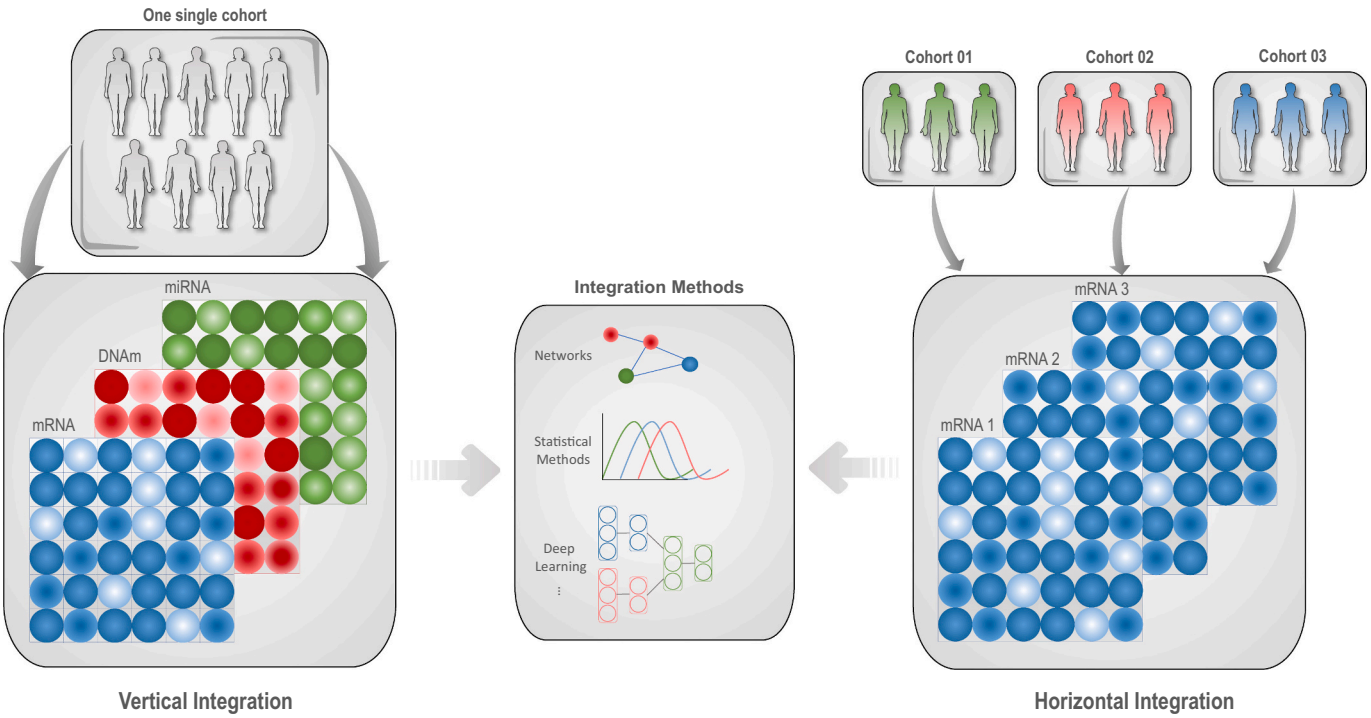


Fig. 02. Data integration approaches. **a.** Horizontal integration: the same molecular data-layer is retrieved from different cohorts. **b.** Vertical integration: multiple omics data are available from the same cohort.

3.1. Horizontally integrated single-omic studies of mood disorders

3.1.1. Genomics

GWAS meta-analyses are commonly applied to capture the shared genetic background of mood disorders. Among the 20 eligible studies listed in Supplementary Table 1, eight focused solely on BD, ten on MDD, one on both BD and MDD, while the last one covered a broader spectrum (ASD, ADHD, BP, MDD, and SCZ). It is worth noting that the latest published GWAS add new subjects to previous ones, which explains their concordance. Regarding BD-associated variants, Stahl et al. (2019) highlighted 30 loci, among which loci within the *ANK3* gene were already identified several times in earlier studies. *ANK3* encodes ankyrin-G, a protein associated with the cell membrane, notably present at nodes of Ranvier and the initial axon segment. In addition, the *ODZ4* gene encoding a transmembrane protein which plays a role in neural development (Ikeda et al., 2018; Psychiatric GWAS Consortium Bipolar, 2011; Mühleisen et al., 2014), and *POU3F2* a DNA-binding transcription factor acting notably in neuronal differentiation (Hou et al., 2016; Mühleisen et al., 2014; Stahl et al., 2019), were also pointed out in several coupled BD GWAS.

Similarly, for MDD, successive GWAS aggregating depressive cohorts identified specific loci, including *DCC*, which encodes a functional transmembrane receptor for NETRIN-1 and mediates neurite outgrowth (Howard et al., 2019; Okbay et al., 2016; Wray et al., 2018); *SORCS3*, a cell surface receptor localized at the postsynaptic density of hippocampal neurons (Howard et al., 2019; Hyde et al., 2016; Wray et al., 2018), both *SORCS3* and *DCC* genes significance associations were replicated in the Coleman et al., 2020's mood disorder GWAS; and *NEGR1,* a cell adhesion molecule promoting neuronal spine plasticity (Li et al., 2018). It is worth mentioning that only the *MAD1L1* gene, which plays a role in the cell cycle, was significantly associated in multiple studies with both MDD and BD (Hou et al., 2016; Howard et al., 2019; Howard et al., 2018; Ikeda et al., 2018). Recently, the *MAD1L1* locus has been significantly associated with anxiety (Levey et al., 2020) and found differentially methylated in post-traumatic stress disorder (Snijders et al., 2020). However, this association was not replicated in a recent mood cohort GWAS meta-analysis (, Coleman et al., 2020). The lack of other loci associated with both MDD and BD, despite common symptoms, likely reflects their heterogeneous and complex pathophysiology. Some authors suggested that inconsistent patient phenotyping may also partly explain this low overlap (Ormel, 2019). Of note, copy number variants (CNVS), single nucleotide variants (SNVs), and other genomics polymorphisms were investigated in MD and are not discussed here further (see Gordovez and McMahon, 2020 and McIntosh et al., 2019 for reviews). Even though these extensive analyses may facilitate identifying genomic variants associated with mood disorders, the post-single-omic GWAS challenges reside in prioritizing functional and causal variants and characterizing their biological mechanisms.

3.1.2. Transcriptomics

The advent of microarrays and NGS technologies has permitted transcriptomic genome-wide investigation of mood disorders. These studies, similar to the other types of functional omic analyses detailed further in this review, focuses mainly on two tissues: i) the brain (with different regions examined, and a gradual shift from whole-tissue to single cell-type, or even single-cell analyses), which is biologically more relevant but difficult to collect in large sample size cohorts, and is currently available for retrospective investigation only; and ii) peripheral tissues, such as blood or saliva, which are easier to collect and can be used in longitudinal designs and for biomarkers identification. The following paragraphs first describe studies on brain and blood separately before concluding on possible convergence between the two tissues.

Among the 4 reported meta-analyses of our systematic screening focusing on peripheral tissues Supplementary Table 2, Hess et al. (2019) have conducted a meta-analysis of seven studies, including BD ($n = 95$) and schizophrenia ($n = 258$) patients. They reported significant enrichments for both differentially expressed genes (DEG) and WGCNA co-expression modules in the immune system's biological function, oxidative stress, and apoptosis pathways. More recently, Wittenberg et al. (2020) reviewed whole-genome data generated from blood samples (whole blood, lymphocytes, or PBMC) in case-control studies (1754 depressed cases and 1145 healthy controls compiled from 10 studies).

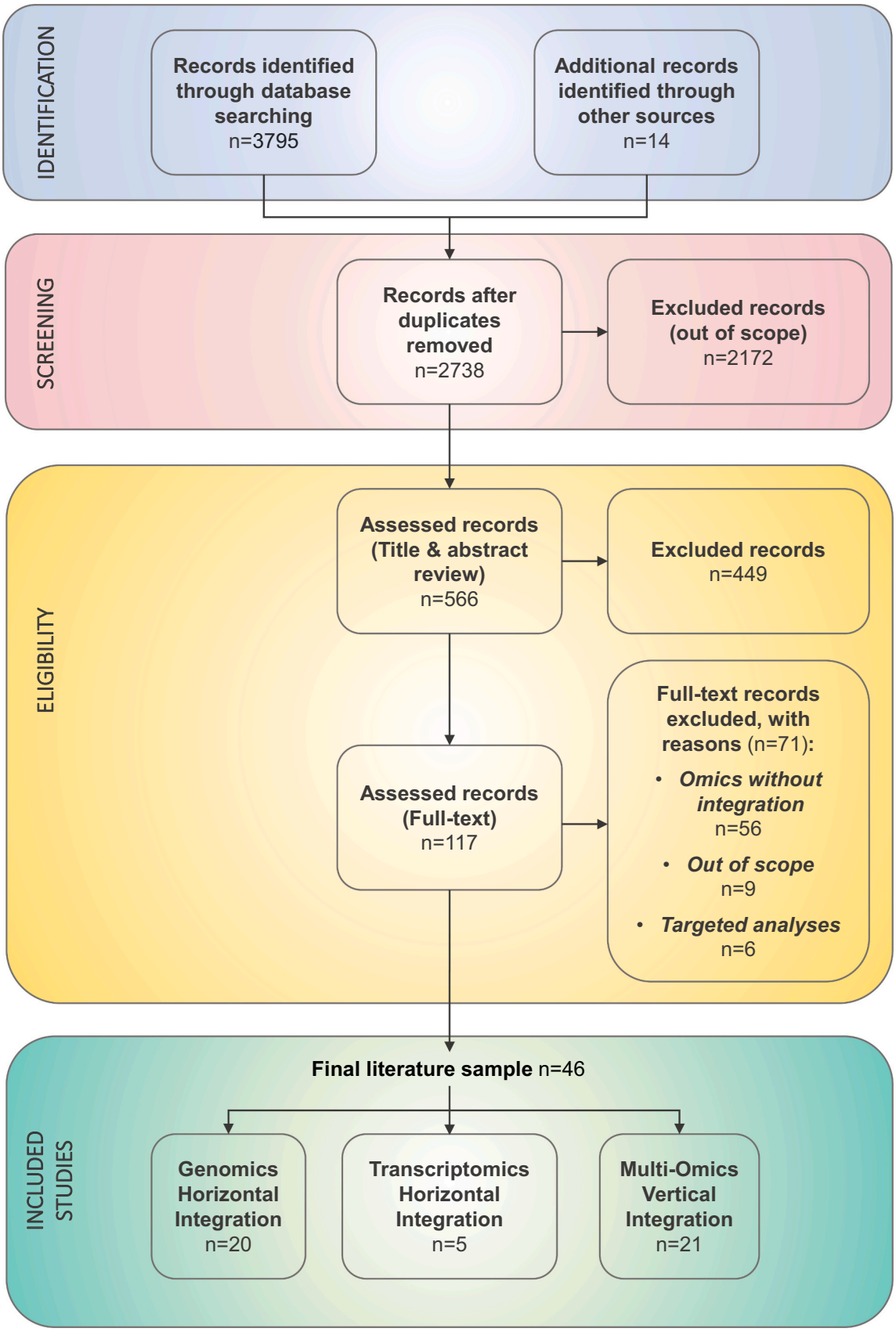


Fig. 03. The PRISMA flow diagram for the systematic review detailing the searched database, the number of abstracts screened and the full texts reviewed.

The meta-analysis was done at two levels: i) by compiling DEGs identified by at least 2 studies, and ii) by conducting a homogenized re-analysis of subject-level data, when such data were available ($n = 8$), or using standardized mean difference otherwise (whole blood studies, $n = 4$). The three approaches highlighted genes associated with the immune response system, especially neutrophil activation. Noteworthy, they also explored sample type (whole blood/PBMC) and data heterogeneity issues with respect to such meta-analysis. Results showed that MDD-related adaptations were more consistent within (e.g. whole blood vs. whole blood) than across (whole blood vs. PBMC) sample type. Importantly, findings were also more concordant among whole blood than among PBMC studies, indicating that neutrophils (not present in the PBMC fraction) and related GO terms exhibit a more robust association with MDD and highlighting the importance of investigating relevant cell types.

Regarding brain tissue-based meta-analyses, Seney et al. (2018) were interested in opposite transcriptional alterations in men and women. They focused on three brain regions (dorsolateral prefrontal cortex, subgenual anterior cingulate cortex, and basolateral amygdala), with a total cohort of 26 MDD cases and 13 controls in men, and 24 MDD cases and 9 controls in women. The authors reported sex-specific changes in MDD, with relatively few overlapping DEG. The altered pathways included opposite transcription gene variations with a decrease in synapse-related genes in male MDD and an increase in female MDD.

In addition to Ciobanu et al. (2016)'s review of replicated finding, only one cross-tissue investigative meta-analysis have been published (Forero et al., 2017). First, in their review, Ciobanu et al. (2016) included transcriptomic datasets obtained from peripheral ($n = 10$ articles) and brain ($n = 15$ articles) tissues in MDD covering adrenal gland, colon transverse, and whole blood along with ten brain structures (dorsal prefrontal cortex, hippocampus, or cerebellum as main examples). The authors reported 57 replicated DEGs in the brain and 21 in peripheral tissues. Their findings suggest more significant overlaps and greater homogeneity among brain regions than peripheral tissues. Second, Forero et al. (2017) investigated brain and blood transcriptomics profiles in MDD. This meta-analysis was of a larger scale than the previous one, as 24 primary datasets were used: 6 from blood, 18 from 4 brain regions (amygdala, cerebellum, prefrontal, and anterior cingulate cortices). The authors identified several enrichment pathways involving synaptic plasticity in the brain datasets but, surprisingly, no significant pathway enrichment in blood.

Altogether, even if MDD-related DEGs vary according to tissue type (blood/brain), results from pathways enrichment appear relatively convergent towards an implication of groups of genes associated with the immune system and/or inflammation. But, one should be concerned that blood-based analyses can be less informative than investigations based on brain tissue, where mood disorders pathophysiological mechanisms primarily occur.

3.1.3. Epigenetics

The mood disorders heritability has been estimated via monozygotic twin studies between 31% to 42% for MDD (Flint and Kendler, 2014), and between 70% to 90% for BD (Gordovez and McMahon, 2020). Despite these significant genetic contributions, it is widely acknowledged that mood disorders and their clinical course involve additional factors, including environmental ones and life events, such as childhood trauma (Aldinger and Schulze, 2017). While mechanisms potentially mediating interactions between genetic determinism and environmental exposure have long remained elusive, their importance is not to neglect. Recently, convincing illustrations of the relevance of G x E interactions have been published in relation to epigenetic consequences of prenatal or early-life factors, which act as distal risk factors for mood disorders (Czamara et al., 2021; Czamara et al., 2019).

Epigenetics refers to physical and chemical substrates that regulate the architecture and activity of the genome without any change in the underlying DNA sequence. They include chromatin conformation, DNA methylation, histone modifications, miRNAs, and other non-coding RNAs (Penner-Goeke and Binder, 2019). In their critical review, Legrand et al. (2021) questioned the epigenetics of BD, retracing the evolution of this emerging field and significant discoveries. Here, we focus on whole-genome epigenetic approaches, mostly related to DNA methylation and miRNAs in the current literature.

3.1.3.1. DNA methylation. DNA methylation (DNAme) is the addition of a methyl group at the 5′ position of cytosine. The modification influences gene expression through multiple mechanisms (Jones, 2012) and globally acts as a repressor on gene expression. Accumulating evidence indicates that DNAme contributes to gene x environment interactions, including for monozygotic twins discordant for psychiatric phenotypes (van Dongen et al., 2014), or following severe traumatic life experiences (Wolf et al., 2018). While initial DNAme studies focused on candidate genes, like brain-derived neurotrophic factor (BDNF) or the glucocorticoid receptor, advance in NGS technologies (MBD-seq, MeDIP-Seq, RRBS, WGBS) and the development of relatively affordable microarrays (450 k and more recently Infinium EPIC arrays), now allow conceiving genome-wide case/control analyses to identify differentially methylated regions (DMRs) or perform MWAS (methylome-wide association study).

As observed for transcriptomics studies, the tissues characterized in mood disorders DNAme studies vary widely, including studies investigating different brain regions and others focusing on peripheral tissues (whole blood). For example, for genome-wide MDD investigations, Li et al. (2019) outlined genes related to neurogenesis, neuroplasticity, and estrogen signaling regardless of the sample's tissue origin. On the other hand, for BD DNAme analyses, Teroganova et al. (2016) highlighted differentially methylated genes in peripheral tissues of BD and schizophrenia patients, including well-known BDNF promoter regions involved in neuroplasticity and dopaminergic pathways.

Of note, Li et al. (2019) underlined a limited number of systematic reviews and meta-analyses on DNAme and MDD, and we observed a similar issue for BD. This lack of quantitative reviews is even more pronounced for genome-wide studies. Thus, we notice an urgent need for genome-wide meta-analyses of DNAme in mood disorders.

3.1.3.2. Micro-RNAs. Recently, non-coding RNAs and particularly micro-RNAs (miRNAs), have become the focus of significant attention. It is now acknowledged that they play an extensive role in regulating gene expression (He and Hannon, 2004). Their canonical mode of action involves silencing mRNAs with complementary target sequences. A miRNA can act on numerous mRNAs, and reciprocally, an mRNA is regulated by various miRNAs. Many public repositories list predicted and experimentally validated targets of miRNAs, such as MiRBase (Kozomara et al., 2019).

Ferrúa et al. (2019) focused on bioinformatic analysis of peripheral miRNAs in patients with depressive symptoms. This systematic review emphasized *AKT, BRAF,* and *PIK3CA* as the target of many differentially expressed miRNAs. The study also inferred potentially affected pathways such as mTOR, phosphoinositide 3-kinase PI3K/Akt, and MAPK signaling pathways.

These pathway enrichment analyses of the miRNA targeted genes concur with those uncovered at other molecular levels mentioned above. Furthermore, miRNAs were also identified as potential biomarkers of BD, schizophrenia (Amoah et al., 2019; Fries et al., 2018), MDD (Gururajan et al., 2016), suicidal behavior in BD patients (Squassina, 2020), and nausea intensity in MDD patients treated with mood stabilizers (Yrondi et al., 2020). Altogether, these results suggest that miRNAs may be potential diagnostic and predictive biomarkers in psychiatric disease (Gibbons et al., 2020).

3.1.3.3. Metabolomics. Metabolomics in mood disorders is an emerging field of research. This omics aims to depict the state of the metabolism's

end products (metabolites) and establish metabolic profiles for various pathways, including carbohydrate, protein, and fatty acid metabolism. MacDonald (2019) performed an extensive systematic review on metabolomics biomarkers of mood disorders, looking at 266 studies measuring metabolites using different techniques: in vivo brain imaging, chromatography, nuclear magnetic resonance, and mass spectrometry. Among the metabolites identified, 122 were reported in two or more BD or MDD studies, and involved mainly in neurotransmission and energy metabolism. Recently, Bot et al. (2020) investigated the metabolic profile of a large cohort of patients with depressive symptoms ($N = 5283$ MDD and 10,145 Controls) in a meta-analysis. Of the 21 metabolites that met the significance criteria (FDR < 0.05), higher blood levels of very-low-density lipoprotein triglycerides, and lower concentration of high-density lipoproteins were related to depression and were consistent with previous findings.

Despite the growing appeal towards metabolomic-based biomarkers in psychiatric diseases, few meta-analyses are to be reported, especially for BD, for which no specific meta-analysis seems to be available. In the same way, no meta-analysis or systematic review on mood disorders and proteomics was identified during the systematic search.

3.1.3.4. Metagenomics. Lately, the influence of the gut microbiota on mood disorders has attracted increasing attention, with the hope that related studies may lead to innovative treatments and diagnosis tools. Despite these promises, a lack of functional and publicly available data has to be deplored. Indeed, the only comprehensive meta-analysis on gut microbiota and depression available at the time of literature screening, from Sanada et al. (2020), focused solely on the gut microbiota composition, with no functional data available, therefore falling outside the scope of the present review.

3.2. Multi-omics integration methods

Understanding complex pathologies such as mood disorders require a holistic vision. These diseases result from an intricate interplay between different actors (genomics, transcriptomics, epigenetics, proteomics, and metabolomics) and environmental factors that regulate the neuronal mechanism underlying cognitive and emotional processes. Although single-omic approaches have brought a better understanding of psychiatric illness, it is now proposed that combining multiple types of data through systems biology may harness further progress. Therefore, in the second part of this review, we also provide a systematic review of the application of multi-omics approaches to mood disorders. Several reviews described multi-omics data integration methodologies and their mathematical aspects. Available approaches can be categorized differently according to various criteria: horizontal, vertical or neither vertical nor horizontal integration methods as illustrated by Eidem et al. (2018). The machine learning criteria distinguishing unsupervised, semi-supervised or supervised approaches are also widely used (Bersanelli et al., 2016; Huang et al., 2017; Wu et al., 2019). This distinction reflects, for example, whether one proceeds in an exploratory manner or applies clinical labels to individual cases. Another categorization criteria distinguish early, intermediate, and late integration approaches (Cantini et al., 2021; Rappoport and Shamir, 2018). For example, step by step combination of results obtained from single omics can be processed concatenating omics before clustering with an "early" integration approach or after clustering using a "late" integration approach (Rappoport and Shamir, 2018). Genuine "intermediate" joint multi-omics integrative methods remain rarer and constitute a highly active area of research (Cantini et al., 2021). Below, we first list the most commonly used methods (with concrete examples from other research fields), and then detail their application to mood disorders.

3.2.1. Step by step methods

These methods simplify data integration by finding correlations and overlaps after analyzing each data type separately while respecting its specificities. Ready-to-use tools or pipelines are not available for this kind of approach. An example of a typical pipeline begins with a differential analysis carried out on each omic, followed by integration via overlaps (identifying entities commonly dysregulated across distinct layers) or correlation (to assess the covariance of omics quantitative data). A major drawback of these methods lies in their over-simplification of the problematic as, among others, known biological interactions between omic types are not taken into account. As the full range of information extractable from the data is not exploited, only strong effects can be unveiled, potentially missing the subtle interactions at stake in psychiatric disease.

3.2.2. Network-based methods

Network theory is widely used at the single-omic level, notably with gene-gene expression correlation methods. Graph theory and network methods are often used in systems biology, and can also help characterize the interconnectivity between multi-omic layers. Various tools have been developed based on the adaptability of network modeling methodologies. Some rely on pre-existing knowledge, such as the Protein-Protein Interaction networks (PPI), as an integration guide. For example, iOmicsPASS has been applied to breast cancer data and revealed a molecular signature for the basal-like subtype that could not be identified using single-omic analyses (Koh et al., 2019). Other tools explore the inherent characteristics of the generated networks, such as their topology features (network connectivity patterns; node degree: number of connections with other nodes; and centrality: number of times a node acts as a bridge between pairs of nodes in the network). Among network-based approaches, similarity network fusion (SNF) was able to identify differential survival profiles among subtypes of 5 cancer types from the TCGA research network, allowing a better patient stratification (Wang et al., 2014).

3.2.3. Dimension reduction methods

Dimensionality reduction methods for the joint analysis of multi-omics datasets were reviewed and benchmarked using simulated data, cancer datasets, and, more recently, single-cell datasets (Cantini et al., 2021; Tini et al., 2018; Wu et al., 2019). Matrix decomposition algorithms offer an answer to the multidimensionality issue of NGS data (a much higher number of features than samples). In order to extract the nature of the data variance, dimension reduction methods aim at decomposing a matrix of N samples by M measures into a matrix product of smaller matrices representing the components that best summarize the data. The most commonly used algorithms include Principal Component Analysis (PCA), and Canonical Correlation Analysis (CCA). Among other examples, the DIABLO method of the mixOmics R package (Rohart et al., 2017) relies on Partial Least Squares (PLS) regression and proposes to integrate several omic matrices by maximizing correlations between the extracted components and phenotypic variables. This package also encapsulates all the algorithms mentioned above (PCA, CCA and PLS) and their variants and has been used in numerous multi-omic studies. Matrix factorization methods are another toolset to circumvent the growing size of data matrices, of which JIVE (Joint and Individual Variation Explained) (Lock et al., 2013) is a popular multi-channel integration method. JIVE decomposes the matrices into a combination of matrix pairs representing each omic's specific and common variation. It has been applied to glioblastoma to uncover miRNA-mRNA interactions and improve the stratification of tumor subtypes (Lock et al., 2013).

3.2.4. Bayesian methods

One major advantage of Bayesian models is the possibility of enhancing the statistical model with a priori information. Thus, these methods are particularly interesting when one wishes to assess the probability of a sample clustering assignment. However, these models often require proper attribution of a prior probability distribution for

each omic type, based on one or more parameters, and a posterior probability distribution of expected results to identify the clustering pattern (Bersanelli et al., 2016). In addition, these models are costly in terms of computing resources. Among these, iRIGS is a framework adapted to identify genes associated with the risk of developing schizophrenia, using GWAS results combined with a gene-gene network integrating multi-omic data (differential expression, de novo mutations, chromatin states (Wang et al., 2019a).

3.2.5. Advanced machine learning

Machine learning and artificial intelligence (AI) have seen remarkable advances over the last decade. The complexity and quantity of NGS data are ideally suited for machine and deep learning algorithms, as these methods rely on extensive data as training sets to improve from experience. Multi-omic integration can be described as a variant of a known AI clustering problem named multi-view clustering. iClusterPlus (Mo et al., 2018), an improvement of the iCluster method (Shen et al., 2009), is among the most commonly used methods. It relies on a generalized linear regression to formulate a joint model, uncover insights on mechanisms driving the phenotype of interest, and was initially applied to colorectal cancer to reveal tumor subtypes.

Further, iClusterBayes (Mo et al., 2018) is a Bayesian computational and statistical enhancement of iClusterPlus, that has been applied to glioblastoma data generated by the TCGA Research Network. Rappoport and Shamir (2018) benchmarked iClusterBayes and other machine learning methods, including those cited previously (JIVE, CCA, and SNF), on 10 cancer datasets. They concluded that no single method systematically outperformed the others, and advocated for their careful use to analyze omics data, as these methods largely disregard their specificities (inter-omic dependency).

3.3. Multi-omic integration applied to mood disorders

Here, we systematically identified and reviewed 21 multi-omic publications described in detail in Table 1. Several factors likely account for this relatively modest number of papers, among which the relative youth of the field and the low amounts of publicly available data, notably for brain tissue. In addition, the majority of these articles have used step-by-step approaches due to the current absence of state-of-the-art methods. Below, we summarize their main findings, subjectively organized from simplest to most advanced approaches.

3.3.1. Step-by-step approach – overlap

Zhao et al. (2015) investigated the relation between DNAme and the transcriptome (RNA-seq) in BA9 tissue from a cohort of 5 SCZ, 7 BD patients, and 6 controls. They found that 11 of the 146 genes with hypermethylation in the promoter region showed a down-regulation at the expression level, and 4 of the 411 hypomethylated showed an up-regulation, indicating that a vast majority of DMRs had no detectable effect on gene expression, at least when analyzing such a small sample size. A high percentage of intronic DMR overlapped with miRNAs that were predicted to target genes identified as differentially expressed, indicating that DNAme may alter gene expression in part via intronic miRNAs. The down-regulated genes showed significant enrichment in neurogenesis and nervous system development. Ju et al. (2019) also overlapped DNAme and gene expression analyses. They investigated molecular predictors of response to antidepressant treatment in peripheral blood samples from 177 MDD (comparing 82 responders to escitalopram and 95 non-responders, sampled before treatment initiation – a predictive approach) and 102 healthy controls. They identified 303 differentially methylated CG sites, among which, again, only a minority ($n = 16$) were located within DEGs (with *CHN2* and *JAK2* the 2 most significant DEGs, involved respectively in neurodevelopmental hippocampal axon pruning and synaptic plasticity). Recently Xie (2021) reprocessed available mRNA and DNAme data from studies by Leday et al. (2018) and Crawford et al. (2018). The analysis included 128 MDD and 64 healthy controls' transcriptomic data and methylomic data for 100 distinct MDD patients and 50 controls. The overlap study identified: 46 hypomethylated and up-regulated genes involved in PI3K-Akt, IL-17, and axon guidance signaling pathways, among others; and 71 hypermethylated and down-regulated genes involved in the MAPK and NF-kappa B signaling pathways. Furthermore, from these overlaps, the authors also developed a random forest-based classifier to discriminate MDD cases and controls, using mRNA and DNAme data separately; comparison of the 2 classifiers showed a better predictive power for the one based on gene expression (AUC > 0.95).

3.3.2. Step-by-step approach – correlation

The integration of different molecular layers by correlation testing has been used mainly to assess relationships between DNAme and gene expression in mood disorders. Chen et al. (2014) used Pearson correlation to identify candidate genes with differential methylation and expression patterns in cerebellar samples from 39 SCZ and 36 BD patients, as well as 43 controls. Among the 20 genes where 204 significantly differentially methylated CpGs were identified, four (*PIK3R1*, *BTN3A3*, *NHLH1*, and *SLC16A7*) were differentially expressed. Fries et al. (2017) combined correlation and pathway analyses to identify molecular differences between PBMCs from 6 young BD patients, 6 unaffected subjects at high-risk of BD, and 6 healthy controls, using DNAme and transcriptomics. Results identified 135 genes with a significant correlation between gene expression and DNAme, while Ingenuity pathway analysis on the 43 genes either differentially expressed ($n = 33$) or methylated ($n = 10$) at individual level implicated circadian rhythms, immune system, synaptic scaffolding, and glucocorticoid receptor signaling. Abdolmaleky et al. (2019) used Spearman correlation to analyze transcriptomic (10 BD, 10 SCZ, and 10 healthy control) and DNAme (27 k CpG sites analyzed in 3–4 individuals from each group) brain data. Among other findings, an anti-correlation between DNAme (increased) and gene expression (decreased) was observed for the *CCND1* locus in SCZ and BD patients. Zhu et al. (2019) used both partial correlation and an overlap test to identify methylation differences associated with changes in gene expression in circulating monocytes in 79 monozygotic twin pairs discordant for MDD. Results indicated that differentially methylated genes were significantly overrepresented in previous GWAS MDD loci (2.32 times, P value $= 2.4 \times 10^{-4}$) and, interestingly, were more likely to be differentially expressed (2.44 times, P value $= 1.1 \times 10^{-4}$). Correlations were also assessed between miRNA and mRNA data from patients with mood disorders. Pisanu et al. (2019) tested for, and identified, negative correlations between miRNA expression and differentially expressed mRNAs in 10 responders and 10 non-responders to lithium BD patients. This resulted in correlations among 30 miRNAs and 277 mRNAs, with one downregulated and one up-regulated sub-cluster network centered on miR-320a and miR-155-3p, respectively, which are involved in the inflammatory response and the regulation of G-protein signaling. Chen et al. (2018) applied correlation analysis to identify miRNA/mRNA co-expressed modules using WGCNA and sequencing data generated using brain tissue from a larger cohort composed of 95 patients with SCZ, 74 with BD, and 225 healthy individuals. They identified a disease-associated module (daM) enriched for rare and de novo variants (but not for GWAS loci), which included 545 mRNAs and 5 miRNAs (miR-320b, miR-320c, miR-320d, miR-320e and miR-585). Among this module's hub genes were 6 transcription factors, including *POU3F2,* which plays a role in brain development and was identified as a key regulator.

3.3.3. Step-by-step approach – regression

Regression approaches were applied several times to mood disorders, often determining the relationship between genomic variations and other molecular layers. Mehta et al. (2014) used a regression approach to uncover 293 eCNVRs (expression-influencing copy number variation regions) that significantly influenced 429 unique transcripts in prefrontal cortex tissue from 35 controls, 34 BD, and 35 SCZ patients, with

Table 1
Selected multi-omic studies applied to mood disorders.

Paper	RNA	DNAme	miR	SNP	ChIP	Cohort	Biological samples	Integration method	Covariate correction	Relevent results	Code available
Abdolmaleky et al., 2019	x	x				mRNA: 10 BD, 10 SCZ, 10 HC DNAm: 4 BD, 4 HC, 4 SCZ	Brain (BA46, DLPFC)	Spearman correlation	Matching samples	Anti-correlation between DNAm and expression of the *CCND1* locus	No
Aberg et al., 2018		x		x		812 MDD, 320 HC	Whole blood	Regression analysis	Quantity of DNAm captured, batches, peak location, age, gender & blood cell composition	23 CpG-SNPs with case/control methylation effect involved in neural function	No
Bhak et al., 2019	x	x				56 MDD SAs, 39 MDD, 87 HC	Whole blood	Random forest*	Not mentionned	Suicidal vs non suicidal MDDs classifer with an accuracy accuracy of 92.6%	Uppon request
Belzeaux et al., 2019	x		x			237 MDD	Whole blood	Multiple logistic regression	Not mentionned	Combined MLRA of treatment-worsening suicidal ideation including miR-5695, *STMN1* mRNA, and *MADRS*	No
Chen et al., 2014	x	x				36 BD,39 SCZ, 43 HC	Brain	Pearson's correlation & linear regression	Age, gender, brain pH & PMI	*PIK3R*, *BTN3A3*, *NHLH1* and *SLC16A7* among 20 significantly correlated CpGs / gene expression showed correlated DEGs and DMGs patterns in psychotic patients brains	No
Chen et al., 2018	x		x	x		26 BD,25 SCZ, 24 HC	Brain	Correlation	Gender age, brain pH, RIN & PMI	SCZ/BD associated module enriched for genes implicated in rare and de novo variants, and for miRNA targets with key regulators including *POU3F2*, *SOX9*, *PAX6*, *SOX5*, *EPAS1*, and *ZNF423*	No
Ciuculete et al., 2020	x	x		x		mRNA: 15 MDD, 15 HC DNAm & meQTL: 23 MDD, 36 HC	Brain & whole blood	Regression analysis, Spearman correlation	Blood cell composition & batches	Lower methylation at cg24627299 was observed in MDD patients, associated with expression levels of the MET ligand HGF in blood	No
Fries et al., 2017	x	x				6 BD, 6 BD HC offsprings, 6 HC	PBMCs	Pathway analysis & Pearson's correlation	Matching samples	42 combines DEGs and DMGS enriched in the circadian rhythm, immune system and synaptic scaffolding	No
Gandal et al., 2018	x			x		222 BD,51 ASD, 559 SCZ, 963 HC	Brain (frontal, temporal cortices)	Regression analysis	RIN, PMI & hidden factors	5 gene/isoform networks enriched in inflammatory and postsynaptic density/ receptor-mediated presynaptic signaling pathways as well as endothelial and pericyte genes	No
Ju et al., 2019	x	x				177 MDD, 102 HC	Whole blood	Overlap test	Age & gender	16 DMPs overlapping with *CHN2* and *JAK2* DEGs involved in the inflammatory response and neurodevelopment	No
Li et al., 2020	x			x		eQTL: $n = 2251$ GWAS: 246.363 MDD & 561.190 HC	Brain	SMR & Bayesian framework *	Diagnosis, gender, institution, AOD, PMI, RIN, Batches & hidden factors	*LRFN5* and *DCC* genes identified as depression risk genes involved in synaptic plasticity	No
Mehta et al., 2014	x			x		23 BD, 23 SCZ, 26 HC	Brain	GLM Likelihood ratio test	Age, gender, ethnicity & PMI	429 transcripts significantly associated with CNVR, enriched in the corticotrophin-releasing hormone signaling pathway	No
Pai et al., 2019	x	x		x		RNA seq:17 BD & SCZ, 17 HC EWAS & Genotypes: 26 BD, 29 SCZ, 27 HC	Brain: PFC neurons	Regression analysis	Age, gender, PMI & ancestry	Concordance between DEG and DMG and enrichments related to embryonic development, synaptic function and immune cell activation	https://github.com/shraddhapai/EpiPsychosis_IGF2

(*continued on next page*)

Table 1 (*continued*)

Paper	RNA	DNAme	miR	SNP	ChIP	Cohort	Biological samples	Integration method	Covariate correction	Relevent results	Code available
Pisanu et al., 2018	x			x		mRNA: 10 BD LR, 10 NR Genotype: 59 LR, 146 NR	LCL	Multiple linear principal components regression	Matching samples	*ZNF429* and *ZNF493* as potential lithium-responsive biomarkers	No
Pisanu et al., 2019	x		x			mRNA: 10 BD LR, 10 NR miRNA: 12 BD LR, 12 NR	LCL	Negative correlation between	Matching samples	A 30 miRNAs and 277 genes correlation network centered on miR-320a and miR-155-3p	No
Wang et al., 2018	x			x	x	H3K27ac: 117 SCZ, 109 HC. mRNA: 259 BD, 630 SCZ, 1094 HC SNP: 172 BD, 497 SCZ, 619 HC	Brain (DLPFC)	WGCNA & Deep Boltzmann Machine*	Matching samples between training and test datasets	An improved model linking regulatory network including immunological, synaptic, and metabolic pathways and higher psychiatric level phenotypes	https://github.com/gersteinlab/PsychENCODE-DSPN
Wang et al., 2019a	x			x		mRNA: 258 SCZ, 279 HC GWAS: 36.989 SCZ, 113.075 HC	Brain & Whole blood	Bayesian framework*	Not mentioned	104 high-confidence risk genes involved in neurogenesis	https://www.vumc.org/cgg/irigs
Wang et al., 2019b	x			x		23 BD, 24 SCZ, 24 HC	Brain	Vertical & horizontal approaches: regression model *	No correction	*MAPK1*, *YWHAE*, *TPH1* and *AKT1* as key features genes in BD and SCZ	https://github.com/shuanggema/VHintegr
Xie, 2021	x	x				mRNA: 128 MDD, 64 HC. DNAme: 100 MDD, 50 HC.	Whole Blood	Overlap	Batch effect	46 Hypomethylated and up-regulated genes involved in -among others- PI3K-Akt IL-17 and axon guidance signaling pathways, 71 Hypermethylated and down-regulated genes involved in the MAPK and NF-kappa B signaling pathways	No
Zhao et al., 2015	x	x				7 BD, 5 SCZ, 6 HC	Brain (BA9)	Overlap Test	Matching samples	7,5% of hypermethylated BD genes showed decreased expression and less than 1% of the hypomethylated ones showed and increase. DMR-related miRNAs down-regulated genes in neurogenesis and nervous system development	No
Zhu et al., 2019	x	x				72 MDD and 72 non MDD twins	PBMCs	Partial correlation	Blood cell composition estimation & batches	DMGs are 2.44 times more likely to be DEGs and are significantly enriched in previous MDD GWAS loci	No

Legends: AOD, alcohol and other drug; ASD, acute stress disorder; BA, Brodmann area; BD, bipolar disorder; CNVR copy number variation regions; DEG, differentially expressed gene; DLPFC, dorsolateral prefrontal cortex; DMG, differentially methylated gene; DMP, differentially methylated probe; DMR, differentially methylated region; DNAme, DNA methylation; eQTL, expression quantitative trait locus; EWAS, epigenome-wide association study; GLM, generalized linear model; GWAS, genome wide association study; HC, healthy controls; LCL, lymphoblastoid cell lines; LR, lithium responder; MDD, major depressive disorder; meQTL, DNA methylation quantitative trait locus; miR, micro RNA; NR, non responder; PBMC, peripheral blood mononuclear cells; PFC, prefrontal cortex; PMI, post mortem intervals; RIN, RNA integrity number; SA, suicidal attempt, SCZ, schizophrenia; SMR, summary data-based Mendelian randomization; SNP, single nucleotide polymorphism; WGCNA, weighted correlation network analysis. * Advanced integration approaches.

enrichment in the corticotrophin-releasing hormone signaling pathway. The PsychENCODE consortium generated brain RNA-seq and genotype data from a large cohort of 51 patients with autism, 559 SCZ, 222 BD, and 963 controls (Gandal et al., 2018). After a comprehensive investigation of this new transcriptome dataset (including alternative splicing, non-coding genes, and co-expression analyses), the authors combined transcriptome with genotype data to compute both gene- and isoform-level expression quantitative trait loci (eQTL). Expression prediction regression models were also tested by conducting a transcriptome-wide association study (TWAS). Taking advantage of these new brain eQTL and large publicly available GWAS in SCZ, autism, and BD, they identified genes whose expression, although not significantly different across disease groups, is regulated in *cis* by disease-associated variants. Pisanu et al. (2018) used a multiple linear principal components regression model to analyze gene expression changes induced by lithium chloride exposure in LCLs from 10 lithium responders (LR) and 10 non-responders (NR) BD patients, as well as genotype data of 56 LR and 141 NR patients. Among the 29 genes whose expression was significantly affected by lithium treatment, two (*ZNF429* and *ZNF493*) were also associated with lithium response using genotype data (analyzed at gene level). Regarding treatment response, regression methods were also applied by Belzeaux et al. (2019) to analyze mRNA, miRNA, and clinical data in relation to antidepressant treatment-worsening suicidal ideation (TWSI) in 237 MDD patients from a double-blinded trial. The authors compared multiple logistic regression models and found that the model that included miR-5695, *STMN1* mRNA, and the MADRS (Montgomery-Asberg Depression Rating Scale) score at baseline significantly outperformed those based on a single omic, or clinical data alone. Regression approaches were also applied for methylome or epigenome-wide association studies (MWAS or EWAS) and for integrating meDNA data with genotype or transcriptome several times. Aberg et al. (2018) concurrently analyzed genotype and DNAme using Methyl-binding domain sequencing (MBD-seq) from blood and brain for the NESDA MDD collection (320 controls and 812 cases). They focused on CpGs created/deleted by SNPs (with minor allele frequency $> 10\%$). Twenty-three CpG showed a significant methylation quantitative trait locus (meQTL) effect. These so-called CpG-SNPs were over-represented among three recent MDD GWAS findings, with genes involved in neural function. Pai et al. (2019) also investigated genetic-epigenetic interactions, focusing on frontal cortex tissue. DNAme was assessed using neuronal nuclei (sorted using FACS in 29 SC, 26 BD patients, and 27 controls), while gene expression was investigated at whole tissue level (in 17 randomly selected cases and 17 controls), and SNPs were also characterized. Results showed a concordance of enriched pathways in both transcriptomic and methylomic analyses, including immune activation, embryonic development, and synaptic transmission. The cis-meQTL analysis (SNP-CpG regression) showed a significant genetic-epigenetic interaction in 13 out of the 18 differentially methylated CpG sites. Ciuculete et al. (2020) considered a step-by-step approach with regressions using blood and brain DNAme data. In their discovery cohort of 59 adolescents, they first identified an increase in blood DNAme at the cg24627299 site that was associated with higher depression scores and suicidal ideation during a 1-year follow-up. Then using genotype data from the same discovery cohort, they identified that rs39748 a SNP located at ~16 kb downstream, was associated with the level of DNAme at this CpG site. Using a replication cohort with brain DNAme data for 45 individuals and multiple published datasets, they found that this site was inversely correlated with *MET* expression in blood and hypomethylated with an increased *MET* expression in MDD subjects' brains.

3.3.4. Advanced approaches

Elaborated approaches have been used to integrate GWAS with eQTL data – see (Reynolds et al., 2021) for a comprehensive review of post-GWAS data integration strategies. In most cases, these studies use data from non-overlapping cohorts of patients. One of the exceptions, Li et al. (2020) identified risk genes for MDD in Europeans and Han Chinese using SMR 'Summary-based Mendelian Randomization' coupled with a 'Bayesian integration risk gene selectors' (iRIGS) algorithms. The authors integrated GWAS data and eQTL, RNA-seq data in the DLPFC, and chromatin conformation data (Hi-C) from multiple sources and very large cohorts. Results identified several genes, the most significant of which was *LRFN5*, involved in the formation and differentiation of synapses, and *DCC*, involved in synaptic plasticity.

Among advanced approaches, machine learning classification was used by Bhak et al. (2019) to distinguish depressive subjects with ($n = 56$) or without ($n = 39$) suicidal behavior and healthy controls ($n = 87$). Using a random forest classifier, the authors integrated blood data on DEGs and differentially methylated CpG sites with leave-one-out cross-validation. This resulted in an MDD suicidal vs. MDD non-suicidal classifier with an accuracy of 92.6%, and 63 features composed of differentially methylated sites only. Strikingly, Wang et al. (2018) was the only study using deep learning algorithms on mood disorder data despite its growing popularity. Both WGCNA and deep Boltzmann machine algorithm were applied to the transcriptome, genome (SNP), and chromatin conformation (Hi-C) PsychENCODE data from DLPFC tissue of BD, SCZ patients, and controls. It resulted in an enhanced model linking psychiatric phenotypes to omics data, with genes involved in immunological and synaptic processes common to both schizophrenia and bipolar disorder.

Because of the rarity of studies leveraging such advanced approaches, we broadened our scope of analysis to SCZ and identified two additional relevant studies. In the first one, Wang et al., 2019a applied the iRIGS framework to infer risk gene probability from the GWAS data generated by the Schizophrenia Working Group of the PGC (36,989 SCZ cases and 113,075 controls). As a result, they built 104 networks derived risk genes (NRGs) and identified 104 high-confidence risk genes (HRGs) collectively involved in neurogenesis, glutamatergic neurotransmission, synaptic plasticity, and calcium channel and signaling, and targeted by miR-137. In the second SCZ publication, Wang et al. (2019b) proposed two integrative approaches of CNV and gene expression data from the same individuals (24 SCZ, 23 BD, and 24 controls from the Stanley foundation). The vertical approach, which aims to override collinearity among different omics, used a "reverse regression" of the CNVs on gene expression, and sparse PCA to reduce the number of selected features within a logistic regression model. The approach, which aims to study the impact of CNV on the expression of genes common to the two diseases, was based on linear regression with LASSO penalties. The sparse PCA in the vertical integration identified previously known SCZ and BD genes (*MAPK1*, *YWHAE*, *TPH1*, and *AKT1*). In contrast, the horizontal one showed *BRCA1* as a potential SCZ biomarker, suggesting that visualizing data using multiple analysis strategies and methods can help identify key disease-associated genes and biomarkers.

4. Discussion

This review first summarized recent studies that used single omic approaches in mood disorders and then looked in more detail at integrative studies that included more than one omic. Cohort size varied a lot among both single and multiple omics studies. We also noted that most recent research focused on the epigenetic aspect of mood disorders in both cases, emphasizing the contribution of DNAme and miRs, followed by their transcriptomic impact.

Up to now, despite the insights that single omic based studies have provided on the etiology of mood disorders, these studies failed at identifying reliable classifiers or molecular signatures that would improve diagnosis or patient stratification, neither for MDD nor BD. Most of these studies have observed small effect sizes, which contributes to the difficulty in uncovering reproducible functional perturbation in mood disorders. This could be explained by the subtlety and heterogeneity of molecular signatures associated with these disorders, among other factors. Accordingly, the well-known complexity of the brain

Table 2
Available multi-omic data on mood affective disorders.

Studies	DOI	Technologies	RNA	DNAm	SNP	Discription	Biological samples	Phenotype data avalaibility	Molecular data availability	Papers
Netherlands Study of Depression and Anxiety (NESDA) & Netherland Twin Register (NTR)	doi:https://doi.org/10.1002/mpr.256 doi:https://doi.org/10.1375/twin.5.5.401	Affymetrix 6.0 Human SNP & Human Genome U219 Array	x		x	SNP Genotypes: 5061 samples mRNA Expression: 5429 samples	Blood & Sliva	MDD Diagnosis, menopausal status, BMI, alcohol consumption, smoking habits, and sociodemographic data	dbGaP Study Accession: phs000486.v1.p1	Aberg et al., 2018
Stanley Medical Research Center (SMRC)	doi:10.1002/ajmg.b.32691	Illumina Infinium Human Methylation 27 k & Human Genome U133 Plus 2.0 array	x	x		mRNA: 10 BD, 10 SCZ, 10 HC DNAm: 4 BD, 4 HC, 4 SCZ	Brain (BA46, DLPFC)	Not publically available	GEO: GSE120342	Abdolmaleky et al., 2019
Pai et al., 2019	doi:10.1038/s41467-019-09786-7	Illumina NextSeq 500 Illumina MethylationEPIC BeadChip Illumina Human PsychArray-24 v1.1	x	x	x	RNA seq: 17 BD & SCZ, 17 HC DNAm & Genotypes: 26 BD, 29 SCZ, 27 HC	Brain: PFC neurons	Age, gender, ethnicity, brain weight, smoking habits, PMI, antipsychotics use, valproate use, and lithium use	GEO: GSE112525	**Wang et al., 2019b and** Pai et al., 2019
CommonMind Consortium	doi:10.1038/s41597-019-0183-6	HiSeq 2000 or HiSeq 4000 HumanOmniExpressExome 8v 1.1b chip,HumanHap650Y, Human1M-Duo, and HumanOmni5M-Quad	x		x	RNA Seq: 353 SCZ, 128 BD/MD, 501 HC. SNP: $n = 1076$	Brain	RIN, library batch, diagnosis, age of death, genetic ancestry, PMI, and gender	https://www.synapse.org//#!Synapse:syn2759792/wiki/69613	**Wang et al., 2019a**and**H. J.** Li et al., 2020
The PsychENCODE project	doi:10.1038/nn.4156	Details for each study available in supplimatry data of https://doi.org/10.1126/science.aat8464	x		x	RNAseq: 259 BD, 630 SCZ and 1094 HC SNP: 172 BD, 497 SCZ and 619 HC	Brain (DLPFC)	Age, gender, ethnicity	http://resource.psychencode.org/#Raw	Gandal et al., 2018

Legends:BA, Brodmann area; BD, bipolar disorder; DLPFC, dorsolateral prefrontal cortex; DNAm, DNA methylation; HC, Healthy Controls; GEO, gene expression omnibus accession; MD; Mood disorders; MDD, major depressive disorder; miR, micro RNA; PFC, prefrontal cortex; PMI, post mortem intervals; RIN, RNA integrity number; SCZ, Schizophrenia; SNP, single nucleotide polymorphism.

histological and structural organization, with multiple intricated glial and neuronal cell types, participates in this heterogeneity, and its underlying genetic architecture. In addition to such variable interactions at each omic level, these studies also likely suffer from their relatively small sample size, diagnostic heterogeneity, as well as insufficient consideration of meaningful additional factors, in particular, patient's medications.

From a systems biology point of view, it, therefore, appears that restricting analyses to a single biological level may not have the capacity to capture the complexity of molecular relationships and regulatory processes involved in mood disorders. With the aim of improving the identification of biomarkers with clinical potential, recent studies have moved towards multi-integrative methods, as reviewed in the second part of the present work. Over the last decade, rapid technical progress and increased availability of large datasets through collaborative and open-source efforts have opened new avenues for multi-omics data integration strategies. Among the 21 multi-omic studies reviewed herein (Table 1), most used step-by-step approaches. Such approaches, however, by definition only allow extracting information whose variance is strong enough to be detected as significant at the level of a single omic layer. Surprisingly, only one publication used a deep machine learning approach (Wang et al., 2018), despite the outstanding progress and popularity that such methods have gained recently, particularly in other research fields such as cancer. This is explained at least in part by technical challenges and higher costs associated with generating multi-omic data on large collections of biological samples. Yet, the application to multi-omics data of approaches based on matrix factorization and dimensionality reduction, as benchmarked recently by Cantini et al. (2021), would help the research community to prioritize genes and molecular biomarkers, and to uncover regulatory principles driving their pathophysiological dysregulation, as well as their potential as targets for drug repurposing programs. Furthermore, the clustering of patients using network-based methods such as SNF (Wang et al., 2014), or best consensus among various clustering methods (Brière et al., 2021), should improve patient stratification, and help better understand disease heterogeneity by identifying more homogeneous sub-groups of patients with MDD or BD.

We note that studies identified in the present systematic review mainly use case-control designs, with a striking paucity of studies based on other potentially relevant designs (e.g. longitudinal, dimensional). This could be explained by the relative novelty of multi-omics tools and integration methods designed around case-control studies. Furthermore, these methods could be used to cluster patients into molecularly homogeneous subgroups, which could potentially help in getting a finer understanding of the mood disorder spectrum. The current lack of statistically significant and reliable molecular signatures for mood disorders could be potentially overcome in the future by studies exploiting larger cohorts, and including groups of individuals at high risk for mood disorders, in prodromal phases of MDD and BD, or with variable symptom severity, using dimensional approaches. Additionally, future studies should seek to gather extensive longitudinal multi-omics profiling, which would likely help acquiring a better understanding of biological processes associated with mood disorders and their interaction with environmental factors at their earliest stage. Despite their substantial financial and logistics requirements, such prospective studies could notably mitigate the impact of significant confounding factors (e.g. medication), and ultimately lead to predictive biomarkers.

Lastly, the limited availability of bioinformatic code used for data analysis is concerning, as only 3 studies made their scripts public. Greater transparency is necessary to encourage the generalized use of multi-omics integration in molecular psychiatry. Even if some of the methods are already released as packages, many of their adaptations applied to molecular psychiatry need to be better documented to facilitate their use, determine their limitations, and ultimately lead to proper applicative recommendations. Besides, almost half of the studies reviewed here did not report corrections for known confounding covariates and were limited to matching sample groups in terms of demographic and clinical variables. Above all, none of the multi-omic studies has taken medications into account (let alone frequent combinations of medications), although available evidence suggests that this factor may be critical.

5. Conclusion

In conclusion, we would like to advocate for the use of multi-omics data in psychiatry. Indeed, applying advanced integration methods could greatly benefit patients with mood disorders by promoting the discovery of biomarkers and, ultimately, precision psychiatry. This would help to overcome the lack of stratification indicators in the current nosology. It is, therefore, necessary to encourage collaboration between clinicians, biologists, and computational biologists. It is also essential to promote data availability, notably through public, freely accessible databases, as performed e.g. in oncology with the cancer genome atlas project (TCGA). Within this line, publicly available multi-omic data discussed in the present review are gathered in Table 2.

Further initiatives should be encouraged to generate multi-omic datasets for higher numbers of individuals, with prospective longitudinal data collection complemented by clinical and phenotypic data (eg brain imaging). The specificities of each data type, their increasing amounts, and analysis of their dynamic interactions raise new challenges for multi-omics and multi-view integration methods. Overall, developing such approaches is nevertheless expected, in the long term, to strongly benefit affected patients and their families.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pnpbp.2022.110520.

Declaration of Competing Interest

None.

Acknowledgments

This work was supported by the Institut National pour la Santé et la Recherche Médicale (INSERM, France), the Centre National de la Recherche Scientifique (CNRS, France), Université de Paris, Université de Strasbourg, Université Sorbonne Paris Nord, the French National Research Agency [ANR-18-CE17-0009-01] (A.D.-D., B.P.) [ANR-18-CE37-0002-03] (A.D.-D., P.-E.L., A.M.), [ANR-18-CE37-0002-02] (E.C.I.), [ANR-19-CE37-0010] (P.E.L.), [ANR-18-CE37-0002-01] (B.E., C.M.-C.), the Fondation de France [N° Engt: 00081244] (P.-E.L., R.B.) [N° Engt:00101850] (B.E., A.D.-D), the Fondation pour la Recherche sur le Cerveau (FRC 2019, P.-E.L.), and the American Foundation for Suicide Prevention (AFSP YIG-1-102-19, P.-E.L.).

References

Abdolmaleky, H.M., Gower, A.C., Wong, C.-K., Cox, J.W., Zhang, X., Thiagalingam, A., Shafa, R., Sivaraman, V., Zhou, J.-R., Thiagalingam, S., 2019. Aberrant transcriptomes and DNA methylomes define pathways that drive pathogenesis and loss of brain laterality/asymmetry in schizophrenia and bipolar disorder. Am. J. Med. Genet. B Neuropsychiatr. Genet. 180, 138–149. https://doi.org/10.1002/ajmg.b.32691.

Aberg, K.A., Shabalin, A.A., Chan, R.F., Zhao, M., Kumar, G., van Grootheest, G., Clark, S.L., Xie, L.Y., Milaneschi, Y., Penninx, B.W.J.H., van den Oord, E.J.C.G., van Grootheest, G., Clark, S.L., Xie, L.Y., Milaneschi, Y., Penninx, B.W.J.H., van den Oord, E.J.C.G., 2018. Convergence of evidence from a methylome-wide CpG-SNP association study and GWAS of major depressive disorder. Transl. Psychiatry 8, 162. https://doi.org/10.1038/s41398-018-0205-8.

Aldinger, F., Schulze, T.G., 2017. Environmental factors, life events, and trauma in the course of bipolar disorder. Psychiatry Clin. Neurosci. 71, 6–17. https://doi.org/10.1111/pcn.12433.

Amoah, S.K., Rodriguez, B.A., Logothetis, C.N., Chander, P., Sellgren, C.M., Weick, J.P., Sheridan, S.D., Jantzie, L.L., Webster, M.J., Mellios, N., 2019. Exosomal secretion of a psychosis-altered miRNA that regulates glutamate receptor expression is affected by antipsychotics. Neuropsychopharmacology. https://doi.org/10.1038/s41386-019-0579–1.

Belzeaux, R., Fiori, L.M., Lopez, J.P., Boucekine, M., Boyer, L., Blier, P., Farzan, F., Frey, B.N., Giacobbe, P., Lam, R.W., Leri, F., MacQueen, G.M., Milev, R., Müller, D. J., Parikh, S.V., Rotzinger, S., Soares, C.N., Uher, R., Foster, J.A., Kennedy, S.H., Turecki, G., 2019. Predicting worsening suicidal ideation with clinical features and peripheral expression of messenger RNA and microRNA during antidepressant treatment. J. Clin. Psychiatry. https://doi.org/10.4088/JCP.18m12556.

Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., Milanesi, L., 2016. Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinforma. 17 https://doi.org/10.1186/s12859-015-0857-9.

Bhak, Y., Jeong, H.-O., Cho, Y.S., Jeon, S., Cho, J., Gim, J.-A., Jeon, Y., Blazyte, A., Park, S.G., Kim, H.-M., Shin, E.-S., Paik, J.-W., Lee, H.-W., Kang, W., Kim, A., Kim, Y., Kim, B.C., Ham, B.-J., Bhak, J., Lee, S., 2019. Depression and suicide risk prediction models using blood-derived multi-omics data. Transl. Psychiatry 9, 262. https://doi.org/10.1038/s41398-019-0595-2.

Bot, M., Milaneschi, Y., Al-Shehri, T., Amin, Najaf, Garmaeva, S., Onderwater, Gerrit L.J., Pool, Rene, Thesing, C.S., Vijfhuizen, L.S., Vogelzangs, N., Arts, I.C.W., Demirkan, Ayse, van Duijn, C., van Greevenbroek, M., van der Kallen, Carla J.H., Köhler, S., Ligthart, L., van den Maagdenberg, Arn M.J.M., Mook-Kanamori, D.O., de Mutsert, R., Tiemeier, H., Schram, M.T., Stehouwer, Coen D.A., Terwindt, Gisela M., Willems van Dijk, K., Fu, Jingyuan, Zhernakova, A., Beekman, Marian, Slagboom, P. E., Boomsma, D.I., Penninx, B.W.J.H., Beekman, M., Suchiman, H.E.D., Deelen, J., Amin, N., Beulens, J.W., van der Bom, J.A., Bomer, N., Demirkan, A., van Hilten, J. A., Meessen, J.M.T.A., Pool, R., Moed, M.H., Fu, J., Onderwater, G.L.J., Rutters, F., So-Osman, C., van der Flier, W.M., van der Heijden, A.A.W.A., van der Spek, A., Asselbergs, F.W., Boersma, E., Elders, P.M., Geleijnse, J.M., Ikram, M.A., Kloppenburg, M., Meulenbelt, I., Mooijaart, S.P., Nelissen, R.G.H.H., Netea, M.G., Stehouwer, C.D.A., Teunissen, C.E., Terwindt, G.M., 't Hart, L.M., van den Maagdenberg, A.M.J.M., van der Harst, P., van der Horst, I.C.C., van der Kallen, C.J. H., van Greevenbroek, M.M.J., van Spil, W.E., Wijmenga, C., Zwinderman, A.H., Zhernikova, A., Jukema, J.W., Sattar, N., 2020. Metabolomics profile in depression: a pooled analysis of 230 metabolic markers in 5283 cases with depression and 10,145 controls. Biol. Psychiatry 87, 409–418. https://doi.org/10.1016/j.biopsych.2019.08.016.

Brière, G., Darbo, É., Thébault, P., Uricaru, R., 2021. Consensus clustering applied to multi-omics disease subtyping. BMC Bioinforma. 22, 1–29. https://doi.org/10.1186/s12859-021-04279-1.

Cantini, L., Zakeri, P., Hernandez, C., Naldi, A., Thieffry, D., Remy, E., Baudot, A., 2021. Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. Nat. Commun. 12, 1–12. https://doi.org/10.1038/s41467-020-20430-7.

Chen, C., Zhang, C., Cheng, L., Reilly, J.L., Bishop, J.R., Sweeney, J.A., Chen, H.-Y., Gershon, E.S., Liu, C., 2014. Correlation between DNA methylation and gene expression in the brains of patients with bipolar disorder and schizophrenia. Bipolar Disord. 16, 790–799. https://doi.org/10.1111/bdi.12255.

Chen, C., Meng, Q., Xia, Y., Ding, C., Wang, L., Dai, R., Cheng, L., Gunaratne, P., Gibbs, R. A., Min, S., Coarfa, C., Reid, J.G., Zhang, C., Jiao, C., Jiang, Y., Giase, G., Thomas, A., Fitzgerald, D., Brunetti, T., Shieh, A., Xia, C., Wang, Y.Y.Y.Y., Wang, Y.Y.Y.Y., Badner, J.A., Gershon, E.S., White, K.P., Liu, C., 2018. The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders. Sci. Transl. Med. 10, 8178. https://doi.org/10.1126/scitranslmed.aat8178.

Ciobanu, L.G., Sachdev, P.S., Trollor, J.N., Reppermund, S., Thalamuthu, A., Mather, K. A., Cohen-Woods, S., Baune, B.T., 2016. Differential gene expression in brain and peripheral tissues in depression across the life span: a review of replicated findings. Neurosci. Biobehav. Rev. 71, 281–293. https://doi.org/10.1016/j.neubiorev.2016.08.018.

Ciuculete, D.M., Voisin, S., Kular, L., Welihinda, N., Jonsson, J., Jagodic, M., Mwinyi, J., Schiöth, H.B., 2020. Longitudinal DNA methylation changes at MET may alter HGF/c-MET signalling in adolescents at risk for depression. Epigenetics 15, 646–663. https://doi.org/10.1080/15592294.2019.1700628.

Coleman, Jonathan R.I., Gaspar, H.A., Bryois, J., Byrne, E.M., Forstner, A.J., Holmans, P. A., de Leeuw, C.A., Mattheisen, M., McQuillin, A., Whitehead Pavlides, J.M., Pers, T. H., Ripke, S., Stahl, E.A., Steinberg, S., Trubetskoy, V., Trzaskowski, M., Wang, Y., Abbott, L., Abdellaoui, A., Adams, M.J., Adolfsson, A.N., Agerbo, E., Akil, H., Albani, D., Alliey-Rodriguez, N., Als, T.D., Andlauer, T.F.M., Anjorin, A., Antilla, V., Van der Auwera, S., Awasthi, S., Bacanu, S.A., Badner, J.A., Bækvad-Hansen, M., Barchas, J.D., Bass, N., Bauer, M., Beekman, A.T.F., Belliveau, R., Bergen, S.E., Bigdeli, T.B., Binder, E.B., Bøen, E., Boks, M., Boocock, J., Budde, M., Bunney, W., Burmeister, M., Buttenschøn, H.N., Bybjerg-Grauholm, J., Byerley, W., Cai, N., Casas, M., Castelao, E., Cerrato, F., Cervantes, P., Chambert, K., Charney, A.W., Chen, D., Christensen, J.H., Churchhouse, C., St Clair, D., Clarke, T.K., Colodro-Conde, L., Coryell, W., Couvy-Duchesne, B., Craig, D.W., Crawford, G.E., Cruceanu, C., Czerski, P.M., Dale, A.M., Davies, G., Deary, I.J., Degenhardt, F., Del-Favero, J., DePaulo, J.R., Derks, E.M., Direk, N., Djurovic, S., Dobbyn, A.L., Dolan, C. V., Dumont, A., Dunn, E.C., Eley, T.C., Elvsåshagen, T., Escott-Price, V., Fan, C.C., Finucane, H.K., Fischer, S.B., Flickinger, M., Foo, J.C., Foroud, T.M., Forty, L., Frank, J., Fraser, C., Freimer, N.B., Frisén, L., Gade, K., Gage, D., Garnham, J., Giambartolomei, C., Goes, F.S., Goldstein, J., Gordon, S.D., Gordon-Smith, K., Green, E.K., Green, M.J., Greenwood, T.A., Grove, J., Guan, W., Hall, L.S., Hamshere, M.L., Hansen, C.S., Hansen, T.F., Hautzinger, M., Heilbronner, U., van Hemert, A.M., Herms, S., Hickie, I.B., Hipolito, M., Hoffmann, P., Holland, D., Homuth, G., Horn, C., Hottenga, J.J., Huckins, L., Ising, M., Jamain, S., Jansen, R., Johnson, J.S., de Jong, S., Jorgenson, E., Juréus, A., Kandaswamy, R., Karlsson, R., Kennedy, J.L., Hassan Kiadeh, F.F., Kittel-Schneider, S., Knowles, J.A., Kogevinas, M., Kohane, I.S., Koller, A.C., Kraft, J., Kretzschmar, W.W., Krogh, J., Kupka, R., Kutalik, Z., Lavebratt, C., Lawrence, J., Lawson, W.B., Leber, M., Lee, P. H., Levy, S.E., Li, J.Z., Li, Y., Lind, P.A., Liu, C., Olde Loohuis, L.M., Maaser, A., MacIntyre, D.J., MacKinnon, D.F., Mahon, P.B., Maier, W., Maier, R.M., Marchini, J., Martinsson, L., Mbarek, H., McCarroll, S., McGrath, P., McGuffin, P., McInnis, M.G., McKay, J.D., Medeiros, H., Medland, S.E., Mehta, D., Meng, F., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mirza, S.S., Mondimore, F.M., Montgomery, G.W., Morris, D.W., Mostafavi, S., Mühleisen, T.W., Mullins, N., Nauck, M., Ng, B., Nguyen, H., Nievergelt, C.M., Nivard, M.G., Nwulia, E.A., Nyholt, D.R., O'Donovan, C., O'Reilly, P.F., Ori, A.P.S., Oruc, L., Ösby, U., Oskarsson, H., Painter, J.N., Parra, J.G., Pedersen, C.B., Pedersen, M.G., Perry, A., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pfennig, A., Pistis, G., Purcell, S.M., Quiroz, J.A., Qvist, P., Regeer, E.J., Reif, A., Reinbold, C.S., Rice, J.P., Riley, B.P., Rivas, F., Rivera, M., Roussos, P., Ruderfer, D.M., Ryu, E., Sánchez-Mora, C., Schatzberg, A.F., Scheftner, W.A., Schoevers, R., Schork, N.J., Schulte, E.C., Shehktman, T., Shen, L., Shi, J., Shilling, P.D., Shyn, S.I., Sigurdsson, E., Slaney, C., Smeland, O.B., Smit, J.H., Smith, D.J., Sobell, J.L., Spijker, A.T., Steffens, M., Strauss, J.S., Streit, F., Strohmaier, J., Szelinger, S., Tansey, K.E., Teismann, H., Teumer, A., Thompson, R.C., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Traylor, M., Treutlein, J., Uitterlinden, A.G., Umbricht, D., Vedder, H., Viktorin, A., Visscher, P.M., Wang, W., Watson, S.J., Webb, B.T., Weickert, C.S., Weickert, T.W., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Xu, W., Yang, J., Young, A.H., Zandi, P., Zhang, P., Zhang, F., Zollner, S., Adolfsson, R., Agartz, I., Alda, M., Arolt, V., Backlund, L., Baune, B.T., Bellivier, F., Berger, K., Berrettini, W.H., Biernacka, J.M., Blackwood, D.H.R., Boehnke, M., Boomsma, D.I., Corvin, A., Craddock, N., Daly, M.J., Dannlowski, U., Domenici, E., Domschke, K., Esko, T., Etain, B., Frye, M., Fullerton, J.M., Gershon, E.S., de Geus, E.J.C., Gill, M., Goes, F., Grabe, H.J., Grigoroiu-Serbanescu, M., Hamilton, S.P., Hauser, J., Hayward, C., Heath, A.C., Hougaard, D.M., Hultman, C.M., Jones, I., Jones, L.A., Kahn, R.S., Kendler, K.S., Kirov, G., Kloiber, S., Landén, M., Leboyer, M., Lewis, G., Li, Q.S., Lissowska, J., Lucae, S., Madden, P.A.F., Magnusson, P.K., Martin, N.G., Mayoral, F., McElroy, S.L., McIntosh, A.M., McMahon, F.J., Melle, I., Metspalu, A., Mitchell, P.B., Morken, G., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Myers, R. M., Neale, B.M., Nimgaonkar, V., Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Oedegaard, K.J., Owen, M.J., Paciga, S.A., Pato, C., Pato, M.T., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Posthuma, D., Potash, J.B., Preisig, M., Ramos-Quiroga, J.A., Ribasés, M., Rietschel, M., Rouleau, G.A., Schaefer, C., Schalling, M., Schofield, P.R., Schulze, T.G., Serretti, A., Smoller, J.W., Stefansson, H., Stefansson, K., Stordal, E., Tiemeier, H., Turecki, G., Uher, R., Vaaler, A.E., Vieta, E., Vincent, J.B., Völzke, H., Weissman, M.M., Werge, T., Andreassen, O.A., Børglum, A.D., Cichon, S., Edenberg, H.J., Di Florio, A., Kelsoe, J., Levinson, D.F., Lewis, C.M., Nurnberger, J.I., Ophoff, R.A., Scott, L.J., Sklar, P., Sullivan, P.F., Wray, N.R., Breen, G., 2020. The genetics of the mood disorder spectrum: genome-wide association analyses of more than 185,000 cases and 439,000 controls. Biol. Psychiatry 88, 169–184. https://doi.org/10.1016/j.biopsych.2019.10.015.

Crawford, B., Craig, Z., Mansell, G., White, I., Smith, A., Spaull, S., Imm, J., Hannon, E., Wood, A., Yaghootkar, H., Ji, Y., Mullins, N., Lewis, C.M., Mill, J., Murphy, T.M., 2018. DNA methylation and inflammation marker profiles associated with a history of depression. Hum. Mol. Genet. https://doi.org/10.1093/hmg/ddy199.

Czamara, D., Eraslan, G., Page, C.M., Lahti, J., Lahti-Pulkkinen, M., Hämäläinen, E., Kajantie, E., Laivuori, H., Villa, P.M., Reynolds, R.M., Nystad, W., Håberg, S.E., London, S.J., O'Donnell, K.J., Garg, E., Meaney, M.J., Entringer, S., Wadhwa, P.D., Buss, C., Jones, M.J., Lin, D.T.S., MacIsaac, J.L., Kobor, M.S., Koen, N., Zar, H.J., Koenen, K.C., Dalvie, S., Stein, D.J., Kondofersky, I., Müller, N.S., Theis, F.J., Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.F.M., Bacanu, S.A., Bækvad-Hansen, M., Beekman, A.T.F., Bigdeli, T.B., Blackwood, D.H.R., Bryois, J., Buttenschøn, H.N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J.H., Clarke, T.K., Coleman, J.R.I., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G.E., Davies, G., Deary, I.J., Degenhardt, F., Derks, E.M., Direk, N., Dolan, C.V., Dunn, E.C., Eley, T.C., Escott-Price, V., Kiadeh, F.F.H., Finucane, H.K., Forstner, A.J., Frank, J., Gaspar, H.A., Gill, M., Goes, F.S., Gordon, S.D., Grove, J., Hall, L.S., Hansen, C.S., Hansen, T.F., Herms, S., Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.J., Hougaard, D.M., Ising, M., Jansen, R., Jorgenson, E., Knowles, J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Li, Y., Lind, P.A., MacIntyre, D.J., MacKinnon, D.F., Maier, R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S.E., Mehta, D., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mondimore, F. M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Quiroz, J.A., Qvist, P., Rice, J.P., Riley, B.P., Rivera, M., Mirza, S.S., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.C.B., Smit, J. H., Smith, D.J., Stefansson, H., Steinberg, S., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van der Auwera, S., van Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., de Geus, E.J.C., DePaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.A.F., Magnusson, P.K., Martin, N.G., McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M., Werge, T., Lewis, C.M., Levinson, D.F., Breen, G., Børglum, A.D., Sullivan, P.F., Räikkönen, K., Binder, E.B., 2019. Integrated

analysis of environmental and genetic influences on cord blood DNA methylation in new-borns. Nat. Commun. 10, 1–18. https://doi.org/10.1038/s41467-019-10461-0.
Czamara, D., Tissink, E., Tuhkanen, J., Martins, J., Awaloff, Y., Drake, A.J., Khulan, B., Palotie, A., Winter, S.M., Nemeroff, C.B., Craighead, W.E., Dunlop, B.W., Mayberg, H.S., Kinkead, B., Mathew, S.J., Iosifescu, D.V., Neylan, T.C., Heim, C.M., Lahti, J., Eriksson, J.G., Räikkönen, K., Ressler, K.J., Provençal, N., Binder, E.B., 2021. Combined effects of genotype and childhood adversity shape variability of DNA methylation across age. Transl. Psychiatry 11. https://doi.org/10.1038/s41398-020-01147-z.
Eidem, H.R., Steenwyk, J.L., Wisecaver, J.H., Capra, J.A., Abbot, P., Rokas, A., 2018. IntegRATE: a desirability-based data integration framework for the prioritization of candidate genes across heterogeneous omics and its application to preterm birth. BMC Med. Genet. 11, 1–13. https://doi.org/10.1186/s12920-018-0426-y.
Ferrari, A., 2015. Formulating a Complete Epidemiological Profile of Major Depressive Disorder: Investigating the Global Distribution, Risk Factors, Outcomes, and Burden of Major Depressive Disorder. University of Queensland Library. https://doi.org/10.14264/uql.2015.147.
Ferrúa, C.P., Giorgi, R., da Rosa, L.C., do Amaral, C.C., Ghisleni, G.C., Pinheiro, R.T., Nedel, F., 2019. MicroRNAs expressed in depression and their associated pathways: a systematic review and a bioinformatics analysis. J. Chem. Neuroanat. https://doi.org/10.1016/j.jchemneu.2019.101650.
Flint, J., Kendler, K.S., 2014. The genetics of major depression. Neuron 81, 484–503. https://doi.org/10.1016/j.neuron.2014.01.027.
Forero, D.A., Guio-Vega, G.P., Gonzalez-Giraldo, Y., 2017. A comprehensive regional analysis of genome-wide expression profiles for major depressive disorder. J. Affect. Disord. 218, 86–92. https://doi.org/10.1016/j.jad.2017.04.061.
Fries, G.R., Bauer, I.E., Scaini, G., Wu, M.-J., Kazimi, I.F., Valvassori, S.S., Zunta-Soares, G., Walss-Bass, C., Soares, J.C., Quevedo, J., 2017. Accelerated epigenetic aging and mitochondrial DNA copy number in bipolar disorder. Transl. Psychiatry 7, 1283. https://doi.org/10.1038/s41398-017-0048-8.
Fries, G.R., Carvalho, A.F., Quevedo, J., 2018. The miRNome of bipolar disorder. J. Affect. Disord. https://doi.org/10.1016/j.jad.2017.09.025.
Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., Shieh, A.W., Haney, J., Parhami, S., Belmont, J., Kim, M., Losada, P.M., Khan, Z., Mleczko, J., Xia, Y., Dai, R., Wang, D., Yang, Y.T., Xu, M., Fish, K., Hof, P.R., Warrell, J., Fitzgerald, D., White, K., Jaffe, A.E., Peters, M. A., Gerstein, M., Liu, C., Iakoucheva, L.M., Pinto, D., Geschwind, D.H., Moran Losada, P., Khan, Z., Mleczko, J., Xia, Y., Dai, R., Wang, D., Yang, Y.T., Xu, M., Fish, K., Hof, P.R., Warrell, J., Fitzgerald, D., White, K., Jaffe, A.E., Consortium, P., Peters, M.A., Gerstein, M., Liu, C., Iakoucheva, L.M., Pinto, D., Geschwind, D.H., Losada, P.M., Khan, Z., Mleczko, J., Xia, Y., Dai, R., Wang, D., Yang, Y.T., Xu, M., Fish, K., Hof, P.R., Warrell, J., Fitzgerald, D., White, K., Jaffe, A.E., Peters, M.A., Gerstein, M., Liu, C., Iakoucheva, L.M., Pinto, D., Geschwind, D.H., 2018. Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science 362. https://doi.org/10.1126/science.aat8127 eaat8127.
Gibbons, A., Sundram, S., Dean, B., 2020. Changes in non-coding RNA in depression and bipolar disorder: can they be used as diagnostic or theranostic biomarkers? Non-Coding RNA 6. https://doi.org/10.3390/ncrna6030033.
Gordovez, F.J.A., McMahon, F.J., 2020. The genetics of bipolar disorder. Mol. Psychiatry 25, 544–559. https://doi.org/10.1038/s41380-019-0634-7.
Gururajan, A., Naughton, M.E., Scott, K.A., O'Connor, R.M., Moloney, G., Clarke, G., Dowling, J., Walsh, A., Ismail, M.F., Shorten, G., Scott, L., McLoughlin, D.M., Cryan, J.F., Dinan, T.G., 2016. MicroRNAs as biomarkers for major depression: a role for let-7b and let-7c. Transl. Psychiatry 6, e862–e869. https://doi.org/10.1038/tp.2016.131.
Hasin, D.S., Sarvet, A.L., Meyers, J.L., Saha, T.D., Ruan, W.J., Stohl, M., Grant, B.F., 2018. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. JAMA Psychiatry 75, 336–346. https://doi.org/10.1001/jamapsychiatry.2017.4602.
He, L., Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. Nat. Rev. Genet. 5, 522–531. https://doi.org/10.1038/nrg1379.
Hess, J.L., Tylee, D.S., Barve, R., de Jong, S., Ophoff, R.A., Kumarasinghe, N., Tooney, P., Schall, U., Gardiner, E., Beveridge, N.J., Scott, R.J., Yasawardene, S., Perera, A., Mendis, J., Carr, V., Kelly, B., Cairns, M., Tsuang, M.T., Glatt, S.J., Unit, N.G., Tsuang, M.T., Glatt, S.J., 2019. Transcriptomic abnormalities in peripheral blood in bipolar disorder, and discrimination of the major psychoses. Schizophr. Res. 217 https://doi.org/10.1016/j.schres.2019.07.036. S0920–9964(19)30316–0.
Hou, L., Bergen, S.E., Akula, N., Song, J., Hultman, C.M., Landén, M., Adli, M., Alda, M., Ardau, R., Arias, B., Aubry, J.-M., Backlund, L., Badner, J.A., Barrett, T.B., Bauer, M., Baune, B.T., Bellivier, F., Benabarre, A., Bengesser, S., Berrettini, W.H., Bhattacharjee, A.K., Biernacka, J.M., Birner, A., Bloss, C.S., Brichant-Petitjean, C., Bui, E.T., Byerley, W., Cervantes, P., Chillotti, C., Cichon, S., Colom, F., Coryell, W., Craig, D.W., Cruceanu, C., Czerski, P.M., Davis, T., Dayer, A., Degenhardt, F., Del Zompo, M., DePaulo, J.R., Edenberg, H.J., Étain, B., Falkai, P., Foroud, T., Forstner, A.J., Frisén, L., Frye, M.A., Fullerton, J.M., Gard, S., Garnham, J.S., Gershon, E.S., Goes, F.S., Greenwood, T.A., Grigoroiu-Serbanescu, M., Hauser, J., Heilbronner, U., Heilmann-Heimbach, S., Herms, S., Hipolito, M., Hitturlingappa, S., Hoffmann, P., Hofmann, A., Jamain, S., Jiménez, E., Kahn, J.-P., Kassem, L., Kelsoe, J.R., Kittel-Schneider, S., Kliwicki, S., Koller, D.L., König, B., Lackner, N., Laje, G., Lang, M., Lavebratt, C., Lawson, W.B., Leboyer, M., Leckband, S.G., Liu, C., Maaser, A., Mahon, P.B., Maier, W., Maj, M., Manchia, M., Martinsson, L., McCarthy, M.J., McElroy, S.L., McInnis, M.G., McKinney, R., Mitchell, P.B., Mitjans, M., Mondimore, F.M., Monteleone, P., Mühleisen, T.W., Nievergelt, C.M., Nöthen, M.M., Novák, T., Nurnberger, J.I.J., Nwulia, E.A., Ösby, U., Pfennig, A., Potash, J.B., Propping, P., Reif, A., Reininghaus, E., Rice, J., Rietschel, M., Rouleau, G.A., Rybakowski, J.K., Schalling, M., Scheftner, W.A., Schofield, P.R., Schork, N.J., Schulze, T.G., Schumacher, J., Schweizer, B.W., Severino, G., Shekhtman, T., Shilling, P.D., Simhandl, C., Slaney, C.M., Smith, E.N., Squassina, A., Stamm, T., Stopkova, P., Streit, F., Strohmaier, J., Szelinger, S., Tighe, S.K., Tortorella, A., Turecki, G., Vieta, E., Volkert, J., Witt, S.H., Wright, A., Zandi, P.P., Zhang, P., Zollner, S., McMahon, F.J., 2016. Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. Hum. Mol. Genet. 25, 3383–3394. https://doi.org/10.1093/hmg/ddw181.
Howard, D.M., Adams, M.J., Shirali, M., Clarke, T.-K., Marioni, R.E., Davies, G., Coleman, J.R.I., Alloza, C., Shen, X., Barbu, M.C., Wigmore, E.M., Gibson, J., Hagenaars, S.P., Lewis, C.M., Ward, J., Smith, D.J., Sullivan, P.F., Haley, C.S., Breen, G., Deary, I.J., McIntosh, A.M., 2018. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nat. Commun. 9, 1470. https://doi.org/10.1038/s41467-018-03819-3.
Howard, D.M., Adams, M.J., Clarke, T.K., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaars, S.P., Ward, J., Wigmore, E.M., Alloza, C., Shen, X., Barbu, M.C., Xu, E.Y., Whalley, H.C., Marioni, R.E., Porteous, D.J., Davies, G., Deary, I.J., Hemani, G., Berger, K., Teismann, H., Rawal, R., Arolt, V., Baune, B.T., Dannlowski, U., Domschke, K., Tian, C., Hinds, D.A., Trzaskowski, M., Byrne, E.M., Ripke, S., Smith, D.J., Sullivan, P.F., Wray, N.R., Breen, G., Lewis, C.M., McIntosh, A. M., 2019. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. 22, 343–352. https://doi.org/10.1038/s41593-018-0326-7.
Huang, S., Chaudhary, K., Garmire, L.X., 2017. More is better: recent progress in multi-omics data integration methods. Front. Genet. https://doi.org/10.3389/fgene.2017.00084.
Hyde, C.L., Nagle, M.W., Tian, C., Chen, X., Paciga, S.A., Wendland, J.R., Tung, J.Y., Hinds, D.A., Perlis, R.H., Winslow, A.R., 2016. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. Nat. Genet. 48, 1031–1036. https://doi.org/10.1038/ng.3623.
Ikeda, M., Saito, T., Kondo, K., Iwata, N., 2018. Genome-wide association studies of bipolar disorder: a systematic review of recent findings and their clinical implications. Psychiatry Clin. Neurosci. https://doi.org/10.1111/pcn.12611.
Isometsä, E., 2014. Suicidal behaviour in mood disorders-who, when, and why? Can. J. Psychiatr. 59, 120–130. https://doi.org/10.1177/070674371405900303.
Jones, P.A., 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492. https://doi.org/10.1038/nrg3230.
Ju, C., Fiori, L.M., Belzeaux, R., Theroux, J.-F., Chen, G.G., Aouabed, Z., Blier, P., Farzan, F., Frey, B.N., Giacobbe, P., Lam, R.W., Leri, F., MacQueen, G.M., Milev, R., Müller, D.J., Parikh, S.V., Rotzinger, S., Soares, C.N., Uher, R., Li, Q., Foster, J.A., Kennedy, S.H., Turecki, G., 2019. Integrated genome-wide methylation and expression analyses reveal functional predictors of response to antidepressants. Transl. Psychiatry 9, 254. https://doi.org/10.1038/s41398-019-0589-0.
Koh, H.W.L., Fermin, D., Vogel, C., Choi, K.P., Ewing, R.M., Choi, H., 2019. iOmicsPASS: network-based integration of multiomics data for predictive subnetwork discovery. NPJ Syst. Biol. Appl. 5 https://doi.org/10.1038/s41540-019-0099-y.
Kozomara, A., Birgaoanu, M., Griffiths-Jones, S., 2019. MiRBase: from microRNA sequences to function. Nucleic Acids Res. 47, D155–D162. https://doi.org/10.1093/nar/gky1141.
Leday, G.G.R.G.R., Vértes, P.E., Richardson, S., Greene, J.R., Regan, T., Khan, S., Henderson, R., Freeman, T.C., Pariante, C.M., Harrison, N.A., Vertes, P.E., Cardinal, R., Richardson, S., Leday, G.G.R.G.R., Freeman, T.C., Regan, T., Hume, D., Wu, Z., Pariante, C.M., Cattaneo, A., Zunszain, P., Borsini, A., Stewart, R., Chandran, D., Carvalho, L., Bell, J., Souza-Teodoro, L., Perry, H., Harrison, N.A., Drevets, W.C., Wittenberg, G.M., Jones, D., Khan, S., Stylianou, A., Henderson, R., Perry, V.H., Drevets, W.C., Wittenberg, G.M., Bullmore, E.T., Greene, J.R., Regan, T., Khan, S., Henderson, R., Freeman, T.C., Pariante, C.M., Harrison, N.A., Perry, V.H., Drevets, W.C., Wittenberg, G.M., Bullmore, E.T., 2018. Replicable and coupled changes in innate and adaptive immune gene expression in two case-control studies of blood microarrays in major depressive disorder. Biol. Psychiatry 83, 70–80. https://doi.org/10.1016/j.biopsych.2017.01.021.
Legrand, A., Iftimovici, A., Khayachi, A., Chaumette, B., 2021. Epigenetics in bipolar disorder: a critical review of the literature. Psychiatr. Genet. https://doi.org/10.1097/YPG.0000000000000267.
Levey, D.F., Gelernter, J., Polimanti, R., Zhou, H., Cheng, Z., Aslan, M., Quaden, R., Concato, J., Radhakrishnan, K., Bryois, J., Sullivan, P.F., Stein, M.B., 2020. Reproducible genetic risk loci for anxiety: results from ~200,000 participants in the Million Veteran Program. Am. J. Psychiatry. https://doi.org/10.1176/appi.ajp.2019.19030256.
Li, X., Luo, Z., Gu, C., Hall, L.S., McIntosh, A.M., Zeng, Y., Porteous, D.J., Hayward, C., Li, M., Yao, Y.G., Zhang, C., Luo, X.J., 2018. Common variants on 6q16.2, 12q24.31 and 16p13.3 are associated with major depressive disorder. Neuropsychopharmacology 43, 2146–2153. https://doi.org/10.1038/s41386-018-0078-9.
Li, M., D'Arcy, C., Li, X., Zhang, T., Joober, R., Meng, X., 2019. What do DNA methylation studies tell us about depression? A systematic review. Transl. Psychiatry 9. https://doi.org/10.1038/s41398-019-0412-y.
Li, H.J., Qu, N., Hui, L., Cai, X., Zhang, C.Y., Zhong, B.L., Zhang, S.F., Chen, J., Xia, B., Wang, L., Jia, Q.F., Li, W., Chang, H., Xiao, X., Li, M., Li, Y., 2020. Further confirmation of netrin 1 receptor (DCC) as a depression risk gene via integrations of multi-omics data. Transl. Psychiatry 10. https://doi.org/10.1038/s41398-020-0777-y.
Lock, E.F., Hoadley, K.A., Marron, J.S., Nobel, A.B., 2013. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. Ann. Appl. Stat. 7, 523–542. https://doi.org/10.1214/12-AOAS597.

MacDonald, K., 2019. Biomarkers for major depressive and bipolar disorders using metabolomics: a systematic review. Am. J. Med. Genet. B Neuropsychiatr. Genet. https://doi.org/10.1002/ajmg.b.32680.

McIntosh, A.M., Sullivan, P.F., Lewis, C.M., 2019. Uncovering the genetic architecture of major depression. Neuron 102, 91–103. https://doi.org/10.1016/j.neuron.2019.03.022.

Mehta, D., Iwamoto, K., Ueda, J., Bundo, M., Adati, N., Kojima, T., Kato, T., 2014. Comprehensive survey of CNVs influencing gene expression in the human brain and its implications for pathophysiology. Neurosci. Res. 79, 22–33. https://doi.org/10.1016/j.neures.2013.10.009.

Mo, Q., Shen, R., Guo, C., Vannucci, M., Chan, K.S., Hilsenbeck, S.G., 2018. A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. Biostatistics 19, 71–86. https://doi.org/10.1093/biostatistics/kxx017.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J. Clin. Epidemiol. 62, 1006–1012. https://doi.org/10.1016/j.jclinepi.2009.06.005.

Mühleisen, T.W., Leber, M., Schulze, T.G., Strohmaier, J., Degenhardt, F., Treutlein, J., Mattheisen, M., Forstner, A.J., Schumacher, J., Breuer, R., Meier, S., Herms, S., Hoffmann, P., Lacour, A., Witt, S.H., Reif, A., Müller-Myhsok, B., Lucae, S., Maier, W., Schwarz, M., Vedder, H., Kammerer-Ciernioch, J., Pfennig, A., Bauer, M., Hautzinger, M., Moebus, S., Priebe, L., Czerski, P.M., Hauser, J., Lissowska, J., Szeszenia-Dabrowska, N., Brennan, P., McKay, J.D., Wright, A., Mitchell, P.B., Fullerton, J.M., Schofield, P.R., Montgomery, G.W., Medland, S.E., Gordon, S.D., Martin, N.G., Krasnow, V., Chuchalin, A., Babadjanova, G., Pantelejeva, G., Abramova, L.I., Tiganov, A.S., Polonikov, A., Khusnutdinova, E., Alda, M., Grof, P., Rouleau, G.A., Turecki, G., Laprise, C., Rivas, F., Mayoral, F., Kogevinas, M., Grigoroiu-Serbanescu, M., Propping, P., Becker, T., Rietschel, M., Nöthen, M.M., Cichon, S., 2014. Genome-wide association study reveals two new risk loci for bipolar disorder. Nat. Commun. 5, 3339. https://doi.org/10.1038/ncomms4339.

Nemeroff, C.B., 2020. The state of our understanding of the pathophysiology and optimal treatment of depression: glass half full or half empty? Am. J. Psychiatry 177, 671–685. https://doi.org/10.1176/appi.ajp.2020.20060845.

Okbay, A., Baselmans, B.M.L., De Neve, J.E., Turley, P., Nivard, M.G., Fontana, M.A., Meddens, S.F.W., Linnér, R.K., Rietveld, C.A., Derringer, J., Gratten, J., Lee, J.J., Liu, J.Z., De Vlaming, R., SAhluwalia, T., Buchwald, J., Cavadino, A., Frazier-Wood, A.C., Furlotte, N.A., Garfield, V., Geisel, M.H., Gonzalez, J.R., Haitjema, S., Karlsson, R., Der Laan, S.W., Ladwig, K.H., Lahti, J., Van Der Lee, S.J., Lind, P.A., Liu, T., Matteson, L., Mihailov, E., Miller, M.B., CMinica, C., MNolte, I., Mook-Kanamori, D., Van Der Most, P.J., Oldmeadow, C., Qian, Y., Raitakari, O., Rawal, R., Realo, A., Rueedi, R., Schmidt, B., Smith, A.V., Stergiakouli, E., Tanaka, T., Taylor, K., Wedenoja, J., Wellmann, J., Westra, H.J., MWillems, S., Zhao, W., Study, L.L.C., Amin, N., Bakshi, A., Boyle, P.A., Cherney, S., Cox, S.R., Davies, G., Davis, O.S.P., Ding, J., Direk, N., Eibich, P., Emeny, R.T., Fatemifar, G., Faul, J.D., Ferrucci, L., Forstner, A., Gieger, C., Gupta, R., Harris, T.B., Harris, J.M., Holliday, E. G., Hottenga, J.J., De Jager, P.L., Kaakinen, M.A., Kajantie, E., Karhunen, V., Kolcic, I., Kumari, M., Launer, L.J., Franke, L., Li-Gao, R., Koini, M., Loukola, A., Marques-Vidal, P., Montgomery, G.W., Mosing, M.A., Paternoster, L., Pattie, A., Petrovic, K.E., Pulkki-R'back, L., Quaye, L., R'ikkönen, K., Rudan, I., Scott, R.J., Smith, J.A., Sutin, A.R., Trzaskowski, M., Vinkhuyze, A.E., Yu, L., Zabaneh, D., Attia, J.R., Bennett, D.A., Berger, K., Bertram, L., Boomsma, D.I., Snieder, H., Chang, S.C., Cucca, F., Deary, I.J., Van Duijn, C.M., Eriksson, J.G., Bültmann, U., Geus, E.J.C., Groenen, P.J.F., Gudnason, V., Hansen, T., Hartman, C.A., Haworth, C. M.A., Hayward, C., Heath, A.C., Hinds, D.A., Hyppönen, E., Iacono, W.G., J'rvelin, M.R., Jöckel, K.H., Kaprio, J., Kardia, S.L.R., Keltikangas-J'rvinen, L., Kraft, P., Kubzansky, L.D., Lehtim'ki, T., Magnusson, P.K.E., Martin, N.G., McGue, M., Metspalu, A., Mills, M., De Mutsert, R., Oldehinkel, A.J., Pasterkamp, G., Pedersen, N.L., Plomin, R., Polasek, O., Power, C., Rich, S.S., Rosendaal, F.R., Den Ruijter, H.M., Schlessinger, D., Schmidt, H., Svento, R., Schmidt, R., Alizadeh, B.Z., SØrensen, T.I.A., DSpector, T., Steptoe, A., Terracciano, A., Thurik, A.R., Timpson, N.J., Tiemeier, H., Uitterlinden, A.G., Vollenweider, P., Wagner, G.G., Weir, D.R., Yang, J., Conley, D.C., Smith, G.D., Hofman, A., Johannesson, M., Laibson, D.I., Medland, S.E., Meyer, M.N., Pickrell, J.K., Esko, T., Krueger, R.F., Beauchamp, J.P., Koellinger, P.D., Benjamin, D.J., Bartels, M., Cesarini, D., 2016. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat. Genet. 48, 624–633. https://doi.org/10.1038/ng.3552.

Ormel, J., 2019. The genetics of depression: successful genome-wide association studies introduce new challenges. Transl. Psychiatry 9. https://doi.org/10.1038/s41398-019-0450-5.

Pai, S., Li, P., Killinger, B., Marshall, L., Jia, P., Liao, J., Petronis, A., Szabó, P.E., Labrie, V., 2019. Differential methylation of enhancer at IGF2 is associated with abnormal dopamine synthesis in major psychosis. Nat. Commun. 10 https://doi.org/10.1038/s41467-019-09786-7.

Penner-Goeke, S., Binder, E.B., 2019. Epigenetics and depression. Dialogues Clin. Neurosci. 21, 397–405. https://doi.org/10.31887/DCNS.2019.21.4/ebinder.

Pisanu, C., Congiu, D., Costa, M., Chillotti, C., Ardau, R., Severino, G., Angius, A., Heilbronner, U., Hou, L., McMahon, F.J., Schulze, T.G., Del Zompo, M., Squassina, A., Del Zompo, M., Squassina, A., 2018. Convergent analysis of genome-wide genotyping and transcriptomic data suggests association of zinc finger genes with lithium response in bipolar disorder. Am. J. Med. Genet. B Neuropsychiatr. Genet. 177, 658–664. https://doi.org/10.1002/ajmg.b.32663.

Pisanu, C., Merkouri Papadima, E., Melis, C., Congiu, D., Loizedda, A., Orrù, N., Calza, S., Orrù, S., Carcassi, C., Severino, G., Ardau, R., Chillotti, C., Del Zompo, M., Squassina, A., 2019. Whole genome expression analyses of miRNAs and mRNAs suggest the involvement of miR-320a and miR-155-3p and their targeted genes in lithium response in bipolar disorder. Int. J. Mol. Sci. 20, 6040. https://doi.org/10.3390/ijms20236040.

Psychiatric GWAS Consortium Bipolar, 2011. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. 43, 977–983. https://doi.org/10.1038/ng.943.

Rappoport, N., Shamir, R., 2018. Multi-omic and multi-view clustering algorithms: review and cancer benchmark. Nucleic Acids Res. 46, 10546–10562. https://doi.org/10.1093/nar/gky889.

Reynolds, T., Johnson, E.C., Huggett, S.B., Bubier, J.A., Palmer, R.H.C., Agrawal, A., Baker, E.J., Chesler, E.J., 2021. Interpretation of psychiatric genome-wide association studies with multispecies heterogeneous functional genomic data integration. Neuropsychopharmacology 46, 86–97. https://doi.org/10.1038/s41386-020-00795-5.

Rohart, F., Gautier, B., Singh, A., Lê Cao, K.A., 2017. mixOmics: an R package for 'omics feature selection and multiple data integration. PLoS Comput. Biol. 13, 1–19. https://doi.org/10.1371/journal.pcbi.1005752.

Sanada, K., Nakajima, S., Kurokawa, S., Barceló-Soler, A., Ikuse, D., Hirata, A., Yoshizawa, A., Tomizawa, Y., Salas-Valero, M., Noda, Y., Mimura, M., Iwanami, A., Kishimoto, T., 2020. Gut microbiota and majore depressive disorder: a systematic review and meta-analysis. J. Affect. Disord. 266, 1–13. https://doi.org/10.1016/j.jad.2020.01.102.

Seney, M.L., Huo, Z., Cahill, K., French, L., Puralewski, R., Zhang, J., Logan, R.W., Tseng, G., Lewis, D.A., Sibille, E., 2018. Opposite molecular signatures of depression in men and women. Biol. Psychiatry 84, 18–27. https://doi.org/10.1016/j.biopsych.2018.01.017.

Shen, R., Olshen, A.B., Ladanyi, M., 2009. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. Bioinformatics 25, 2906–2912. https://doi.org/10.1093/bioinformatics/btp543.

Snijders, C., Maihofer, A.X., Ratanatharathorn, A., Baker, D.G., Boks, M.P., Geuze, E., Jain, S., Kessler, R.C., Pishva, E., Risbrough, V.B., Stein, M.B., Ursano, R.J., Vermetten, E., Vinkers, C.H., Smith, A.K., Uddin, M., Rutten, B.P.F., Nievergelt, C.M., 2020. Longitudinal epigenome-wide association studies of three male military cohorts reveal multiple CpG sites associated with post-traumatic stress disorder. Clin. Epigenetics 12, 1–13. https://doi.org/10.1186/s13148-019-0798-7.

Squassina, A., 2020. MicroRNA expression profiling of lymphoblasts from bipolar disorder patients who died by suicide, pathway analysis and integration with postmortem brain findings. Eur. Neuropsychopharmacol. 34, 39–49. https://doi.org/10.1016/j.euroneuro.2020.03.005.

Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., de Leeuw, C.A., Steinberg, S., Pavlides, J.M.W., Trzaskowski, M., Byrne, E.M., Pers, T.H., Holmans, P. A., Richards, A.L., Abbott, L., Agerbo, E., Akil, H., Albani, D., Alliey-Rodriguez, N., Als, T.D., Anjorin, A., Antilla, V., Awasthi, S., Badner, J.A., Bækvad-Hansen, M., Barchas, J.D., Bass, N., Bauer, M., Belliveau, R., Bergen, S.E., Pedersen, C.B., Bøen, E., Boks, M.P., Boocock, J., Budde, M., Bunney, W., Burmeister, M., Bybjerg-Grauholm, J., Byerley, W., Casas, M., Cerrato, F., Cervantes, P., Chambert, K., Charney, A.W., Chen, D., Churchhouse, C., Clarke, T.K., Coryell, W., Craig, D.W., Cruceanu, C., Curtis, D., Czerski, P.M., Dale, A.M., de Jong, S., Degenhardt, F., Del-Favero, J., DePaulo, J.R., Djurovic, S., Dobbyn, A.L., Dumont, A., Elvsåshagen, T., Escott-Price, V., Fan, C.C., Fischer, S.B., Flickinger, M., Foroud, T.M., Forty, L., Frank, J., Fraser, C., Freimer, N.B., Frisén, L., Gade, K., Gage, D., Garnham, J., Giambartolomei, C., Pedersen, M.G., Goldstein, J., Gordon, S.D., Gordon-Smith, K., Green, E.K., Green, M.J., Greenwood, T.A., Grove, J., Guan, W., Guzman-Parra, J., Hamshere, M.L., Hautzinger, M., Heilbronner, U., Herms, S., Hipolito, M., Hoffmann, P., Holland, D., Huckins, L., Jamain, S., Johnson, J.S., Juréus, A., Kandaswamy, R., Karlsson, R., Kennedy, J.L., Kittel-Schneider, S., Knowles, J.A., Kogevinas, M., Koller, A.C., Kupka, R., Lavebratt, C., Lawrence, J., Lawson, W.B., Leber, M., Lee, P.H., Levy, S.E., Li, J.Z., Liu, C., Lucae, S., Maaser, A., MacIntyre, D. J., Mahon, P.B., Maier, W., Martinsson, L., McCarroll, S., McGuffin, P., McInnis, M. G., McKay, J.D., Medeiros, H., Medland, S.E., Meng, F., Milani, L., Montgomery, G. W., Morris, D.W., Mühleisen, T.W., Mullins, N., Nguyen, H., Nievergelt, C.M., Adolfsson, A.N., Nwulia, E.A., O'Donovan, C., Loohuis, L.M.O., Ori, A.P.S., Oruc, L., Ösby, U., Perlis, R.H., Perry, A., Pfennig, A., Potash, J.B., Purcell, S.M., Regeer, E.J., Reif, A., Reinbold, C.S., Rice, J.P., Rivas, F., Rivera, M., Roussos, P., Ruderfer, D.M., Ryu, E., Sánchez-Mora, C., Schatzberg, A.F., Scheftner, W.A., Schork, N.J., Shannon Weickert, C., Shehktman, T., Shilling, P.D., Sigurdsson, E., Slaney, C., Smeland, O.B., Sobell, J.L., Søholm Hansen, C., Spijker, A.T., St Clair, D., Steffens, M., Strauss, J.S., Streit, F., Strohmaier, J., Szelinger, S., Thompson, R.C., Thorgeirsson, T.E., Treutlein, J., Vedder, H., Wang, W., Watson, S.J., Weickert, T.W., Witt, S.H., Xi, S., Xu, W., Young, A.H., Zandi, P., Zhang, P., Zöllner, S., Adolfsson, R., Agartz, I., Alda, M., Backlund, L., Baune, B.T., Bellivier, F., Berrettini, W.H., Biernacka, J.M., Blackwood, D.H.R., Boehnke, M., Børglum, A.D., Corvin, A., Craddock, N., Daly, M. J., Dannlowski, U., Esko, T., Etain, B., Frye, M., Fullerton, J.M., Gershon, E.S., Gill, M., Goes, F., Grigoroiu-Serbanescu, M., Hauser, J., Hougaard, D.M., Hultman, C. M., Jones, I., Jones, L.A., Kahn, R.S., Kirov, G., Landén, M., Leboyer, M., Lewis, C.M., Li, Q.S., Lissowska, J., Martin, N.G., Mayoral, F., McElroy, S.L., McIntosh, A.M., McMahon, F.J., Melle, I., Metspalu, A., Mitchell, P.B., Morken, G., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Myers, R.M., Neale, B.M., Nimgaonkar, V., Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Oedegaard, K.J., Owen, M.J., Paciga, S.A., Pato, C., Pato, M.T., Posthuma, D., Ramos-Quiroga, J.A., Ribasés, M., Rietschel, M., Rouleau, G.A., Schalling, M., Schofield, P.R., Schulze, T.G., Serretti, A., Smoller, J.W., Stefansson, H., Stefansson, K., Stordal, E., Sullivan, P.F., Turecki, G., Vaaler, A.E., Vieta, E., Vincent, J.B., Werge, T., Nurnberger, J.I., Wray, N.R., Di Florio, A., Edenberg, H.J., Cichon, S., Ophoff, R.A., Scott, L.J., Andreassen, O.A., Kelsoe, J., Sklar, P., 2019. Genome-wide association study identifies 30 loci

associated with bipolar disorder. Nat. Genet. 51, 793–803. https://doi.org/10.1038/s41588-019-0397-8.
Teroganova, N., Girshkin, L., Suter, C.M., Green, M.J., 2016. DNA methylation in peripheral tissue of schizophrenia and bipolar disorder: a systematic review. BMC Genet. https://doi.org/10.1186/s12863-016-0332-2.
Tini, G., Marchetti, L., Priami, C., Scott-Boyer, M.P., 2018. Multi-omics integration-a comparison of unsupervised clustering methodologies. Brief. Bioinform. 20, 1269–1279. https://doi.org/10.1093/bib/bbx167.
van Dongen, J., Ehli, E.A., Slieker, R.C., Bartels, M., Weber, Z.M., Davies, G.E., Slagboom, P.E., Heijmans, B.T., Boomsma, D.I., 2014. Epigenetic variation in monozygotic twins: a genome-wide analysis of DNA methylation in buccal cells. Genes (Basel) 5, 347–365. https://doi.org/10.3390/genes5020347.
Wang, B., Mezlini, A.M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., Goldenberg, A., 2014. Similarity network fusion for aggregating data types on a genomic scale. Nat. Methods 11, 333–337. https://doi.org/10.1038/nmeth.2810.
Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., 2018. Comprehensive functional genomic resource and integrative model for the human brain.
Wang, Q., Chen, R., Cheng, F., Wei, Q., Ji, Y., Yang, H., Zhong, X., Tao, R., Wen, Z., Sutcliffe, J.S., Liu, C., Cook, E.H., Cox, N.J., Li, B., 2019a. A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. Nat. Neurosci. 22, 691–699. https://doi.org/10.1038/s41593-019-0382-7.
Wang, S., Shi, X., Wu, M., Ma, S., 2019b. Horizontal and vertical integrative analysis methods for mental disorders omics data. Sci. Rep. 1–12 https://doi.org/10.1038/s41598-019-49718-5.
Wittenberg, G.M., Greene, J., Vértes, P.E., Drevets, W.C., Bullmore, E.T., 2020. Major depressive disorder is associated with differential expression of innate immune and neutrophil-related gene networks in peripheral blood: a quantitative review of whole-genome transcriptional data from case-control studies. Biol. Psychiatry 88, 625–637. https://doi.org/10.1016/j.biopsych.2020.05.006.
Wolf, E.J., Maniates, H., Nugent, N., Maihofer, A.X., Armstrong, D., Ratanatharathorn, A., Ashley-Koch, A.E., Garrett, M., Kimbrel, N.A., Lori, A., Mid-Atlantic, V.A., Workgroup, M.I.R.E.C.C., Aiello, A.E., Baker, D.G., Beckham, J.C., Boks, M.P., Galea, S., Geuze, E., Hauser, M.A., Kessler, R.C., Koenen, K.C., Miller, M. W., Ressler, K.J., Risbrough, V., Rutten, B.P.F., Stein, M.B., Ursano, R.J., Vermetten, E., Vinkers, C.H., Uddin, M., Smith, A.K., Nievergelt, C.M., Logue, M.W., 2018. Traumatic stress and accelerated DNA methylation age: a meta-analysis. Psychoneuroendocrinology 92, 123–134. https://doi.org/10.1016/j.psyneuen.2017.12.007.
Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., Bacanu, S.A., Bækvad-Hansen, M., Beekman, A.F.T., Bigdeli, T.B., Binder, E.B., Blackwood, D.R.H., Bryois, J., Buttenschøn, H.N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J.H., Clarke, T.K., Coleman, J.I.R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G.E., Crowley, C.A., Dashti, H.S., Davies, G., Deary, I.J., Degenhardt, F., Derks, E.M., DIrek, N., Dolan, C.V., Dunn, E.C., Eley, T. C., Eriksson, N., Escott-Price, V., Kiadeh, F.H.F., Finucane, H.K., Forstner, A.J., Frank, J., Gaspar, H.A., Gill, M., Giusti-Rodríguez, P., Goes, F.S., Gordon, S.D., Grove, J., Hall, L.S., Hannon, E., Hansen, C.S., Hansen, T.F., Herms, S., Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.J., Hougaard, D.M., Hu, M., Hyde, C.L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Lane, J.M., Li, Yihan, Li, Yun, Lind, P.A., Liu, X., Lu, L., MacIntyre, D.J., MacKinnon, D.F., Maier, R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S.E., Mehta, Di., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mill, J., Mondimore, F.M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Purcell, S.M., Quiroz, J.A., Qvist, P., Rice, J.P., Riley, B.P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.B.C., Smit, J.H., Smith, D.J., Stefansson, H., Steinberg, S., Stockmeier, C.A., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van Der Auwera, S., Van Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., De Geus, E.C.J., Depaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Hinds, D.A., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.F.A., Magnusson, P.K., Martin, N. G., McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M., Werge, T., Winslow, A.R., Lewis, C.M., Levinson, D.F., Breen, G., Børglum, A.D., Sullivan, P.F., 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. 50, 668–681. https://doi.org/10.1038/s41588-018-0090-3.
Wu, C., Zhou, F., Ren, J., Li, X., Jiang, Y., Ma, S., 2019. A selective review of multi-level omics data integration using variable selection. High-Throughput 8, 1–25. https://doi.org/10.3390/ht8010004.
Xie, Y., 2021. Integrated analysis of methylomic and transcriptomic data to identify potential diagnostic biomarkers for major depressive disorder. Genes (Basel) 12. https://doi.org/10.3390/genes12020178.
Yrondi, A., Fiori, L.M., Frey, B.N., Lam, R.W., MacQueen, G.M., Milev, R., Müller, D.J., Foster, J.A., Kennedy, S.H., Turecki, G., 2020. Association between side effects and blood microRNA expression levels and their targeted pathways in patients with major depressive disorder treated by a selective serotonin reuptake inhibitor, escitalopram: a CAN-BIND-1 report. Int. J. Neuropsychopharmacol. 23, 88–95. https://doi.org/10.1093/ijnp/pyz066.
Zhao, H., Xu, J., Pang, L., Zhang, Y., Fan, H., Liu, L., Liu, T., Yu, F., Zhang, G., Lan, Y., Bai, J., Li, X., Xiao, Y., 2015. Genome-wide DNA methylome reveals the dysfunction of intronic microRNAs in major psychosis. BMC Med. Genet. 8, 62. https://doi.org/10.1186/s12920-015-0139-4.
Zhu, Y., Strachan, E., Fowler, E., Bacus, T., Roy-Byrne, P., Zhao, J., 2019. Genome-wide profiling of DNA methylome and transcriptome in peripheral blood monocytes for major depression: a Monozygotic Discordant Twin Study. Transl. Psychiatry 9, 215. https://doi.org/10.1038/s41398-019-0550-2.

Supplementary table 01: Mood disorders associated genome wide association studies.

Study	Cohort	Disease	Ethnicity	Number of significant genome-wide identified risk loci	Nearest candidate genes to the GWAS SNP*
Charney et al. 2017	13 903 BD, 19 279 HC	BD	EU	8	*ADD3, ANK3, CACNA1C, ITIH1, LMAN2L, LOC100505933, RHEBL1/DHH, SYNE1, TRANK1, XPNPEP1*
Chen et al. 2013	5568 BD EUR, 7187 HC EU; 2000 BD EAS 1000, 1000 HC EAS	BD	EU and EA	5	*ANK3, LMAN2, PRM1M, PTGFR/IFI44L/ADGRL4L, TRANK1*
CONVERGE consortium. 2015	5303 MDD, 5337 HC	MDD	EA ♀□	2	*LHPP, SIRT1*
Direk et al. 2016	9240 MDD, 9519 HC	MDD	EU	1	*FHIT*
Ferreira et al. 2008	4387BD, 6209 HC	BD	EU	1	*ANK3*
Hall et al. 2018	3852 MDD, 16 034 HC	MDD	EU	1	*CRTAP*
Hou et al. 2016	7647 BD, 27 303 HC	BD	EU	6	*DDN, ERBB2, MAD1L1, MIR2113, POU3F2, TRANK1*
Howard et al. 2018	8276 MDD, 209 308 HC	MDD	EU	1	*MAD1L1*
Howard et al. 2019	246 363 MDD, 561 190 HC	MDD	EU	101	*AREL1, ASTN2, ATP2A2, B3GLCT, BAG5, BEND4, C3orf84, CACNA1E, CELF4, CTTNBP2, CYP7B1, DAGLA, DCC, DCDC1, DENND1A, DENND1B, ELAVL2, ELAVL4, EMILIN3, EP300, ESR2, FAM172A, FHIT, GRM5, HIST1H1B, HIST1H2AL, HIST1H3I, HTT, KIAA1109, KLHDC8B, LINC01876, LOC102724552, LRFN5, LTBP3, MAD1L1, METTL9, MIR1281, NRDC, PAX5, PCDH9, PCLO, POU5F2, RBFOX1, RERE, RFWD2, RNF103/CHMP3, RSRC1, RTN1, SEMA6D, SGIP1, SHANK2, SHISA9, SLC12A5, SORCS3, SOX5, SPPL3, STK24, SYNE2, TCF4, TLR4, TMEM106B, TMEM161B, ZNF536*
Hyde et al. 2016	130 620 MDD, 347 620 HC	MDD	EU	15	*HACE1/LIN28B, KIAA0020/RFX3, L3MBTL2, MEIS2/TMCO5A, NEGR, NEGR1, OLFM4, PAX5, RERE, RSRC1/MLF1, SLC6A15, SORCS3, TMEM161/MEF2C, TMEM161B/MEF2C, VRK2*
Ikeda et al. 2018	2964 EAS BD, 61 887 EAS HC; 7481 EU BD, 9250 EU HC	BD	EA and EU	6	*C11orf10, C11orf9, DAGLA, FADS1, FADS2, FEN1, MAD1L1, NFIX, ODZ4, TRANK1*
Li et al. 2018	90 150 MDD, 246 603 HC	MDD	EU and EA ♀□	10	*LINC00461, NEGR1, PCDH8P1, RSRC1, SPPL3*
Mühleisen et al. 2014	9747 BD, 14 278 HC	BD	EU	5	*ADCY2, ANK3, MIR2113, ODZ4, POU3F2, TRANK1*
Okbay et al. 2016	16 471 MDD,58 835 HC	MDD	EU	2	*DCC, KSR2*
PGC BD Group. 2011	7481 BD, 9250 HC	BD	EU	2	*CANCNA1C, ODZ4*
Power et al. 2017	4000 MDD,9519 HC	MDD	EU	1	*EHHADH-AS1*
Stahl et al. 2019	20 352 EU BD, 31 358 HC	BD	EU	31	*ACOT12, ADCY2, ADD3, ALPK3, ANK3, ANP32E, APH1A, ASB16, ASB16-AS1, ATXN7L3, BANF1, BRAF, BRMS1, C11orf80, C17orf53, C1orf51, C1orf54, CA14, CACNA1C, CATSPER1, CD248, CD47, CNIH2, CSRNP3, CST6, DNM1P41, EIF1AD, FADS1, FSTL5, G6PC3, GAL3ST3, GOLGA6L4, GOLGA6L5, GRIN2A, HAUS2, HDAC5, IFT57, ITIH1, ITIH3, KCNS1, KCNS1, KLC2, LMAN2L, LOC100506874, LOC388152, LOC440300, LOC642423, LRFN4, LRRC57, LSM12, MRPS33, NCAN, NMB, OTUD7B, PABPC1L, PACS1, PC, PI3, PLEKHO1, POU3F2, PUS7, RAB1B, RCE1, RIN1, RNU5D-1, RNU5E-1, RPS6KA2, SART1, SCAND2P, SCN2A, SEC11A, SF3B2, SHANK2, SHANK2-AS1, SLC28A1, SLC4A1, SNAP23, SPTBN2, SRPK2, SSBP2, STARD9, STK4, STK4-AS1, THSD7A, TMEM151A, TMEM178B, TMUB2, TOMM34, TRANK1, UBE2Q2P1, UBTF, VPS45, WDR73, WFDC12, WFDC5, YIF1A, ZCCHC2, ZNF592, ZSCAN2*
Wray et al. 2018	135 458 MDD, 344 901 HC	MDD	EU	44	*APOPT1, ASTN2, BAG5, C6orf168, CCDC122, CCDC68, CHADL, CPPED1, CRYBA1, DCAF4L1, DCC, DENND1A, DHX38, DKFZp686K1684, DLST, ELP4, ENOX1, EP300-AS1, ESR2, FBXL4, GPD2, L3MBTL2, LACC1, LHX2, LINC00461, LINC00682, LINC01065, LINC01231, LINC01360, LINC01876, LOC101927421, LRFN5, MEF2C, MIR4528, MIR4529, MIR548H1, MIR924HG, MYO18A, NEGR1, NR4A2, NUFIP2, OLFM4, PAUPAR, PAX6, PMFBP1, PROX2, PUM3, RAB27B, RBFOX1, RERE, RPS6KL1, SHISA9, SLC30A9, SLC45A1, SORCS3, SOX5, SYNE2, TCAIM, TCF4, TENM2, TOPAZ1, VRK2, VWDE, ZNF445*
Coleman et al. 2020	185,285 MDD and BD, 439,741 HC	MDD and BP	EU	73	*AL513166.1, LOC105378797, COP1, AL136114.1, LINC01645, CNNM2, SORCS3, SOX6, PAUPAR, AL035078.4, PAX6-AS1, CCDC73, TMEM258, MYRF, FEN1, MIR611, FADS2, MARK2, GRM5, CNTN5, CACNA1C, SOX5, AC069234.1, KDM2B, ENOX1, OLFM4, LRFN5, AL121821.2, LOC105370469, SYNE2, ESR2, AREL1, AC007956.2, BAG5, RNU7-160P, AL139300.2, AC021351.1, RBFOX1, AC005774.2, SHISA9, LOC105371092, PMFBP1, PIPOX, LOC107985157, MIR924HG, SETBP1, SETBP1-DT, LOC101927943, DCC, RAB27B, TCF4, ZNF536, AC096570.1, ACTG1P22, LINC01876, PLCL1, ZMYND8, L3MBTL2, CHADL, L3MBTL2-AS1, NFAM1, HDAC10, TUBGCP6, MAPK12, MYH15, RSRC1, SLC30A9, SCAMP1, MEF2C-AS2, FAM172A, AC106818.2, AC099520.1, LINC02240, AC109466.1, TENM2, AL133338.2, LOC107984041, LIN28B-AS1, TMEM106B, AC003093.1, LOC107986779, PCLO, AL355672.1, LOC105375974, ZCCHC7, AL512604.1, ASTN2, LOC107987122, DENND1A, AL513166.2*
PGC Cross disorder Group , 2013	33,332 cases (among which 6990 BP and 9227 MDD) 27,842 HC	ASD, ADHD, BP, MDD, SCZ	EU	4	*ITIH3, CACNB2, AS3MT, CACNA1C*

Legends: BD, bipolar disorder; EA, east asia; EU, europe; HC, healthy controls; MDD, major depressive disorder; ♀, female individuals only; * SNPs to gene assignement was based on the nearest gene if available retrived from Ensembl Variant Effect Predictor Tool (VEP: http://www.ensembl.org/Multi/Tools/VEP?db=core)

Supplementary table 02: Recent and main genome-wide transcriptomics meta-anlyses studies on mood affective disorders.

Study	Cohort	Biological sample	Technologies	Main analysis method	Covariate correction	Relevent results
Forero et al. 2017	358 MDD and 385 HC	Blood and Brain (4 regions)	Micoarrays	Meta-analysis	Not mentionned	Significant alterations of pathways detected only in the brain datasets involving neural plasticity and signaling .
Hess et al. 2019	95 BD & 111 HC and 258 SCZ & 241 HC	Whole blood	Micoarrays and RNA-seq	Meta-analysis and WGCNA	Age, gender, study site and significant surrogate variables	DEG and co-expression modules from WGCNA described alterations of the immune system, oxidative stress and apoptosis in BD patients.
Seney et al. 2018	26 MDD men and 24 MDD women	Brain (3 regions)	Illumina HiSeq 2000	Meta-analysis	Age, gender, PMI, brain pH, RIN, death by suicide, medication and alcohol dependence	Opposite transcriptional alteration in man and woman with major depressive disorder.
Wittenberg et al. 2020	1754 MDD and 1145 HC	Peripheral tissue	Micoarrays and RNA-seq	Meta-analysis	Age, gender and batch	Common genes associated with the immune response system and especially neutrophil activation.

Legends :DEG, differentially expressed gene; HC, healthy control; MDD, major depressive disorder; PMI, post mortem interval; RIN, RNA integrity number; WGCNA, weighted correlation network

Chapter 3

Chapter 3: Multi-Omics data integration approaches

3.1 Multi-omics nomenclature and classification 64
3.1.1 Step-wise Integration 64
3.1.2 Horizontal and Vertical Integration 64
3.1.3 Simultaneous Data Integration 67
3.1.4 Unsupervised and Supervised Integration 67
3.2 Integration strategies 68
3.2.1 Network based 68
3.2.2 Machine learning based 78
3.3 Challenges in multi-omics integration 90
3.3.1 Data heterogenity 90
3.3.2 Multi-collinearity 90
3.3.3 Missing data 91
3.3.4 Cross-Validation in unsupervised analysis 91
3.3.5 Validation 91

Previously, in chapter 2, we systematically reviewed the use of multi-omic data in relation to MD. In the present chapter, we detail the different tools and concepts underlying multi-omics integrative methods.

Recent technological advances now allow a thorough investigation of the biological intricacies that come at play in disease onset and development. However, no single data type has been able to fully grasp this complexity. Taking advantage of multiple modalities, including clinical and biological data (genomics, transcriptomics, epigenomics, proteomics, and metabolomics), integrative methodologies appear as a potential solution to capture the relationships between these different measurements, and phenotypes of interest.

Multi-omic data integration methods are used to analyze and interpret data that cover different forms of genomic layers or processes. These methods can be used to identify patterns and relationships between different types of data, and to generate hypotheses about the underlying biology. Multi-omic data integration is therefore a powerful tool for understanding the complex biology of disease and for developing new therapies. Yet, those datasets are multidimensional, noisy, and prone to biases due to technical limitations or biological heterogeneity. As a result, generating relevant data and combining several omics into coherent or predictive models is challenging. Various recent research methodologies and tools have been developed to address this problem. The ideal multi-omic integration framework should be able to: i) extract biologically and medically relevant features (i.e. omic molecular measurements, phenotypic traits, and/or surrogates based on one or both of these variables), ii) recapitulate the co-variation of inputted modalities, and iii) predict classes and phenotypes. Over the last decade, many strategies have been implemented to achieve such goals, varying from step-wise approaches (which mostly consists in the identification of overlaps among omics) to more elaborated frameworks, including machine and deep learning algorithms.

Unfortunately, no gold standard has been identified so far among available methods. Instead, many classification criteria have been suggested in the literature, some based on the experimental design itself (Zitnik et al., 2019) (vertical and horizontal integration, see below), others focused on the analytical strategy (early, mixed, intermediate or late) (Picard et al., 2021). The remaining ones are centered around tools, including the type of classification that is performed (supervised, semi-supervised or unsupervised), or the mathematical field from which the algorithms are derived from (Bersanelli et al., 2016) (machine learning, deep learning, graph theory, or Bayesian statistics).

In this context, in this chapter, I will first recapitulate the proposed taxonomies for multi-omics approaches (with tools and application examples for each strategy). I will then

elaborate on shared key concepts of multimodal integration, and provide a detailed explanation of most popular tools. Finally, I will itemize most frequently encountered challenges during such integrations.

3.1 Multi-omics nomenclature and classification

Choosing an effective approach is demanding and not straightforward. This task is significantly dependent on data availability and research aims. Furthermore, the implementation of any given approach is hampered by challenges that are detailed below (section 3.3). As no gold standard taxonomy for multi-omics methods had been yet established, here we briefly describe the classification that are most frequently used, with a primary distinction between step-wise and horizontal & vertical integration techniques, as suggested by Wörheide et al. (2021).

3.1.1 Step-wise Integration

Step-wise, or late, integration (Cai et al., 2022; Picard et al., 2021; Wani & Raza, 2019) distinguishes properties across multiple molecular levels that are used to construct independent models. This form of integration is driven by layer-specific hypothesis and statistical proprieties, as well as methods for combining models in a comprehensive manner, while keeping data-specific attributes. Despite its simplistic formulation, step-wise integration presents heavy consequences. Indeed, while independent data modeling for each omic layer protects from a predominance of the omic layer with the highest dimensionality, as a counterpart, some cross-omic interactions may be neglected.

Step-wise analyses are by definition sequential, and include analyses such as eQTL and correlations between epigenomic regulators and gene expression. Yet, the core of this technique focuses on one omic at a time, with integration performed only afterwards. Hence, these techniques are not optimal for combining simultaneously the same kind of omic datasets from different cohorts or multiple distinct omics from the same cohort. To do so, one would prefer to employ holistic approaches such as horizontal and vertical integration.

3.1.2 Horizontal and Vertical Integration

With the growing number of integration tools and strategies, classification criteria were needed. In their extensive review, Zitnik et al. (2019) proposed a conceptual organization, and drew a distinction between horizontal integration, which consists in combining the same data type across multiple cohorts or studies, from the vertical one, in which different types of data from the same samples are integrated.

3.1.2.1 Horizontal Data Integration

One of the limitations of omic studies lies in their sample sizes.Indeed, in order to capture a significant biological signal, the number of samples required increases with the complexity of the phenotype under study. Unfortunately, limited resources and budgets, combined with constrained accessibility to biological samples, especially for psychiatric disorders, complicate the identification of statistically significant associations. Therefore, horizontal meta-analyses approaches have been developed to overcome the cohort size limitation, by combining studies and increasing statistical power.

Several horizontal approaches have been developed (Richardson et al., 2016), and meta-analyses can be considered a special case among them.Indeed, the overall goal of meta-analyses is to leverage a prior corpus of published studies to assess the statistical significance and effect sizes of their combined results, as well as their consistency. Hence, meta-analyses imposes selection criteria, where general horizontal approaches do not. One of the most common applications of horizontal meta-analyses are in GWAS. These methods are aggregating different statistics from formal GWAS studies and combine either their p-values or effect sizes (Zeggini, 2009).

Combining p-values. When the supporting data of the studies are either impossible to integrate or no longer accessible, combining p-values from independent statistical tests is a common method in meta-analyses. Since the initial idea proposed by Fisher (1934), the literature has seen a wide variety of p-value combination strategies (Heard & Rubin-Delanchy, 2018). Here, we present the three most commonly used :

- **Fisher's method (Fisher, 1934)** defines the test statistic as the sum of log-transformed p-values ($S_F = \sum_{i=1}^{n} \log p_i$).
- **Pearson's method (Pearson, 1933)** defines the test statistic as the sum of log-transformed 1 minus p-values ($S_p = -\sum_{i=1}^{n} \log (1 - p_i)$)
- **Stouffer's method (Stouffer, 1949)** defines the test statistic as the inverse of ϕ, the cumulative standard normal distribution ($S_s = \sum_{i=1}^{n} \phi^{-1} (p_i)$).

Combining effect sizes. The average of the effect sizes across the studies might theoretically be used to represent the total, or combined, effect size in a meta-analysis. Such over-simplification is, however, prohibited by the analyses' heterogeneity. Therefore, a weighted mean that favors accurate research over less trustworthy ones should be included, which raises the question of how to determine mean's weights. Choi et al. (2003) formulated the effect size model as follows: Let μ be the parameter of interest, the overall mean, and y_i be

the observed effect size for the ith study, knowing that there were k independent studies. The general model is given as:

$$\begin{cases} y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\left(0, s_i^2\right) \\ \theta_i = \mu + \delta_i, \quad \delta_i \sim \mathcal{N}\left(0, \tau^2\right) \end{cases}, \quad i = 1, \ldots, k. \tag{3.1}$$

where $\mathcal{N}\left(0, s_i^2\right)$ stands for the normal distribution centered in 0 and of variance s_i^2. Here, s_i^2 and τ^2 represent respectively the within and between study variance. To estimate the overall effect, two mains models were suggested and can be described as follows :

- **A Fixed Effect Models** (FEM) aggregates the effect size by applying a linear model with an underlying true effect size μ plus a random error ε_i in each study Chang et al., 2013. By doing so, it makes the assumption that in (3.1), the inter-study variance τ^2 is null and thus that variance between effects is only explained by sampling errors. Such a model can be simplified as $y_i \sim \mathcal{N}\left(\mu, s_i^2\right),\ i = 1, \ldots, k$, and the overall mean can be estimated as Choi et al. (2003):

  $$\hat{\mu}_{FEM} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i},$$

  where $w_i = s_i^{-2}$. Here, the fixed effect model implies a common basic effect upon which all of our observed effect sizes are predicated. Yet, the inter-study heterogeneity due to multiple covariates including, amongst others, age, ethnicity and methodological discrepancies could considerably impact the true effects. To overcome this issue, random effect models REM were developed.

- **A Random Effect Model** (REM) extends FEM assuming that each study is a random sampling with a specific mean θ_i and variance s_i^2 of a super population with an overall mean μ and variance τ^2, and the combined effects estimates the mean effect in this distribution. Choi et al. (2003) gives the estimation of the overall mean that can again be expressed as a weighted mean of the observed effects :

  $$\hat{\mu}_{REM} = \frac{\sum_{i=1}^{n} w_i^* y_i}{\sum_{i=1}^{n} w_i^*},$$

  where $w_i^* = \left(s_i^2 + \tau^2\right)^{-1}$. Again we can see that FEM is a special case of REM where the inter-study heterogeneity $\tau^2 = 0$.

Determining which method is appropriate for a body of studies comes down to estimating their heterogeneity. For this purpose, the total variance statistic Q, referred as Cochran's

Q, was proposed. Under the (null) hypothesis of homogeneity, it follows a χ^2_{k-1} distribution with k-1 degrees of freedom (where k is the number of studies).

3.1.3 Simultaneous Data Integration

3.1.3.1 Vertical Data Integration

Biological properties and activities, as well as disease-causing processes, are the result of intricate and complex interactions between cellular elements. As a result, a comprehensive understanding of these processes is challenging, and single-omic models (mono- or multi-omic horizontal integration) can only focus on certain biological mechanisms, and would benefit from being complemented by other biological information nested in a vertical integration strategy.

Vertical data integration is a powerful method to leverage massive amounts of data to get insights and produce new hypotheses on intricate etiologies. Vertical integration differs from the horizontal one in that it collects several levels of biological modalities on the same collection of samples. This method combines the complementary information obtained by each omic, while reducing their inherent noise, and may reveal a common signal associated with the phenotype under study. The enhanced data produced may subsequently be utilized for a variety of applications, cinluding for biomarker identification and patient stratification.

According to the desired output, integration approaches can be divided in 2 types: classification algorithms, which predict qualitative system outputs such as discrete classes (e.g patients from controls) or distinct states (e.g disease subtypes); and regression algorithms, which predict quantitative continuous values (e.g a clinical severity score). Furthermore, the algorithms are divided into 2 types based on their learning behavior: supervised or unsupervised (Picard et al., 2021). In addition, Bersanelli et al. (2016) suggested a classification of these methods according to their mathematical aspects. In the following section, we list the main classification criteria found in the literature and milestone algorithms, on which further sophisticated tools were developed.

3.1.4 Unsupervised and Supervised Integration

3.1.4.1 Unsupervised analyses

Unsupervised studies allow an unbiased appreciation of the data and can detect trends in omics, such as sample or variable clusters, without considering prior knowledge about either the sample groups or outcome information. The obtained classification is provided with an impartial representation of the data and explains how samples innately form sub-

groups according to their expression patterns, where similarity between samples is appreciated through a statistical criterion (Rappoport & Shamir, 2018). Initially, multi-omics methods are unsupervised methods aiming at deciphering the complex relation existing between multiple omics modalities. Several studies have conducted comprehensive comparisons or benchmarks of multi-omic approaches. For example, Meng et al. (2016) described their mathematical properties, while Tini et al. (2018) compared their classification performances, and Pucher et al. (2019) assessed the overlap between relevant predictors for classification models and known biological pathways. Recently, Cantini et al. (2021) released an extensive benchmark of joint dimensionality reduction methods, which is of particular interest in the context of our own work (see Chapter 4).

3.2 Integration strategies

3.2.1 Network based

Various strategies have been investigated to perform the integration and interpretation of multi-omic data. Besides statistical methods, network-based integration approaches show representation and interpretability advantages that captured the attention of many investigations. In comparison, statistical integration seeks to uncover overlapping patterns or common fingerprints across several modalities, without the need for prior information. Network-based integration, on the other hand, sees the biological system as an interconnected network of molecular entities and needs prior information or assumptions about its organization. Several network approaches have been developed, including knowledge-based, calculated co-expression and similarity networks.

3.2.1.1 Knowledge Based networks

The main concept behind network-based integration is to translate multi-omic modalities in the context of a knowledge framework as various interacting networks. Proteins, for example, can bind to DNA or RNA to control gene expression, and interactions between proteins, metabolites, and lipids can influence molecular machinery and signaling cascades. Over the last decade, tremendous effort has gone into developing appropriate databases on molecular annotations, systems biology, and molecular interaction networks (Yan et al., 2017). Currently, curated metabolic pathways can be found in several comprehensive databases including KEGG, Reactome, StringDB, and WikiPathways.

Typically, building context-specific multi-omic subnetworks using experimental data is accomplished by linking key molecular findings in a particular omic layers with other layers, using known relationships in aforementioned databases. For example, a list of DEGs

can be mapped into a Protein-Protein Interaction (PPI) databases via their corresponding proteins, forming PPI sub-networks. Nevertheless, this approach presents a fundamental disadvantage, which resides in a lack of direct interactions between molecules across omic layers, resulting in sparse subnetworks linking only a few of its components. To overcome this, researchers frequently construct sub-networks that include all omic features, their linkage, and direct interacting partners (also known as first-order interaction network). However, such networks might become too large to harness relevant information, or enable practical visualization.

3.2.1.2 Co-expressed gene networks

A co-expression gene network is an undirected graph where each node is associated with a gene and 2 genes are connected by an edge if their expression are considered as significantly similar. Co-expression gene networks are often constructed using transcriptome and interactome data. One of the main analyses made out of co-expression gene networks is to detect genes sub-networks (also called modules), followed by the identification, within these modules, of key genes (defined as "hub" genes) that are well-connected within its host module. Questioning the association between modules and a phenotype under study is also an important step in this type of analysis. Finally, using knowledge databases, overlaps, and correlations, additional omic features can be mapped to modules of interest. Gene co-expression analyses can be achieved using several tools and approaches that include Weighted Gene Co-expression Network Analysis (WGCNA) CoXpress (Watson, 2006), DiffCoEx (Tesson et al., 2010), TopoFun (Janbain et al., 2021) and DINGO (Ha et al., 2015). Weighted Gene Co-expression Network Analysis (WGCNA) is the most frequently used method.

WGCNA WGCNA was developed by Horvarth's team (Langfelder & Horvath, 2008) to enable the identification of groups of densely interacting genes based on their expression profiles, providing biological insights on the phenotype under study. WGCNA also enables retrieval of relevant hub genes within each module. This method has become the gold standard for constructing weighted co-expression modules. It starts by computing a correlation matrix, used as a measure of the co-expression between pairs of genes, followed by generating the Topological Overlap Matrix (TOM), used to organize genes into modules by performing hierarchical clustering. For better understanding of the WGCNA algorithm, the following sections describe its key notions.

Correlation Matrix This first step consists in assessing co-expression between genes by applying a measure of similarity. In WGCNA, this is performed through correlation. More

specifically, let $\mathbf{S} = (s_{ij}) \in \mathbb{R}^{p \times p}$ be the similarity matrix that subsumes all two-by-two similarities between the p considered genes and let the vectors $\mathbf{x}_i \in \mathbb{R}^n$ and $\mathbf{x}_j \in \mathbb{R}^n$ respectively represent the expression profile of the genes i and j over the n considered samples. In order to have a similarity coefficient c_{ij} in the range $[0; 1]$ associated with the co-expression of these 2 genes, WGCNA proposes 2 measures based on correlation. One that do not keep the sign of the correlation:

$$s_{ij} = |cor(\mathbf{x}_i, \mathbf{y}_j)|,$$

and one that preserves the information on the correlation sign:

$$s_{ij} = \frac{1 + cor(\mathbf{x}_i, \mathbf{y}_j)}{2}. \tag{3.2}$$

In the R package of WGCNA, several ways of computing the correlation are proposed, amongst which Pearson's correlation, Spearman's correlation, and Bi-weight mild-correlation (BiCor). As Spearman's correlation is computed between 2 ranked variables, it usually provides an estimation of correlation that is more robust to outliers (if one sample of two variables, for example ranked as the highest value, is multiplied by two, it could considerably change Pearson's, but not Spearman's correlation). BiCor is also an interesting way of estimating the link between 2 variables. If 2 gene expressions measurements composed of n samples are considered, $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$, BiCor can be defined, similarly to the Pearson correlation, as:

$$cor^{Bi}(\mathbf{x}, \mathbf{y}) = \frac{cov^{Bi}(\mathbf{x}, \mathbf{y})}{\sqrt{cov^{Bi}(\mathbf{x}, \mathbf{x}).cov^{Bi}(\mathbf{y}, \mathbf{y})}}. \tag{3.3}$$

where the covariance is defined as:

$$cov^{Bi}(\mathbf{x}, \mathbf{y}) = \sum_{i}^{n} \left([x_i - med(\mathbf{x})]w_{x_i}\right) \left([y_i - med(\mathbf{y})]w_{y_i}\right), \tag{3.4}$$

where $med(\mathbf{x})$ and $med(\mathbf{y})$ correspond to the median associated with each gene expression. In (3.4), compared to the usual definition of the covariance, the mean is replaced by the median and each observation x_i (resp. y_i) is weighted by w_{x_i} (resp. w_{y_i}) computed as follows:

$$w_{x_i} = [1 - (RS(x_i))^2]^2 I\left(RS(x_i)\right),$$

$$\text{where } RS(x_i) = \frac{x_i - med(\mathbf{x})}{9.med(|\mathbf{x} - med(\mathbf{x})\mathbb{1}_n|)} \text{ and } I(RS(x_i)) = \begin{cases} 0 & \text{if } |RS(x_i)| \geq 1 \\ 1 & \text{otherwise} \end{cases}.$$

where $\mathbb{1}_n$ stands for column vector of size n made only of ones. Furthermore, $w_{x_i} = 1$ in a case where a sample equals its median and $w_{x_i} = 0$ in cases where the difference between a sample and its median is greater than 9 times its Median Absolute Deviation (MAD). In this last case, the sample is so different from its median that it can be considered an outlier, which explain why RS is sometimes called the Robust Scale, and how, similarly to Spearman's correlation, BiCor is considered an estimation of correlation that is robust to outliers (Langfelder & Horvath, 2012).

Adjacency Matrix In graph theory, a graph $G = (V, E)$ is represented by a set of vertices V, containing all the nodes (in our cases, a node corresponds to a gene), and a set of edges E that connect the nodes together. Furthermore, an adjacency matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p \times p}$, where p is the number of nodes (genes), can be associated with this graph and encodes the connection strength between each pairs of nodes. A graph is said to be unweighted if the adjacency matrix is only made of 0 and 1 to describe if a connection is present or not. When it is weighted, the adjaceny matrix is made of real positive values that weights all the edges in the graph.

The next step in WGCNA consists in estimating the adjacency matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{n \times n}$ out of the similarity matrix $\mathbf{S}$.In the unweighted scenario, hard-thresholding is usually applied to estimate which genes are connected or not. It consists in defining a threshold τ for similarities for which a connection is considered to exist:

$$a_{ij} = \begin{cases} 0, & \text{if } s_{ij} \leq \tau \\ 1, & \text{if } s_{ij} > \tau \end{cases}$$

This threshold is arbitrary, and does not take into account the intra-organization of genes in a biological system.

In the weighted scenario, WGCNA proposed a soft-thresholding of the similarity matrix:

$$a_{ij} = |s_{ij}|^{\beta} \tag{3.5}$$

where the power parameter $\beta \in \mathbb{N}^{\star}$. The higher β, the more small similarity values are shrunk down to zero, making the estimation of the weighted adjacency matrix potentially more robust to background noise.

In Zhang and Horvath, 2005, hard and soft-thresholding has been thoroughly compared. They strongly advocate for the use of soft-thresholding as it leads to more relevant adjacency matrices with regard to the underlying biological systems. Indeed, as explained in Zhang and Horvath, 2005, biological networks are well described by scale-free networks. They are defined as networks where the probability that a node is connected to k other nodes decay as the power of $1/k$. Such networks are extremely heterogeneous and are dominated by highly connected nodes (called hubs) which link the rest of the less connected nodes to the system. Indeed, the genes present in the human genome are the results of a long evolution and a preferential selection that favors interactions among genes (a gene may play a role in several functions). Thus, the appearance of modules (densely connected sub-networks) occurs. Soft-thresholding tends to produce adjacency matrices that better fit the scale-free networks'propriety.

Setting the parameter β As mentioned above, a biological network is often well described by a scale-free network which has the property that the degree of a node in this graph follows a power law distribution. Thus, if we introduce the probability density function $p : \mathbb{R}^{+} \mapsto [0, 1]$ of the degree of a node, it can be written as:

$$p\left(k\left(\beta\right)\right) = c.\left[k\left(\beta\right)\right]^{-\gamma} \tag{3.6}$$

where $c \in \mathbb{R}^{+}$ and $\gamma \in \mathbb{R}^{+}$ are two positive constants. Furthermore, the degree k is written as a function of β as it depends on this parameter. From (3.6), a linear dependency between the probability density and the degree can be explicited as follows:

$$\log\left[p\left(k\left(\beta\right)\right)\right] = log\left(c\right) - \gamma \log\left[k\left(\beta\right)\right] \tag{3.7}$$

Langfelder and Horvath (2008) advocates to choose a β power that respects the scale-free topology by verifying that the linear link between $log\left[p\left(k\left(\beta\right)\right)\right]$ and $\log\left[k\left(\beta\right)\right]$ is supported by a $R^2 > 0.8$.

Topological Overlap At this stage, gene modules could be estimated through a hierarchical clustering algorithm applied on the adjacency matrix. However, the adjacency matrix could be limited as it only describes the similarity between two genes by a function of their correlation, which does not take into account their surroundings in the graph, and which

also could be subject to biases due to background noises. In order to remedy this, Li and Horvath (2007) proposed a measure of similarity based on the graph's topological characteristics, allowing to compute the Topological Overlap Matrix (TOM) noted $\Omega = (\omega_{ij}) \in \mathbb{R}^{p \times p}$. In the TOM, the similarity between two genes traduces the level of overlapping between the neighboring of the two considered genes. Thus, it not only allows anymore to group genes with high level of adjacency, but it also could cluster them together if their communities highly overlap. Hence, the construction of modules based on TOMs is more biologically significant. It can be computed as:

$$\omega_{ij} = \frac{l_{ij} + a_{ij}}{min(k_i, k_j) + 1 - a_{ij}} \tag{3.8}$$

where :

$$l_{ij} = \sum_{u=1}^{p} a_{iu} a_{uj} \qquad \& \qquad k_i = \sum_{u=1}^{p} a_{iu}$$

ω_{ij} is comprised in the range $[0; 1]$. By convention $a_{ii} = 0$, $\forall i \in \{1; \ldots; n\}$. For the sake of interpretation, let us consider the case of an adjacency matrix computed by hard-thresholding. In this context, k_i corresponds to the degree of gene i (number of genes connected to this gene) and l_{ij} to the number of genes both connected to i and j. Hence, on one hand, $\omega_{ij} = 0$ when the two genes are not connected and do not share any common neighbor. On the other hand, $\omega_{ij} = 1$ when they are connected and when the gene associated with the smallest number of neighbors has all of them in common with the other gene. This last situation is associated with a high level of overlapping between the topology of the two genes, hence a high value of ω_{ij}.

Module Construction

Hierarchical Clustering In order to group genes into clusters (the so-called modules) WGCNA relies on hierarchical clustering performed on the distance/dissimilarity matrix $\mathbf{D}_\omega = (1 - \omega_{ij}) \in \mathbb{R}^{p \times p}$. An important part in hierarchical clustering is to define the linkage criterion in order to compute the distance between two modules. In WGCNA, the Unweigthed Average Linkage Clustering is used. If 2 set of genes M_i and M_j are considered, the distance between them, $d(M_i, M_j)$, can be computed as follows:

$$d(M_i, M_j) = \frac{1}{|M_i|.|M_j|} \sum_{k \in M_i} \sum_{l \in M_j} (1 - \omega_{kl}) \tag{3.9}$$

where $|M_i|$ and $|M_j|$ are associated with the cardinal of M_i and M_j respectively. We can

notice that, as all elements of the TOM are between 0 and 1, these distances also are in the range $[0; 1]$.

Cut Tree Relying on the clustering results, WGCNA performs a cut from a constant height (set here to cutHeight = 0.2), and will only keep modules of a size larger than N (min size = 100 for our study). DynamicCutTree, an alternative top-down method that relies on an iterative process of cluster decomposition using floating cut heights until it converges to a stable number of cluster (see (Langfelder & Horvath, 2008)), is also implemented in the R package, and was also tested but is not presented in this report. In favor of less redundant and more biologically significant modules, Langfelder and Horvath (2008) recommends the merging of modules whose first principal component Pearson correlation r > 0.75.

WGCNA consensus A variant of the classical WGCNA is WGCNA Consensus, which differs from the original one only by the plural number of datasets (associated with the same modality) used for the construction of the correlation matrix. Indeed, WGCNA consensus allows the study of the commonalities in the transcriptional organization of the datasets. One of its main advantages emerges intrinsically, as the generated consensus network is common to several datasets even if they originate from different studies or analyses.

Module's preservation Module's preservation aims at quantifying the robustness of a module. A series of metrics that well describe a module were defined. In order to conclude if this module is well-preserved or if it is data specific, these metrics are evaluated for the same module, computed on an external dataset. Langfelder and Horvath (2008) proposed a composite preservation statistics (Z-summary), summarizing several conservation metrics based on graph theory. A total of 8 preservation measurements is calculated and summarized in Z-summary (for a detailed overview, see (Ray et al., 2017)). Based on simulation studies, Langfelder and Horvath (2008) advocated the following standard thresholds:

$$\begin{cases} 2 > & \text{Z-summary} & & : \text{Not preserved.} \\ 2 < & \text{Z-summary} & < 10 & : \text{Moderately preserved.} \\ & \text{Z-summary} & > 10 & : \text{Highly preserved.} \end{cases}$$

In addition to the Z-summary, two other principles are key in the evaluation of a module's preservation:

- **Density:** $D = \frac{1}{|M|(|M|-1)} \sum_{i \in M} \sum_{j \in M} a_{ij}$, where M is a set of nodes belonging to the considered module. Density based preservation estimates whether a module remains highly connected in a network.

- **Connectivity:** scaled connectivity $k_i^{(s)}$:

$$k_i^{(s)} = \frac{k_i}{\max_{j \in M} k_j}$$

where k_i is the number of nodes connected to gene i in a hard-thresholding context. In a soft-thresholding case, as defined before, $k_i = \sum_{u \in M} a_{iu}$. Here, the node connectivity (degree) is scaled by the maximum connectivity in the graph considered. Connectivity based preservation measures how well the node's connection is preserved in the tested dataset.

Eigengenes The correlation of modules with categorical and quantitative phenotypes were achieved by summarizing the module's expression into an eigen vector called eigengene. This eigengene is a representative vector extracted from Singular Value Decomposition (SVD) of the WGCNA fed data matrix, allowing a great reduction in the number of statistical tests.

Assuming that $\mathbf{X} \in \mathbb{R}^{m \times n}$ represents the $m \times n$ data matrix, and introducing $r = \min(n, m)$ and $\top$ the transpose operator, its singular value decomposition is given by :

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^\top \tag{3.10}$$

where $\mathbf{U} \in \mathbb{R}^{m \times r}$ and $\mathbf{V} \in \mathbb{R}^{n \times r}$ denotes respectively the left and right matrix of singular vectors. Both matrices are semi-orthogonal, $\mathbf{U}^\top\mathbf{U} = \mathbf{V}^\top\mathbf{V} = \mathbf{I}_r$, and the square one is even orthogonal, $\mathbf{V}\mathbf{V}^\top = \mathbf{I}_r$ (in the case where $r = n$). $\mathbf{D}$ is a non-negative diagonal matrix of size r whose elements are the singular values. In a WGCNA context, where $\mathbf{X}$ is a gene expression matrix, $\mathbf{U}$ and $\mathbf{V}$ are interpreted as the summary of respectively the gene and sample's profiles.

3.2.1.3 SNF

As shown in the above section, networks are a convenient way of representing complex datasets. Such type of integration has grown in popularity in biological and medical fields. With the assumption that networks generated on multiple datasets and looking at the same phenomenon in different ways are in fact different representations of a common underlying phenotype coded in a hidden common network, one may wish to fuse the generated networks on each dataset to discover the structure of the underlying, unknown, one (Tyler et al., 2021).

In this context, Wang et al. (2014) proposed Similarity Network Fusion (SNF), an unsupervised transformation-based approach that first computes for each dataset a similar-

ity network. These networks are computed at the sample level and thus have the same number of nodes (one per sample). They describe how close pairs of samples are in each dataset. Then, SNF iteratively propagates the information of a network to the others until all networks converge to a single aggregated one. The efficiency of SNF had initially been demonstrated on glioblastoma by integrating miRNA, mRNA and DNAm data to derive three cancer subtypes (Wang et al., 2014).

Let us consider L data matrices, $\mathbf{X}_1, \ldots, \mathbf{X}_L$, that all have the same number of samples but potentially a different number of variables: $\forall l \in \{1, \ldots, L\}$, $\mathbf{X}_l = \left[\mathbf{x}_{l1}^\top \ldots \mathbf{x}_{ln}^\top\right]^\top \in \mathbb{R}^{n \times p_l}$, where $\mathbf{x}_{li} \in \mathbb{R}^{p_j}$ actually describes individual i over all its p_l variables measured in dataset l. For each data matrix, after normalizing data columns to have mean equal to zero and standard deviation equal to one, SNF begins by calculating for each omics dataset a similarity network $G_l = (V_l, E_l)$ where each vertex of V_l corresponds to a sample $\mathbf{x}_{li}$.

SNF starts by computing, for each dataset, a similarity matrix $\mathbf{W}_l = \left(w_{ij}^{(l)}\right) \in \mathbb{R}^{n \times n}$ that corresponds to the weights of all edges in graph G_l. According to Wang et al. (2014), a scaled exponential similarity kernel is used to compute each element of $\mathbf{W}_l$:

$$w_{ij}^{(l)} = \exp\left(\frac{-\|\mathbf{x}_{li} - \mathbf{x}_{lj}\|_2^2}{\alpha \varepsilon_{ij}^{(l)}}\right) \tag{3.11}$$

where $\|.\|_2$ is the ℓ_2-norm (the square root of the sum of squares of all the elements), α is a hyper-parameter, usually chosen in the interval $[0.3, 0.8]$, that acts as a variance parameter, and $\varepsilon_{ij}^{(l)}$ describes how close sample i and j are to each other and also to their respective K-nearest neighbors. It can be computed as:

$$\varepsilon_{ij}^{(l)} = \frac{\|\mathbf{x}_{li} - \mathbf{x}_{lj}\|_2 + \frac{1}{|N_i^l|}\sum_{k \in N_i^l}\|\mathbf{x}_{li} - \mathbf{x}_{lk}\|_2 + \frac{1}{|N_j^l|}\sum_{k \in N_j^l}\|\mathbf{x}_{lj} - \mathbf{x}_{lk}\|_2}{3}, \tag{3.12}$$

where N_i^l and N_j^l are the sets of K nearest neighbors (estimated with a K-nearest-neighbors (kNN) algorithm on the original standardized dataset) of samples i and j respectively, and $|N_i^l|$, $|N_j^l|$ defines their cardinal. Based on $\mathbf{W}_l$, two additional matrices are computed, $\mathbf{P}_l^{(0)} = \left(p_{ij}^{(l,0)}\right) \in \mathbb{R}^{n \times n}$, a normalized version, and $\mathbf{S}_l = \left(s_{ij}^{(l)}\right) \in \mathbb{R}^{n \times n}$ that focuses more on local similarities. Normalization is performed as follows:

$$p_{ij}^{(l,0)} = \begin{cases} \frac{w_{ij}^{(l)}}{2\sum_{k\neq i} w_{ik}^{(l)}} & j \neq i \\ 1/2 & j = i \end{cases} \tag{3.13}$$

This normalization step allows for each line to sum to one. Moreover, it forces self-similarities to be all equal across samples, with any given sample being always more similar to itself than to other samples (which, if not controlled for, could lead to instabilities) (Wang et al., 2014). Local similarities are estimated as:

$$s_{ij}^{(l)} = \begin{cases} \frac{w_{ij}^{(l)}}{2\sum_{k\in N_i^l} w_{ik}^{(l)}} & j \neq N_i^l \\ 0 & \text{otherwise} \end{cases} \tag{3.14}$$

Local similarity matrices S_l are modified similarity matrices that only consider the nodes that are in their K nearest neighboring. Otherwise, the similarity is null. SNF's procedure is described in algorithm 1.

If we introduce $\bar{\mathbf{P}}_l^{(t)} = \left(\bar{p}_{km}^{(l,t)}\right) = \frac{1}{L-1}\sum_{\substack{r=1 \\ r\neq l}}^{L} \mathbf{P}_r^{(t)}$, then, the merging step (3.15) can be rewritten as :

$$\tilde{p}_{ij}^{(l,t+1)} = \sum_{k\in N_i^l} \sum_{m\in N_j^l} s_{ik}^{(l)} s_{jm}^{(l)} \bar{p}_{km}^{(l,t)}. \tag{3.17}$$

This merging step can simply be seen as a weighted sum of the elements of $\bar{\mathbf{P}}_l^{(t)}$ associated with a pair of samples that respectively belongs to the K nearest neighbors of samples i and j. In this weighted sum, the weights depend on the local similarity that these samples have with i and j in graph G_l.

Despite numerous advantages, SNF suffers from limitations, including the number of hyper-parameters (the number of iteration T_{max}, the number of K neighbors and the hyper-parameter α) that must be given by the user or tweaked (which can be time-consuming). Furthermore, these parameters have an influence on the produced clusters. Another important limitation is that the method uses all the variables in the similarity network computation, therefore some coherent but yet subtle interactions on each omic could be overlooked. Hence, variable selection and supervised learning could be beneficial to SNF.

Algorithm 1: Algorithm for Similarity Network Fusion (SNF)

Data: $\mathbf{P}_1^{(0)}, \ldots, \mathbf{P}_L^{(0)}, \mathbf{S}_1, \ldots, \mathbf{S}_L, T_{max}$
Result: $\mathbf{P}^{(c)}$
Init: $t \leftarrow 0$;
while $t < T_{max}$ **do**
  for $l = 1, \ldots, L$ **do**
    ▷ Merging Step

$$\tilde{\mathbf{P}}_l^{(t+1)} = \mathbf{S}_l \left(\frac{1}{L-1} \sum_{\substack{k=1 \\ k \neq l}}^{L} \mathbf{P}_k^{(t)} \right) \mathbf{S}_l^{\top} \tag{3.15}$$

  end
  for $l = 1, \ldots, L$ **do**
    ▷ Normalization Step

$$p_{ij}^{(l,t+1)} = \begin{cases} \frac{\tilde{p}_{ij}^{(l,t+1)}}{2 \sum_{k \neq i} \tilde{p}_{ij}^{(l,t+1)}} & j \neq i \\ 1/2 & j = i \end{cases} \tag{3.16}$$

  end
  $t \leftarrow t + 1$
end
$\mathbf{P}^{(c)} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{P}_l^{(T_{max})}$

3.2.2 Machine learning based

Next generation sequencing technologies generate very high dimensional data, where the number of variables p is much higher than the number of samples n ($p >> n$). Such a situation is prone to overfitting, as the more predictor you have, the more likely it will be possible to perfectly predict your outcome. Overfitting leads to statistical models with poor generalization performances with regard to their capacity to correctly predict the outcome of a new, unseen, sample.

To mitigate the impact of overfitting, several strategies have been proposed. Amongst them, regularization is used to enforce regularity in the models, potentially preventing it from perfectly fitting the outcome. For example, the well-known ℓ_1 and ℓ_2 regularizations are often imposed through penalizations or constraints. The ℓ_1 regularization also leads to sparse models, which means they only need a small subset of the variables to predict the

outcome. This is also called variable selection, and it can help to reduce overfitting. Finally, a last strategy consists in projecting the data into a lower dimensional space. Variables are combines in a meaningful way (for example such that the aggregated variable explains a high portion of the variability in the dataset), which, similarly to variable selection, reduces the number of predictors.

In a sense, all these strategies consist in putting more constraints on the statistical models (enforce regularity, with a few number of predictors, leaving in a low-dimensional space). However, these models have to be combined with a procedure that evaluates to what extent they indeed avoid overfitting the outcome. In this context, Cross-Validation (CV) was proposed. In a nutshell, CV consists in separating the data into a train set and a test set, fit the model on the train set and then, apply it on the test set in order to have an insight on the generalization capacity of the model. Starting from there, several re-sampling techniques are possibles for CV: K-fold CV (separate the dataset into K folds, train the model onto $K-1$ folds, test it on the remaining one and finally iterates onto every possible combination of $K-1$ folds), repeated K-folds (same procedure as before, but the initial splitting of the dataset into K-folds is repeated multiple times), Monte-Carlo CV (randomly sample the train test, without replacement, and use the remaining samples as the test set. Repeat the procedure a high number of times)...

Thus combining constraint models and CV, the goal is to find the constraint parameters (the regularization parameters, the number of selected variable, the size of the lower-dimensional space) that leads to the models that perform best in terms of generalization capacity.

Dimension reduction methods have not only been proposed to avoid overfitting. They are also a well-established tool to investigate the intricate links that may exist between a collection of datasets and, as such, they are the focus of the next section.

3.2.2.1 Dimension Reduction Methods

Dimension reduction methods aims at finding a lower dimensional representation of a data matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$, where n is the number of samples (rows) and p is the number of variables (columns). For example, Matrix factorization, as illustrated in the above figure 3.1, is a multivariate method that decomposes an original matrix into 2 sub-matrices. The first one, called components or factors matrix, of dimension $n \times k$, summarizes the information at the sample level and is designed to aggregate the information from all p variables into k factors. The second one, called loadings matrix, of size $k \times p$, summarizes the data at the variable space (p). Given $k << p$, the original matrix is now described using a much smaller dimensional space.

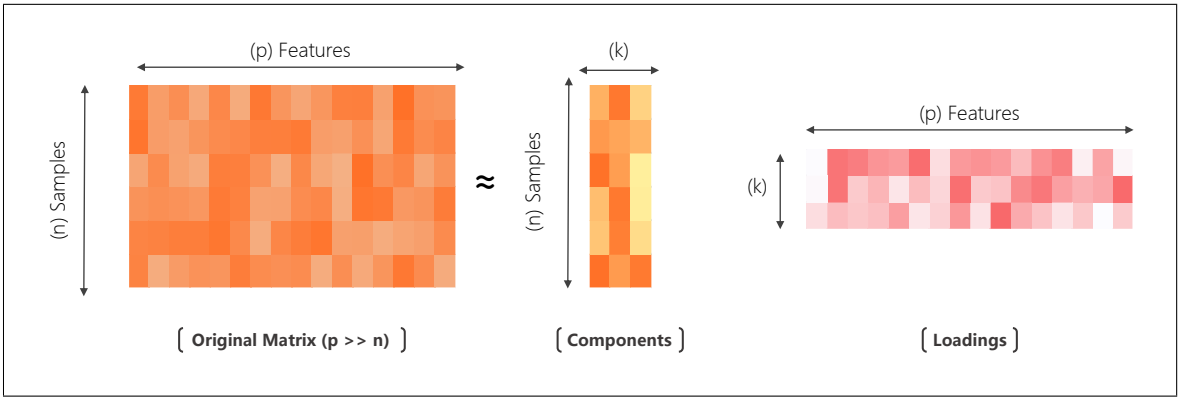


Figure 3.1: Matrix factorization: example of a data matrix with n samples and p features decomposed into a set of k components and k corresponding loading vectors.

Amongst the classical dimension reduction methods for a single matrix, Singular Value Decomposition (SVD) and Independent Component Analysis (ICA) are linear models while non-negative matrix factorization (NMF) is non-linear. Furthermore, Canonical Correlation Analysis (CCA) and Co-Inertia Analysis (CIA) propose to jointly reduce the dimension of two matrices in a linear way. Such methods are often used as the first step for the development of more elaborated algorithms (Cantini et al., 2021).

CCA Canonical Correlation Analysis (CCA) aims at identifying canonical components (linear combination of variables), from each considered data matrix, such that their correlation is maximized. Let us considered two datasets $\mathbf{X}_1 \in \mathbb{R}^{n \times p_1}$ and $\mathbf{X}_2 \in \mathbb{R}^{n \times p_2}$, measured on the same set of samples (all variables are assumed to be centered). The optimization problem of CCA can be written as follow:

$$(\widehat{\mathbf{w}}_1, \widehat{\mathbf{w}}_2) = \underset{\mathbf{w}_1, \mathbf{w}_2}{\operatorname{argmax}} \operatorname{cor}\left(\mathbf{X}_1 \mathbf{w}_1, \mathbf{X}_2 \mathbf{w}_2\right)$$

Such optimization criterion is associated with the extraction of te first canonical component from each block. Actually, a more general optimization problem could be stated in order to extract simultaneously an arbitrary number of canonical components from each block (matrix), with the additional constraint that, similarly to PCA, they are orthogonal to one another.

RGCCA Regularized Generalized Canonical Correlation Analysis (RGCCA) (Tenenhaus & Tenenhaus, 2011) is a multi-block component method that was proposed as a generalization of CCA to handle more than 2 high-dimensional matrices. Let us consider a collection of L data matrices $\mathbf{X}_1, \ldots, \mathbf{X}_L$. We assume that all variables are centered. RGCCA aims at finding block components $\mathbf{y}_l = \mathbf{X}_l \mathbf{w}_l$, $l = 1, \ldots, L$ (where $\mathbf{w}_l \in \mathbb{R}^{p_l}$) that best sum-

marize the relevant information between and within the blocks. RGCCA is defined with the following optimization problem Tenenhaus and Tenenhaus (2011) and Tenenhaus et al. (2017):

$$\begin{aligned} &\underset{\mathbf{w}_1,\ldots,\mathbf{w}_L}{\operatorname{argmax}} \sum_{k,l=1}^{L} c_{kl} g\left(\operatorname{Cov}\left(\mathbf{X}_k\mathbf{w}_k, \mathbf{X}_l\mathbf{w}_l\right)\right) \\ &\text{s.t.} \left(1-\tau_l\right) \operatorname{Var}\left(\mathbf{X}_l\mathbf{w}_l\right) + \tau_l \left\|\mathbf{w}_l\right\|_2^2 = 1,\ l = 1,\ldots,L. \end{aligned} \tag{3.18}$$

where $\|.\|_2$ is the ℓ_2-norm (the square root of the sum of squares of all the elements) and:

- The scheme function g is a continuously differentiable convex function. Among the commonly used scheme functions are: the Horst scheme which maximizes the sum of covariances between block components. The centroid scheme maximizes the sum of covariances' absolute values; the factorial scheme maximizes the sum of squared covariances; and the m-scheme maximizes the sum of the m power of covariances.
- The design matrix $\mathbf{C} = (c_{kl}) \in \mathbb{R}^{L\times L}$ is a symmetric matrix composed on non-negative elements that encode the connections between the L different blocks that the user wants to take into account. Usually $c_{lk} = 1$ for two connected blocks and 0 otherwise. Figure 3.2 presents the most used designs, amongst which, the complete (FULL) design, when all possible connections between blocks are taken into accounts. The hierarchical (NULL) design, when all blocks are connected to a single matrix/vector. An intermediate design with weighted links (Weighted)
- τ_l is a shrinkage parameter belonging to the interval $[0, 1]$. It allows a compromise between covariance $(\tau_l = 1)$ and correlation $(\tau_l = 0)$.

These parameters allow for RGCCA to be a very flexible framework that encompasses several classical multi-block methods (see Tenenhaus et al., 2017 for a detailed overview). Sparse Generalized Canonical Correlation Analysis (SGCCA) (Tenenhaus et al., 2014) is an extension of RGCCA that enforces feature selection through the combination of the ℓ_1 (the sum of the absolute values) and the ℓ_2 norm penalties on the weight vectors associated with each block.

MCIA Co-inertia analysis (CIA) (Dolédec & Chessel, 1994) was first developed in order to comprehend the environment-species interaction. The term inertia refers to the sum of variances in a dataset, and co-inertia to the sum of square covariances between distinct data matrices. CIA aims at finding co-inertia axes such that the square covariances between each dataset projected onto these axes is maximized (Dray et al., 2003). A generalization of CIA, called multiple co-inertia analysis (MCIA), was proposed in (Chessel &

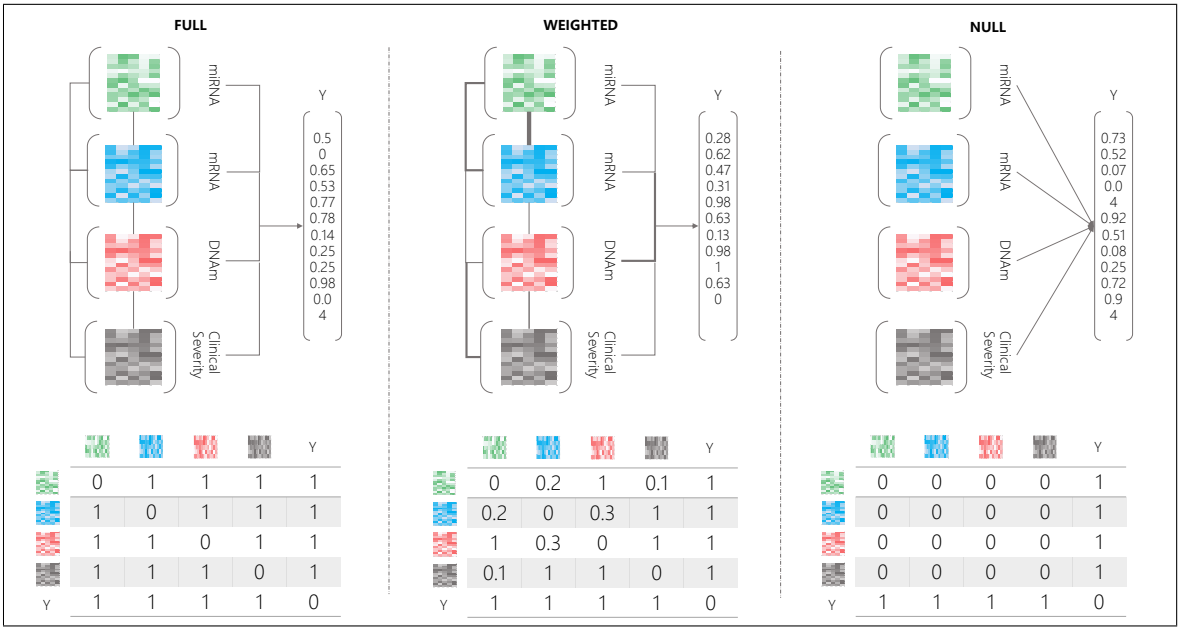


Figure 3.2: Different design matrices in RGCCA: multiomics data integration.

Hanafi, 1996) in order to study the co-inertia between an arbitrary number of datasets. Let us consider L datasets $\tilde{\mathbf{X}}_1, \ldots, \tilde{\mathbf{X}}_l$ that may have been through an appropriate transformation (amongst ordination methods such as for example PCA, Correspondence Analysis (CA), Non-Symmetric CA (NSCA) or amongst different centering/scaling strategies; for a detailed overview, the reader is mainly referred to Dray et al., 2003 and also to Culhane et al., 2003; Meng et al., 2014). The original optimization problem associated with MCIA can be stated as:

$$\begin{aligned} &\underset{\tilde{\mathbf{v}},\tilde{\mathbf{w}}_1,\ldots,\tilde{\mathbf{w}}_L}{\operatorname{argmax}} \sum_{l=1}^{L} \omega_l \left(\tilde{\mathbf{v}}^\top \mathbf{D}\tilde{\mathbf{X}}_l \mathbf{Q}_l \tilde{\mathbf{w}}_l\right)^2 \\ &\text{s.t.} \begin{cases} \tilde{\mathbf{w}}_l^\top \mathbf{Q}_l \tilde{\mathbf{w}}_l &= 1 \quad , l = 1, \ldots, L. \\ \tilde{\mathbf{v}}^\top \mathbf{D}\tilde{\mathbf{v}} &= 1 \end{cases} \end{aligned} \tag{3.19}$$

where $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix, composed of non-null and non-negative elements such that $\mathrm{Tr}(\mathbf{D}) = 1$ (where Tr is the trace operator, i.e. the sum of the diagonal elements of a matrix), used to measure distances between variables in the sample space and $\mathbf{Q}_l \in \mathbb{R}^{p_l \times p_l}$, $l = 1, \ldots, L$ are symmetric positive definite matrices used to measure distances between samples in the variable space. Most of the time, $\mathbf{D}$ and $\mathbf{Q}_l$ are defined as identity matrices, but in specific applications, more expert matrices can be used that emphasize the importance of a group of observations or a group of variables for the computation of distances (see (Min & Long, 2020) for more explanations). Furthermore, as specified in (Min

& Long, 2020), $\omega_l, l = 1, \ldots, L$ are non-null, non-negative elements specified by the user, such that $\sum_l \omega_l = 1$ and that a priori define the importance of each block in the inference of the common co-inertia axes.

Similarly to Min and Long (2020), let us consider the following change of variables: $\forall l \in \{1, \ldots, L\}$, $\mathbf{X}_l = n\sqrt{\omega_l}\mathbf{D}^{1/2}\tilde{\mathbf{X}}_l\mathbf{Q}_l^{1/2}$, $\mathbf{w}_l = \mathbf{Q}_l^{1/2}\tilde{\mathbf{w}}_l$ and $\mathbf{v} = \mathbf{D}_l^{1/2}\tilde{\mathbf{v}}$, which would led to the new optimization criterion:

$$\begin{aligned} &\underset{\mathbf{v},\mathbf{w}_1,\ldots,\mathbf{w}_L}{\operatorname{argmax}} \sum_{l=1}^{L} \left(\frac{1}{n}\mathbf{v}^\top \mathbf{X}_l \mathbf{w}_l\right)^2 \\ &\text{s.t.} \begin{cases} \|\mathbf{w}_l\|_2^2 &= 1 \quad , l = 1, \ldots, L. \\ \|\mathbf{v}\|_2^2 &= 1 \end{cases} \end{aligned} \tag{3.20}$$

Based on the algorithm derived in Chessel and Hanafi, 1996 to solve (3.20), it can be shown that, if each block $\mathbf{X}_l$ is centered for each variable, this last optimization problem is equivalent to a much more common one:

$$\begin{aligned} &\underset{\mathbf{w}_1,\ldots,\mathbf{w}_L,\mathbf{w}_{L+1}}{\operatorname{argmax}} \sum_{l=1}^{L} \operatorname{Cov}^2\left(\mathbf{X}_l\mathbf{w}_l, \mathbf{X}_{L+1}\mathbf{w}_{L+1}\right) \\ &\text{s.t.} \begin{cases} \|\mathbf{w}_l\|_2^2 &= 1 \quad , l = 1, \ldots, L. \\ \operatorname{Var}\left(\mathbf{X}_{L+1}\mathbf{w}_{L+1}\right) &= 1 \end{cases} \end{aligned} \tag{3.21}$$

where $\mathbf{X}_{L+1} = [\mathbf{X}_1 \ldots \mathbf{X}_L]$, which is called a superblock in the multi-block literature. This optimization problem goes by many names: MAXVAR-A, Covariance criterion, Split principal component, Multiple Factor Analysis, Consensus PCA, CPCA-W (see Tenenhaus et al., 2017 for a detailed overview). It can be shown that it is a special case of RGCCA (see Tenenhaus et al., 2017 for more explanations).

In Meng et al. (2014), MCIA was introduced to the literature of multi-omics, with the particular emphasis on a combination of steps: (i) PCA, CA or NSCA is applied separately on each dataset to project them onto comparable lower dimensional spaces, (ii) a specific centering/scaling strategy is employed (considering a generic dataset $\mathbf{Z} = (z_{ij}) \in \mathbb{R}^{n\times p}$ and $\tilde{\mathbf{Z}}$ its centered/scaled version, then $\tilde{z}_{ij} = z_{ij}/z_{i+} - z_{+j}/z_{++}$, where $z_{i+} = \sum_j z_{ij}$, $z_{+j} = \sum_i z_{ij}$ and $z_{++} = \sum_{i,j} z_{ij}$) and (iii) MCIA is applied with $\mathbf{D}$ defined as an identity matrix and $\mathbf{Q}_l$ as a diagonal matrix where each element correspond to the sum of the associated variable through all samples (before the centering/scaling procedure).

JIVE Joint and Individual Variation Explained (JIVE) (Lock et al., 2013), is a component-based integration method for multi-block data analysis. Let us consider L data block $\mathbf{X}_1, \ldots, \mathbf{X}_L$, where each block is scaled by its Frobenius norm (the square-root of the sum of squares of all elements in the matrix). JIVE proposes to decompose each block $\mathbf{X}_l \in \mathbb{R}^{n \times p_l}$ as the sum of a joint structure $\mathbf{J} \in \mathbb{R}^{n \times (\sum_l p_l)}$ that bears commonalities between all blocks, and an individual structure $\mathbf{A}_k \in \mathbb{R}^{n \times p_l}$ that is specific to block l:

$$\mathbf{X}_l = \mathbf{J}_l + \mathbf{A}_l + \mathbf{R}_l, \quad l = 1, \ldots, L. \tag{3.22}$$

where $\mathbf{R}_l \in \mathbb{R}^{n \times p_l}$ is a residual noise matrix and $\mathbf{J}_l \in \mathbb{R}^{n \times p_l}$ denotes a sub-matrix of the joint matrix $\mathbf{J}$ that can be written as:

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 & \ldots & \mathbf{J}_L \end{bmatrix}.$$

JIVE constraints the whole joint matrix $\mathbf{J}$ to be of low-rank, whereas each individual matrix $\mathbf{A}_k$ are constrained to be of low-rank separately. This allows for $\mathbf{J}$ to represent shared information between all blocks. By introducing $\mathbf{X} = [\mathbf{X}_1 \ldots \mathbf{X}_L]$, the following optimization criterion can be derived:

$$\underset{\mathbf{J}, \mathbf{A}_1, \ldots, \mathbf{A}_L}{\operatorname{argmin}} \; \| \mathbf{X} - \mathbf{J} - [\mathbf{A}_1 \ldots \mathbf{A}_L] \|_F^2 \tag{3.23}$$

$$\text{s.t.} \begin{cases} rank(\mathbf{J}) &= R & \\ rank(\mathbf{A}_l) &= R_l & , \; l = 1, \ldots, L. \\ \mathbf{J}^\top \mathbf{A}_l &= \mathbb{0}_{(\sum_l p_l) \times p_l} & , \; l = 1, \ldots, L. \end{cases} \tag{3.24}$$

where $\|.\|_F$ refers to the Frobenius norm (as mentioned earlier) and $\mathbb{0}_{(\sum_l p_l) \times p_l}$ corresponds to an all zero matrix of size $(\sum_l p_l) \times p_l$. The last constraint correspond to orthogonality constraints imposed to ensure that joint structures are different from individual structures.

The fact that the joint matrix $\mathbf{J}$ is constrained to be of rank R implies that its SVD can be written as $\mathbf{J} = \mathbf{Y}\mathbf{\Delta}\mathbf{V}^\top$, where $\mathbf{Y} \in \mathbb{R}^{n \times R}$ and $\mathbf{V} \in \mathbb{R}^{(\sum_l p_l) \times R}$ are two semi-orthogonal matrices (i.e. $\mathbf{Y}^\top \mathbf{Y} = \mathbf{V}^\top \mathbf{V} = \mathbf{I}_R$) and $\mathbf{\Delta} \in \mathbb{R}^{R \times R}$ is a diagonal matrix of non-negative elements arranged in decreasing order. By introducing $\mathbf{W} = \mathbf{V}\mathbf{\Delta}$ and the fact that $\mathbf{W} = \begin{bmatrix} \mathbf{W}_1^\top \ldots \mathbf{W}_L^\top \end{bmatrix}^\top$, decomposition (3.22) can be rewritten as:

$$\mathbf{X}_l = \mathbf{Y}\mathbf{W}_l^\top + \mathbf{A}_l + \mathbf{R}_l, \;\; l = 1, \ldots, L.$$

with this last formulation, we can clearly see that the factor matrix $\mathbf{Y}$ act as a joint structure shared across all blocks. Similar factorization could be written for each individual structure $\mathbf{A}_l$, however, as they are only considered and constrained separately, they are only describing the data matrix they are associated with.

MOFA Multi-Omics Factor Analysis (MOFA) (Argelaguet et al., 2018) is a matrix factorization methods which decomposes the L omics matrices into a common factor matrix $\mathbf{Y} \in \mathbb{R}^{n \times R}$, with R latent components, and L weight matrices $\mathbf{W}_l \in \mathbb{R}^{p_l \times R}$, $l = 1, \ldots, L$, specific to each omic:

$$\mathbf{X}_l = \mathbf{Y}\mathbf{W}_l^\top + \varepsilon_l, \; l = 1, \ldots, L,$$

where ε_l refers to the residual noise associated with block l.

As opposed to all other methods presented in this section, MOFA uses a probabilistic Bayesian framework with prior distributions associated with $\mathbf{Y}$, $\mathbf{W}_l$ and ε_l. In their model, Argelaguet et al. (2018) suggested a standard normal prior for the factors $\mathbf{Y}$, and sparsity priors for the weight matrices $\mathbf{W}_l$. Furthermore, MOFA employs a 2-level regularization scheme to provide interpretable factors: the first fosters factor-wise sparsity, allowing direct identification of which factor is active in which omic; the second fosters feature-wise sparsity, resulting in a minimal number of relevant weighted features. Furthermore, MOFA provides versatile application by allowing to combine different noise models in order to be able to include different types of data, such as continuous, binary and count data. This is performed by combining multiple likelihood models, which also allows handling imputation of missing values.

intNMF Non-negative Matrix Factorization (NMF) is a particular application of matrix factorization, where the original data $\mathbf{X} \in \mathbb{R}^{n \times p}$ is decomposed into a non-negative factor matrix $\mathbf{W} \in \mathbb{R}^{n \times R}$ and a omic-specific weight matrix $\mathbf{H} \in \mathbb{R}^{R \times p}$. This is consistent with the notion that a biological perturbations pattern across samples is the result of the weighted sum of the implied features. Furthermore, in contrast to PCA, where the principal components are sorted by the amount of variance they explain, factors in NMF are equally important (Lee & Seung, 1999). Classical NMF proposes to solve the following optimization problem:

$$\underset{\mathbf{W},\mathbf{H}\geq 0}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{WH}\|_F^2 ,$$

where the constraints impose for all the elements of both $\mathbf{W}$ and $\mathbf{H}$ to be non-negative. Thus, the goal is to estimate two low-rank non-negative matrices $\mathbf{W}$ and $\mathbf{H}$ such that:

$$\mathbf{X} \approx \mathbf{WH}.$$

In Chalise and Fridley (2017) intNMF was proposed as extension to NMF in order to integrate an arbitrary number of omics data $\mathbf{X}_l \in \mathbb{R}^{n\times p_l},\ l = 1, \ldots, L$. Building on NMF, intNMF aims at finding a non-negative factor matrix common to all omics modality, where each factor is associated with omics-specific non-negative weights. The following optimization problem can be written for intNMF:

$$\underset{\mathbf{W},\mathbf{H}_1,\ldots,\mathbf{H}_L\geq 0}{\operatorname{argmin}} \sum_{l=1}^{L} \theta^l \|\mathbf{X}_l - \mathbf{WH}_l\|_2^2$$

where $\theta^l > 0$ is the user specified weight for the lth dataset. As exposed in Chalise and Fridley, 2017, a typical choice is $\theta^l = \{\max_k \|\mathbf{X}_k\|_F\}/\|\mathbf{X}_l\|_F$, which can be seen as a way to normalize the Frobenius norm of the datasets accross modalities.

As opposed to MOFA, intNMF does not make any assumptions on the statistical distribution of the data, which make it more flexible.

Data Fusion: scikit-fusion Data Fusion (DF) enhances traditional factorization formulation to accommodate multiple relation and multiple object-type data without requiring significant data transformation. DF, implemented in the scikit-fusion python package (Žitnik & Zupan, 2015b), is a penalized matrix tri-factorization approach that decomposes each data matrix into a product of 3 latent matrices.

Scikit-Fusion proposes to study the interaction between a collection of L modalities, referred to as object types in the framework and denoted as $\mathcal{E}_1, \ldots, \mathcal{E}_L$. There is no concrete dataset behind these objects types, which are more conceived of as concepts. However, Scikit-Fusion study them through their interactions. To do so, the user has to provide a relation matrix that describes interactions between each pair of object types. This relation matrix can be written as follows:

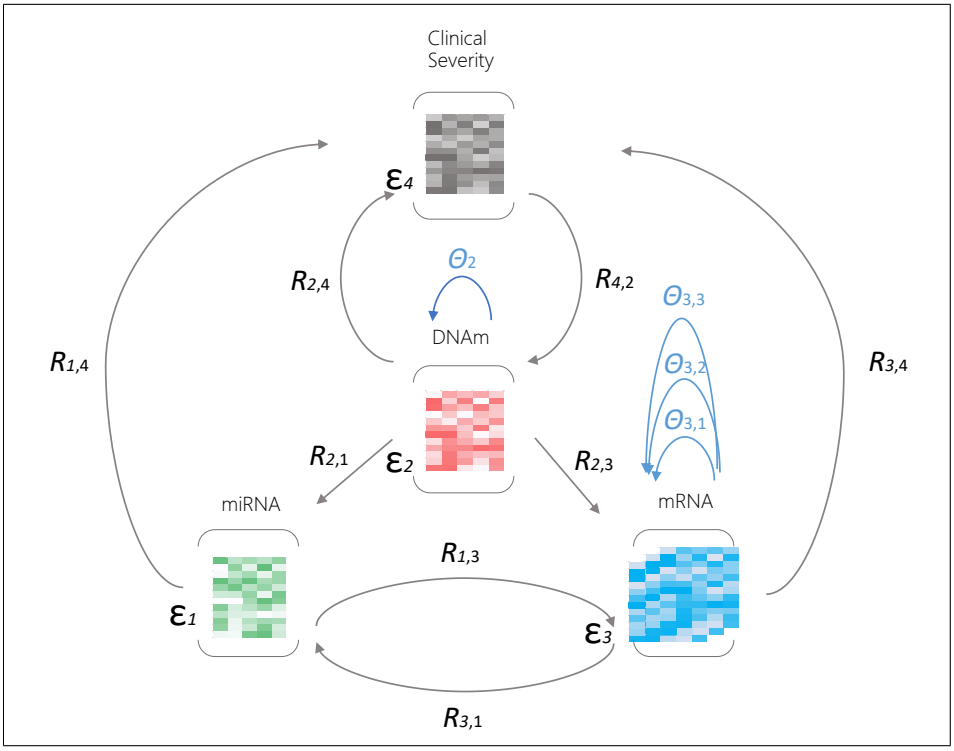


Figure 3.3: Conceptual fusion configuration for four dataset types: $\varepsilon_1 \ldots \varepsilon_4$, equivalently represented by the graph of relations between dataset types. Adapted from Žitnik and Zupan (2015a)

$$\mathbf{R} = \begin{bmatrix} \mathbf{0} & \mathbf{R}_{12} & \ldots & \mathbf{R}_{1L} \\ \mathbf{R}_{21} & \mathbf{0} & \ldots & \mathbf{R}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{L1} & \mathbf{R}_{L2} & \ldots & \mathbf{0} \end{bmatrix} \tag{3.25}$$

where the interaction between object types $\mathcal{E}_l$ and $\mathcal{E}_k$ is embodied by the block matrix $\mathbf{R}_{lk} \in \mathbb{R}^{p_l \times p_k}$, not necessarily symmetric, describing the interaction between the p_l and p_k objects of $\mathcal{E}_l$ and $\mathcal{E}_k$ respectively. Scikit-Fusion proposed to decompose $\mathbf{R}_{lk} \in \mathbb{R}^{p_l \times p_k}$ by the following tri-factorization model:

$$\mathbf{R}_{lk} \approx \mathbf{G}_l \mathbf{S}_{lk} \mathbf{G}_k^{\top}, \tag{3.26}$$

where $\mathbf{S}_{lk}^{r_l \times r_k}$ defines the relation between two object types $\mathcal{E}_l$ and $\mathcal{E}_k$. $\mathbf{G}_l \in \mathbb{R}^{p_l \times r_l}$ and $\mathbf{G}_k \in \mathbb{R}^{p_k \times r_k}$ are two non-negative matrices specific to each object type and composes of r_l and r_k latent components, respectively. Such factorization take into account inner-block interaction as well as inter-block relations in its factorization method.

Moreover, each factor matrix can be penalized by multiple constraint matrices $\mathbf{\Theta}_l^{(t)} \in \mathbb{R}^{p_l \times p_l},\ t = 1, \ldots, T_l$. By introducing T_{max}, the number of constraint matrices associated with the object type having the highest number of constraint matrices, $T_{max} = \max_{l \in \{1,\ldots,L\}} T_l$, constraints can be summarized by the following collection of block diagonal matrices :

$$\mathbf{\Theta}^{(t)} = \begin{bmatrix} \mathbf{\Theta}_1^{(t)} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta}_2^{(t)} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{\Theta}_L^{(t)} \end{bmatrix} \tag{3.27}$$

where $t = 1, \ldots, T_{max}$ and $\forall l \in \{1, \ldots, L\},\ \mathbf{\Theta}_l^{(t)}$ is only made of zeros if $t > T_l$. Thus, by introducing:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{G}_L \end{bmatrix};\ \mathbf{S} = \begin{bmatrix} \mathbf{0} & \mathbf{S}_{12} & \ldots & \mathbf{S}_{1L} \\ \mathbf{S}_{21} & \mathbf{0} & \ldots & \mathbf{S}_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_{L1} & \mathbf{S}_{L2} & \ldots & \mathbf{0} \end{bmatrix}, \tag{3.28}$$

and combining equations (3.25), (3.26) and (3.27), the following optimization problem can be derived for Scikit-Fusion:

$$\underset{\mathbf{G} \geq 0, \mathbf{S}}{\operatorname{argmin}} \|\mathbf{R} - \mathbf{G}\mathbf{S}\mathbf{G}^\top\|_2^2 + \sum_{t=1}^{T_{max}} \operatorname{Tr}\left(\mathbf{G}^\top \mathbf{\Theta}^{(t)} \mathbf{G}\right), \tag{3.29}$$

where Tr is the trace operator, as explained above. For example, if $\mathbf{\Theta}_l^{(1)} = \mathbf{I}_{p_l}$, the identity matrix of size p_l, and $\mathbf{G}_l = [\mathbf{g}_{1l} \ldots \mathbf{g}_{r_l l}]$, then $\operatorname{Tr}\left(\mathbf{G}_l^\top \mathbf{\Theta}_l^{(1)} \mathbf{G}_l\right) = \|\mathbf{g}_{1l}\|_2^2 + \ldots + \|\mathbf{g}_{r_l l}\|_2^2$, where $\|.\|_2$ is the ℓ_2-norm, the sum of squares of all the elements of a specific vector.

A multi-omics application of this model can be found in the Multi-Omics Mix (Momix) framework (Cantini et al., 2021), where the first object type is associated with the followings are associated with each omic modality. Each omic data matrix thus describes the relation between the first object (subjects) and the corresponding omic. No constraints are considered. Such settings would give results close to a Principal Component Analysis (PCA) on the concatenated omic matrices.

3.2.2.2 Deep learning based

Deep learning (DL) refers to a class of complex computer models to solve problems that were difficult with machine learning. DL relies mainly on Artificial Neural Network ANN that mimics the human brain neurons. DL models typically contain more than 2 ANN layers. Deep learning have lately been employed to achieve significant advances in medical research, especially for imagining data, achieving expert-level performance. DL has emerged as the dominant method for computer vision applications due to breakthroughs in computing techniques and algorithms, the availability of graphics processing units, and the collection of enormous training datasets (Levine et al., 2019).

Moreover, DL based multimodal representation learning has received a lot of interest in recent years because of its powerful representation capacity with several levels of abstractions. Multimodal representation learning aims to close the heterogeneity gap across different representation of the phenotypes under study. Despite satisfying outcomes from matrix reduction methods, some critical features may be filtered out before training the classifier due to a lack of supervision throughout the dimensionality reduction process, influencing the final performance. Hence, the rationale for employing deep learning data is that the complementary information may be retrieved from each modality for a particular learning task in a non-linear way, providing a richer representation that can be utilized to create clustering and classification models that achieve significantly better performance than using only one modality.

Several deep learning architectures have been proposed (Ramachandram & Taylor, 2017), among which Variational Autoencoders VAE has seen growing popularity lately (Poirion et al., 2021; Ronen et al., 2019; Simidjievski et al., 2019; Wissel et al., 2022). An auto-encoder is a network of neurons whose output layer is identical to the input layer after passing through a smaller intermediate layer. This intermediary layer's output subsequently provides a new, more compact representation of the data. Ronen et al. (2019) created an autoencoder-based approach for combining data from multi-omics research. VAE were used to estimate latent factors that summarize molecular trends from RNA-seq, single nucleotide polymorphisms, and copy number variations data sets (CNVs) from TCGA datasets.

Furthermore, Simidjievski et al. (2019)'s subsequent work on variational autoencoder architectures in the context of multi-omic integration showed that not all designs and integration strategies perform equally well. This study specifically looked at breast cancer patients and examined the performance of several VAE designs and integration strategies (including a hierarchical autoencoder), and 3 downstream classification approaches. The authors described hierarchical VAE as one of the most well-considered architectures.

Hierarchical Variational Autoencoder H-VAE extends standard meta-learning methodologies to the integration of many individual models. H-VAE is composed of numerous low-level VAEs that relate to each data source independently, with the result being integrated into a high-level VAE. In addition, Wissel et al. (2022) revealed that utilizing a modality selection approach to focus on the most informative modalities for each cancer type improved the performances of H-VAE models.

VAE generates latent variables that capture patterns explaining variance across data modalities, including some critical features in cancer biology, such as various gene expression programs, mutational profiles, copy number profiles, and their interactions. Several studies have shown that these latent variables predict previously identified cancer subtypes, and outperform linear approaches. Unfortunately, the use of such techniques is still scarce in molecular psychiatry, and needs to be further investigated.

3.3 Challenges in multi-omics integration

Achieving a comprehensive understanding of complex biological phenotypes through integrating multi-omics data sets is difficult. These problems must be evaluated, ideally before data collection, in order to create trustworthy, valid, and relevant interpretations (Krassowski et al., 2020). Among these problems are:

3.3.1 Data heterogenity

Omics measurements are acquired through ever-advancing technologies, exhibiting heterogeneous information encoded in a variety of formats and dimensions. Each modality encapsulates inherent technical biases and analytical obstacles. Thus, different bioinformatics tools are developed to keep apace. Hence, data heterogeneity hampers data modeling under a common analytical framework.

3.3.2 Multi-collinearity

As demonstrated in the section 3.2.2, when exploring connections between risk variables and complex phenotypes, the most commonly used strategy is a multivariate analysis. However, its effectiveness heavily depends on the correlation pattern among predictive variables. In addition, covariates in multi-omics models are rarely independent to one another, mainly due to it high dimensionality (small n samples and large p variables). Hence, multi-collinearity issues arise (Yoo et al., 2014), resulting in a biased estimation of inter- and intra-omic variance.

3.3.3 Missing data

Due to biological, technical and even financial limitations, dealing with missing data when integrating multiple omics is a common issue. Thus, some samples may not appear in all data tables. This is particularly challenging for the majority of multi-omic integration methods that cannot handle missing data, thus lowering the overall statistical power of the study. Subsequently, imputation is often employed to infer missing values. Song et al. (2020) recently provided a comprehensive review of imputation methods that have been developed for mult-omic studies.

3.3.4 Cross-Validation in unsupervised analysis

The information included in multivariate datasets is considerable. Unfortunately, due to their inherent noise, applying sophisticated models and machine learning algorithms on these data is tricky. Since these datasets contain more features than samples, methods are vulnerable to overfitting. As explained in section 3.2.2, Cross-Validation is a well established tool to assess if a model actually overfits or not. Especially in a supervised context when we aim at predicting an outcome. However, the unsupervised nature of most of the multi-omic methods hampers the distinction between relevant biological models from overfitted ones. Moreover, assessing the efficiency of unsupervised techniques on experimental datasets in terms of stability and degree of overfitting is challenging. McCabe et al. (2019) proposed an evaluation framework based on Cross-Validation. McCabe et al. (2019) concluded, after examining a limited number of methods, that results were satisfying for dataset with large sample size, but showed limited performance and inconsistencies on smaller ones. This reinforces the importance of the notion of sample size in multi-omic integration analyses.

3.3.5 Validation

One of the main goals behind multi-omic data integration is the generation of new hypotheses. Thus, in most circumstances, no ground truth is accessible, and the model's guard rails can be limited to quality control statistics and/or the enrichment of the resulted outputs on known pathways.

Part II

Contribution

Chapter 4

Chapter 4: Sex-specific and multi-omic integration to enhance accuracy of peripheral blood biomarkers of major depressive disorder

Foreword: This chapter corresponds to a first finalized version of our article on blood biomarkers of MDD, which we plan to submit for publication within the next 2 months. This article represents the main original research contribution emanating from the present PhD work. It results from a collaboration between clinicans (Pr Bruno Etain, Paris; Pr Raoul Belzeaux, Montpellier), neurobiologists (Dr Cynthia Marie-Claire, Paris; Dr El-Cherif Ibrahim, Marseille), computational scientists (Dr Arnaud Gloaguen) and bioinformaticians (Dr David Cohen, Mrs Margot Derouin). My personal contribution to the article was to process raw sequencing data, and to conduct all bioinformatic and statistical analyses presented below (with the exception of EPIC arrays, which were primarily processed by Dr David Cohen). In addition, multi-omic integration was conducted in tight collaboration with Dr Arnaud Gloaguen.

The approaches and results presented below were the subject of an oral communication and a poster:

- **Comprehensive multi-omics integration of major depressive disorder**; Oral presentation by Amazigh MOKHTARI; Journées Ouvertes en Biologie, Informatique et Mathématiques - France (JOBIM 2022); Session 3, Systems biology metabolomics.
- **A multi-omics integration framework applied to majordepressive disorders**; Poster - Amazigh MOKHTARI; European Human Genetics Conference (EHGC) 2022.

4.1 Introduction . . . 96
4.2 Material and methods . . . 98
4.2.1 Human samples . . . 98
4.2.2 DNA methylation arrays . . . 100
4.2.3 RNA isolation . . . 100
4.2.4 miRNA-sequencing . . . 101
4.2.5 RNA-sequencing . . . 101
4.2.6 Raw data processing . . . 101
4.2.7 Analyses of covariates . . . 102
4.2.8 Differential analyses . . . 102
4.2.9 Step-wise multi-omic integration . . . 103
4.2.10 Advanced integration . . . 103
4.2.11 Cross-validation . . . 104
4.2.12 Bootstrapping . . . 105
4.3 Results . . . 105
4.3.1 Single-omic analyses reveal mRNA, micro-RNA and DNA methylation changes in males and females with MDD . . 105
4.3.2 Comparison of genome-wide changes in males and females with MDD reveals shared and sex-specific single-omic signatures . . . 107
4.3.3 Step-wise integration identified male and female MDD-associated gene co-expression modules enriched for MDD-related transcriptional and epigenetic dysregulations and mood-related SNPs . . . 109
4.3.4 Advanced integration with supervised feature selection and non-supervised clustering prioritize mRNAs and DNA methylation CpGs as more efficient biomarkers of MDD . 114
4.3.5 Different omics analyses and integration strategies shows common features serving as potential biomarkers and molecular signature for sex-specific MDD . . . 117
4.4 Discussion . . . 117
4.4.1 Sex specific molecular signatures of MDD . . . 117
4.4.2 Added value and limitations of our multi-omic data integration framework . . . 120
4.4.3 Sample size limitations . . . 122
4.4.4 Is blood a relevant tissue for mood disorder research ? . . 123
4.4.5 Confounding factors . . . 123
4.5 Perspectives . . . 124
4.6 Conclusion . . . 125
4.7 Supplementary material . . . 125
4.7.1 Adjusting data for batch effects with ComBat while preserving group differences may lead to an over estimation of the significance of the differential analyses. . . . 125
4.7.2 Label Propagation . . . 127
4.8 Supplementary Figures and Tables . . . 128

Abstract: Major depressive disorder (MDD) is a leading cause of disability worldwide, with comorbidities that include increased risk of suicide, drug dependence, and reduced life expectancy. Women are nearly twice as likely as men to be diagnosed with MDD. In addition to symptomatic heterogeneity, the variety of environmental factors that trigger and aggravate MDD complexify the identification of molecular biomarkers that may guide clinical management. Here, we implemented a multi-omic integrative framework to tackle these challenges. It was designed to extract peripheral blood molecular signatures of MDD, and to evaluate the added value of combining several types of omics data when classifying patients and controls. To do so, we combined several strategies to analyze transcriptomic (RNA-Sequencing) and epigenomic (microRNA-Sequencing and DNA methylation arrays) mechanisms in blood samples from a well-characterized cohort of individuals with MDD (n=80), and healthy controls (n=89). First, we applied gene co-expression network and annotation enrichment analyzes. We identified gene modules that significantly associated with SNPs implicated in the risk for MDD, and also showed significant enrichment for differentially expressed genes, differential DNA methylation, and regulation by differentially expressed miRNAs. Importantly, these results showed strong sex differences in the biological signature of MDD. Second, we compared several methods for joint dimensionality reduction, in order to identify features that better discriminate between patients and controls. This allowed prioritizing the subset of features (individual protein-coding genes, micro-RNAs, and methylation probes) that maximized accuracy when clustering MDD patient and controls. In both the pooled and male only cohort, clustering based on the combination of mRNA and DNAm outperformed results obtained using a single omic layer. In females, classifications using the mRNA and DNAm combination, or mRNA only, reached the highest accuracy. Altogether, these results support the hypothesis that advanced multi-omic integration, compared to the analysis of single omic profiles, has the potential to significantly improve patient clustering in MDD. This work therefore has significant implication for the development of clinically useful biomarkers.

4.1 Introduction

Major depressive disorder (MDD) is a systemic disease defined by recurrent and clinically significant changes in mood, energy, cognitive functions, sleep, or appetite. It is very frequent (Hasin et al., 2018), with a prevalence of 20.6% in the general population. Starting at puberty, its female:male ratio is around 2:1 (Albert, 2015), suggesting significant sex differences. Although multiple treatments are available, their efficacy remains disappointing (Cipriani et al., 2011), with more than 30% of patients showing resistance to antidepressant pharmacotherapy, among others (Jaffe et al., 2019). Also, due to its recurrence and associ-

ated comorbidities, such as substance use or anxiety disorders, MDD places a significant burden on patients and their families. Therefore, developing molecular biomarkers to improve diagnosis, and ultimately personalize treatment strategies, is critical towards better care.

The discovery of diagnostic biomarkers for MDD has proved challenging. This is likely explained by the fact that multiple factors, ranging from genetic predisposition to environmental influences, are contributing to the disease, and hamper the understanding of its molecular pathophysiology. Over the last decade, family studies and genome-wide association studies (GWAS) have quantified the proportion of the risk for MDD that can be attributed to genetic factors (with heritability estimated around ∼37%) (Sullivan et al., 2000). In parallel, epidemiological studies indicate that environmental factors, including stressful life events, also significantly contribute to this risk. At the interplay of these factors, recent advances point towards a critical role for epigenetics, defined as mechanisms that regulate genomic activity. Therefore, characterizing epigenetic processes in MDD patients has the potential to better capture interactions among underlying risk factors, thereby leading to more predictive biomarkers (Penner-Goeke & Binder, 2019).

The investigation of genome-wide epigenetic regulation has been facilitated by high-throughput technologies that assess an ever-growing number of omic types, including DNA methylation, small non-coding RNAs, and others. Available post-mortem brain and peripheral blood studies have now established a robust overview of most frequently dysregulated pathways in MDD: immune system, stress response, brain neurotrophic factors and neurotransmitters (Mariani et al., 2021). Importantly, these studies also provide emerging evidence that functional dysregulations underlying MDD may significantly vary among male and females (Seney et al., 2022), urging to directly examine sex differences, rather than simply control for it in MDD case/control comparisons (Parel & Peña, 2022). Another crucial factor to consider is the numerous molecular levels and regulatory systems that modulate gene expression. Characterizing differences between MDD patients and controls at the level of single, individual omic data is unlikely to capture the full chain of molecular events that lead to MDD, in part due to inter-individual variability in disease mechanisms, as well as technical limitations inherent to each methodology. To overcome this, integrating several molecular layers is a strategy that has proved powerful across a variety of other medical conditions (Li et al., 2018; Rappoport & Shamir, 2018). In relation to MDD, however, available studies have mainly used "step-by-step" integration, whereby significant discrete differences are identified at the level of each single-omic layer, and then aggregated or intersected across layers (using concatenations, overlaps, correlation, or regression approaches; (see Bhak et al. (2019) and Mokhtari et al. (2022) for review). These approaches

do not leverage the potential of available data. In contrast, a fully multi-omic and genome-wide integration is expected to yield significant progress for the identification of diagnostic biomarkers, and for patient stratification.

Based on these considerations, in the present work, we characterized and integrated transcriptomic data with two layers of epigenomic regulation (DNA methylation and micro-RNAs), generated using peripheral blood samples from MDD patients (n=80) and healthy controls (n=89). All analyses were conducted separately for each sex. First, each omic was analyzed individually in order to extract genome-wide signatures of MDD, which were then validated by comparison to external studies or meta-analyses. Second, we defined the network organization of gene co-expression. This identified gene modules that were associated with MDD, and that showed enrichment for differential epigenomic regulation. Finally, an advanced genome-wide and multi-omic integration was developed, combining the Momix (Multi-Omics Mix) framework (Cantini et al., 2021) with Similarity Network Fusion (SNF), to prioritize blood-derived signatures in MDD patients. SNF computes similarity matrices at the sample level, for each omic modality, before merging them. While the method was designed to use all available features, only a subset are likely of interest in the context of MDD. Hence, here we first selected features that best represented each omic's variance, taking into consideration their shared variance with MDD severity. Compared to its previous implementation, and to avoid overfitting, Momix was encapsulated in a Cross-Validation (CV) procedure. Finally, stability of selected features was evaluated through a bootstrap procedure. Overall, this resulted in the identification of subsets of features (i.e., individual mRNAs, micro-RNAs or DNA methylation probes) that most efficiently clustered MDD cases and controls. Collectively, our results support the hypotheses that MDD biomarkers are partly sex-specific, and that multi-omic integration predicts MDD with better accuracy than previously used single-omic or step-by-step methodologies. These results therefore have implications towards developing clinically useful multimodal biomarkers.

4.2 Material and methods

A general overview of the study is presented in Fig4.1.

4.2.1 Human samples

Data were generated using peripheral blood samples collected from a naturalistic and longitudinal multi-centric cohort registered at ClinicalTrials.gov (ID:NCT02209142), composed of n=80 MDD patients and n=89 euthymic controls (see Fig4.1). Cases and controls were matched for age, sex and body mass index (BMI). All experiments on human subjects

a

Cohort Description

Sex	n .max	Group	mean.Age	mean.BMI	mean.HDRS	n (mRNA)	n (miRNA)	n (DNAm)
Female	55	Control	45	23.97	1.44	55 (100%)	51 (92.7%)	50 (90.9%)
	50	Patient	42	24	23.16	50 (100%)	39 (78%)	39 (78%)
Male	34	Control	45	24.02	1.18	34 (100%)	34 (100%)	31 (91.1%)
	30	Patient	43	25.51	23.47	30 (100%)	30 (100%)	28 (93.3%)
Total	**169**							

b

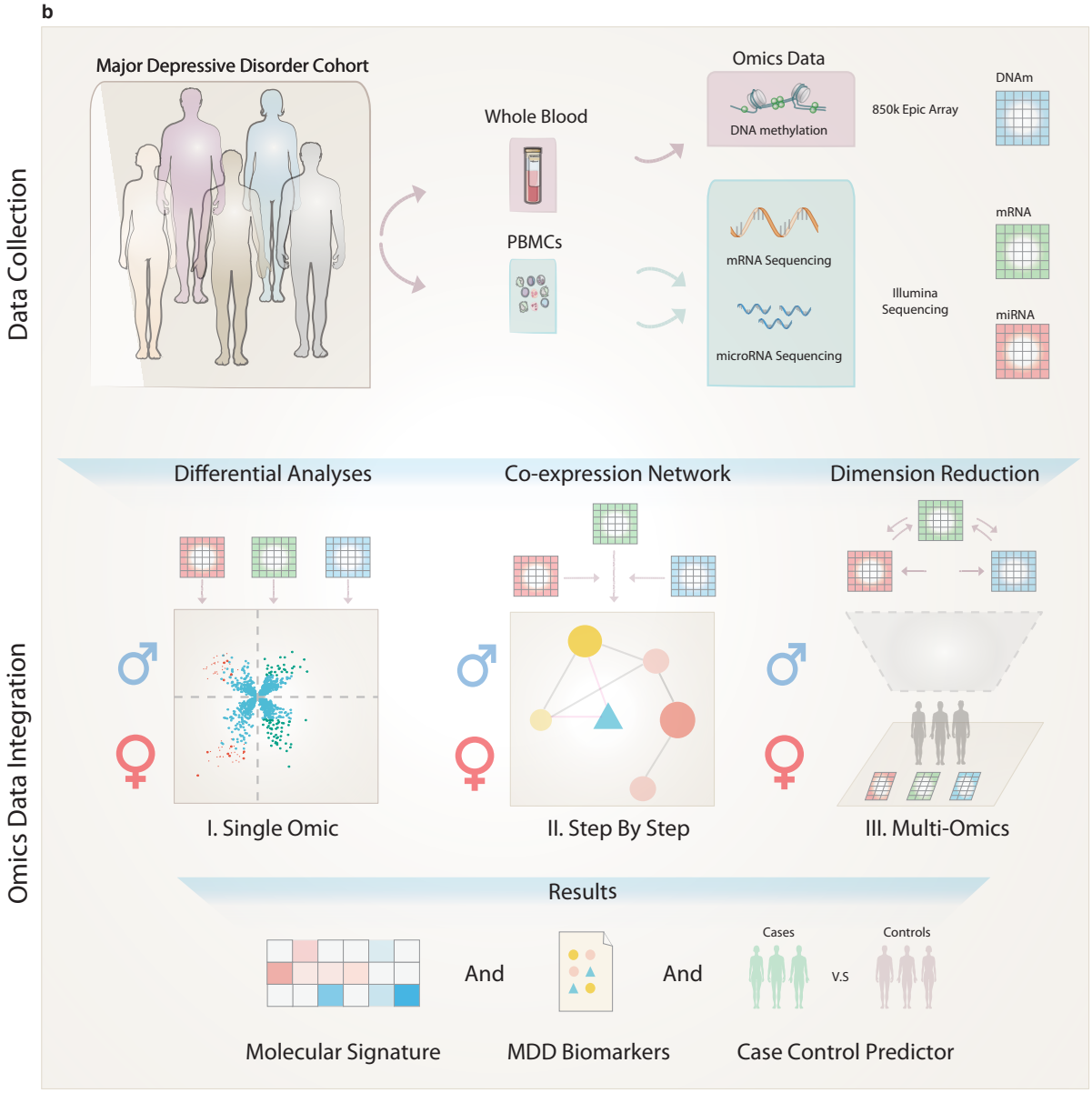


Figure 4.1: Study summary. **a** Descriptive statistics of the major depressive cohort. **b.** Graphical abstract.

were conducted in accordance with the Declaration of Helsinki. The project was approved by the local ethics committee (CPP Sud Méditerranée II, Marseille, France, study registered under number 2011-A00661-40), and written informed consent was obtained from all subjects. All patients met DSM-IV-TR criteria for at least severe MDD episode (17-item Hamilton Depression Rating Scale, HDRS, score 19) at the time of blood sampling. Patients were excluded if they suffered from bipolar disorder, schizophrenia or substance use disorder. All patients were treated at inclusion with treatment as usual, upon discretion of the treating psychiatrist. Healthy controls were free of any psychiatric disorder according to semi-structured interviews.

4.2.2 DNA methylation arrays

Genomic DNA was extracted from EDTA-tubes venous blood using PureLink TM genomic DNA mini kit(Invitrogen), according to the manufacturer's instructions. DNA was deaminated using the EZ-96 DNA Methylation Kit (Zymo Research). Bisulfite conversion efficiency was controlled by qPCR, with one assay targeting a methylated region of DNAJC15 and two assays targeting the GNAS locus (one assay for the unmethylated allele and one assay for the methylated allele). Deaminated DNA derived from blood was amplified in parallel and served as positive controls. A sample passed quality control when Ct-values for either the two GNAS loci, or the DNAJC15 locus, reached the threshold no later than 5 amplification cycles compared to the positive control. All samples passed quality control. Genome-wide DNA methylation was assessed using Infinium MethylationEPIC BeadChip microarrays, following Illumina's recommendations (Illumina, San Diego, CA).

4.2.3 RNA isolation

EDTA-tubes venous blood was passed through LeukoLOCK™ filters (Life Technologies, Ambion) to eliminate red blood cells, platelets, and plasma. Leukocytes trapped on LeukoLOCK filters were lysed with TRI reagent (Ambion) and mixed with Bromo-3-chloropropane (Sigma-Aldrich, St. Louis, MO, USA). After centrifugation, total RNA from the aqueous phase was precipitated with ethanol, purified on spin cartridge, washed, and eluted with 0.1 mM EDTA. RNA quantity and quality were assessed using a NanoDrop-1000 (Thermo Fischer Scientific) and 2100 Bioanalyzer (Agilent). Total RNA was used for both miR-sequencing and RNA-sequencing.

4.2.4 miRNA-sequencing

MicroRNA libraries were prepared using the Bioo Scientific NEXTflex Small RNA-Seq kit v3, following the manufacturer's instructions. Briefly, 500ng of total RNA was used as input, and NEXTflex 4N adenylated adapters were ligated to the 3'- and 5'-ends of the RNA. After adapters ligation and cleanup (with magnetic beads), first-strand cDNA was synthesized, cleaned (with magnetic beads), isolated and amplified by PCR (16 cycles). Successful library preparation was verified using a Fragment analyzer (Advanced analytical technologies). Quantification of libraries was performed using the Qubit dsDNA HS assay (Life Technologies). Libraries were normalized, pooled, denatured and sequenced on an Illumina NextSeq 500 at a depth of 9.69 $\pm$ 0.05 million 50 bp single-end reads per sample for males and 7.5 $\pm$ 0.01 million reads for females.

4.2.5 RNA-sequencing

RNA-seq libraries were prepared using 500 ng of total RNA and poly-A capture and the KAPA mRNA Hyper Prep Kit (Roche, ref. KK8581/KK8581, Basel, Switzerland) and TruSeq Stranded mRNA LT Sample Preparation Kit (Illumina, San Diego, CA) for female and male samples, respectively. Libraries were prepared separately for each sample, amplified by PCR, quantified by microcapillary electrophoresis, and then pooled at equimolar concentrations for sequencing using Illumina chemistry: in females, 100-base pair (b) paired-end on a NovaSeq6000, SP flowcell, at the ICM (Institut du Cerveau et de la Moelle Epiniere) and generated a mean of 35.12 $\pm$ 0.56 million reads ; in males, 50-bp single-end on a Hiseq 4000 at the GenomEast platform generating a mean of 45.1 $\pm$ 1.07 million reads.

4.2.6 Raw data processing

DNA methylation data were analyzed using the R package ChAMP (v2.16.2). Probes were filtered out when detection p-value was above 0.01 in one or more samples, or bead counts less than 3 in at least 5% of the samples. Probes corresponding to non-CpG methylation sites, identified as SNPs by Zhou et al., 2016, aligning to multiple locations, or to the X or Y chromosomes, were removed. For miR-seq data, reads were trimmed using cutadapt (v1.18), following the QuickMIRseq pipeline. First, a small RNA and mRNA database was generated with sequences collected from the GRCh38.p12 (hg38) human reference genome. Then, a miRNA and hairpin database was generated with sequences collected from miRBase (v22-www.mirbase.org) (Kozomara et al., 2018). Reads were aligned and quantified using bowtie (v1.2.2) and default parameters. For RNA-seq, male and female raw reads were trimmed with bbduk (bbmap package v38.73) (Bushnell, 2014), and aligned to the

GRCh38.p12 (hg38) human reference genome using STAR (v2.5.3a) (Dobin et al., 2012), and Gencode (v29) annotations. Gene expression was quantified using HTSeq (v0.11.2) (Anders et al., 2014).

4.2.7 Analyses of covariates

To identify sources of biological or technical variation, we applied the variancePartition R package (Hoffman & Schadt, 2016) to miR- and RNA-sequencing data, and the singular value decomposition (SVD) function of the ChAMP R package (Tian et al., 2017) to EPIC arras. According to ChAMP's (Tian et al., 2017) recommendation, blood cell composition was inferred from DNAm using Houseman's method (Houseman et al., 2012). For miR- and RNA-sequencing, white blood cells counts were used in final models (only polynuclear neutrophils and lymphocyte were included, due to high collinearity among white blood cell subtypes, as expected; see Supplementary Figure 4.2). Because the smoking status showed high correlation with the phenotype of interest (Chi-squared test p-value = 2.24×10^{-3}), it was not included in final models.

4.2.8 Differential analyses

Sex-specific differential analyses were conducted on each omic to compare MDD cases and controls. For miRNA-seq and RNA-seq data, the R package DESeq2 (Love et al., 2014) was used, after filtering out lowly expressed miRNAs (number of reads ≥ 1 in more than 60% of either cases or controls) and RNAs (mean raw expression < 10 reads). For EPIC arrays, the ChAMP package identified differentially methylated probes (DMP, corresponding to individual CpG sites). Based on our analyses of variance (see previous section 4.2.7) and the literature, the following models were used for male and female data (with the addition of sex for pooled analyses of both sexes):

- **DNAm** $\sim$ Age + Slide + Array + BMI + Blood Cells Composition (Houseman's method (Houseman et al., 2012))
- **miRNAseq** $\sim$ Batch + Age + BMI + RNA integrity number (RIN) + Lymphocyte + Polynuclear neutrophile
- **RNAseq** $\sim$ Age + BMI + RNA integrity number (RIN) + Lymphocyte + Polynuclear neutrophile

Functional enrichments were computed using the Webgestalt implementation of GSEA (with 10^4 permutations) on GO, KEGG and Reactome databases and the fgsea R package

(Korotkevich et al., 2016) for previously published lists of differentially expressed genes (DEGs).

4.2.9 Step-wise multi-omic integration

For step-wise and advanced integration (see below), the DNAm dataset was adjusted for covariates using the R package Combat (Leek et al., 2012), while RNA-seq and miRNA-seq data were adjusted using linear models, after a variance stabilization normalization (Love et al., 2014). Then, for step-wise integration, Weighted Gene Co-expression Network Analysis (WGCNA) was performed in male and female separately on adjusted gene counts. WGCNA parameters were optimized to maximize overlaps with biological networks from the Reactome database (Abbassi-Daloii et al., 2020). Gene modules were tested for preservation in other blood transcriptome datasets using z-summaries of the WGCNA's *modulePreservation* function, as well as for enrichment in: i) DEGs identified in males and females, using GSEA; ii) targets of Differentially Expressed miRNA (DEmiR) and DMPs, using a Fisher overlapping test; and iii) SNPs associated with psychiatric disorders, using MAGMA (de Leeuw et al., 2015) and summary statistics of from the Psychiatric Genomics Consortium GWAS, assigned to their nearest gene (within 1 Mb). Enrichment for these SNP-associated genes was computed for each module. Finally, associations between each module's eigengene and MDD status, as well as scores from the Childhood Trauma Questionaire (CTQ), were evaluated using Spearman correlations.

4.2.10 Advanced integration

4.2.10.1 Pre-processing steps

First, as mentioned above, Combat was used to adjust for covariates, with two slight modifications: (i) neuroComBat, an improved version of ComBat dedicated to CV procedures (v1.0.5, available at https://github.com/Jfortin1/ComBatHarmonization/tree/master/R, Fortin et al., 2018; Johnson et al., 2007), was used, and (ii) in DNAm experiments: for the pooled and female cohorts, only the *Slide* covariate was corrected for separately in the train and test sets, as it was composed of too many categories for proper representation (in train/test sets) and correction across all CV folds; while for similar reasons, in the male cohort, both *Slide* and *Array* were corrected for separately. Then, due to computational limitations, a pre-filtering was applied to the DNAm block, whereby 10% of most variable CpGs were selected for downstream analyses.

4.2.10.2 Feature selection with the Momix framewok

Building on the Momix benchmark (Cantini et al., 2021), we encapsulated 6 joint Dimension Reduction (jDR) methods (RGCCA, JIVE, MCIA, MOFA2, intNMF and SciKit-Fusion) in a CV procedure, in order to extract common variance between HDRS matrix scores and every possible combination of either one (mRNA, miRNA, DNAm), two (mRNA/miRNA, mRNA/DNAm, miRNA/DNAm) or three (mRNA/miRNA/DNAm) omic blocks (7 possible combinations). For each jDR method and each combination of blocks, a factor matrix, $\mathbf{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_R] \in \mathbb{R}^{n \times R}$ representing the shared variance across all blocks, was derived (where n is the number of samples and $R = 10$ the number of factors extracted). Among all 10 computed factors, $\mathbf{f}_{r^\star}$, the one that correlated the most (in absolute value) with MDD status, was kept. Then, weight vectors $\mathbf{w}_l^{(r^\star)} \in \mathbb{R}^{p_l},\ l = 1, \ldots, L$ (where l refers to one modality among the L considered composed of p_l variables), used to compute the kept factor, were retrieved (except for the weights associated with the HDRS matrix). These weights represent the contribution of each variable to the factor to which they belong. Thus, for each modality, only the top 10%-ranking elements of $\mathbf{w}_l^{(r^\star)}$ (in absolute value) were retained.

4.2.10.3 Multi-Omics clustering with SNF

To evaluate the ability of our selected features to estimate meaningful clusters of subject with regard to MDD status, we used Similarity network fusion (SNF) (Wang et al., 2014), an unsupervised integrative clustering technique. Several sets of SNF parameters were evaluated through CV (every possible combination between $neighbours \in \{10, 20, \ldots, 50\}$, $iters \in \{10, 20, \ldots, 60\}$ and $\alpha \in \{0.3, 0.4, \ldots, 0.8\}$). The best set of parameters was selected according to the accuracy of prediction between predicted and the true status. Finally, cluster labels were transferred from the train to the test set using label propagation. This consists in iterative propagation, to the test set, of labels defined during training, based on a measure of similarity between samples (see supplementary material 4.7.2).

4.2.11 Cross-validation

A repeated 5-fold CV (5 repetitions) with stratification was used to prevent overfitting. For each re-sampling to be representative of the original dataset, stratification was systematically undertaken based on MDD status, the *Slide* covariate (DNAm experiments) and sex (in the pooled cohort only). For each train set (4/5 of the samples), pre-processing, feature selection and clustering were performed as described in 4.2.10.1, 4.2.10.2 and 4.2.10.3. For the test set (remaining 1/5), variables selected on the train set were extracted, pre-processed by applying the different models fitted on the train set (with a small subtlety for Combat, as described in section 4.2.10.1) and, finally, label propagation was used to transfer inferred clusters from the train to the test.For comparison with features obtained by jDR methods,

the same CV procedure was applied to features corresponding to differential analyses results (with DESeq2 for miRNA and mRNA and CHAMP for DNAm).

4.2.12 Bootstrapping

A bootstrap re-sampling strategy (Efron, 1979; Efron, 1987) was employed to evaluate the stability of variables that were selected. $B = 1000$ bootstraps of the same size as the original cohort were repeatedly sampled with replacement from the original data. For the pooled cohort, each re-sampling was stratified according to the sex variable to ensure that the sex ratio is representative of the original cohort. For each re-sampling, preprocessing and jDR based feature selection were performed as described above. Two metrics were computed for each variable, each omic combination and each jDR method through the 1000 samplings: first, the occurrence of selection i.e. the number of times a variable was associated with the highest discriminative factor in regard to MDD status, and was in the top 10%-ranking weights (in absolute value) of its omic modality. Second, the sign ratio of the weights $w_{kl}^{(b),(r^\star)}$ obtained across all bootstraps, i.e. the number of positive over negative weights (or the other way around, with the less frequent sign at the numerator), where k, b, l and $r^\star$ stand respectively for the variable, the sample, the omic and the best factor index (according to correlation with the status). This ratio can be seen as a non-parametric way of estimating the probability for a weight $w_{kl}^{(r^\star)}$ to change sign across bootstraps. It ranged between $1/B$ (as B samplings were undertaken) and 1 (as many negative as positive estimates). The lowest is this probability, the most stable is the considered weight across bootstraps.The probabilities were adjusted by the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). Most reliable features were selected using a stability criterion of an adjusted probability < 0.05 and an occurrence of selection $> 80\%$.

4.3 Results

4.3.1 Single-omic analyses reveal mRNA, micro-RNA and DNA methylation changes in males and females with MDD

We first identified DEG, DEmiR and DMP between MDD cases and controls. Regarding gene expression, a higher number of genome-wide significant DEG (FDR<0.05) were observed in females than in males, with 1824 (1034 up, 790 down) and 944 (671 up, 273 down) dysregulated genes, respectively (Fig4.2a). Among these, only 3.82% (n=106 DEG) were common to both males and females (with 58 up and 20 down-regulated genes in both sexes), although a stronger sex concordance emerged using a threshold-free approach (see section 4.3.2 below). This is reminiscent of recent work in the field that reported a signifi-

cant degree of sex differences in relation to MDD ((Labonté et al., 2017a; Seney et al., 2018)). For external validation of our data, we compared these results with those from the quantitative meta-analysis of MDD peripheral blood studies recently conducted by (Wittenberg et al. (2020), Fig4.2d). Using GSEA for threshold-free comparison (see Methods and Subramanian et al. (2005)), we found that up- or down-regulated DEG in our male cohort were significantly enriched for dysregulation in similar directions in Wittenberg's results (Up: Normalized Enrichment score, NES =1.53 - p_{adj} (adjusted p-value) $= 1.6 \times 10^{-3}$; Down: NES = -1.60 – $p_{adj} = 1.6 \times 10^{-3}$), while no significant enrichment was found in females.

Regarding micro-RNAs, miR-124-3p, a widely studied candidate in MDD, was the only genome-wide significant hit in males (FDR<0.05; $log_2(FoldChange, FC) = 1.88$, $p_{adj} = 1.36 \times 10^{-5}$). This miR was similarly up-regulated in females, along with 4 other miRs (miR-3613-3p, miR-550a-3p, miR-1299 and miR-6511b-3p), while 5 miRs were down-regulated (miR-5585-3p, miR-340-3p, miR-142-3p, miR-652-5p and miR-3611; Fig4.2b). Genome-wide comparison with previous MDD miR studies van den Berg et al. (2020)) showed significant enrichment for up-regulated DEmiR in both males and females (Females: NES = 1.55 - $p_{adj} = 4.3 \times 10^{-3}$; Males: NES = 1.63 - $p_{adj} = 4.5 \times 10^{-3}$; Fig4.2e), indicating that, similar to our transcriptomic data, these miR analyses capture part of a reproducible biological signal previously associated with MDD.

For DNA methylation, among the >700k sites investigated (Fig4.2c), only a scarce number met a loosened 0.1 FDR threshold in female (81 hypo-methylated and 74 hyper-methylated), while no DMPs were identified in males. This was expected as, for consistency with results from other omics, we conducted group comparisons without using the preservation function from Combat/ChAMP. Indeed, as previously noted by Zindler et al. (2020), this function strictly preserves DMP ranking but artificially inflates p-values (see Supplementary materials 4.7.1; for reference, results with preservation are presented in Supplementary Fig 4.1). Despite this low number of DMPs, concordance with previous studies was detectable at genome-wide level (again using GSEA; Fig4.2f, (Hüls et al., 2020; Roberson-Nay et al., 2020; Walker et al., 2015)). Accordingly, we observed enrichment for DMPs identified by Walker et al. in relation to a 20-Mb haplotype associated with MD Fig4.2f ((Walker et al., 2015)). This was detected for hypo-methylated DMPs in males (ULH: NES = -1.07 - $p_{adj} = 3.9 \times 10^{-3}$; ALH: NES = -1.10 - $p_{adj} = 3.9 \times 10^{-3}$) and for hyper-methylated DMPs in females (ULH: NES = 1.21 - $p_{adj} = 2.2 \times 10^{-4}$; ALH: NES = 1.17- $p_{adj} = 2.2 \times 10^{-4}$). Overall, these analyses provide external validation of our results for each single omic layer investigated in our cohort.

4.3.2 Comparison of genome-wide changes in males and females with MDD reveals shared and sex-specific single-omic signatures

Because recent work suggests a significant degree of sex-specificity in molecular adaptations underlying MDD, we then systematically compared male and female data. Using RRHO2 for threshold-free comparison across the whole genome (rank-rank hypergeometric overlay; Cahill et al., 2018 and (Fig4.2d)), we identified statistically significant overlaps among groups of DEGs, DEmiRs and DMPs that were dysregulated in similar directions in males and females (bottom-left and upper-right quadrants, respectively; Fig4.2 g-j). These overlaps were strongly significant for each omic layer: (bottom-left quadrant's best -log10 (p-values) : miRNA $= 8.53$, mRNA$= 29.07$, and DNAm$= 410.33$) In addition to these adaptations in similar directions, the RRHO2 analysis also identified groups of genes, exclusively in RNA-Seq data, that showed opposite changes in males and females with MDD (upper-left and bottom-right quadrants). While a global pattern of MDD-related molecular concordance in male and female is intuitive, and has already been described ((Labonté et al., 2017a)), opposite dysregulations are more surprising. A recent post-mortem brain investigation, however, brings support to this notion Seney et al., 2018. Combining data spanning 3 brain regions, Seney et al. prioritized a list of 52 genes exhibiting sex discordance, which we opted to confront with our own results. Out of these 52 genes, 19 were reliably quantified in blood, among which 6 were dysregulated in opposite directions in males and females with MDD (*C2CD2L, OGFR, KIAA2013, MLF2, ZBTB46, ADRM1*; Fig4.2k), while the remaining 13 (*PTPRF, SUN1, ADCY3, RCCD1, SMAD3, ADCY9, ICMT, ZDHHC8, RBM15, EIF5A2, NR2C2, PRR7, ELP2*) showed concordant dysregulation. These results therefore suggest that, even across brain and blood tissues, MDD may associate with opposite gene expression changes affecting specific genes as a function of sex.

Finally, we conducted functional enrichment analyses in males and females, using GSEA and KEGG pathways, Fig4.2l. Female DEGs were significantly enriched for Aldosterone (NES=1.84, FDR=0.02) and Cortisol synthesis (NES=1.84, FDR=0.02), as well as the Glutamatergic synapse (NES=1.86, FDR=0.02), while DEGs in males showed enrichment for the immune response system through IL-17 signaling pathway (NES=1.75, FDR=7.58 10E-03) and neurodegenerative diseases (including Huntington, Parkinson and Alzheimer disease - NES=1.86, FDR=1.71 10E-03), again emphasizing sex differences in biological pathways most significantly impacted in each sex.

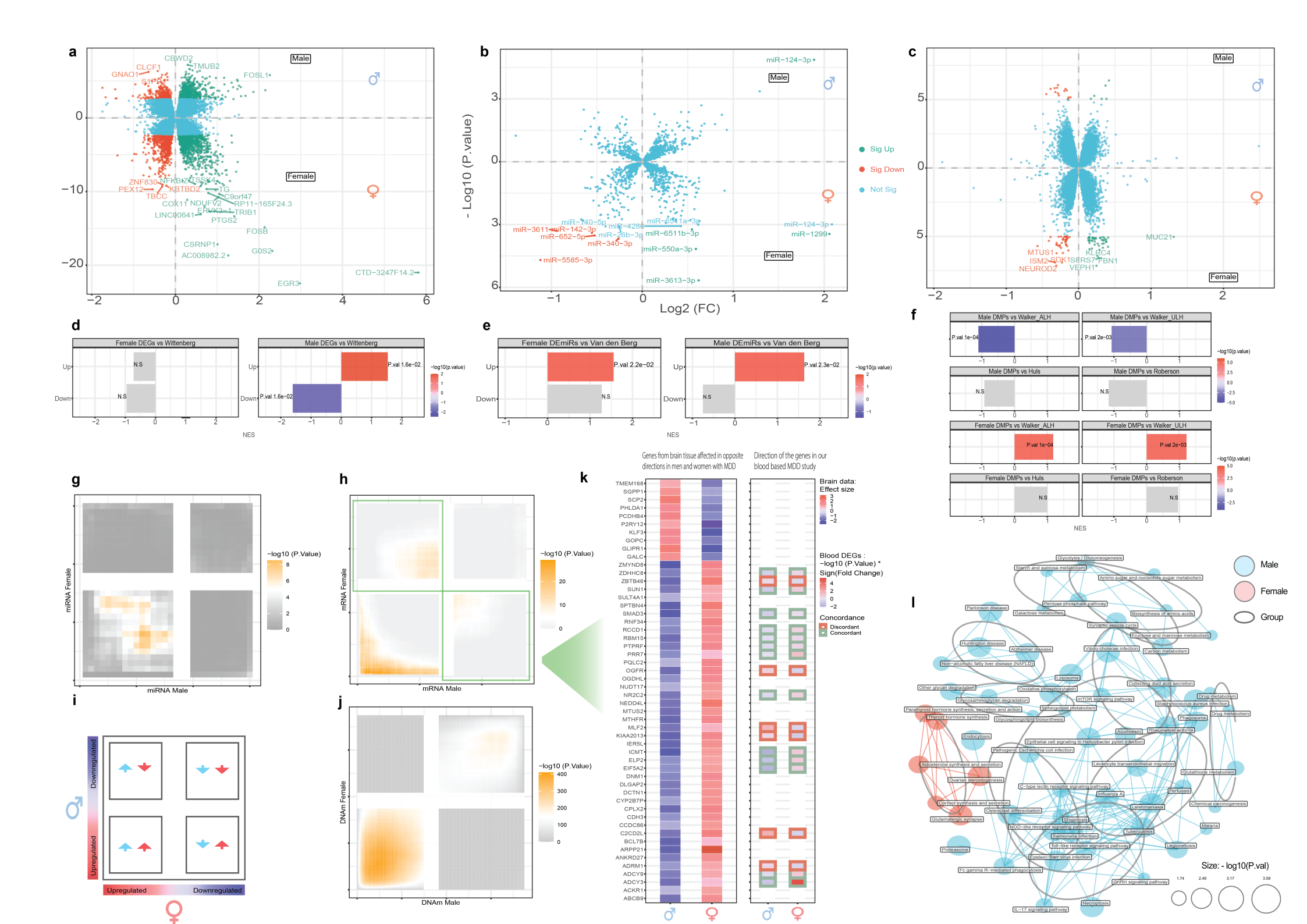

a
Male
Female
CBWD2
CLCF1
TMUB2
GNAO1
FOSL1
S1PR1
ZNF830
NFKBIZ
TSSK4
PEX12
KBTBD2
TG
TBCC
C9orf47
COX11
NDUFV2
RP11−165F24.3
ERVK3−1
TRIB1
LINC00641
PTGS2
FOSB
CSRNP1
AC008982.2
G0S2
CTD−3247F14.2
EGR3
b
Male
Female
miR−124−3p
miR−1299
miR−5585−3p
miR−3613−3p
miR−550a−3p
miR−6511b−3p
miR−6511a−3p
miR−4286
miR−26b−3p
miR−340−3p
miR−652−5p
miR−3611
miR−142−3p
miR−140−5p
- Log10 (P.value)
Log2 (FC)
Sig Up
Sig Down
Not Sig
c
Male
Female
MUC21
MTUS1
ISM2
SDK1
SRRS7
PBN1
NEUROD2
VEPH1
KLEC4
d
Female DEGs vs Wittenberg
Male DEGs vs Wittenberg
Up
Down
N.S
P.val 1.6e−02
−log10(p.value)
NES
e
Female DEmiRs vs Van den Berg
Male DEmiRs vs Van den Berg
P.val 2.2e−02
P.val 2.3e−02
f
Male DMPs vs Walker_ALH
Male DMPs vs Walker_ULH
Male DMPs vs Huls
Male DMPs vs Roberson
Female DMPs vs Walker_ALH
Female DMPs vs Walker_ULH
Female DMPs vs Huls
Female DMPs vs Roberson
P.val 1e−04
P.val 2e−03
N.S
−log10(p.value)
NES
g
miRNA Female
miRNA Male
−log10 (P.Value)
h
mRNA Female
mRNA Male
−log10 (P.Value)
i
Upregulated
Downregulated
j
DNAm Female
DNAm Male
−log10 (P.Value)
k
Genes from brain tissue affected in opposite directions in men and women with MDD
Direction of the genes in our blood based MDD study
Brain data: Effect size
Blood DEGs : −log10 (P.Value) * Sign(Fold Change)
Concordance
Discordant
Concordant
TMEM168
SGPP1
SCP2
PHLDA1
PCDHB4
P2RY12
KLF3
GOPC
GLIPR1
GALC
ZMYND8
ZDHHC8
ZBTB46
SUN1
SULT4A1
SPTBN4
SMAD3
RNF34
RCCD1
RBM15
PTPRF
PRR7
PQLC2
OGFR
OGDHL
NUDT17
NR2C2
NEDD4L
MTUS2
MTHFR
MLF2
KIAA2013
IER5L
ICMT
ELP2
EIF5A2
DNM1
DLGAP2
DCTN1
CYP2B7P
CPLX2
CDH3
CCDC86
C2CD2L
BCL7B
ARPP21
ANKRD27
ADRM1
ADCY9
ADCY3
ACKR1
ABCB9
l
Male
Female
Group
Size: - log10(P.val)
Glycolysis / Gluconeogenesis
Starch and sucrose metabolism
Amino sugar and nucleotide sugar metabolism
Parkinson disease
Pentose phosphate pathway
Galactose metabolism
Biosynthesis of amino acids
Synaptic vesicle cycle
Fructose and mannose metabolism
Huntington disease
Alzheimer disease
Vibrio cholerae infection
Carbon metabolism
Non−alcoholic fatty liver disease (NAFLD)
Lysosome
Collecting duct acid secretion
Other glycan degradation
Oxidative phosphorylation
Drug metabolism
Glycosaminoglycan degradation
mTOR signaling pathway
Staphylococcus aureus infection
Parathyroid hormone synthesis, secretion and action
Sphingolipid metabolism
Thyroid hormone synthesis
Glycosphingolipid biosynthesis
Phagosome
Alcoholism
Rheumatoid arthritis
Endocytosis
Epithelial cell signaling in Helicobacter pylori infection
Pathogenic Escherichia coli infection
Aldosterone synthesis and secretion
Leukocyte transendothelial migration
Glutathione metabolism
Ovarian steroidogenesis
C−type lectin receptor signaling pathway
Influenza A
Pertussis
Cortisol synthesis and secretion
Osteoclast differentiation
Leishmaniasis
Chemical carcinogenesis
Glutamatergic synapse
Shigellosis
NOD−like receptor signaling pathway
Malaria
Tuberculosis
Salmonella infection
Toll−like receptor signaling pathway
Legionellosis
Proteasome
Epstein−Barr virus infection
Fc gamma R−mediated phagocytosis
GnRH signaling pathway
Necroptosis
IL−17 signaling pathway

Figure 4.2: Single-omics differential analyses results and enrichements: a. Male and female volcano plots of differentially expressed genes (threshold: $p_{adj} < 0.05$). **b.** Male and female volcano plots of differentially expressed micro-RNAs (threshold: $p_{adj} < 0.05$). **c.** Male and female volcano plots of differentially methylated probes (threshold: $p_{adj} < 0.1$). **d.** GSEA results of male and female DEG comparison with Wittenberg's quantitative review. **e.** GSEA results of male and female DEmiRs comparison with van den Berg's results. **f.** GSEA results of male and female DMPs comparison with Huls, Roberson and Walker's results. Rank-rank hypergeometric overlap results: **g.** miRNA male vs female comparaison; **h.** mRNA male vs female comparaison; **i.** interpretation of the RRHO results; **d.** DNAm male vs female comparaison; **k.** Overlap of male/female discordant gene with Seney's sex specific depression signature. **l.** Graphical representation of the KEGG enrichement pathways for male and female respective DEGs.

4.3.3 Step-wise integration identified male and female MDD-associated gene co-expression modules enriched for MDD-related transcriptional and epigenetic dysregulations and mood-related SNPs

We next sought to perform multi-omic integration, using first a step-by-step approach, similar to previous work in the field Mokhtari et al. (2022). To do so, WGCNA was applied to RNA-Seq data in order to construct co-expression gene networks, separately in males and females. A total of 111 male and 135 female modules were identified. To validate these modules, we assessed their preservation in external blood transcriptome datasets. As expected, most exhibited satisfying preservation (using a classical threshold of z-summary > 2; Fig4.3a): 96.3%/88.8 % were preserved in whole blood male/female data from the GTEX consortium, respectively, while 99.0%/91.8 % were also preserved in PBMC data from (Krebs et al., 2020). Interestingly, the maximum similarity of each module with a module of the other sex remains low whatever is its association to MDD (Supplementary Figure4.4), suggesting that modular gene co-expression captured in our analyses could reflect significant sex-specificity gene regulations in peripheral blood., suggesting that modular gene co-expression exhibited significant sex-specificity in peripheral blood. Next, we prioritized most relevant modules for the identification of MDD biomarkers, using the following criteria: i) enrichment in DEG, DEmiR and DMP, ii) association with MDD severity (using the HDRS score) or past histories of childhood trauma (CTQ score), a potent risk factor for MDD (see also Chapter 6), and iii) enrichment in genetic variation associated with psychiatric disorders (using PGC GWAS data; see 4.7 for details).

Multiple modules were found significantly enriched for gene expression changes and targets of DEmiR in MDD patients. Accordingly, among the 111 male and 135 female mod-

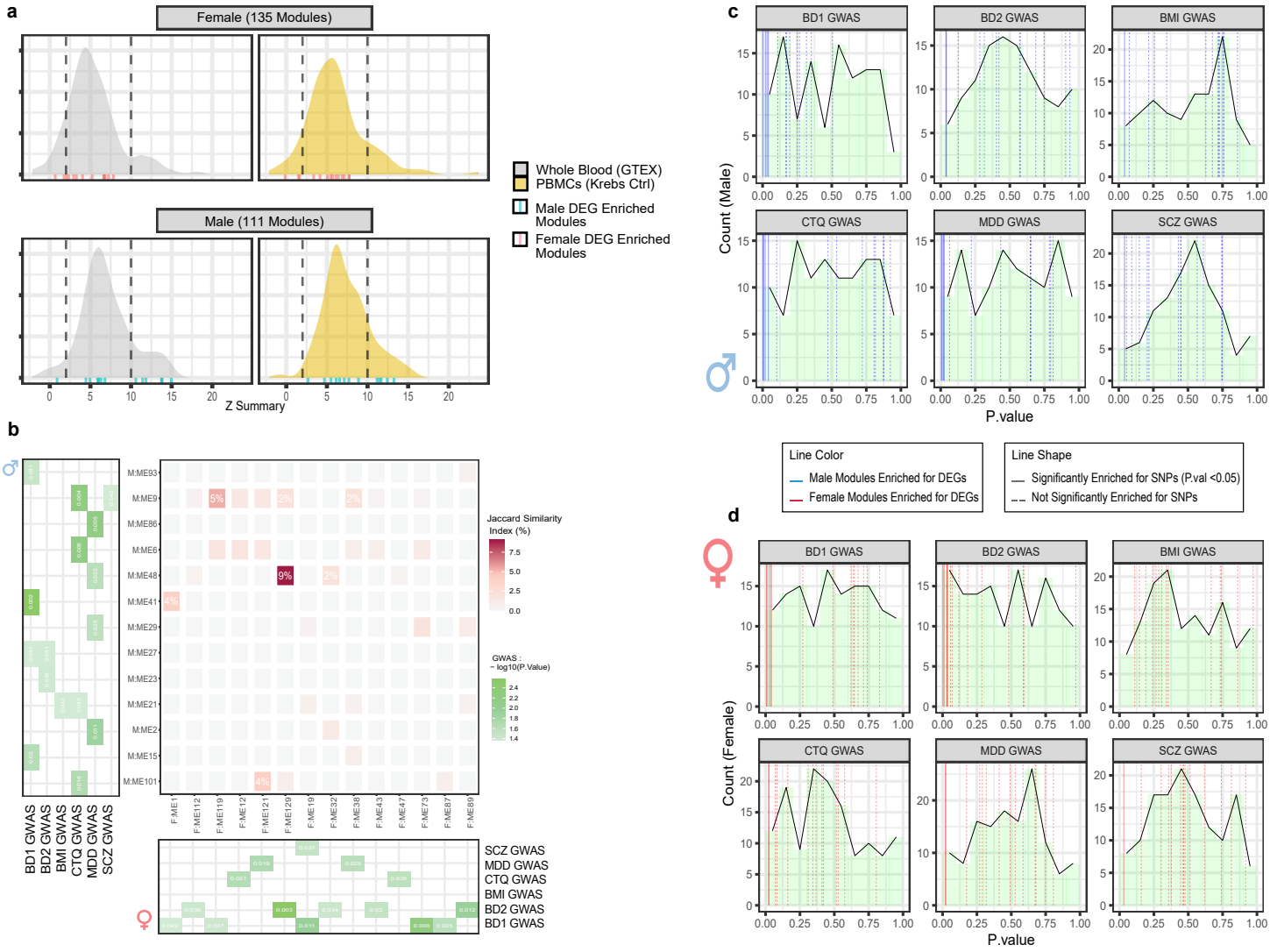


Figure 4.3: WGCNA module analysis: a. Distribution of WGCNA's Z-summary statistics, the module's preservation statistics, in an external whole blood and PBMCs datasets from GTEX database and Krebs's et al. controls (Krebs et al., 2020), respectively. Male and female DEG enriched module Z-summary values are represented respectively in blue and red. Modules are preserved if Z-summary > 2. **b.** Jaccard similarity matrix of the DEG enriched modules, the p-values of the module's enrichment for SNPs associated with BMI and psychiatric phenotypes are represented at the bottom and left side of the matrix. **c and d** Distribution of the p-values of the enrichment for BMI and psychiatric phenotypes GWASs in all the generated modules, the values of the DEG enriched modules are represented with vertical lines.

ules, 13 and 14 were significantly enriched for up- or down-regulation in MDD. Similar to what was observed for all modules, those that associated with MDD showed limited overlap across males and females, with the highest Jaccard index of 9% for the F:ME129 and M:ME48 modules (Fig4.3b). In addition, among the 27 MDD-related modules, several exhibited sig-

nificant associations with GWAS SNPs identified in relation to MDD or BD, in both males (Fig4.3d) and females (Fig4.3d). As expected, these associations were more frequent than those observed for SNPs related to BMI (used as negative control), or to schizophrenia (consistent with the notion that the genetic architecture of MDD is closer to that of BD than of psychosis).

To better appreciate how multiple types of module annotation together reveal those that are most relevant to MDD, we use a circle plot representation (Fig4.4), and below describe its successive concentric circles, from exterior to interior. Among these modules, 8 exhibited enrichment for the targets of at least one DEmiR (1 male module and 7 female modules). No enrichment, however, was significant for DMPs, perhaps reflecting the much larger dimensionality of DNA methylation data. Furthermore, of the 27 modules, two-thirds (18 modules, 5 in males, 13 in females) showed significant correlation between their eigengene (a measure of overall co-expression among each module) and the HDRS total depression score. As expected, this correlation was much less frequent for other modules not enriched for MDD-related transcriptional dysregulation, consistent with the notion that WGCNA co-expression and differential expression analyses capture only partly overlapping signals. In addition, 10 modules (1 in males, 9 in females) showed significant correlation with the overall CTQ childhood trauma score. Finally, regarding associations with GWAS data, 5 modules (3 in male, 2 in females modules) were enriched for genes associated with MDD-related SNPs, and 5 (3 in males, 2 in females) with childhood trauma-related SNPs. Again, the later association was more frequent than for other, non-MDD related modules, altogether supporting the validity of the modular genetic architecture captured by our WGCNA approach.

As illustration of these results, we then focused on the 2 most consistently annotated modules, M:ME48 (men module) and F:ME129 (female module), for more in-depth analyses. Specifically, M:ME48 (181 genes, men module) and F:ME129 (66 genes, female module) exhibited enrichment for male MDD up-regulated genes (NES = 2.10, $p_{adj} = 1.0 \times 10^{-4}$) and female MDD down-regulated genes (NES = -2.11, $p_{adj} = 4.0^{-4}$), respectively. Furthermore, while both M:ME48 and F:ME129 were both enriched for validated gene targets of 6 miRs (miR-124-3p, miR-532-3p, miR-92a-3p, miR-1270, miR-181d-5p and miR-4286), enrichments for miR-550a-3p and miR-320b were specific to M:ME48. Interestingly, despite the low similarity observed, regardless of MDD, across all male and female modules, these 2 modules showed the best Jaccard similarity value among the 27 modules of interest (J(M:ME48, F:ME129) $\sim$ 0.09). This further illustrates the notion that, although sex differences are meaningful, a certain degree of similarity also appears across males and females with MDD.

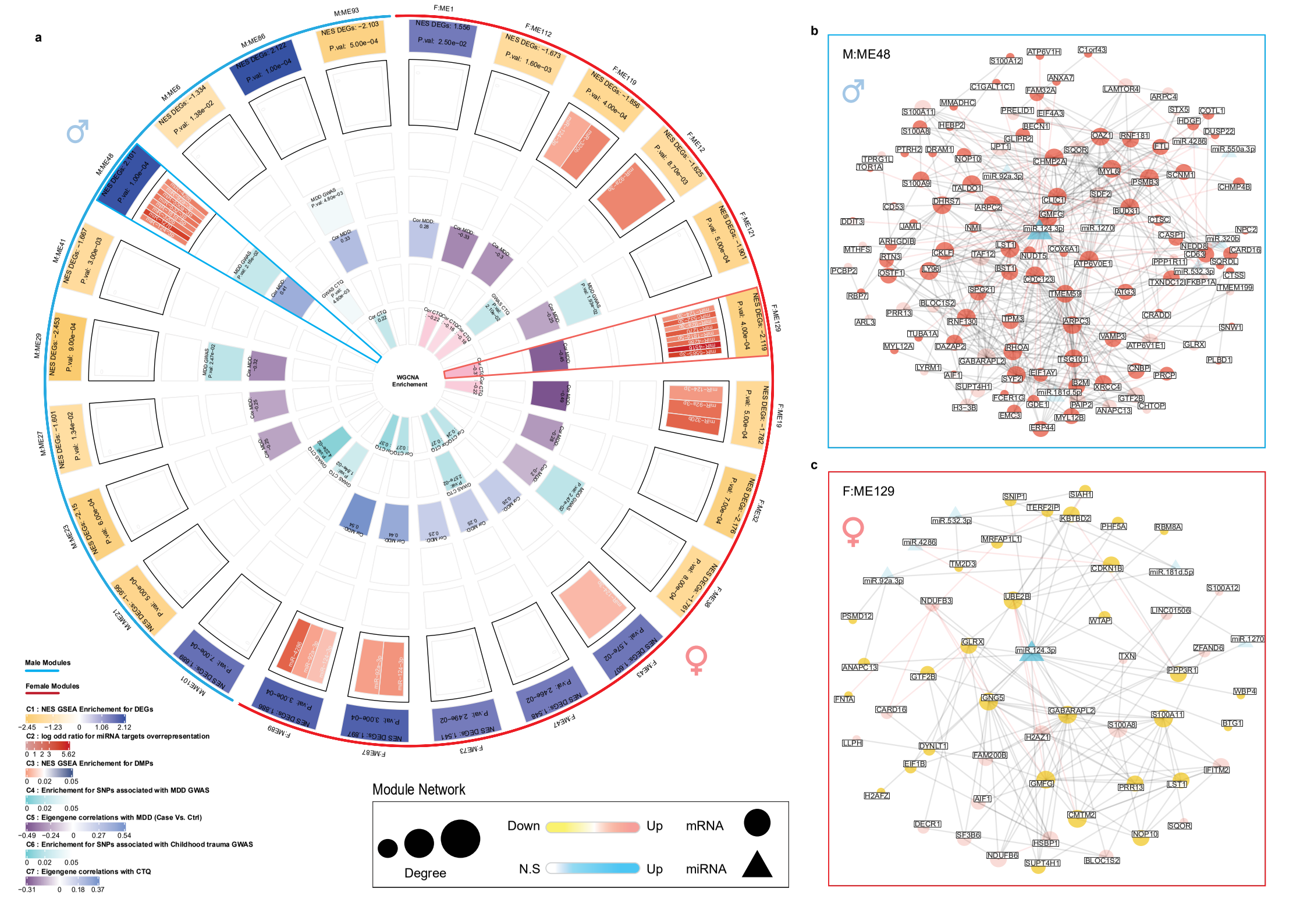

a
b
c
M:ME48
F:ME129
WGCNA Enrichment
Male Modules
Female Modules
C1 : NES GSEA Enrichment for DEGs
C2 : log odd ratio for miRNA targets overrepresentation
C3 : NES GSEA Enrichment for DMPs
C4 : Enrichment for SNPs associated with MDD GWAS
C5 : Eigengene correlations with MDD (Case Vs. Ctrl)
C6 : Enrichment for SNPs associated with Childhood trauma GWAS
C7 : Eigengene correlations with CTQ
Module Network
Degree
Down
Up
mRNA
N.S
Up
miRNA

Figure 4.4: Visualization of the modules of interest a. Circos plot: Concentric circles representing the different levels of the modules enriched for MDD-related transcriptional dysregulation, each ring represents a different level of enrichment and each triangle represents a module. Male modules are highlighted in blue and female ones in red. A color gradient is applied for each enrichment according to its direction or p-value if the module's enrichement met statisitcal significance criteria (p-value < 0.05). The first circle from the outside to the inside represents the results of the GSEA with the DEGs, the second the overlaps of a module's genes with the biologically validated genes target of mirs, the third for the GSEAs with the DMPs, the 4th and 6th are enrichment for SNPs associated with the risk of developing MDD and childhood trauma, respectively, and finally the 5th and 7th are for the respective correlations of module's eigengenes with depression and childhood trauma score phenotypes. **b and c** Graphical representations of the M:ME48 and F:ME129 networks: The vertices in the form of circles represent the co-expressed genes in each module and the triangles are miRs from the enrichement analysis, the edges colored in grey are the weighted co-expressions calculated by WGCNA between the module's genes and those in red represent the epigenetic regulations of miR resulting from crossing with the mirBase database. The size of the vertices is proportional to their degree of connection

A graphical representation of the M:ME48 and M:F129 networks is presented in (Fig4.4b-c), and depicts both DEGs and validated targets of differentially expressed miR as hub genes within each network. Interestingly, the differentially expressed miR-124-3p (males: $p_{adj} = 9.9 \times 10^{-3}$, females (N.S) : $p_{adj} = 5.37 \times 10^{-2}$) appeared centrally located in both modules, with significant enrichment for its target genes (M:ME48 :$p_{adj} = 4.5 \times 10^{-27}$, F:ME129 $p_{adj} = 2.31 \times 10^{-13}$), indicating this miR may potentially act as a key regulator of these modules. In addition, 16 genes were commonly found in both modules (*GABARAPL2, ANAPC13, GMFG, SQOR, GTF2B, S100A8, S100A11, GLRX, NOP10, CARD16, AIF1, LST1, PRR13, BLOC1S2, SUPT4H1, S100A12*), among which *GMFG, S100A8, NOP10, CARD16, AIF1, LST1* and *S100A12* were significantly up-regulated in males and *GABARAPL2, ANAPC13, GMFG, GTF2B, S100A11, GLRX, NOP10, LST1, PRR13* and *SUPT4H1* significantly down-regulated in females. Gene ontology, KEGG and Reactome pathways enrichments in supp figure 4.3 showed involvement of M:ME48 genes in immune response, including neutrophil-mediated immunity and leukocyte degranulation, while F:ME129 was related to WNT signaling pathway, which is under intense investigation in both preclinical models of MDD (Hussaini et al., 2014) and peripheral biomarkers studies of antidepressant response (Lopez et al., 2017). Overall, this step-by-step network approach provides a coherent and sex-specific description of the disruption of modular gene co-expression in MDD. To go beyond this and other approaches previously used in the field, and with the goal of fully leveraging the genome-wide and multi-omic nature of our dataset, we finally implemented a supervised framework, as detailed below.

4.3.4 Advanced integration with supervised feature selection and non-supervised clustering prioritize mRNAs and DNA methylation CpGs as more efficient biomarkers of MDD

Because biomarker discovery programs aims at identifying clinically useful features, we opted for a framework designed, in a diagnostic approach, to select the set of features (miRNA, mRNA and CpG probes) that could best discriminate MDD patients and controls. As illustrated in figure 4.5, in order to maximize their stability and robustness, their identification was encapsulated in a 5-fold CV. First, for each training set, the 6 joint dimension reduction (jDR) methods of Momix Cantini et al., 2021 were used to identify the top 10% features that most contributed to co-covariance among the 3 omic blocks and MDD severity (HDRS score). Then, a multi-omic-based similarity matrix, generated using SNF label propagation, was used to predict MDD status of each test sample.

The resulting accuracy of predictions for each omic combination, in each male-only, female-only, or pooled cohort, are displayed in Fig.4.6.

Regarding different omic combinations, in the pooled male and female cohort (Fig.4.6.c), the combination of mRNA and DNAm features identified by JIVE yielded the highest accuracy (ACC = 0.67). Similarly, in the male cohort (Fig.4.6.b), the same mRNA and DNAm combination, this time using RGCCA, outperformed other combinations (ACC = 0.71). In contrast, in the female cohort (Fig.4.6.b), the mRNA block alone, using either RGCCA or JIVE (ACC = 0.77 and ACC = 0.76, respectively), performed slightly better than the combination of the 3 omics analyzed using RGCCA (ACC = 0.73).

Regarding relative performances among the 6 jDR approaches that we used, we note that JIVE and RGCCA systematically outperformed the others, in all 3 cohorts (see section 4.4.2 for a discussion). Interestingly, DEGs, DEmiRs or DMPs, when obtained for each training set by single-omic differential analyses, and applied to the test datasets, performed worse than jDR-predicted features (with a maximal mean accuracy at 0.64 for the male cohort, using both mRNA and DNAm).

Finally, the stability of features derived from the appreciation of ACC results was assessed using bootstrapping (n=1000). The steps of this procedure are summarized in section 4.2.12. Female features identified in the female cohort exhibited, on average, higher levels of stability, followed by features identified in males only (see supp Table4.6). Overall, in the pooled cohort, the number of stable features were lower than those obtained with the sex separated cohorts, with better results obtained for the mRNA block only, or the DNAm - miR combination (see supp Table4.6).This lower stability of features in the pooled cohort is

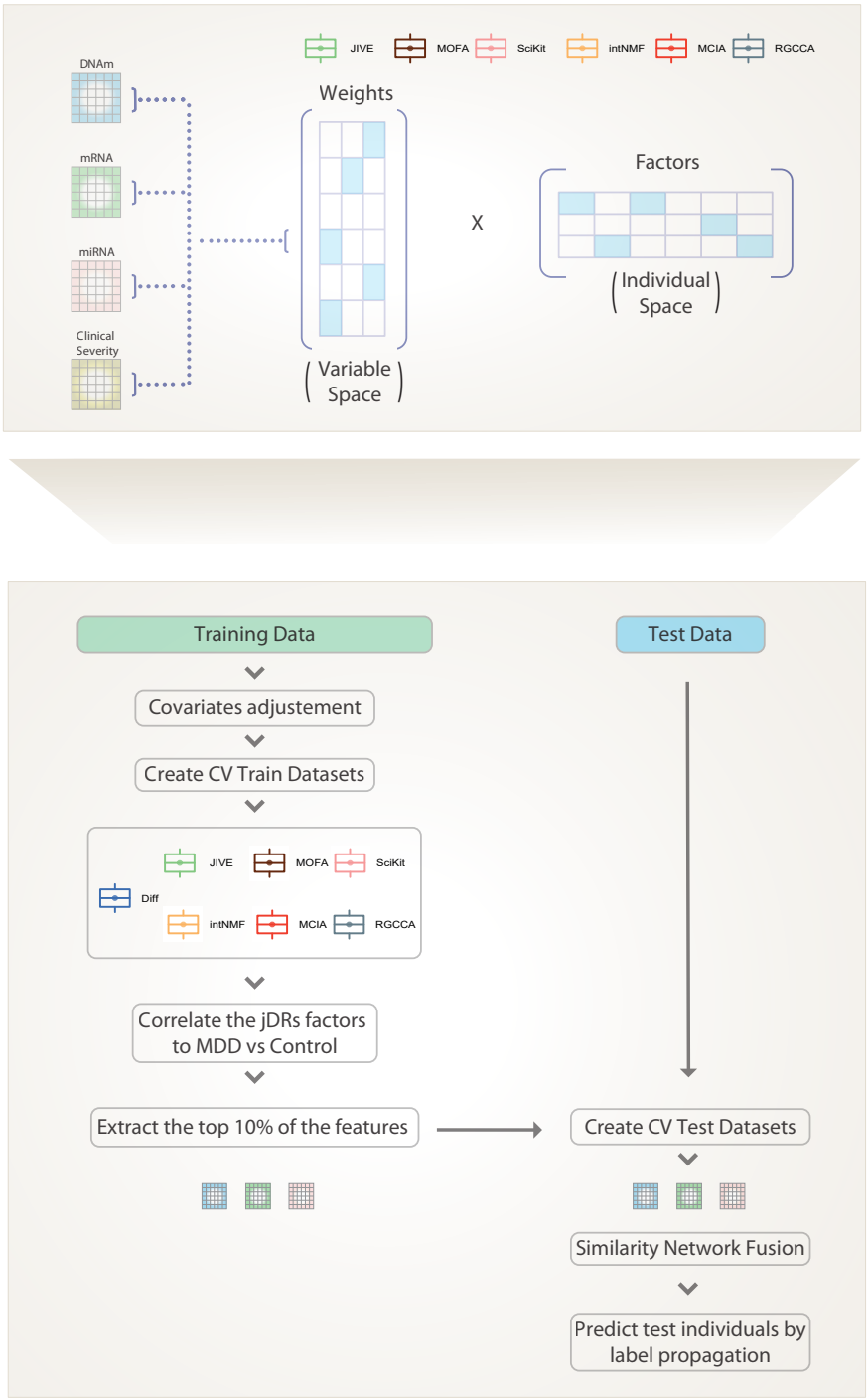


Figure 4.5: Graphical visualization of the multi-omic integration and classification framework: For each of the 25 splits of the 5-fold cross validation, all the combinations of the omics data and the clinical severity matrix of depression (represented by the score of the 17 questions from Hamilton) are given as inputs to 6 different jDR methods, the resulting factor matrices are then correlated with the MDD vs Controls phenotype and the top 10% of the omics factors contributing to the best correlation are extracted to reconstruct a new set of matrices on the train dataset. A similarity matrix is generated from the new train matrices using SNF, and a label propagation procedure is applied to predict the group of each sample.

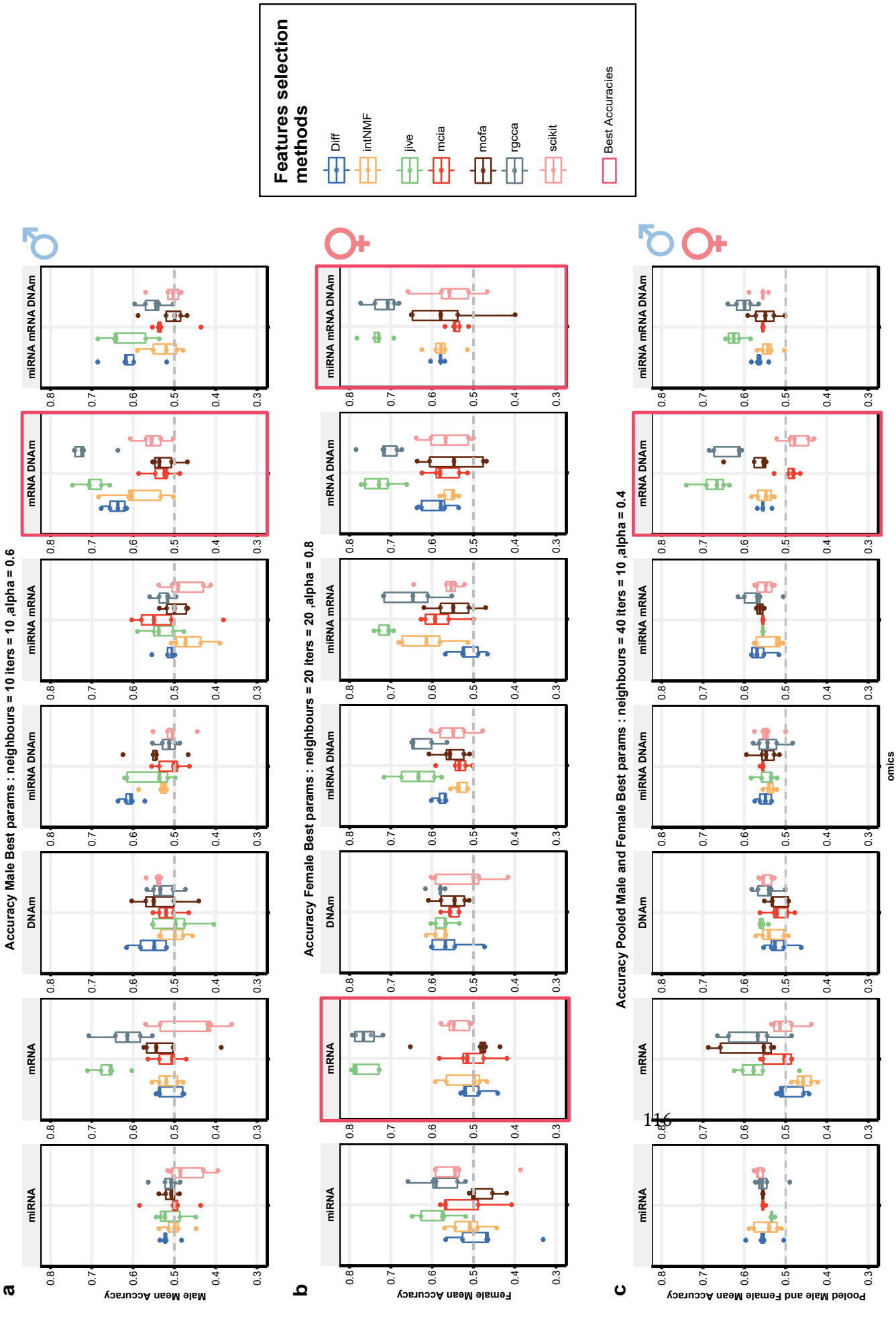

Features selection methods
Diff
intNMF
jive
mcia
mofa
rgcca
scikit
Best Accuracies
a
Accuracy Male Best params : neighbours = 10 iters = 10 ,alpha = 0.6
Male Mean Accuracy
b
Accuracy Female Best params : neighbours = 20 iters = 20 ,alpha = 0.8
Female Mean Accuracy
c
Accuracy Pooled Male and Female Best params : neighbours = 40 iters = 10 ,alpha = 0.4
Pooled Male and Female Mean Accuracy
miRNA
mRNA
DNAm
miRNA mRNA
miRNA DNAm
mRNA DNAm
miRNA mRNA DNAm
omics

Figure 4.6: Comparison of prediction accuracy performances of the selection feature methods and omics combination: In the boxplots, the mean accuracy of the 5 splits of each iteration of the cross validation is represented with a single point. The boxes are colored according to the 6 jDR and differential analysis methods. The results are divided according to the investigated omics combinations. Boxplots **a and b** respectively represent the results of the framework applied to the male and female datasets and **c** those of the pooled cohort.

consistent with the notion of different MDD-related biomarker profiles across sexes.

4.3.5 Different omics analyses and integration strategies shows common features serving as potential biomarkers and molecular signature for sex-specific MDD

In the present work, we opted to apply sequentially 3 analytical strategies (single-omic analysis, network modeling, and multi-omic integration), and finally investigated their potential convergence. Overlap analyses were used to assess significant enrichment in stable mRNA features (from multi-omic classification and boostrapping) among the co-expressed MDD-related modules (identified by WGCNA). Over the 246 modules, both M:ME48 and F:ME129 appeared systematically enriched for stable mRNA features. Moreover, as shown in supplementary Fig4.5, when crossed with results of the differential expression analysis and the modules of interest from the step-wise multi-omic integration, the overlapping genes corresponded to DEGs in their vast majority. Regarding miRNA, miR-124-3P was simultaneously differentially expressed in males, as a key regulator in the two prioritized modules, and also appeared as a stable miR feature in males (in the JIVE combined analysis of all 3 omics). Among the stable mRNA and DNAm features, many were also found using the single omic differential analyses in a case/control design. Importantly, mRNA and DNAm features selected using JIVE or RGCCA outperformed the prediction performance obtained with DEGs (alone or in combination with DMPs), whatever the cohort used (sex-specific, or pooled). This, therefore, strongly argues for the added value, when predicting MDD status, of integrative methods that take into consideration the covariance between multi-omic measures and clinical severity.

4.4 Discussion

4.4.1 Sex specific molecular signatures of MDD

Depression is a complex disorder with substantial sex heterogeneity, whose molecular underpinnings are poorly understood. This question of sex-differences in MDD has only

recently started to be investigated at molecular and genome-wide levels, with available evidence coming from transcriptomic studies conducted on brain tissue (Eid et al., 2019). In the present work, we conducted sex-stratified single-omic and multi-omic integrative analyses, with the aim of characterizing blood biomarkers of MDD.

The comparison of molecular differences associated with MDD were to some extent concordant for each omic across males and females, but also revealed significant discrepancies, particularly for the mRNA layer. Indeed, regarding gene expression (see RRHO2 results, Fig4.2g-j), these differences were subtle, but appeared nevertheless partly consistent with those previously identified in the brain by Seney et al. (2018). Seney and colleagues, who investigated 3 brain regions (in a cohort of N=26 males and 24 females with MDD, as well as sex-matched healthy controls), found that among 73 DEGs commonly observed in males and females, 52 genes exhibited transcriptional changes in opposite directions across sexes. Beyond those individual genes, at a broader level their GO analyses further revealed that synapse-related genes were decreased in depressed males, but elevated in females. These findings raise the intriguing possibility that MDD pathophysiology may recruit distinct, and possibly partly opposite, gene expression changes. In our own blood-based study, results from KEGG pathway enrichment analysis also differed significantly: in males, terms were related to immune system response, IL-7 signaling pathway and toll like receptors, while, in females, they implicated the production and secretion of thyroid and parathyroid hormone, aldosterone and ovarian steroidogenesis, as well as glutamatergic signaling. Interestingly, sex variations in the immune response to infection and inflammatory diseases have already been documented (Klein & Flanagan, 2016). In the context of MDD, Ramsey et al. (2016) found greater levels of several immune-related proteins, including the acute phase reactant C-reactive protein (CRP) in males compared to females. In addition, (Eid et al., 2019), Niles et al. (2018) previously reported sex differences in longitudinal relationships between depression, anxiety symptoms and inflammation in an elderly population, using high sensitivity C-reactive protein dosing. In their study, Niles and colleagues found that over four years, depressive symptoms predicted increasing inflammation for men but not women, and inflammation predicted worsening depression for women but not men.

Moreover, Ernst et al. (2021) showed that inflammatory indicators were better predictors of MDD in males than in women. On the other hand, in their recent literature review of human studies, Bekhbat et al. (2019) observed that the pro-inflammatory response in MDD was more prominently observed in women, and was accompanied by decreased glucocorticoid sensitivity. One of the molecular mechanisms that may underlie these immune sex difference involve steroid hormones. Indeed, (Kokras et al., 2019) showed sex differences in HPA axis regulation in depression, with gonadal steroids acting on the sensitivity to glu-

cocorticoid negative feedback. In addition, Frokjaer et al. (2015) demonstrated that net reduction in estradiol (steroid hormone) increased depressive symptoms in healthy women. Despite these first evidence, future studies will be necessary to go deeper in the understanding of sex differences that may drive distinct transcriptomic adaptations in males and females with MDD.

To go beyond these analyses at the level of individual genes, we next adopted a systems biology approach, with a step-wise multi-omic integration. The first step consisted of independently identifying groups of genes in each sex cohort, defined as modules, which significantly co-varied in peripheral blood. Results uncovered small overlaps among sex-specific co-expression modules whose eigengenes significantly correlated with MDD status, highlighting substantial sex divergence in MDD. Among these modules, we then prioritized the 2 that appeared most relevant to MDD as they exhibited several enrichment in other omic layers, as well as with loci previously associated with the risk of MDD (in males: module M:ME48; in females, module F:ME129).

Using the same WGCNA approach, Labonté et al. (2017b) previously characterized 143 and 167 modules in 6 brain areas of male and female MDD brains, among which only 55 male modules showed high levels of overlap with female modules (corrected Fisher's Exact Test < 0.05 ; Fold enrichment > 15). Furthermore, out of these 55 conserved modules, only 10 showed shared network connectivity with their female homologues. Hence, Labonté et al. (2017b)'s findings suggest the presence of sex-specific transcriptional signatures associated with MDD. Interestingly, GO analyses of Labonté and colleagues modules showed cell-type specific enrichment including neurons, glia, and endothelial cells. In contrast, in our blood-derived results, the GO terms of the upregulation-enriched M:ME48 module included mainly the response of the innate immune system, where the downregulation-enriched F:ME129 module showed mostly involved *IL*-1 and *WNT* signaling pathways (see supp Fig5.3). Of note, the WNT signaling pathway has previously been associated with antidepressant treatment response in depression (Hussaini et al., 2014) (see also the discussion of limitations, section 4.4.5).

Interestingly, despite the low similarity among these 2 modules, both exhibited miR-124-3P as a potential regulator, with gene targets of this miR more frequent in those modules than expected by chance. Moreover, miR-124 was significantly dysregulated in males (FDR < 0.05), and nominally significant in females. These results appear intuitive considering accumulating evidence of a significant contribution of miRNAs to depression (Mokhtari et al., 2022; Yuan et al., 2018; Żurawek & Turecki, 2021b), most notably miR-124. Indeed, this miR has been suggested as a putative epigenetic biomarker of the onset of MDD (Roy et al., 2016; Roy et al., 2017), and as a regulator of immune response and inflammatory

disorders (Inflammatory bowel diseases, Rheumatoid arthritis, and Ankylosing spondylitis) in peripheral tissues (Qin et al., 2016). Yet, given smoking's correlation with miR-124 expression (de Ronde et al., 2017), our results should be considered with caution. In light of these findings, further investigation of the sex-dependent dysorganization of gene modules related to inflammatory and immune responses should be carried out in MDD.

4.4.2 Added value and limitations of our multi-omic data integration framework

In parallel to single omic analyses and step-by-step integration discussed above, we also implemented an original framework for multi-omic integration. This framework first included a supervised feature selection (jDR), followed by unsupervised cluster prediction (network similarity). To our knowledge, this is the first time that jDR integration approaches are applied to MDD. Previously, Bhak et al. (2019) had proposed a classifier to differentiate 56 suicide attempters, 39 MDD patients, and 87 controls based on gene expression and methylation data using a supervised machine learning approach (random forest). They reported accuracies of near 0.9 for their prediction models. However, they selected features corresponding solely to differentially expressed and methylated loci, which were tested for performance in a cross-validation procedure. Based on our framework evaluation, we suspect this may have resulted in data overfitting.

In our framework, we carefully included CV and bootstrap procedures to systematically examine the selected features' generalization, as well as their stability, while avoiding overfitting. Importantly, because this methodology was applied to a cohort of samples assessed for three omic layers, we could systematically assess the information gain potentially attainable for predicting MDD status when analyzing 1, 2, or 3 omic types. For the pooled males/females cohort, or males only, and although the accuracy that was achieved remained relatively modest (Acc $\sim$ 0.7), our results provide evidence that the combined use of DNAm and mRNA may provide better and more stable biomarkers for MDD prediction. However, this informational gain was not significant for the woman dataset, likely because the mRNA single omic already enabled high accuracy in that cohort (reflecting greater mRNA case-control variance).

Remarkably, regardless of the used combination of omics or cohort, JIVE and RGCCA near systematically outperformed the remaining 4 Momix jDR methods. Two hypotheses may explain the high accuracy yielded by these two methods only. First, we can notice that Scikit-Fusion, MOFA and intNMF are based on matrix factorization, thus constraining all blocks to share a common factor matrix, which may be too restrictive. On the other hand, JIVE allows for a common and an individual factor matrix to both represent a block. This

allows for strong individual structure, to have less influence on the determination of the common structure as they are estimated separately. For RGCCA, its covariance criterion might give more flexibility to the factor matrix that can be estimated. MCIA is closely related to RGCCA, however, in the implementation proposed by Meng et al., 2014, as a pre-processing step, each block is separately projected on a lower dimensional space, which might lose information on their interactions. Second, specific limits to each jDR methods can be found. For instance, in Scikit-Fusion and MOFA, it seems that there is no scaling across blocks, which might lead to unfair representation of the omic modalities by the factor matrix. Furthermore, Scikit-Fusion in the Momix framework is applied without constraint on the weight matrices, which result in an analysis close to a PCA on the concatenated blocks. Finally, the non-negative constraint, specific to intNMF, might not be relevant for this particular dataset.

Our benchmark of jDR approaches with different omic combinations allowed direct comparison of their respective accuracy for the prediction of MDD status. Results indicated that the combination of mRNA and DNAm performed, on average, better than single omics alone in that task. However, the mRNA dataset alone performed better for female data, with near-equivalent prediction accuracy to multi-omic ones. In addition, globally better accuracy and higher number of stable features were obtained in each sex-specific cohort than in the pooling one. This again underscores the importance of stratifying for sex when investigating epigenetics and gene expression to characterize the blood molecular signatures associated to MDD.

Several approaches for unsupervised clustering based on multi-omic data have been used in the present work. However, there is room for improvement in our framework. Indeed, for jDR methods, we applied default parameters for the sake of comparison. This does not fully accommodate for the effectiveness of each method. Along this line, parameter optimization should be considered in future studies, despite the fact that it would require significantly more computing resources. One of the most important parameters for multi-block integration is the design matrix that describes the relationships between blocks C. That is, in our case, the relationships between the different omic blocks and MDD severity (HDRS scores). Indeed, one might evaluate the optimal strength of these links, or alternatively impose higher or lower link, depending on prior hypotheses on the nature of relationships among those blocks (e.g. prioritize the covariance captured between the epigenetic blocks and the variable of interest over the one captured by transcriptomics).

While all our approaches were linear, molecular interactions leading to MDD may be nonlinear. Thus, techniques of deep learning, particularly those based on auto-encoders recently applied in the cancer field (Franco et al., 2021), might be an interesting alternative.

The major drawbacks of these approaches are the large number of samples required, and their "black box" component, which complicates biological interpretation.

Finally, the addition of a sparsity constraints might help to maintain model interpretability (linear and nonlinear), and prevent overfitting. Indeed, while in its current implementation our feature selection is based on an arbitrary threshold (top 10%), adding sparsity would force the estimation of feature weights to contain only a small set of non-zeros. Several methods to incorporate such sparsity into multi-block integration exist, including SGCCA (Tenenhaus et al., 2014) and PathME (Lemsara et al., 2020). Interestingly, the latter takes biological interactions (signaling pathways and groups of genes of interest) into consideration, and may lead to improved prediction and a reduced number of potential biomarkers.

4.4.3 Sample size limitations

Our analysis is not devoid of limitations. In the following paragraphs, we briefly elaborate on these, with a distinction between biological limitations, inherent to molecular psychiatry, and technical ones, which are more specific to multi-omic integration.

The first limitation is our cohort's size. Although the investigations described above include analyses of more than 150 individuals and 3 distinct molecular mechanisms, these numbers do not allow drawing definitive conclusions about MDD biomarkers. For example, even though multi-omic integration appeared to improve MDD status prediction, the accuracy that was achieved remained relatively modest. Among other considerations, the etiological diversity and symptomatic heterogeneity (Ormel et al., 2019) of MDD are 2 factors that contribute to explain why a larger sample size will be necessary for future studies. Furthermore, our results suggest that sex represents a source of variation that should be carefully considered for biomarker development, with poor overlap among features identified among men and women during multi-omic integration. This sample size limitation is not a surprise as, for example, GWAS analyses of several psychiatric disorders, including MDD, have now clearly demonstrated that the number of genome-wide significant findings steadily increase with cohort size (which now typically counts in hundreds of thousands of individuals in the latest PGC publications (Giannakopoulou et al., 2021)). Because the kind of functional genomic mechanisms that we investigated here can be considered "closer" to the expression of symptoms than genetic vulnerability, it is possible to speculate that a lower sample size could be sufficient for the identification of robust biomarkers, although this putative lower threshold remains to be determined. Recruiting and analyzing larger cohorts will inevitably lead to serious financial, clinical and logistical challenges, which will likely require international collaborations.

In order to establish larger cohorts, particularly for GWAS research, minimal phenotyping is often used instead of professional-base diagnostics. However, Cai et al. (2020) have shown that a strategy relying on self-reported MDD tends to provide findings that are not unique to MDD. Indeed, heritability for MDD with minimal phenotyping was lower than that for DSM5-defined MDD. In addition, GWAS hits using minimal MDD phenotyping were shared with those from neuroticism and smoking. These results suggest that building larger cohorts through minimal phenotyping may lead to less specific results due to higher clinical heterogeneity.

4.4.4 Is blood a relevant tissue for mood disorder research ?

Despite the fact that the brain is at the center of mental disorders pathophysiology, for obvious reasons the majority of MD biomarker research is performed on more easily accessible tissues, such as blood and saliva (Mokhtari et al., 2022). This raises the question of the extent to which molecular differences measured e.g. in blood may have any relationship to pathological processes occurring in the brain. In this regard, research indicates that the correlation between molecular profiles assessed across peripheral tissues and the brain is limited, even in the absence of any consideration of psychopathology. In their review, Tylee et al. (2013) examined these comparisons for transcriptomic, epigenetic, and proteomic data. This review revealed a poor assessment of cross-tissue proteomics variability, a gene specific correlation between transcriptomics expression levels in brain and blood (ρ cross-tissue correlation $\sim$ [0.25 - 0.64]), and finally a high correlation among CpG sites ($\rho = 0.90$). Furthermore, Hannon et al. (2015) and Walton et al. (2015) investigated the relationship between DNA methylation patterns across brain and blood. Their findings revealed that among 450k CpG sites investigated in the human genome, a strong association ($\rho > 0.5$) between the 2 tissues was only reported for a tiny proportion (7.9%), which could be those that might be of interest in a psychiatric context. Additionally, Braun et al. (2019) and Smith et al. (2014) showed that, on average, brain-peripheral tissues correlations are higher for saliva than blood while, for individual CpGs, blood had the highest proportion of CpGs correlating with those in the brain. Finally, Edgar et al. (2017) introduced BEcon as a method to examine simultaneously multiple metrics of the concordance of blood and brain DNAm.

4.4.5 Confounding factors

In the present work, careful analysis of possible confounding factors was carried out in order to identify the variables that most substantially impacted variance, and to adjust for such effects. It is known that quantitative measurements of gene expression and methyla-

tion are correlated with gender, age, BMI, and blood cell composition. We are confident in our analyses for these confounding variables, as they were adjusted for. Despite previous report of an effect of the smoking status on blood omic data (Charlesworth et al., 2010; Gao et al., 2015), we found only a weak effect in our data. Given the strong correlation between smoking status and MDD (Fluharty et al., 2016), the distinction of their respective variances is challenging (Barbu et al., 2020). For these reasons, we chose not to account for smoking. For the confounding factors of medications and substance abuse, the additional difficulty was the high rate of missing data for these factors in our data. Although it has already been shown that patient's medication, including antidepressants, may have strong influence (Belzeaux et al., 2018; Lisoway et al., 2018; Webb et al., 2020), the unavailability and incompleteness of the medication reports is often an issue. To circumvent these missing data and possible recall default on these questions, other designs were applied, such as longitudinal studies to identify antidepressant signatures and biomarkers of antidepressant response (da Silva et al., 2021). One could also compare the molecular signatures revealed by our analyses with those of medications accessible in the Library of Integrated Network-based Cellular Signatures (LINCs) databases (Lamb et al., 2006).

4.5 Perspectives

As shown in the present work, multi-omic data integration and separation of male and female data enhance the likelihood of distinguishing MDD patients from controls using peripheral blood molecular measures. Based on these results, it is possible to speculate that applying such methods may help identify subgroups of MDD patients that may better respond to distinct types of pharmacotherapies. Indeed, recent efforts to identify biomarkers of antidepressant response have left some stones unturned. The development of such biomarkers was mainly confined to single-omic approaches, and suffered from limited cohort size (Amasi-Hartoonian et al., 2022). A more radical approach will likely be necessary to ensure the future success of the response to treatment biomarker research. Building on previously accumulated knowledge and new multiomic study design and methodologies, one could explore the differential impact of antidepressants on biological pathways simultaneously on patients' transcriptome, epigenome, and metabolome. A second approach could focus on advanced multi-omic integrative methods to predict the response to multiple treatments in a longitudinal study and evaluate the predictive power of a generalized multi-omic model.

4.6 Conclusion

In conclusion, our study provides evidence for sex-specific peripheral molecular signatures of MDD, and advocates for the use of multi-omic integration strategies for MDD prediction, and the development of personalized treatments.

4.7 Supplementary material

4.7.1 Adjusting data for batch effects with ComBat while preserving group differences may lead to an over estimation of the significance of the differential analyses.

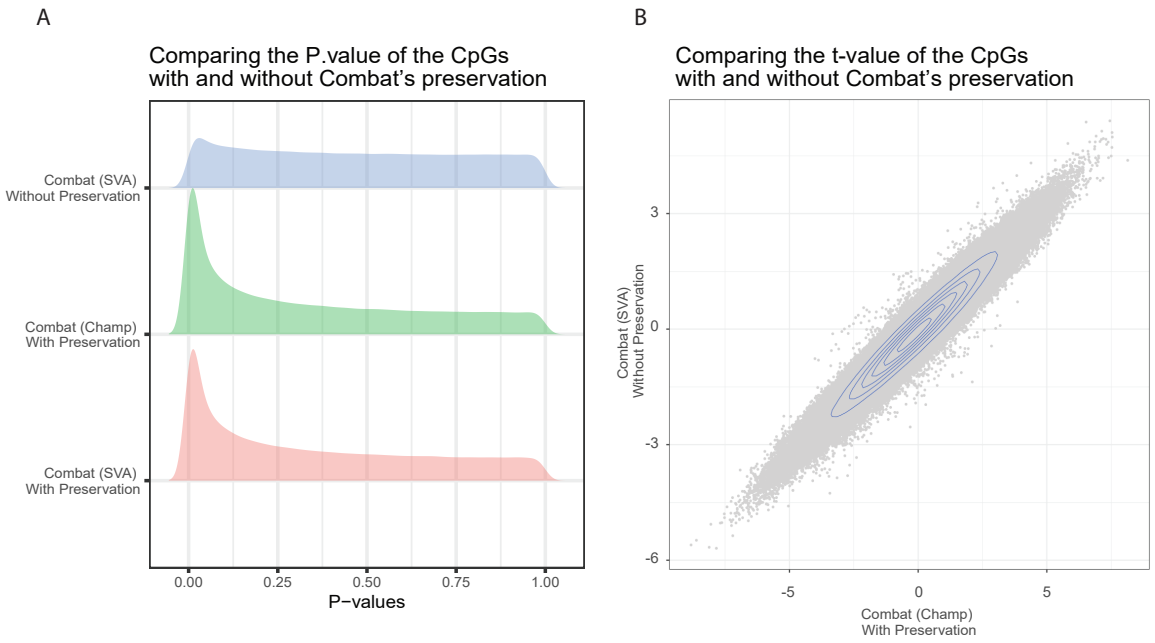
The analysis of variance presented in section 4.2.7 demonstrated the presence of technical batches, inherent to the methylation chips (covariates Array and Slide), and other biological batches (covariates Age, Sex, BMI, and the blood cell's composition) in our DNAm datasets. These confounding effects must then be addressed during the differential analysis in order to avoid estimating spurious association or missing meaningful ones. ComBat (Johnson et al., 2006), originally implemented in the sva package, was developed to remove such batch effects from microarry experiments by making use of an empirical Bayesian approach to adjust data and avoid over-correction for small cohort sizes. ComBat progressively became one of the most frequently used tool for adjusting data, and more particularly for DNAm EpicArray data. This gave birth to a popular framework for DNAm differential analysis, Champ, Tian et al. (2017) where a version of ComBat (Leek et al., 2012) is implemented.

One of the particularity of ComBat is that it allows to preserve the effect of a specific variable. In practice, it consists in estimating this effect, removing it, correcting for the different batch effets and finally adding back by the preserved effect. In the sva package, this is presented as an option, however, when used in the Champ framework, the variable of interest (i.e., cases vs. controls) is automatically passed to ComBat for preservation.

Previous studies showed that when used with the preservation option, in a differential analysis context, ComBat results in an inflation of the number of type I errors with abnormally smaller p-values, which is particularly true for analyses with smaller cohort sizes and unbalanced groups (i.e. number of controls is much higher than the number of cases) (Nygaard et al., 2015; Price & Robinson, 2018; Zindler et al., 2020).

In order to test this reported impact of the preservation option on the results of the differential analysis of the DNA methylation, we conducted an experiment on the female co-

hort composed of $N = 51$ MDD patients and $N = 39$ matching controls. Three adjustment scenarios were compared: i) sva's implementation of ComBat without the preservation option, ii) sva's implementation of ComBat with the preservation option of the variable of interest (MDD vs. Control) and iii) Champ's implementation of ComBat which automatically preserves the variable of interest.



Supplementary 4.1: The impact of preserving group differences in ComBat batch effect correction: A. Density distribution of the p-values from the differential analysis with and without the preservation option. **B.** Scatter plot of the 850K t-values from the differential analyzes with and without the preservation option.

The resulting distributions of the p-values corresponding to the three scenarios are presented in figure 4.1A, where we can see that the peak associated with p-values lower than 0.05 is particularly higher for the two implementations that preserve the variable of interest, whereas this peak is much more attenuated without the preservation. This experiment, undertaken on our cohort, confirms the observations made originally by Nygaard et al. (2015) and later on by Price and Robinson (2018) and Zindler et al. (2020).

Nonetheless, the ranking of the CpG sites in regard to their level of association with the status case/control may be preserved across the different adjustments. In order to investigate this question, we retrieved the t-values (statistic evaluated during a Student's Test) computed for each CpG site and each adjustment case. In Fig4.1B the t-values for the non-preserved case are plotted against the preserved one. First, we realized that the impact of the preservation on the extreme t-values is of almost two folds in terms of magnitude. Then, the shape of the cloud of points suggest a strong correlation between the two cases (Spear-

man correlation coefficient ρ = 0.96, p-value $= 2.2 \times 10^{-16}$) confirming that the relative level of association with the diagnosis is not altered.

In light of these results, we cannot directly conclude if there are indeed inflation of the type I errors, as we do not know the true genes associated with the status. Further simulation or permutation tests would be required to confirm this inflation. However, these observations still strongly advocates for avoiding the preservation option on the case/control status as designed in Champ. The p-values of the differential analyzes resulting from non-preservation, although less significant, remain more reliable.

4.7.2 Label Propagation

Let us introduce n_t and $n - n_t$, the number of samples in the train and test set respectively, and $\mathbf{Y}^0_{n_t} = \left(y^{0,n_t}_{ik}\right) \in \mathbb{R}^{n_t \times K}$, the disjunctive table associated with the assigned labels in the train set, in other words $y^{0,n_t}_{ik} = 1$ if sample i is in group k and $y^{0,n_t}_{ik} = 0$ otherwise. Label propagation provides a way to estimate $\hat{\mathbf{Y}}_{n-n_t} \in \mathbb{R}^{(n-n_t)\times K}$ that characterize the probability for each test sample to belong to each group. This estimation is performed through algorithm 2.

Algorithm 2: Label Propagation principle

Data: $\mathbf{Y}^0_{n_t}$, $\mathbf{W} = (w_{ij}) \in \mathbb{R}^{n\times n}$, T_{max}
Result: $\hat{\mathbf{Y}}^{T_{max}}_{n-n_t} \in \mathbb{R}^{(n-n_t)\times K}$
Init: $\mathbf{P} = (p_{ij}) \in \mathbb{R}^{n\times n}$ such that $\forall (i,j), p_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$;
$\mathbf{Y}^0_{n-n_t} \leftarrow \mathbb{0}_{(n-n_t)\times K}$;
$t \leftarrow 0$;
while $t < T_{max}$ **do**

$$\begin{bmatrix} \mathbf{Y}^{t+1}_{n_t} \\ \mathbf{Y}^{t+1}_{n-n_t} \end{bmatrix} = \mathbf{P} \begin{bmatrix} \mathbf{Y}^{t}_{n_t} \\ \mathbf{Y}^{t}_{n-n_t} \end{bmatrix}; \tag{4.1}$$

$\mathbf{Y}^{t+1}_{n_t} \leftarrow \mathbf{Y}^0_{n_t}$;
$t \leftarrow t + 1$;

end
$\hat{\mathbf{Y}}_{n-n_t} \leftarrow \mathbf{Y}^{T_{max}}_{n-n_t}$

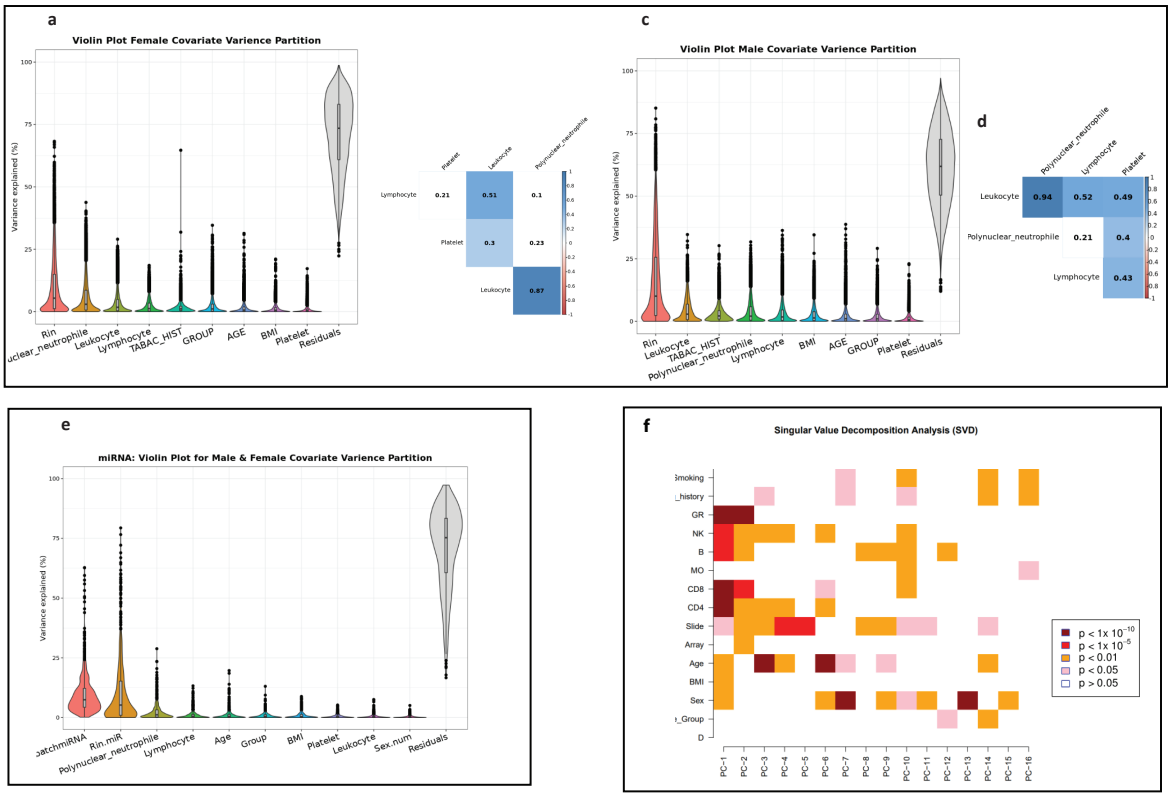
Where $\mathbf{W}$ denotes the pairwise similarity matrix between all samples (both from the train and test sets) computed by SNF, T_{max} is the maximum number of iterations and $\mathbb{0}_{(n-n_t)\times K}$ is a matrix of dimension $(n - n_t) \times K$ only made of 0. From equation (4.1), $y^{t+1,n-n_t}_{ik}$, the estimation at step $t + 1$ of the probability of test sample i to belong to group

k, can be re-written as:

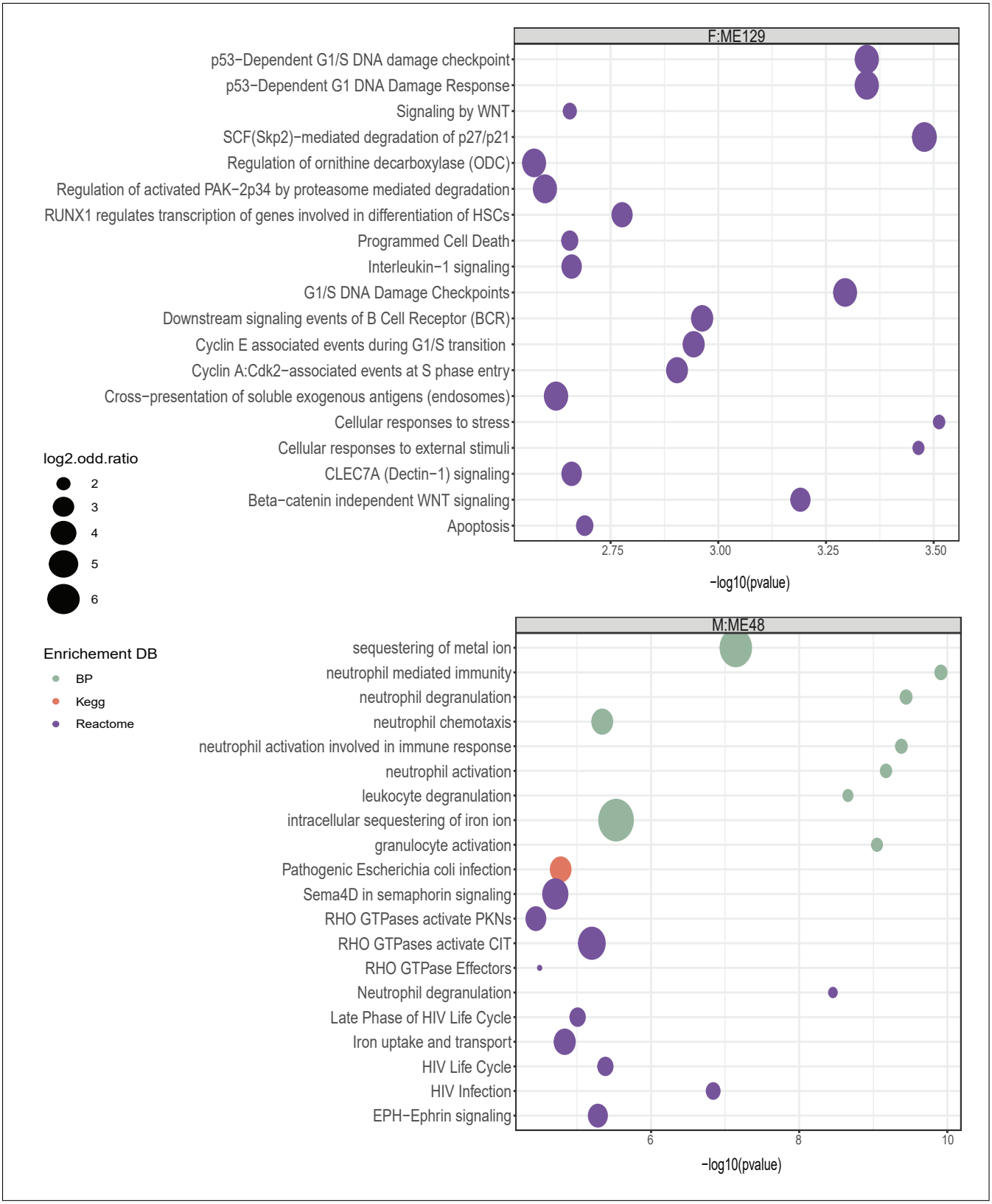
$$y_{ik}^{t+1,n-n_t} = \frac{\sum_{l=1}^{n_t} w_{il} y_{lk}^{0,n_t} + \sum_{l=n-n_t+1}^{n} w_{il} y_{lk}^{t,n-n_t}}{\sum_{j=1}^{n} w_{ij}} \tag{4.2}$$

It can be interpreted as a weighted mean of the probability of each sample to belong to the group k (if samples comes from the train set, this probability is either 0 or 1), where the weights depend on the similarity between sample i and all the others. In (4.2), the only element that depends on step t is $y_{lk}^{t,n-n_t}$ and at $t = 0$, they are all equals to zero. Thus, at the first iteration, what only matters is the distance to the train samples. Then, progressively, as the probability of each test sample increases in a specific group and decreases in all the others, this will modify the test samples' neighborhood, hence the notion of propagation.

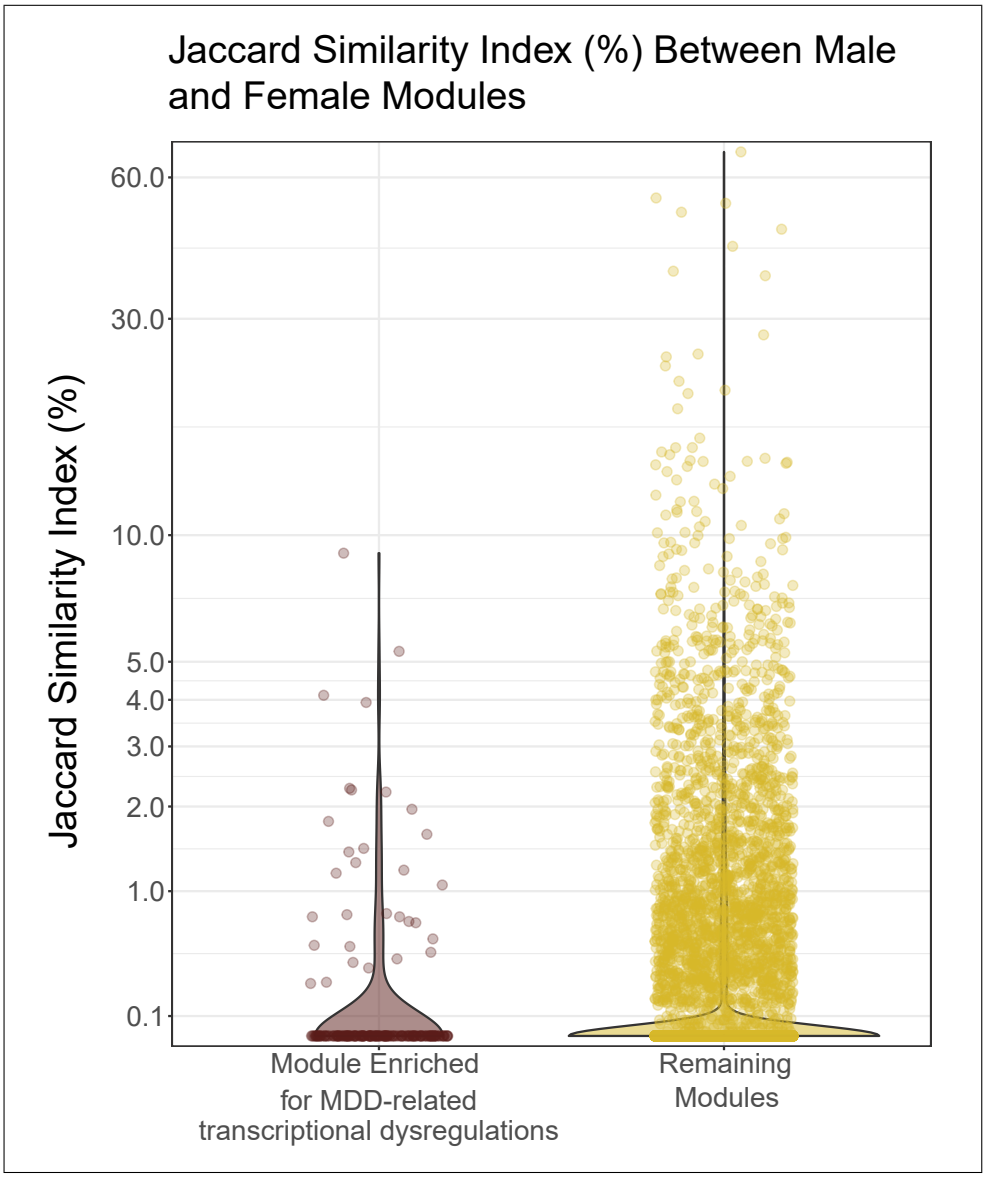
4.8 Supplementary Figures and Tables



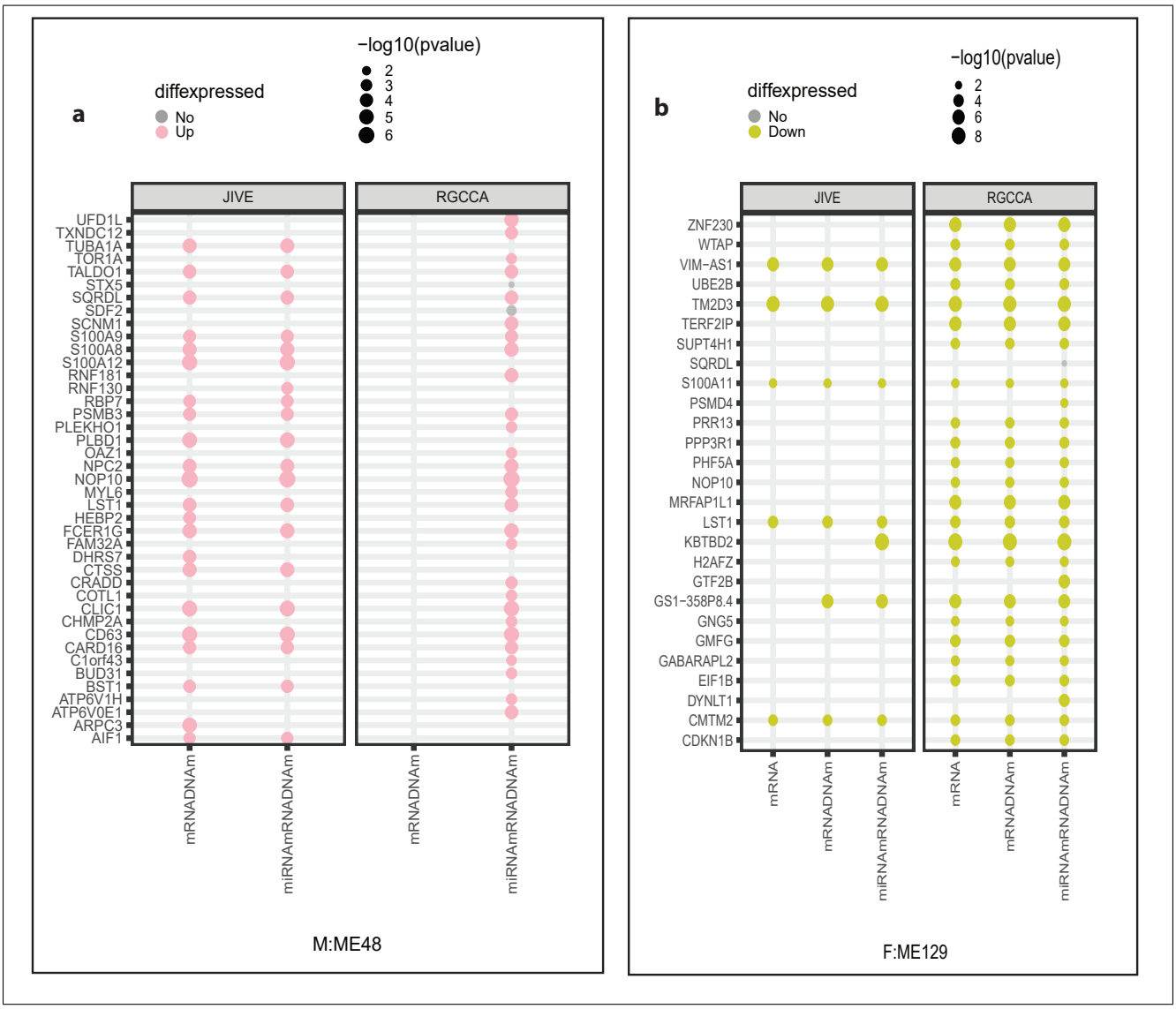
Supplementary 4.2: Variance partition analyses results: The violin plots **a,c** indicate the explained variances in overall gene expression in females and males, respectively. The correlograms **b,d** show the correlation between the blood cell composition measurements in females and males, respectively. The violin plots **e** indicate the explained variances in overall microRNA expression in the pooled cohort. The heatmap **e** indicates the explained variances in the overall DNAm methylation of the pooled cohort following a singular value decomposition analysis.



Supplementary 4.3: Dot plot visualization: GO, Kegg, and Reactome pathways enrichement analysis results for the female F:ME129 and male M:ME48 modules. The size of the dots indicates the $log2(oddratios)$ and the y-axis represent the $-log10(p-value)$.



Supplementary 4.4: Violin Plots: Distribution of the pairwise Jaccard similarity indices for each male and female module combination. MDD-related modules are depicted in brown.



Supplementary 4.5: Dot plot visualization: Crossing overlaps between the stable mRNA features of the predictive model and the DEGs found in the modules M:ME48 and F:ME129.

a Male Dataset

Omics	miRNA mRNA & DNAm			mRNA & DNAm	
jDR Methods	miRNA	mRNA	DNAm	mRNA	DNAm
JIVE	1	198	0	137	0
RGCCA	0	438	229	0	0

b Female Dataset

Omics	miRNA mRNA & DNAm			mRNA & DNAm		mRNA
jDR Methods	miRNA	mRNA	DNAm	mRNA	DNAm	mRNA
JIVE	1	191	124	206	85	195
RGCCA	28	198	132	301	493	347

c pooled Dataset

Omics	miRNA mRNA & DNAm			mRNA & DNAm	
jDR Methods	miRNA	mRNA	DNAm	mRNA	DNAm
JIVE	1	124	0	114	0
RGCCA	0	0	0	0	0

Supplementary 4.6: Count tables: Number of significantly (adjusted p-value < 0.05 and ranking $> 80\%$) stable features in the male, female and pooled cohorts.

GWAS Phenotype	First Author	N	Article
MDD (excluding 23andMe data)	Wray et al. (2018)	135,458 MDD and 344,901 controls	Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression
BD I and **BD2**	Stahl et al. (2019)	BD1 (n=14,879 cases), BD2 (n=3,421 cases) 31,358 controls	Genome-wide association study identifies 30 loci associated with bipolar disorder
PTSD: Childhood Trauma	Dalvie et al. (2020)	124,711 samples with available childhood maltreatment data	Genomic influences on self-reported childhood maltreatment
SCZ (Eur)	Lam et al. (2019)	EUR samples (n = 33,640 cases; 43,456 controls)	Comparative genetic architectures of schizophrenia in East Asian and European populations
BMI	Shungin et al. (2015)	224,459 samples	New genetic loci link adipose and insulin biology to body fat distribution

Supplementary 4.7: Size of the Genome Wide Association Studies used for module enrichment

Chapter 5

Chapter 5: A multi-Omics based childhood trauma signature in bipolar disorder

Bipolar disorder (BD) is a mental illness characterized by extreme changes in mood. These changes can include periods of mania, when a person may feel excessively euphoric or irritable, and periods of depression, when a person may feel sad or hopeless. BD can also affect a person's energy, sleep patterns, and concentration ability. Furthermore, BD imposes a heavy burden on both patients and caregivers (Kupfer, 2005; Perlick et al., 2007). Indeed, BD is a critical mental illness and a leading cause of disability worldwide that can significantly impair a person's ability to function in everyday life. In addition, it is associated with more frequent obesity, cardiovascular diseases, and a higher risk of committing suicide (Fiedorowicz et al., 2008; Pompili et al., 2013)

Several studies have shown that genetic and environmental factors influence the onset of BD. Moreover, a correlation between early life traumatic events (childhood trauma) and the onset of BD has been established consistently (Aas et al., 2016; Agnew-Blais & Danese, 2016; Palmier-Claus et al., 2016). Other studies showed that early-life adversity, such as physical or sexual abuse, is more frequent in patients with BD (Etain et al., 2013; Garno et al., 2005). In addition, BD patients who have suffered from Childhood Trauma (CT) are more prone to have severe symptoms, to be bad responders to lithium-based treatment, and at higher risk of suicide attempts (Etain et al., 2017). CT may also increase the likelihood of developing other mood-related conditions, such as anxiety and depression (Heim & Nemeroff, 2001).

The mechanisms underlying childhood trauma and its association with the onset and

development of BD are unclear (De Bellis & Zisk, 2014). However, a combination of transcriptomic and epigenetic alterations following trauma exposure is suspected to be at the core of these processes (Parade et al., 2021; Thumfart et al., 2022). Indeed, a growing body of evidence indicates that trauma is associated with changes in methylation patterns and gene expression levels in individuals with BD (Aas et al., 2016; Comes et al., 2020). However, no single, specific biological mechanism has been identified as underlying the association between childhood trauma and the development of BD. Instead, multiple mechanisms likely contribute to this association.

Hence, based on these considerations, in this chapter, we will report early findings of single and stepwise integration studies of gene expression and epigenetic regulators variations associated with childhood trauma in bipolar disorder patients. We characterized both transcriptomics (gene expression), and epigenetic regulation (DNA methylation and microRNA) from whole blood samples of BD patients (N=156), using the same pipeline as in Chapter 5, for MDD.

To associate these peripheral omic data with CTs, we first extracted single omic features (mRNA, microRNA, and methylation probes) that significantly correlated with Childhood Trauma Questionnaire (CTQ) scores. In parallel, at a systems biology level, we constructed networks of co-expressed genes and identified those associated with CTQ scores and enriched with single-omics features previously identified as correlated with CT. Additionally, these gene modules were tested for enrichment in SNPs related to psychiatric disorders (GWAS data). Finally, based on all of these findings, we prioritized a module that might potentially serve as a molecular signature for childhood trauma in BD patients.

5.1 Cohort Description

The cohort consists of n=158 BD patients who have been assessed for CT as part of an extensive clinical characterization (ClinicalTrials.org: NCT02572674). Individuals were recruited from 8 academic psychiatric departments in France. Patient inclusion criteria were as follows: aged between 18 years and 55 years, diagnosed with BD (either type I or II, according to DSM-IV criteria), and clinically euthymic at the time of inclusion (Young Mania Rating Scale (YMRS) < 12 and Montgomery Asberg Depression Rating Scale (MDRS) < 10).

Sex	BD Type	n.BD.Type	mean.Age	mean.BMI	mean.CTQ.Total	mean.RIN	n (mRNA)	n (miRNA)	n (DNAm)
Female (n=94)	BD I	57	40	25.8	38.1	7.38	57(100%)	55 (96.5%)	48 (84.2%)
	BD II	37	40.7	25.2	44.83	7.48	37 (100%)	37 (100%)	31 (83.7%)
Male (n=64)	BD I	42	37	26.8	36.7	7.49	42 (100%)	42 (100%)	40 (95.2%)
	BD II	22	38.8	25.4	36.04	7.67	22(100%)	22(100%)	18(81.8%)
Total (n=158)		158	**39.1**	**24.25**	**39.12**	**7.48**			
									n all omics = 137 (86.7%)

Table 4: Descriptive statistics of the Bipolar Disorder cohort.

5.1.1 Human Samples

5.1.1.1 RNA extraction and library preparation

Total RNA (including miRNA) was extracted with the PAXgene Blood miRNA Kit (Qiagen) from RNA blood collection tubes. Following the manufacturer's recommendations, total RNA was globin- and rRNA-depleted. The quantity and quality of RNA were assessed using a NanoDrop-1000 (Thermo Fischer Scientific) and a 2100 Bioanalyzer (Agilent). mRNA Libraries were prepared using the KAPA mRNA Hyper Prep Kit (Roche, ref. KK8581/KK8581, Basel, Switzerland), separately for each sample, and then amplified by PCR, quantified by microcapillary electrophoresis, and then pooled at equimolar concentrations for sequencing using Illumina chemistry. Sequencing was performed on a NovaSeq6000, SP flowcell, at the ICM (Institut du Cerveau et de la Moelle Epiniere), with 100 base pair paired-end sequencing, at a mean depth of 29.3 $\pm$ 0.49 million reads. Regarding miRNA, libraries were normalized, pooled, denatured, and sequenced on the Illumina NextSeq 500 to generate 8.1 $\pm$ 0.12 million reads on average.

5.1.1.2 DNAm extraction

According to Illumina's protocol, DNA was extracted from each sample and deaminated using the EZ-96 DNA Methylation Kit (Zymo Research). Bisulfite conversion was controlled by qPCR. Genome-wide DNA methylation was assessed by an investigator blind to group allocation using Infinium MethylationEPIC BeadChip microarrays (Illumina, San Diego, CA), detecting methylation status at over 850,000 loci in the human genome.

5.1.2 Data Analysis

5.1.2.1 Raw Data Processing

Sequencing data were processed using the same pipelines as for our MDD cohort (detailed in section 4.2.6).

5.1.2.2 Adjustment for confounding factors

We used the R package variancePartition (Hoffman & Schadt, 2016) to determine the impact of clinical and technical factors on data variability. The variancePartition function uses a linear mixed model technique to characterize the contribution of specified factors to transcriptional variability. The following models were used:

- **DNAm** ~ Age + Slide + Sex + Array + BMI + Blood Cells Composition (Houseman's method Houseman et al. (2012))
- **miRNAseq** ~ Batch + Age + BMI + Sex + RNA integrity number (RIN)
- **RNAseq** ~ Age + BMI + Sex + RIN

Of note, because blood cell counts were unavailable for the BD cohort, we performed a deconvolution to impute cellular composition using the Cibersort method (Chen et al., 2018). However, the final models did not consider these results due to significant correlations with CTQ scores. Both mRNA and miRNA data were adjusted for their covariable using a linear model after variance normalization. Regarding DNAm data, the adjustment was performed using Combat. The residuals of these models were then used for the following analyses.

5.1.3 The Childhood Trauma Questionnaire Scores

Childhood trauma was assessed using the 28-item short version of the Bernstein et al. (2003) Childhood Trauma Questionnaire (CTQ) (see Annexe 6.3). The CTQ is one of the most widely used tools to assess childhood trauma in adulthood retrospectively. It uses a Likert scale, with 5 response options, ranging from 1 *"Totally disagree"* to 5 *"Totally agree"*. Five types of trauma are assessed by the CTQ: physical neglect, emotional neglect, physical abuse, emotional abuse and sexual abuse). Each type of trauma is interrogated by 5 items, the remaining 3 questions being used to evaluate denial. The score can thus vary from 5 to 25 for each type of trauma. The total CTQ score, defined by the sum of the 5 types of abuse and neglect, varies from 25 to 125. Additionally, sub-scores for neglect or abuse can be calculated by summing their respective items. The CTQ has been extensively validated across a variety of clinical settings as reflecting the traumatic burden of childhood. A high score on this test indicates a high level of childhood trauma. The distribution of the CTQ score in our BD cohort is depicted in Fig5.2.

5.1.3.1 Study Design

To extract molecular signatures associated with childhood trauma in BD, we first took advantage of the transcriptomic (RNA-seq) and epigenetic (miRNA and DNAm) datasets at the single-omic level. Using correlation analyses, we extracted CTQ-associated omics features. Then at the systems biology level, we extracted from the mRNA data co-expressed gene modules that were simultaneously associated with the i) CT phenotype, ii) previously identified CT-correlated mRNA miRNA and DNAm CpG and external SNPs data. A general overview of the study is presented in Fig5.1

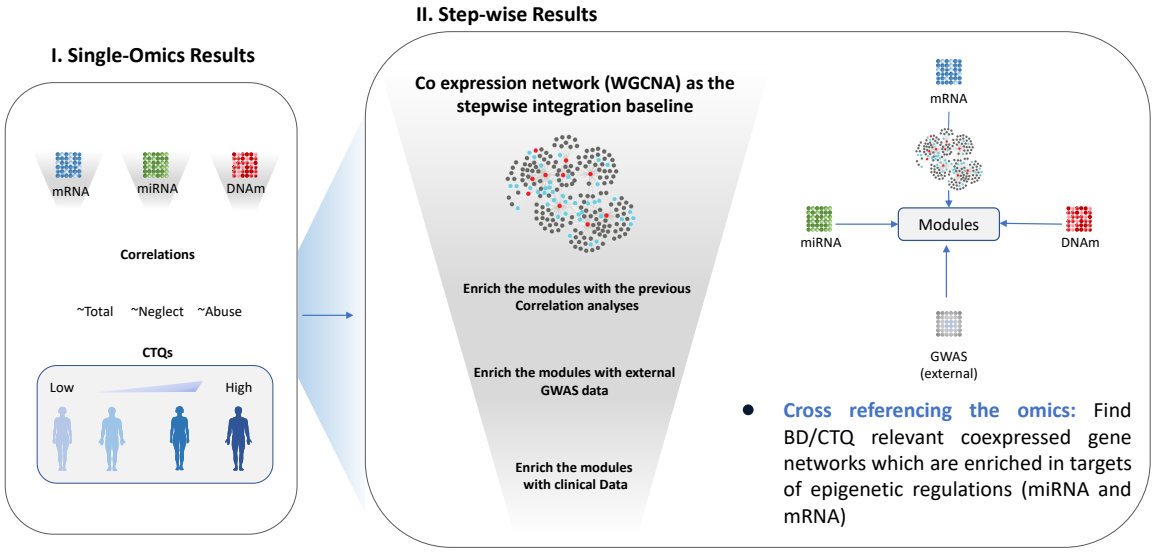


Figure 5.1: Graphical summary: single and step-wise omic association with childhood trauma questionnaire.

5.2 Preliminary Results

5.2.1 The distribution of the CTQ scores in the BD cohort

The BD cohort showed a significant difference in the distribution of CTQ sub-scores, as depicted in Fig5.2A. Indeed, patients scored higher for emotional subtypes than for abuse ones. Furthermore, the distribution of sexual abuse scores was bimodal, with most patients scoring very low on this trauma and a small number showing very high scores, as expected. On the other hand, looking at the abuse and neglect sub-scores, or the total CTQ score in Fig5.2B, distributions followed a normal curve, with a small number of patients exhibiting extremely high scores (5 patients with CTQ > 70). These scores were used in our down-stream analyses, representing the phenotype of interest.

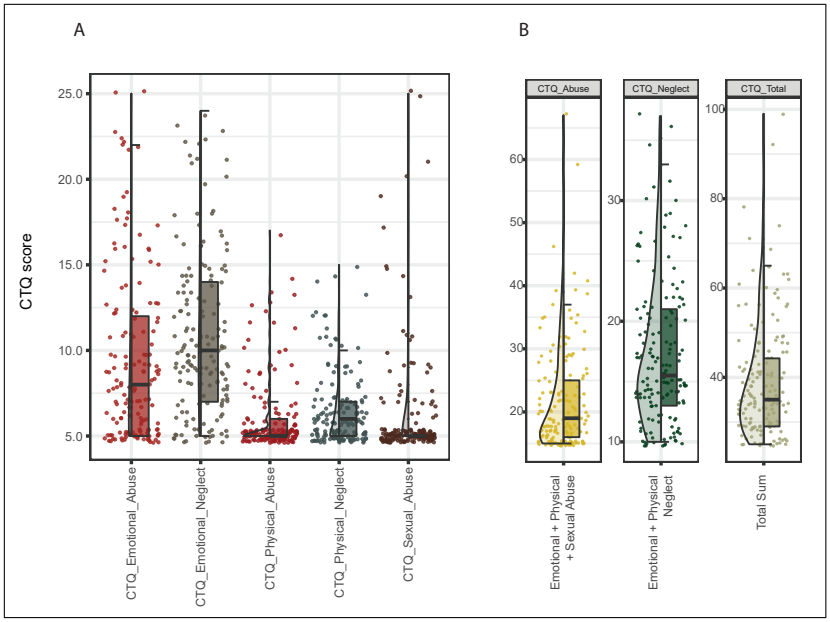


Figure 5.2: Box plots of the Childhood Trauma Questionnaire Scores A. The distribution of the CTQ score according to its 5 items. **B.** The distribution of the cumulative scores of the abuse, neglect and total items

5.2.2 Correlations of mRNA, miRNA and DNAm with CTQ scores

Associations between omic features (gene and microRNA expression, DNA methylation probes) and CTQ scores were assessed using Pearson's correlation. Significance thresholds of correlation coefficients $|\rho|>0.2$, and p-value < 0.01 were applied.

Transcriptome-CT-phenotype correlations : Genes that met the significance threshold, described above, were then tested for enrichment in Gene Ontology, KEGG, and Reactome pathways using Webgestalt's function for over-representation analyses with all the input mRNAs as a background (Liao et al., 2019).

CTQ	mean (ρ)	n: total	Standard Error Mean
CTQ Total	0.2333	1155	9e-04
CTQ Neglect	0.2517	3041	7e-04
CTQ Abuse	0.2297	467	1e-03
CTQ Physical Neglect	0.2392	1920	7e-04
CTQ Emotional Neglect	0.2474	2650	7e-04
CTQ Physical Abuse	0.2414	1757	8e-04
CTQ Emotional Abuse	0.2308	430	1e-03
CTQ Sexual Abuse	0.2307	895	9e-04

Table 5: Descriptive statistics of the top correlation coefficients,($|\rho| > 0.2$, p-value < 0.01), with each CTQ sub-score. Mean of the coefficients, Number of genes, and Their Standard Error Mean .

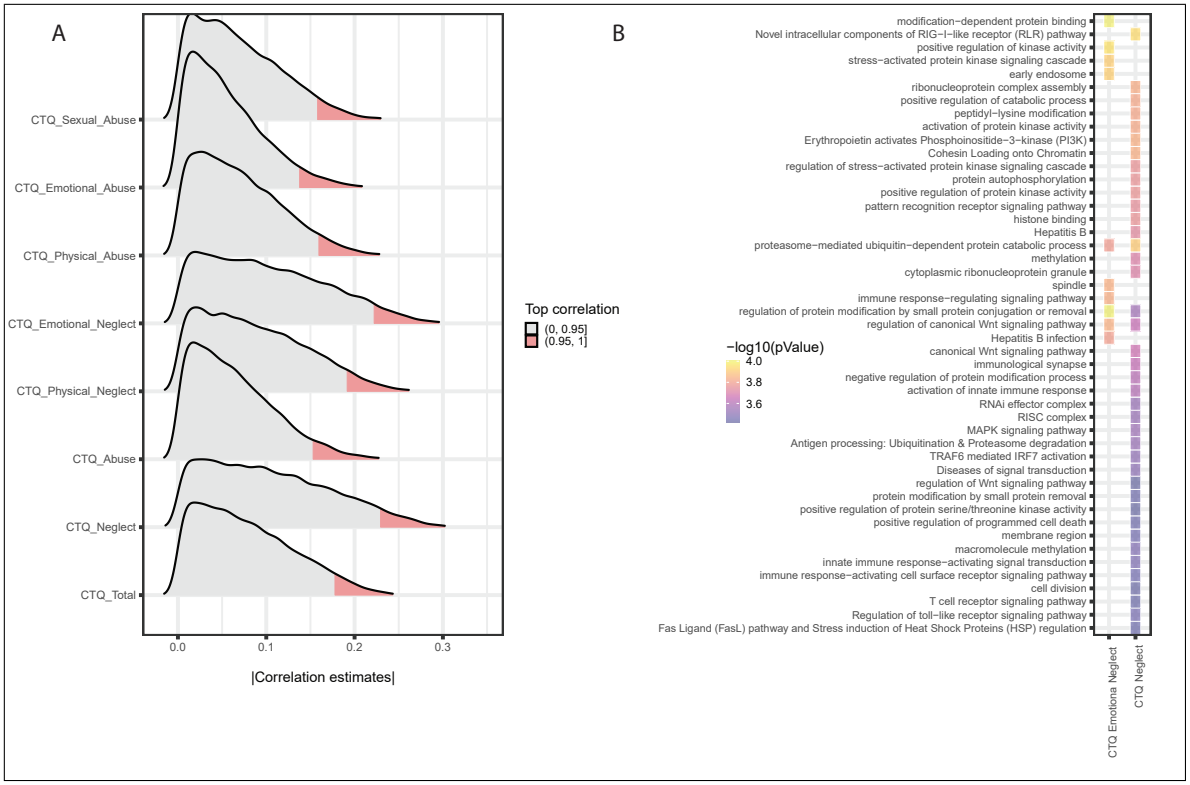


Figure 5.3: Linear associations between CTQ scores and gene expression in the BD cohort. A. Density distribution of the absolute values of the Pearson's correlation coefficients (ρ) between CTQ scores and adjusted mRNA expression. Top correlations (|coefficient| > 0.2) are depicted in red. **B.** Top results for Gene Set Enrichment Analyses of most significant correlations (|coefficient| > 0.2, p-value < 0.01)

Results of these linear associations are presented in Fig5.3. Comparisons of density distributions of Pearson's coefficients ρ across CTQ sub-scores showed higher correlation for items related to neglect (physical, emotional and their combination) than those related to abuse. Indeed, as shown in Table5, correlations between CTQ and neglect exhibited both the highest numbers of significantly correlated mRNAs that met the threshold criteria (n = 3041), and the highest mean ρ coefficient value $0.25 \pm 7.10\text{-}4$. Moreover, enrichment analyses for highly correlated mRNAs shown in Fig5.3B identified terms related to the innate immune response system, regulation of toll-like receptors (TLR), and *MAPK* signaling pathway. These preliminary results concur with the hypothesis of an immune dysregulation by CT in BD. Indeed, Oliveira et al. (2015), Oliveira et al. (2014) previously showed that the *TLR2* rs3804099 risk genotype acts as an amplifier of the CT negative effect in BD patients. Furthermore,do Prado et al. (2017) and Schiweck et al. (2020a) respectively hypothesized the implication of *MAPK* genes in CT and suicidal ideation in MDD patients who suffered from child abuse.

Epigenome-CT-phenotype correlations: After controlling for confounding variables, linear associations between epigenomic mechanisms (microRNA and DNA methylation) were performed following the mRNAs' procedure. Best correlations' are presented in Fig5.4.

Regarding microRNA correlations with CTQ scores, the miR-142-3p exhibited significant association with all sub-scores, except physical abuse. Interestingly, this high association of miR-142 expression with CTQ scores in whole blood samples from BD patients is consistent with results from previous studies. Indeed, both Ceylan et al. (2020) and Lee et al. (2020) reported associations between this miR's dysregulation and BD type II or type I, respectively. Furthermore, Nie et al. (2021) showed that miR-142 silencing decreased PTSD-like symptoms in rats exposed to a single prolonged stress. Such miR-142 silencing is also associated with reduced amounts of proinflammatory cytokines, including *IL-1* and *IL-6*. These results, therefore, support a role for miR-142 in shaping the risk for BD and its potential association with stress and traumatic life events (Clausen et al., 2022).

Regarding DNA methylation, several studies provide evidence that childhood trauma may associate with detectable changes in both brain (Labonté et al., 2017b; Lutz et al., 2021; Lutz et al., 2017) and blood (Bustamante et al., 2018; Frach et al., 2020) tissues in patients with MD. Yet, research on this association has been carried out mainly using peripheral tissues such as saliva and blood. For example, Mehta et al. (2013) showed a two-fold higher association between altered DNA methylation patterns and gene expression at the same loci in the blood of patients with post-traumatic stress disorder who had undergone childhood trauma, compared to patients with PTSD who did not. Notably, these epigenome-wide studies have generally failed to identify classical candidate genes, such as *FKBP5, BDNF*, and *NR3C1* (Thumfart et al., 2022). Examining the top CpG probes most strongly correlated to CTQ in our dataset revealed a scarce overlap with the literature, requiring further exploration of these data.

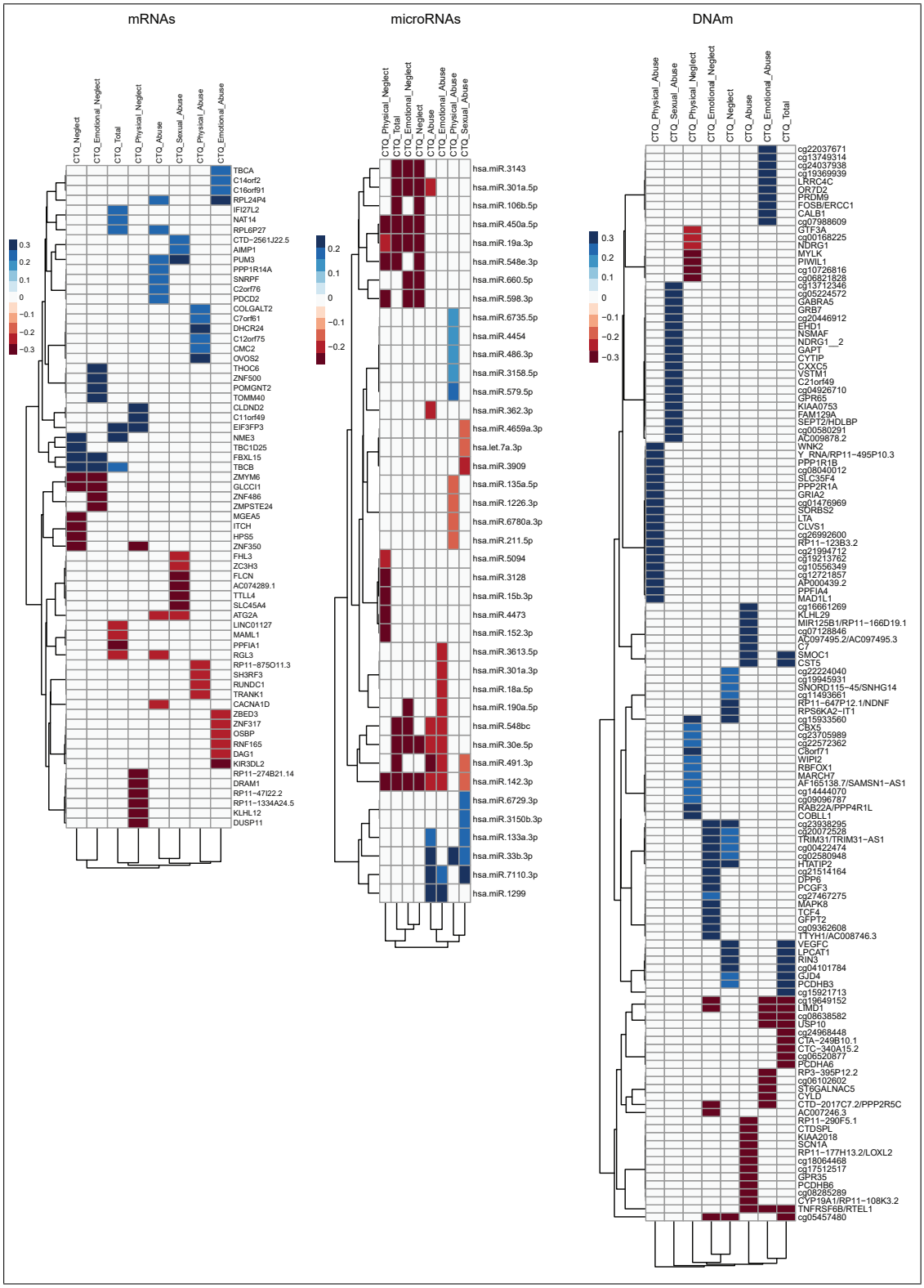


Figure 5.4: Heatmaps with hierarchical clustering dendrogram: most significant Pearson's correlations (Top 2% of absolute ρ values) between omic features and CTQ scores in the BD cohort.

5.2.3 Step-Wise Integration

In order to identify gene modules potentially associated with childhood trauma, we then performed a co-expression network analysis using WGCNA (Horvath, 2013). Similar to the procedure detailed in section 4.2.9, WGCNA parameters were optimized in order to maximize overlaps between gene modules and pathways from the Reactome database: power threshold of 11, a deep split of 4, minimum module size of 15, and cut height of 0.14 (see section 3.2.1.2 for further details on the role of these parameters). The resulting gene network was then analyzed for enrichment with features most significantly correlated with CTQ scores, as of previous single-omic analyses. In addition, module eigengene values were also correlated with the 8 CTQ sub-scores. Finally, using MAGMA (de Leeuw et al., 2015), gene modules were tested for significant enrichment with SNPs previously implicated in the risk for psychiatric disorders (see Supplementary Table 4.7 for details).

Results from these analyses are displayed in Fig5.5. A total of 65 modules were identified and tested for preservation in external datasets from (Krebs et al., 2020) and (Lonsdale et al., 2013). Analysis of these modules identified MEred as the most promising molecular signature of childhood trauma, based on the criterion that this is the unique module that exhibited at least one significant result (p-value < 0.05) for each of the 5 types of enrichment that were computed Fig5.5. In addition, it contained genes that were highly correlated with CTQ items related to neglect. Moreover, the eigengene of the MEred module significantly associated with neglect-related CTQ sub-scores, and was enriched for SNPs associated with the risk of developing a major depressive disorder. Further Gene Ontology analysis of the module found enrichment for terms related to toll-like receptor pathways, the innate immune response system, *IL-16, NOTCH4* signal and MAP kinase activation.

The MEred module included 827 genes. To visualize interactions among this network, we next performed a Protein-Protein Interaction (PPI) analysis using Stringdb V10 (Szklarczyk et al., 2020), filtering out low confidence gene co-expression relationships. Non-connected genes were filtered out, retaining only the biologically validated interactions in the final network with at least a medium confidence > 0.4. Results are presented in Fig5.6.

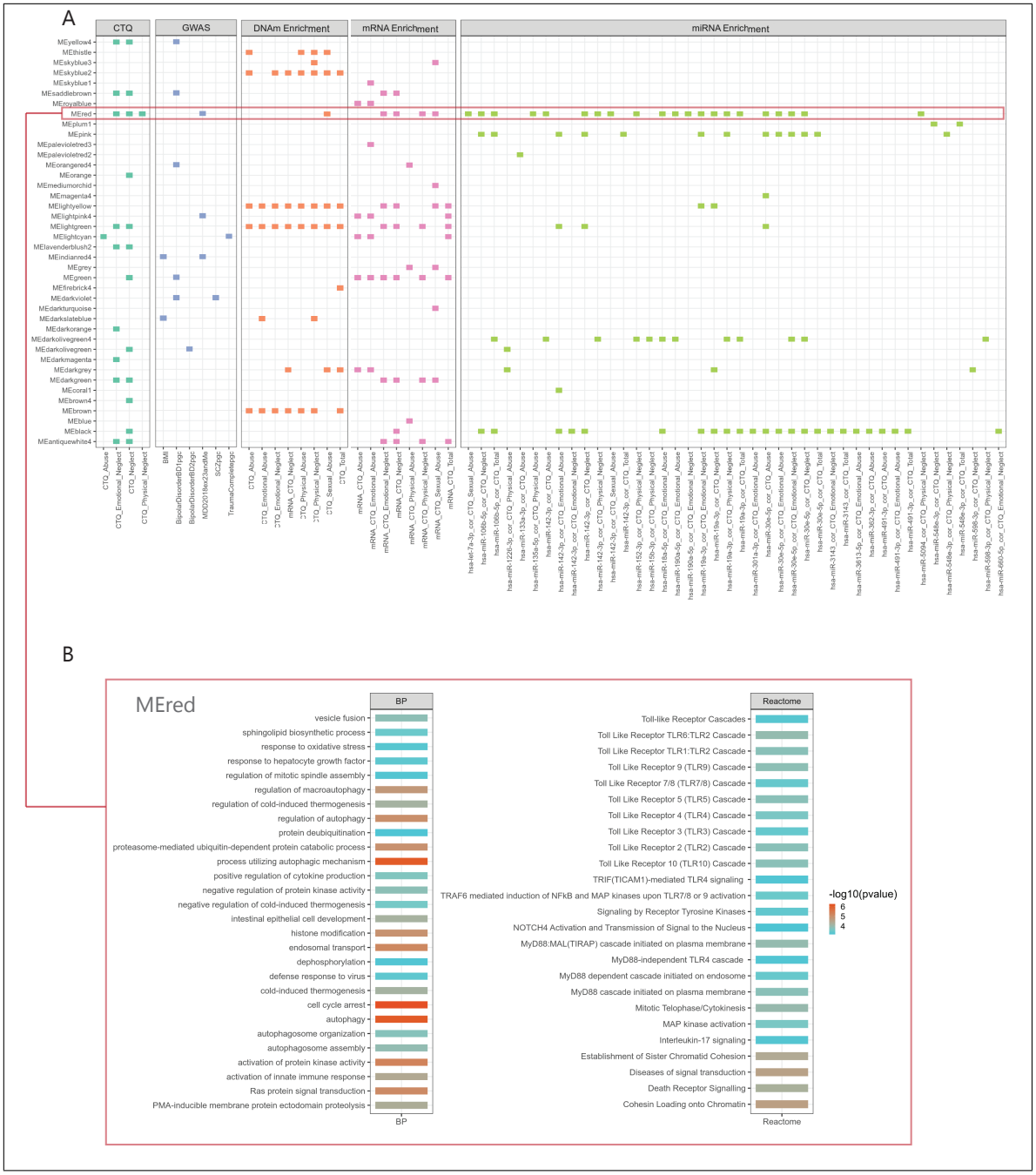


Figure 5.5: Network-based Step-Wise multi-omic integration: A. Tile plot of the enrichment of the 65 BD WGCNA modules with CT scores, genes enriched for GWAS's SNPs associated with psychiatric phenotype, methylated CpG probes, mRNAs and microRNAs correlated with the CT score. Significant enrichments (p-value < 0.05) are represented in colored tiles. **B.** Gene ontology biological processes and Reactome pathways enrichment results for the module of interest MEred.

In addition to the PPI interaction network, Fig5.6 also depicts results from enrichment analyses. Interestingly, miR-142-3p was found as a key regulator of the MEred module, and predicted to regulate several genes significantly correlated with CTQ. Moreover, the

network comprised 12 genes whose expression, and nearby CpG probes, significantly correlated with CTQ scores (*PTPRC, EXOC1, STX12, STAG1, TNFAIP3, ATXN1, RC3H1, NAB1, FLI1, MAP3K2, AFF1, OXSR1*). Further annotation of these 12 genes revealed that 2 of them, *FLI158* and *PTPRC*, were explicitly associated with anxious phenotypes in a study on PTSD in N=355 World Trade Center first-responders (Marchese et al., 2022). In the latter study, *PTPRC* was found differentially expressed (p-value < 0.05) from an analysis comparing gene expression between PTSD cases (defined as meeting DSM-4 criteria for PTSD and a total Clinician-Administered PTSD Scale score or CAPS > 40) and controls (all the remaining samples). Moreover, *TNFAIP3* was identified by Breen et al. (2015) as a master regulator of a PTSD-associated module involved in innate immune response (Torshizi & Wang, 2017), as well as with higher prevalence of a history of CT in patients with MDD (Schiweck et al., 2020b).

These preliminary results advocate further exploration of the MEred module, which could serve as a potential molecular signature for childhood trauma.

5.2.4 sex-specific effect of Childhood trauma

Although sex differences in BD have only recently been considered and a greater prevalence of BD shown in women (Dell'Osso et al., 2021; Gordovez & McMahon, 2020), sex differences in the way trauma is experienced and expressed are well-documented (Garcia et al., 2016; Tiwari & Gonzalez, 2018). Indeed, Pruessner et al. (2019) recently showed that male patients reported higher rates of physical or emotional neglect, whereas female patients indicated significantly higher rates of emotional abuse. However, the sex differences in the effects of CT on adult mood disorders are still debated. Both Gallo et al. (2018) and Garcia et al. (2016) showed higher association levels between CT and depression/anxiety symptoms in women. However, these differences were not statistically significant.

In addition, the sex-specific molecular impact of CT on mood disorders in human studies is scarcely studied. However, the literature on the subject suggests that sex is potentially associated with epigenetic modifications in response to childhood trauma (Thumfart et al., 2022).

Based on these considerations, in our BD cohort, we assessed whether sex is associated with a difference in correlation between the CTQ and genome-wide transcriptomics and epigenetics measurements. To do so, we compared in males and females: i) the mRNA correlations with the CTQ score, and ii) the gene ontology enrichment for the significantly correlated mRNAs (coefficient > 0.2 and p-value < 0.05) and, iii) the total CTQ correlation based geometrical overlaps for the mRNA, miRNA and DNAms CpGs features.

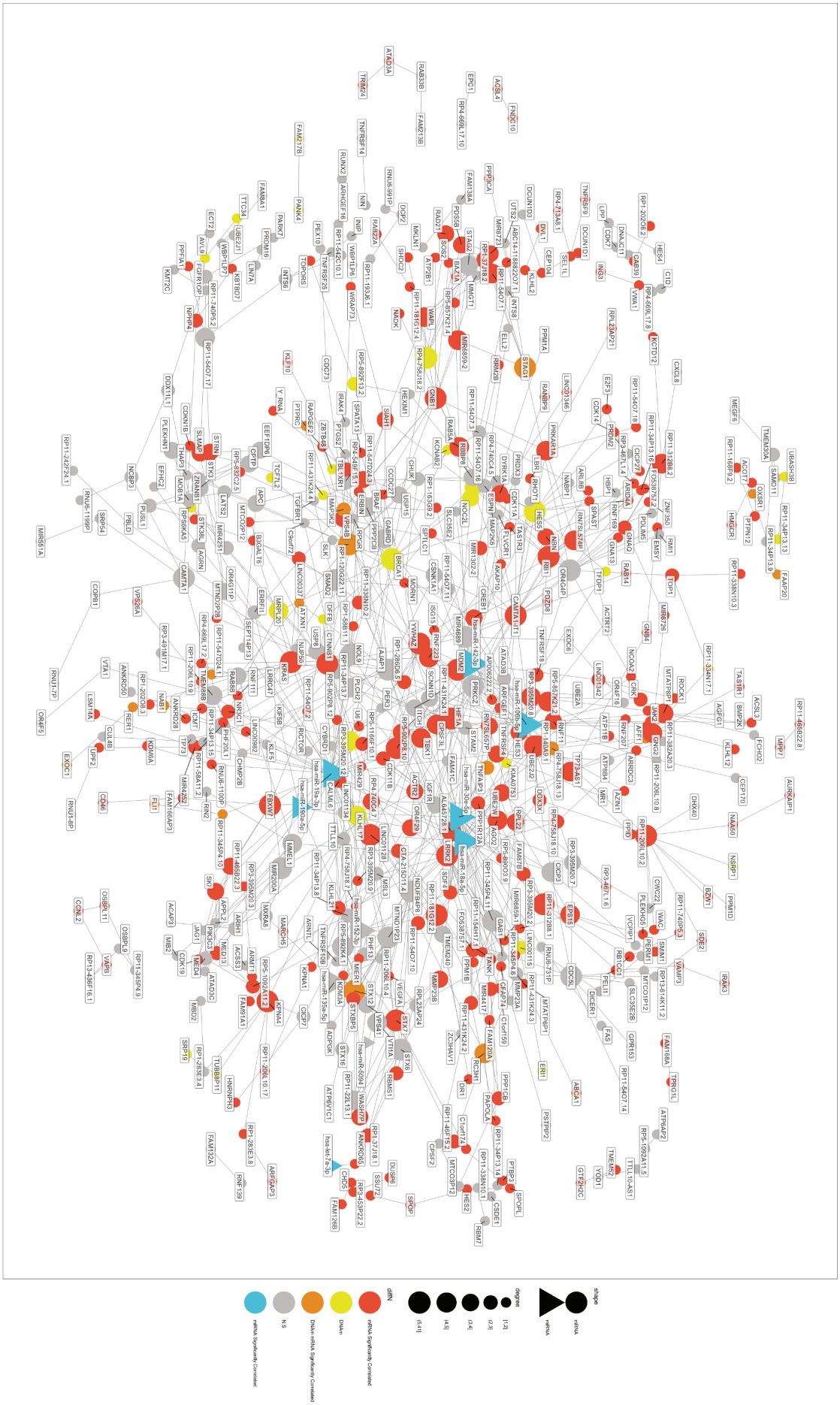

shape
miRNA
mRNA
degree
[1-2]
[2-3]
[3-4]
[4-5]
[5-6]
diffN
mRNA Significantly Correlated
DNAm-mRNA Significantly Correlated
DNAm
N.S
mRNA Significantly Correlated

Figure 5.6: Network visualization of the protein-protein interactions among the genes of the CT associated MEred module: Genes and microRNA are respectfully represented in circular and triangular shapes, genes of which only their: i) expressions are significantly correlated with CTQ are cored in red, ii) associated CpG methylation probe are correlated with the CTQ in yellow. Genes which exhibited significant CTQ correlations of both their expression and associated CpG methylation probe are colored in orange. The size of the module are proportional to their number of connection (degree) in the network.

The results of these comparisons are presented in Fig.5.7. Overall, the linear association of gene expression variability with the different CT intensity scores were similar in amplitude and magnitude across males and females (see Fig5.7A). However, noticeable divergence in the abuse CT score and two of its three components (Physical and Emotional) were noted. Indeed, males reported higher numbers of significant gene correlations (coefficient > 2, p-value < 0.05) with both combined and emotional abuse scores, while females presented more predominant correlations with physical abuse (see Fig5.7B).

Furthermore, both males and females, exhibited common enrichments for pathways involving the immune response system (adaptive immune response, neutrophil migration, *IL-2*, and *NF-Kappa*-pathways) for mRNA correlated with emotional and physical abuse. These observations concur with Breen et al. (2017)'s results on the shared inflammatory pathways across males and females in a PTSD cohort. Yet, males presented an abuse-specific enrichment involving the ribonucleoside monophosphate metabolic process, and total CTQ-specific terms involving *SEMA4D* in semaphorin signaling and mitochondrial large ribosomal subunits (see Fig5.7C). However, these differences could be attributed to the difference in the sizes of the gene lists of interest.

Finally, we investigated total-CTQ-related sex differences in the three omics levels using RRHO2 (see section 4.3.2). The heatmaps presented in Fig5.7D revealed no significant disparities in the geometric overlapping.

Altogether, our results speculate a potential molecular sexual dimorphism of abuse CT in BD. However, these analyses are preliminary and need further investigation for statistically significant differences. Indeed, to expand on these initial findings, we are considering including a bootstrapping procedure in the analysis of linear associations between the phenotypic scores of childhood trauma and transcriptomic and epigenetic variations. This will allow us to approximate the amplitude and variability of the correlation coefficient, and provide a better estimate of their stability. In terms of the step-by-step integration, a more detailed annotation is being explored, and the modules' co-expressions are cross-referenced with the MDD data processed in the preceding chapter 04.

A
CTQ_Total
CTQ_Neglect
CTQ_Abuse
CTQ_Physical_Neglect
CTQ_Emotional_Neglect
CTQ_Physical_Abuse
CTQ_Emotional_Abuse
CTQ_Sexual_Abuse
Male
Female
0.0
0.1
0.2
0.3
0.4
|Correlation estimates|
Top correlation
(0, 0.95]
(0.95, 1]
B
Trauma Sex mean n se
CTQ Abuse Female 0.2297 467 1.00E-03
Male 0.2403 970 1.00E-03
CTQ Physical Abuse Female 0.2414 1757 8.00E-04
Male 0.2361 485 2.00E-03
CTQ Emotional Abuse Female 0.2308 430 1.00E-03
Male 0.2507 1473 1.00E-03
C
U2−type catalytic step 2 spliceosome
Regulation of lipolysis in adipocytes
regulation of peptidyl−lysine acetylation
positive regulation of gene expression, epigenetic
Sema4D in semaphorin signaling
ribonucleoside monophosphate metabolic process
peptidyl−tyrosine phosphorylation
peptidyl−lysine trimethylation
negative regulation of secretion
NF−kappa B signaling pathway
IL−2 Signaling Pathway
organellar large ribosomal subunit
mitochondrial large ribosomal subunit
regulation of protein stability
regulation of nuclear division
positive regulation of stress−activated MAPK cascade
positive regulation of mRNA catabolic process
organelle membrane fusion
T cell receptor signaling pathway
Antigen processing: Ubiquitination & Proteasome degradation
ERK1 and ERK2 cascade
adaptive immune response
protein deacetylation
neutrophil migration
−log10 (pValue)
3.1
3.0
2.9
2.8
2.7
Female
Male
CTQ_Total
CTQ_Physical_Abuse
CTQ_Neglect
CTQ_Emotional_Neglect
CTQ_Emotional_Abuse
D
Female
Male
and
RRHO2
mRNA Female
mRNA Male
CTQ Total
-log10(P.Value)
DNAm Female
DNAm Male
miRNA Female
miRNA Male

Figure 5.7: Linear associations of sex specific molecular variations with the CTQ scores in the bipolar disorder cohort: A. Density distribution of the ρ correlation coeffetients of the mRNAs with the CT score in male and females. **B.** Descriptive table of the differences in the top n mRNA correlations with the abuse CTs in males and females (Trauma: CTQ score item, Sex: Male and Female, mean: mean of the top correlation's coefficients ,n: number of significantly correlated genes, SE: standard mean error) . **C.** Pathway enrichment analysis of the significant mRNA correlation ($\mid \rho \mid > 0.2$, p-value < 0.05 in males and females. **D.** Sex based ranked geometric overlaps between mRNA, miRNA and DNAm correlations with total CT scores

As a result, we expect that by integrating the MDD, Control, and BD samples, and by associating them with CT scores, we shall be able to better define the biological interplay between CT and MD. Finally, an upgraded version of the multi-omic integration pipeline reported in the previous chapter 4.2.10.2 is being developed to incorporate continuous phenotypes, allowing for more accurate extraction of biomarkers and predicting characteristics of childhood trauma in patients with MD.

Part III

Conclusion

Chapter 6

Chapter 6: General Discussion and Conclusion

6.1 General Discussion

Throughout the studies reported in this thesis, we were interested in characterizing the pathophysiology of mood disorders by identifying their blood-derived molecular signatures, and pathways associated with their risk factors such as sex and CT.

First, we conducted a comprehensive review of the available literature to report on state-of-the-art genome-wide research on MDD and BD from both a single omic and a multi-omic perspective. Despite the increased interest in genome-wide genomic, transcriptomic, and epigenetic data in molecular psychiatric research, their use is still in its early stages, and multi-omic integrative approaches remain scarce. This literature review is presented in chapter 02 and was published in Progress in Neuro-Psychopharmacology and Biological Psychiatry (Mokhtari et al., 2022). In chapter 03, we further discussed popular multi-omic integrative tools and their application to multimodal disorders, mainly oncological diseases, which yielded promising results.

In chapter 04, we narrowed the focus of our analyses to samples from MDD and controls to characterize omic-based biomarkers and signatures of MDD. We have set up 3 analysis strategies: i) a first one consists of classical differential analyses; ii) a second strategy consists of a step-wise integrative framework with a baseline of gene co-expression networks enriched with epigenomics data and external GWAS SNPs. Lastly, iii) we developed an original multi-omics framework based on joint dimension reduction methods. We measured and compared the information gained for a dichotomous prediction of MDD and controls using one or more combinations of omics modalities.

All the above analyses were stratified according to sex to investigate the involvement of sex in the molecular signatures associated with MDD. Indeed, our results pointed towards a sex-specific peripheral molecular signature of MDD at the transcriptomic and epigenomic levels. The differentially expressed genes in our data with discordant directionality between men and women converged on the sex-specific molecular signature previously observed in depressive patients' brains (Seney et al., 2018).

Enrichment tests between co-expressed gene modules, MDD dysregulated genes, and association signals from PGC GWASs let us prioritize two modules of interest associated with the depressed phenotype (one in males and the other in females). While they exhibited a limited overlap, both had miR-124-3P as a shared key regulator. The same mir-124 was interestingly found to be highly differentially expressed in males, but failed to reach the p-value adjustment requirement for significance in females. Throughout the differential analyses and step-by-step integration, the differentially expressed genes and the modules of interest showed enrichments in pathways of interest involving immune system response. These results also support the hypothesis of immune cells involvement in the etiology of mood disorders. Regarding the blood DNA methylation, despite few DMPs passing the multiple testing threshold, the role of some CpGs as predictive biomarker of MDD was confirmed stable and generalizable by the bootstrap procedure in our multi-omic integration framework. The low number of significant DMPs in our data is accentuated by the strict use of Combat, without the option of preserving the variable of interest. As previously demonstrated by Zindler et al. (2020) and Nygaard et al. (2015) we observed in our data that preserving the variable of interest, as implemented in the Champ package Tian et al. (2017), results in an inflation of the false positive rate (see annex 4.7.1 for more details). Hence, we carefully processed the differential methylation analyses without this preservation.

The benchmark of joint multi-omics dimensionality reduction approaches with different combinations of omics data in our framework allowed us to compare the accuracy of MDD prediction for obtained selected features with each combination. Our results indicated that the combination of mRNA and DNAm performed, on average, better than single omics alone in the case/control prediction task. However, the mRNA dataset alone performed better for female data, with near-equivalent prediction accuracy. Note that in our framework, both the feature selection and the prediction steps were included in the cross-validation procedures to avoid data over-fitting. Moreover, the stability of the features extracted using the two approaches that outperformed systematically (RGCCA and JIVE) were tested by a bootstrapping approach. Overfitting prediction results may arise if the feature selection step is not included in the CV procedures. For instance, Bhak et al. (2019) proceeded to cross-validations to verify their prediction model performance. However, the model al-

ways relied on the same input features based on the output of differential expression and/or methylation analyses on all samples. Hence, this could partly explain the high accuracy of 87.3% obtained using a random forest classifier distinguishing MDD from controls using blood methylome and transcriptome of only 39 MDD patients and 87 controls.

Even though the AUC was not evaluated yet, the accuracy obtained within our framework (0.75) reached a relatively high level for psychiatric phenotypes. For instance, Clark et al. (2019) showed an AUC of 0.571 using 75k CpG sites and a DNAm risk score to predict depressive status. The authors also used the same number of CpGs in a feature selection procedure (Elastic-Net) to predict depressive status over 6 years, with an AUC of 0.724. However, Barbu et al. (2020), aiming at reproducing Clark's approach, showed an AUC of 0.58 on their 363 cases and 1417 control dataset. Finally, despite having a smaller cohort size, we achieved satisfying prediction rates compared to those observed in the molecular psychiatry literature by combining DNAm and transcriptomics or using mRNA only. One could suggest adding in our framework additional optimization steps for the parameters of the Momix wrapped methods. This could significantly improve the performances and obtained accuracies, but would require substantial computational resources and time.

In chapter 05, we aimed to expand the analytical strategies we developed on the MDD cohort to examine the association between retrospectively self-reported child trauma and genome-wide transcriptomic and epigenomic patterns in peripheral blood of BD patients. At this stage of the project, the two first analysis strategies were adapted to the context of this new cohort of BD patients without healthy controls. The initial project included longitudinal clinical data evaluating patients for two years to record mood recurrences (ClinicalTrials.org: NCT02572674), but the timeline did not let us integrate the clinical follow-up data which have been only recently completed but have not been released by the clinical research unit in charge yet. Hence, in chapter 05, we presented exploratory analyses.

As first strategy, we investigated each omic separatly testing the correlation between multi-omic data and the CTQ score of 158 BD patients. These preliminary results showed interesting enrichments for immune-related GO terms among the top correlating genes ($|\rho| > 0.2$, p-value < 0.01) with CTQ neglect items. The GO terms include, but not exclusively, innate immune response, toll-like and T-cell receptor signaling, and the Wnt pathway. Interestingly, previous studies associated CT severity with activation of immune markers in patients with BD and psychosis (Aas et al., 2016; Baumeister et al., 2015; Corsi-Zuelli et al., 2019), suggesting a potential involvement of CT in the onset and development of psychiatric disorder through its impact on the innate and adaptive immune response systems. Indeed, in their review, Mondelli and Vernon (2019) pointed in experimental studies on rodants an increased activation of T cells following early life stress.

Moreover, in regard to the correlation of epigenetic regulators with CT, our study showed miR-142-3p as significantly associated with all the CTQ items scores (except physical abuse). Interestingly, miR-142 is thought to be implicated in the risk of BD onset (Clausen et al., 2022).

Although our preliminary exploration of these newly generated data showed noteworthy correlations between DNA methylation levels of individual CpG sites and CTQ, the GO term enrichments of their associated genes did not reach significance (FDR < 0.05). Likewise, Løkhammer et al. (2022) showed scarce significant association (FDR < 0.05) between genome-wide DNA methylation and CTQ, in their Epigenome-Wide Association Study (EWAS) of childhood trauma in a cohort of 602 patients with a current severe mental illness. Indeed, only CpG (cg07625619) associated with physical neglect, located in the body of the gene *TANGO6*, was identified as differently methylated and associated with physical neglect. Regrettably, these unsatisfactory results are similar to those of candidate gene approaches (Parade et al., 2021), and could be explained by a high heterogeneity in the CT phenotype. While larger cohort sizes (thousands of samples for an efficient EWAS) investigating the DNA methylation signature of CT in psychiatric diseases could help to circumvent this issue, the integration of other omics data can also leverage smaller datasets. For instance, combining genotype and DNA methylation data of five independent cohorts (a total of 1074 individuals) Czamara et al. (2021) showed that genetics and gene-by-CT interactions account for the majority of variance in blood and saliva DNA methylation compared to CT alone.

Here, we chose other strategies that could take full advantage of the three omic layers available for our data. As second strategy, the step-wise multi-omic integration strategy taking into account gene-co-expression network analysis was applied following the procedure described in section 4.2.9. The first results of this still ongoing work identified blood-derived gene co-expression modules associated with CTQ. Among these, we prioritized the MEred one because of its enrichments for CTQ-correlated genes, miRNAs, and CpG (through their associated genes). This module included immune-related terms in their GO enrichment results (mainly toll-like receptor signaling) and miR-142-3p as a key regulator. All modules were also tested for enrichment for association with MDD, BP, and self-reported CT using publically GWAS data (see Supp Table 4.7). If the influence of genetic variation on CT seems paradoxical, as CT is by definition an environmental exposure, several studies brought evidence of association of genetic variations with CT (Czamara et al., 2021; Pezzoli et al., 2019; Warrier et al., 2021). We used in our step-wise framework the GWAS of Dalvie et al. (2020) (124,000 samples) that reported two genome-wide significant loci associated with CT (rs142346759, p-value $= 4.35 \times 10^{-8}$, *FOXP1*; rs10262462, and p-

value $= 3.24 \times 10^{-8}$, *FOXP2*), and showed a SNP heritability of 6% for CT. More recently, a GWAS meta-analysis of CT in 185 414 individuals identified 13 novel loci and found a causal effect of CT on depression using Mendelian randomization analyses (Warrier et al., 2021). Altogether, these studies emphasize the importance to take into account genotype data to investigate CT despite its low heritability and despite the fact that it is an environmental exposure. Enrichment tests using additional GWAS and GWAS meta-analyses for CT and other psychiatric traits could give new insight to our data and redirect the prioritised CT-associated module. The addition of the genotype omic layer is discussed with our collaborators, as it could leverage the jDR methods that we plan to apply.

Overall, the results of the preliminary study presented above, if validated, are encouraging and advocate using advanced multi-omic methods for patient status prediction and stratification.

6.2 Highlighted contributions:

- A thorough evaluation of the use of omic data and multi-omic techniques in mood disorders provides state of the art.
- The existence of shared and sex-specific molecular signatures of depression in blood epigenetic and transcriptome data.
- An implementation of an original framework that performs a feature selection phase based on dimension reduction methods and a prediction phase based on network similarity and label probation algorithms.

6.3 Perspectives and conclusion:

Psychiatric genomics and multiomics approaches are both extremely highly active areas of research. The results presented in chapter 4 "Sex-specific and multiomic integration to enhance accuracy of peripheral blood biomarkers of major depressive disorder" are going to be submitted for publication within the next 2 months. Overlap tests using additional recent GWAS, DNA methylation or other newly available datasets will be discussed and prioritized to avoid postponing the submission.

Concerning other approaches that could also be added, tests investigating the difference in gene co-expression between men and women and between MDD and controls might be evaluated (DiffCoEx Tesson et al., 2010, and DGCA (McKenzie et al., 2016)). The outputs could point-out other gene modules based on their specifically differential co-expression in one sex compared to the other sex, or in MDD compared to controls.

In the analytical integration framework, all dimension reduction techniques that we have assessed are linear. Additional non-linear approaches, such as unsupervised machine learning methods, auto-encoders, might be investigated, and could be the object of another study. Methodologically, another way to adapt our framework, would be to direct the jDR step using previous knowledge of known biological pathways based on databases such as REACTOME or PathwayCommons (Cerami et al., 2010). For instance, group lasso sparsity using these pathways as groups constraints could improve the prediction rates and simplify the biological interpretation of the selected predictive features. Furthermore, the application of our framework to continuous phenotype (CTQ score) is considered for future work.

In conclusion, the advent of personalized psychiatry, particularly mood disorders, will necessitate more than a finer clinical diagnosis. Indeed, the contribution of biological approaches, particularly omics, in discovering prognostic and diagnostic biomarkers is critical, and particular attention must be paid to the development of integrative multi-omics approaches in discovering these biomarkers. Our study above shows that combining several modalities might substantially impact the identification of depressed patients and, in the long run, could aid in the stratification of individuals with psychiatric disorders.

Bibliography

Aas, M., Henry, C., Andreassen, O. A., Bellivier, F., Melle, I., & Etain, B. (2016). The role of childhood trauma in bipolar disorders. https://doi.org/10.1186/s40345-015-0042-0

Abbassi-Daloii, T., Kan, H. E., Raz, V., & Hoen, P. '. (2020). Recommendations for the analysis of gene expression data to identify intrinsic differences between similar tissues. *Genomics*, *112*(5), 3157–3165. https://doi.org/10.1016/j.ygeno.2020.05.026

Aberg, K. A., Dean, B., Shabalin, A. A., Chan, R. F., Han, L. K., Zhao, M., van Grootheest, G., Xie, L. Y., Milaneschi, Y., Clark, S. L., Turecki, G., Penninx, B. W., & van den Oord, E. J. (2020). Methylome-wide association findings for major depressive disorder overlap in blood and brain and replicate in independent brain samples. *Molecular Psychiatry*, *25*(6), 1344–1354. https://doi.org/10.1038/s41380-018-0247-6

Agnew-Blais, J., & Danese, A. (2016). Childhood maltreatment and unfavourable clinical outcomes in bipolar disorder: A systematic review and meta-analysis. *The Lancet Psychiatry*, *3*(4), 342–349. https://doi.org/10.1016/s2215-0366(15)00544-1

Akbarian, S., Liu, C., Knowles, J. A., Vaccarino, F. M., Farnham, P. J., Crawford, G. E., Jaffe, A. E., Pinto, D., Dracheva, S., Geschwind, D. H., Mill, J., Nairn, A. C., Abyzov, A., Pochareddy, S., Prabhakar, S., Weissman, S., Sullivan, P. F., State, M. W., Weng, Z., … Sestan, N. (2015). The PsychENCODE project. *Nature Neuroscience*, *18*(12), 1707–1712. https://doi.org/10.1038/nn.4156

Akiskal, H. S., Bourgeois, M. L., Angst, J., & Post, R. (2000). Re-evaluating the prevalence of and diagnostic composition within the broad clinical spectrum... - Abstract - UK PubMed Central. *59*. http://ukpmc.ac.uk/abstract/MED/11121824/reload=0;jsessionid=tmriRkjAILNJtDhdn3Qr.2

Akula, N, Barb, J, Jiang, X, Wendland, J. R., Choi, K. H., Sen, S. K., Hou, L, Chen, D. T. W., Laje, G, Johnson, K, Lipska, B. K., Kleinman, J. E., Corrada-Bravo, H, Detera-Wadleigh, S, Munson, P. J., & McMahon, F. J. (2014). RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Molecular psychiatry*, *19*(11), 1179–1185. https://doi.org/10.1038/mp.2013.170

Albert, P. R. (2015). Why is depression more prevalent in women? *Journal of Psychiatry and Neuroscience*, *40*(4), 219–221. https://doi.org/10.1503/jpn.150205

Alladi, C. G. (2018). Dna methylation as a biomarker of treatment response variability in serious mental illnesses: A systematic review focused on bipolar disorder, schizophrenia, and major depressive disorder. https://doi.org/10.3390/ijms19103026

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, *17*(8), 487–500. https://doi.org/10.1038/nrg.2016.59

Alonso, J., Petukhova, M., Vilagut, G., Chatterji, S., Heeringa, S., Üstün, T. B., Alhamzawi, A. O., Viana, M. C., Angermeyer, M., Bromet, E., Bruffaerts, R., De Girolamo, G., Florescu, S., Gureje, O., Haro, J. M., Hinkov, H., Hu, C. Y., Karam, E. G., Kovess, V., … Kessler, R. C. (2011). Days out of role due to common physical and mental conditions: Results from the WHO World Mental Health surveys. *Molecular Psychiatry*, *16*(12), 1234–1246. https://doi.org/10.1038/mp.2010.101

Álvarez Ruiz, E. M., & Gutiérrez-Rojas, L. (2015). Comorbidity of bipolar disorder and eating disorders. *Revista de Psiquiatría y Salud Mental (English Edition)*, *8*(4), 232–241. https://doi.org/10.1016/j.rpsmen.2015.05.001

Amasi-Hartoonian, N., Pariante, C. M., Cattaneo, A., & Sforzini, L. (2022). Understanding treatment-resistant depression using "omics" techniques: A systematic review. *Journal of Affective Disorders*, *318*, 423–455. https://doi.org/10.1016/j.jad.2022.09.011

Amoah, S. K., Rodriguez, B. A., Logothetis, C. N., Chander, P., Sellgren, C. M., Weick, J. P., Sheridan, S. D., Jantzie, L. L., Webster, M. J., & Mellios, N. (2019). Exosomal secretion of a psychosis-altered miRNA that regulates glutamate receptor expression is affected by antipsychotics. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 10.1038/s41386–019–0579–1. https://doi.org/10.1038/s41386-019-0579-1

Anders, S., Pyl, P. T., & Huber, W. (2014). HTSeq–a python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Antunes, A. S., & Martins-de Souza, D. (2022). Single-Cell RNA Sequencing and Its Applications in the Study of Psychiatric Disorders. *Biological Psychiatry Global Open Science*, (20), 1–11. https://doi.org/10.1016/j.bpsgos.2022.03.013

Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., Buettner, F., Huber, W., & Stegle, O. (2018). Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology*, *14*(6), 0–13. https://doi.org/10.15252/msb.20178124

Arion, D., Huo, Z., Enwright, J. F., Corradi, J. P., Tseng, G., & Lewis, D. A. (2017). Transcriptome Alterations in Prefrontal Pyramidal Cells Distinguish Schizophrenia From Bipolar and Major Depressive Disorders. *Biological psychiatry*, *82*(8), 594–600. https://doi.org/10.1016/j.biopsych.2017.03.018

Association, A. P. (2013, May). *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890425596

Bagot, R. C., & Meaney, M. J. (2010). Epigenetics and the biological basis of gene × environment interactions. *Journal of the American Academy of Child and Adolescent Psychiatry*, *49*(8), 752–771. https://doi.org/10.1016/j.jaac.2010.06.001

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Barbu, M. C., Amador, C., Kwong, A. S., Shen, X., Adams, M. J., Howard, D. M., Walker, R. M., Morris, S. W., Min, J. L., Liu, C., van Dongen, J., Ghanbari, M., Relton, C., Porteous, D. J., Campbell, A., Evans, K. L., Whalley, H. C., & McIntosh, A. M. (2022). Complex trait methylation scores in the prediction of major depressive disorder. *eBioMedicine*, *79*, 104000. https://doi.org/10.1016/j.ebiom.2022.104000

Barbu, M. C., Shen, X., Walker, R. M., Howard, D. M., Evans, K. L., Whalley, H. C., Porteous, D. J., Morris, S. W., Deary, I. J., Zeng, Y., Marioni, R. E., Clarke, T.-K., & McIntosh, A. M. (2020). Epigenetic prediction of major depressive disorder. *Molecular psychiatry*. https://doi.org/10.1038/s41380-020-0808-3

Baumeister, D, Akhtar, R, Ciufolini, S, Pariante, C. M., & Mondelli, V. (2015). Childhood trauma and adulthood inflammation: A meta-analysis of peripheral c-reactive protein, interleukin-6 and tumour necrosis factor-. *Molecular Psychiatry*, *21*(5), 642–649. https://doi.org/10.1038/mp.2015.67

Bekhbat, M., Howell, P. A., Rowson, S. A., Kelly, S. D., Tansey, M. G., & Neigh, G. N. (2019). Chronic adolescent stress sex-specifically alters central and peripheral neuro-immune reactivity in rats. *Brain, Behavior, and Immunity*, *76*, 248–257. https://doi.org/10.1016/j.bbi.2018.12.005

Belzeaux, R., Lin, R., Ju, C., Chay, M.-A., Fiori, L. M., Lutz, P.-E., & Turecki, G. (2018). Transcriptomic and epigenomic biomarkers of antidepressant response. *Journal of Affective Disorders*, *233*, 36–44. https://doi.org/10.1016/j.jad.2017.08.087

Benjamini, Y., & Hochberg, Y. (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), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., Stokes, J., Handelsman, L., Medrano, M., Desmond, D., & Zule, W. (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse and Neglect*, *27*(2), 169–190. https://doi.org/10.1016/S0145-2134(02)00541-0

Berrettini, W. (2004). Genetics of major mood disorders. *Psychiatry (Edgmont (Pa. : Township))*, *1*(2), 38–48. http://www.ncbi.nlm.nih.gov/pubmed/21197377{\%}0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3012618

Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., & Milanesi, L. (2016). Methods for the integration of multi-omics data: Mathematical aspects. *BMC Bioinformatics*, *17*(2). https://doi.org/10.1186/s12859-015-0857-9

Berthelot, N., Garon-Bissonnette, J., Jomphe, V., Doucet-Beaupré, H., Bureau, A., & Maziade, M. (2022). Childhood Trauma May Increase Risk of Psychosis and Mood Disorder in

Genetically High-risk Children and Adolescents by Enhancing the Accumulation of Risk Indicators. *Schizophrenia Bulletin Open*, *3*(1), 1–10. https://doi.org/10.1093/schizbullopen/sgac017

Bhak, Y., oh Jeong, H., Cho, Y. S., Jeon, S., Cho, J., Gim, J.-A., Jeon, Y., Blazyte, A., Park, S. G., Kim, H.-M., Shin, E.-S., Paik, J.-W., Lee, H.-W., Kang, W., Kim, A., Kim, Y., Kim, B. C., Ham, B.-J., Bhak, J., & Lee, S. (2019). Depression and suicide risk prediction models using blood-derived multi-omics data. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-019-0595-2

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L., Schroth, G. P., Gunderson, K. L., Fan, J. B., & Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

Bibikova, M., Le, J., Barnes, B., Saedinia-Melnyk, S., Zhou, L., Shen, R., & Gunderson, K. L. (2009). Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics*, *1*(1), 177–200. https://doi.org/10.2217/epi.09.14

Bierut, L. J., Heath, A. C., Bucholz, K. K., Dinwiddie, S. H., Madden, P. A. F., Statham, D. J., Dunne, M. P., & Martin, N. G. (2015). Major Depressive Disorder in a Community-Based Twin Sample. *56*(June 1999), 557–563.

Binder, E. B. (2017). Disección de los mecanismos moleculares de las interacciones gen x entorno: Implicaciones para el diagnóstico y tratamiento de trastornos psiquiátricos relacionados con el estrés. *European Journal of Psychotraumatology*, *8*(5). https://doi.org/10.1080/20008198.2017.1412745

Braun, P. R., Han, S., Hing, B., Nagahama, Y., Gaul, L. N., Heinzman, J. T., Grossbach, A. J., Close, L., Dlouhy, B. J., Howard, M. A., Kawasaki, H., Potash, J. B., & Shinozaki, G. (2019). Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-019-0376-y

Breen, M. S., Maihofer, A. X., Glatt, S. J., Tylee, D. S., Chandler, S. D., Tsuang, M. T., Risbrough, V. B., Baker, D. G., O'Connor, D. T., Nievergelt, C. M., & Woelk, C. H. (2015). Gene networks specific for innate immunity define post-traumatic stress disorder. *Molecular Psychiatry*, *20*(12), 1538–1545. https://doi.org/10.1038/mp.2015.9

Breen, M. S., Tylee, D. S., Maihofer, A. X., Neylan, T. C., Mehta, D., Binder, E. B., Chandler, S. D., Hess, J. L., Kremen, W. S., Risbrough, V. B., Woelk, C. H., Baker, D. G., Nievergelt, C. M., Tsuang, M. T., Buxbaum, J. D., & Glatt, S. J. (2017). PTSD blood transcriptome mega-analysis: Shared inflammatory pathways across biological sex and modes of trauma. *Neuropsychopharmacology*, *43*(3), 469–481. https://doi.org/10.1038/npp.2017.220

Brennan, G. P., & Henshall, D. C. (2020). MicroRNAs as regulators of brain function and targets for treatment of epilepsy. *Nature Reviews Neurology*, *16*(9), 506–519. https://doi.org/10.1038/s41582-020-0369-8

Bushnell, B. (2014). Bbmap: A fast, accurate, splice-aware aligner. https://www.osti.gov/biblio/1241166

Bustamante, A. C., Armstrong, D. L., & Uddin, M. (2018). Epigenetic profiles associated with major depression in the human brain (2017/12/14). *Psychiatry research*, *260*, 439–442. https://doi.org/10.1016/j.psychres.2017.12.010

Cahill, K. M., Huo, Z., Tseng, G. C., Logan, R. W., & Seney, M. L. (2018). Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-27903-2

Cai, H., Xie, X. M., Zhang, Q., Cui, X., Lin, J. X., Sim, K., Ungvari, G. S., Zhang, L., & Xiang, Y. T. (2021). Prevalence of Suicidality in Major Depressive Disorder: A Systematic Review and Meta-Analysis of Comparative Studies. *Frontiers in Psychiatry*, *12*(September). https://doi.org/10.3389/fpsyt.2021.690130

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F. M., Breen, G., Byrne, E. M., Clarke, T.-K., Forstner, A. J., Grabe, H. J., Hamilton, S. P., Levinson, D. F., Lewis, C. M., Lewis, G., Martin, N. G., Milaneschi, Y., Mors, O., Müller-Myhsok, B., Penninx, B. W. J. H., … Flint, J. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature Genetics*, *52*(4), 437–447. https://doi.org/10.1038/s41588-020-0594-5

Cai, Z., Poulos, R. C., Liu, J., & Zhong, Q. (2022). Machine learning for multi-omics data integration in cancer. *iScience*, *25*(2), 103798. https://doi.org/10.1016/j.isci.2022.103798

Campbell, J., & Ehlert, U. (2012). Acute psychosocial stress: Does the emotional stress response correspond with physiological responses? *Psychoneuroendocrinology*, *37*(8), 1111–1134. https://doi.org/10.1016/j.psyneuen.2011.12.010

Cantini, L., Zakeri, P., Hernandez, C., Naldi, A., Thieffry, D., Remy, E., & Baudot, A. (2021). Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nature Communications*, *12*(1), 1–12. https://doi.org/10.1038/s41467-020-20430-7

Cerami, E. G., Gross, B. E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., Schultz, N., Bader, G. D., & Sander, C. (2010). Pathway commons, a web resource for biological pathway data. *Nucleic Acids Research*, *39*(Database), D685–D690. https://doi.org/10.1093/nar/gkq1039

Ceylan, D., Tufekci, K. U., Keskinoglu, P., Genc, S., & Özerdem, A. (2020). Circulating exosomal microRNAs in bipolar disorder. *Journal of Affective Disorders*, *262*, 99–107. https://doi.org/10.1016/j.jad.2019.10.038

Chalise, P., & Fridley, B. L. (2017). Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. *PLoS ONE*, *12*(5), 1–18. https://doi.org/10.1371/journal.pone.0176278

Chan, R. F., Turecki, G., Shabalin, A. A., Guintivano, J., Zhao, M., Xie, L. Y., van Grootheest, G., Kaminsky, Z. A., Dean, B., Penninx, B. W., Aberg, K. A., & van den Oord, E. J. (2019). Cell Type–Specific Methylome-wide Association Studies Implicate Neurotrophin and Innate Immune Signaling in Major Depressive Disorder. *Biological Psychiatry*, *87*(5), 431–442. https://doi.org/10.1016/j.biopsych.2019.10.014

Chang, L.-C., Lin, H.-M., Sibille, E., & Tseng, G. C. (2013). Meta-analysis methods for combining multiple expression profiles: Comparisons, statistical characterization and an application guideline. *BMC Bioinformatics*, *14*(1). https://doi.org/10.1186/1471-2105-14-368

Charlesworth, J. C., Curran, J. E., Johnson, M. P., Göring, H. H., Dyer, T. D., Diego, V. P., Kent, J. W., Mahaney, M. C., Almasy, L., MacCluer, J. W., Moses, E. K., & Blangero, J. (2010). Transcriptomic epidemiology of smoking: The effect of smoking on gene expression in lymphocytes. *BMC Medical Genomics*, *3*(1). https://doi.org/10.1186/1755-8794-3-29

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. In: In *Methods in molecular biology*. Springer New York, 2018, pp. 243–259. https://doi.org/10.1007/978-1-4939-7493-1_12.

Chen, R., & Snyder, M. (2012). Promise of personalized omics to precision medicine. *WIREs Systems Biology and Medicine*, *5*(1), 73–82. https://doi.org/10.1002/wsbm.1198

Chessel, D., & Hanafi, M. Analyses de la co-inertie de k nuages de points. In: 1996.

Choi, J. K., Yu, U., Kim, S., & Yoo, O. J. (2003). Combining multiple microarray studies and modeling interstudy variation. *Bioinformatics*, *19*(SUPPL. 1). https://doi.org/10.1093/bioinformatics/btg1010

Chrousos, G. P. (2009). Stress and disorders of the stress system. *Nature Reviews Endocrinology*, *5*(7), 374–381. https://doi.org/10.1038/nrendo.2009.106

Ciobanu, L. G., Sachdev, P. S., Trollor, J. N., Reppermund, S., Thalamuthu, A., Mather, K. A., Cohen-Woods, S., & Baune, B. T. (2016). Differential gene expression in brain and peripheral tissues in depression across the life span: A review of replicated findings. *Neuroscience and Biobehavioral Reviews*, *71*, 281–293. https://doi.org/10.1016/j.neubiorev.2016.08.018

Ciobanu, L. G., Sachdev, P. S., Trollor, J. N., Reppermund, S., Thalamuthu, A., Mather, K. A., Cohen-Woods, S., Stacey, D., Toben, C., Schubert, K. O., & Baune, B. T. (2018). Co-

expression network analysis of peripheral blood transcriptome identifies dysregulated protein processing in endoplasmic reticulum and immune response in recurrent MDD in older adults. *Journal of psychiatric research*, *107*, 19–27. https://doi.org/10.1016/j.jpsychires.2018.09.017

Ciobanu, L. G., Sachdev, P. S., Trollor, J. N., Reppermund, S., Thalamuthu, A., Mather, K. A., Cohen-Woods, S., Stacey, D., Toben, C., Schubert, K. O., & Baune, B. T. (2020). Down-regulated transferrin receptor in the blood predicts recurrent MDD in the elderly cohort: A fuzzy forests approach. *Journal of Affective Disorders*, *267*(February), 42–48. https://doi.org/10.1016/j.jad.2020.02.001

Cipriani, A., Barbui, C., Salanti, G., Rendell, J., Brown, R., Stockton, S., Purgato, M., Spineli, L. M., Goodwin, G. M., & Geddes, J. R. (2011). Comparative efficacy and acceptability of antimanic drugs in acute mania: A multiple-treatments meta-analysis. *The Lancet*, *378*(9799), 1306–1315. https://doi.org/10.1016/s0140-6736(11)60873-8

Clark, S. L., Hattab, M. W., Chan, R. F., Shabalin, A. A., Han, L. K. M., Zhao, M., Smit, J. H., Jansen, R., Milaneschi, Y., Xie, L. Y., van Grootheest, G., Penninx, B. W. J. H., Aberg, K. A., & van den Oord, E. J. C. G. (2019). A methylation study of long-term depression risk. *Molecular psychiatry*, 10.1038/s41380–019–0516–z. https://doi.org/10.1038/s41380-019-0516-z

Clausen, A. R., Durand, S., Petersen, R. L., Staunstrup, N. H., & Qvist, P. (2022). Circulating miRNAs as potential biomarkers for patient stratification in bipolar disorder: A combined review and data mining approach. *Genes*, *13*(6), 1038. https://doi.org/10.3390/genes13061038

Comes, A. L., Czamara, D., Adorjan, K., Anderson-Schmidt, H., Andlauer, T. F. M., Budde, M., Gade, K., Hake, M., Kalman, J. L., Papiol, S., Reich-Erkelenz, D., Klöhn-Saghatolislam, F., Schaupp, S. K., Schulte, E. C., Senner, F., Juckel, G., Schmauß, M., Zimmermann, J., Reimer, J., … Heilbronner, U. (2020). The role of environmental stress and DNA methylation in the longitudinal course of bipolar disorder. *International journal of bipolar disorders*, *8*(1), 9. https://doi.org/10.1186/s40345-019-0176-6

Corsi-Zuelli, F., Loureiro, C. M., Shuhama, R., Fachim, H. A., Menezes, P. R., Louzada-Junior, P., Mondelli, V., & Del-Ben, C. M. (2019). Cytokine profile in first-episode psychosis, unaffected siblings and community-based controls: The effects of familial liability and childhood maltreatment. *Psychological Medicine*, *50*(7), 1139–1147. https://doi.org/10.1017/s0033291719001016

Culhane, A. C., Perrière, G., & Higgins, D. G. (2003). *BMC Bioinformatics*, *4*(1), 59. https://doi.org/10.1186/1471-2105-4-59

Czamara, D., Tissink, E., Tuhkanen, J., Martins, J., Awaloff, Y., Drake, A. J., Khulan, B., Palotie, A., Winter, S. M., Nemeroff, C. B., Craighead, W. E., Dunlop, B. W., May-

berg, H. S., Kinkead, B., Mathew, S. J., Iosifescu, D. V., Neylan, T. C., Heim, C. M., Lahti, J., … Binder, E. B. (2021). Combined effects of genotype and childhood adversity shape variability of DNA methylation across age. *Translational Psychiatry*, *11*(1). https://doi.org/10.1038/s41398-020-01147-z

da Silva, L. R., Carlson, P. T., Silva-Costa, L. C., de Souza, D. M., & de Almeida, V. (2021). Molecular mechanisms associated with antidepressant treatment on major depression. *Complex Psychiatry*, *7*(3-4), 49–59. https://doi.org/10.1159/000518098

Dalvie, S., Maihofer, A. X., Coleman, J. R. I., Bradley, B., Breen, G., Brick, L. A., Chen, C.-Y., Choi, K. W., Duncan, L. E., Guffanti, G., Haas, M., Harnal, S., Liberzon, I., Nugent, N. R., Provost, A. C., Ressler, K. J., Torres, K., Amstadter, A. B., Austin, S. B., … Nievergelt, C. M. (2020). Genomic influences on self-reported childhood maltreatment. *Translational Psychiatry*, *10*(1). https://doi.org/10.1038/s41398-020-0706-0

Danese, A. (2020). Annual Research Review: Rethinking childhood trauma-new research directions for measurement, study design and analytical strategies. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *61*(3), 236–250. https://doi.org/10.1111/jcpp.13160

Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W., & Kelley, K. W. (2008). From inflammation to sickness and depression: When the immune system subjugates the brain. *Nature Reviews Neuroscience*, *9*(1), 46–56. https://doi.org/10.1038/nrn2297

de Leeuw, C. A., Mooij, J. M., Heskes, T., & Posthuma, D. (2015). MAGMA: Generalized gene-set analysis of GWAS data (H. Tang, Ed.). *PLOS Computational Biology*, *11*(4), e1004219. https://doi.org/10.1371/journal.pcbi.1004219

de Ronde, M. W., Kok, M. G., Moerland, P. D., den Bossche, J. V., Neele, A. E., Halliani, A., van der Made, I., de Winther, M. P., Meijers, J. C., Creemers, E. E., & Pinto-Sietsma, S.-J. (2017). High miR-124-3p expression identifies smoking individuals susceptible to atherosclerosis. *Atherosclerosis*, *263*, 377–384. https://doi.org/10.1016/j.atherosclerosis.2017.03.045

De Bellis, M. D., & Zisk, A. (2014). The biological effects of childhood trauma. *Child and Adolescent Psychiatric Clinics*, *23*(2), 185–222.

Dell'Osso, B., Cafaro, R., & Ketter, T. A. (2021). Has bipolar disorder become a predominantly female gender related condition? analysis of recently published large sample studies. *International Journal of Bipolar Disorders*, *9*(1). https://doi.org/10.1186/s40345-020-00207-z

Dempster, E. L., Wong, C. C. Y., Lester, K. J., Burrage, J., Gregory, A. M., Mill, J., & Eley, T. C. (2014). Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. *Biological psychiatry*, *76*(12), 977–983. https://doi.org/10.1016/j.biopsych.2014.04.013

Diflorio, A., & Jones, I. (2010). Is sex important Gender differences in bipolar disorder. *International Review of Psychiatry*, *22*(5), 437–452. https://doi.org/10.3109/09540261.2010.514601

do Prado, C. H., Grassi-Oliveira, R., Daruy-Filho, L., Wieck, A., & Bauer, M. E. (2017). Evidence for immune activation and resistance to glucocorticoids following childhood maltreatment in adolescents without psychopathology. *Neuropsychopharmacology*, *42*(11), 2272–2282. https://doi.org/10.1038/npp.2017.137

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2012). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dolédec, S., & Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater Biology*, *31*(3), 277–294. https://doi.org/10.1111/j.1365-2427.1994.tb01741.x

Dome, P., Rihmer, Z., & Gonda, X. (2019). Suicide risk in bipolar disorder: A brief review. *Medicina (Lithuania)*, *55*(8), 8–10. https://doi.org/10.3390/medicina55080403

Dong, M., Zeng, L. N., Lu, L., Li, X. H., Ungvari, G. S., Ng, C. H., Chow, I. H., Zhang, L., Zhou, Y., & Xiang, Y. T. (2019). Prevalence of suicide attempt in individuals with major depressive disorder: A meta-analysis of observational surveys. *Psychological Medicine*, *49*(10), 1691–1704. https://doi.org/10.1017/S0033291718002301

Dray, S., Chessel, D., & Thioulouse, J. (2003). Co-inertia analysis and the linking of ecological data tables. *Ecology*, *84*(11), 3078–3089. https://doi.org/10.1890/03-0178

Edgar, R. D., Jones, M. J., Meaney, M. J., Turecki, G, & Kobor, M. S. (2017). BECon: A tool for interpreting DNA methylation findings from blood in the context of brain. *Translational Psychiatry*, *7*(8), e1187–e1187. https://doi.org/10.1038/tp.2017.171

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, *7*(1), 1 –26. https://doi.org/10.1214/aos/1176344552

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American statistical Association*, *82*(397), 171–185.

Egan, M. F., Kojima, M., Callicott, J. H., Goldberg, T. E., Kolachana, B. S., Bertolino, A., Zaitsev, E., Gold, B., Goldman, D., Dean, M., Lu, B., & Weinberger, D. R. (2003). The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell*, *112*(2), 257–269. https://doi.org/10.1016/S0092-8674(03)00035-7

Eid, R. S., Gobinath, A. R., & Galea, L. A. (2019). Sex differences in depression: Insights from clinical and preclinical studies. *Progress in Neurobiology*, *176*, 86–102. https://doi.org/10.1016/j.pneurobio.2019.01.006

Ernst, M., Brähler, E., Otten, D., Werner, A. M., Tibubos, A. N., Reiner, I., Wicke, F., Wiltink, J., Michal, M., Nagler, M., Münzel, T., Wild, P. S., König, J., Pfeiffer, N., Borta, A., Lackner, K. J., & Beutel, M. E. (2021). Inflammation predicts new onset of depression in men, but not in women within a prospective, representative community cohort. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-81927-9

Etain, B., Lajnef, M., Brichant-Petitjean, C., Geoffroy, P. A., Henry, C., Gard, S., Kahn, J. P., Leboyer, M., Young, A. H., & Bellivier, F. (2017). Childhood trauma and mixed episodes are associated with poor response to lithium in bipolar disorders. *Acta Psychiatrica Scandinavica*, *135*(4), 319–327. https://doi.org/10.1111/acps.12684

Etain, B., Aas, M., Andreassen, O. A., Lorentzen, S., Dieset, I., Gard, S., Kahn, J.-P., Bellivier, F., Leboyer, M., Melle, I., et al. (2013). Childhood trauma is associated with severe clinical characteristics of bipolar disorders. *The Journal of clinical psychiatry*, *74*(10), 2585.

Fabbri, C. (2022). Genetics in psychiatry: Methods, clinical applications and future perspectives. *Psychiatry and Clinical Neurosciences Reports*, *1*(2), 1–13. https://doi.org/10.1002/pcn5.6

Fernandes, B. S., Williams, L. M., Steiner, J., Leboyer, M., Carvalho, A. F., & Berk, M. (2017). The new field of 'precision psychiatry'. *BMC Medicine*, *15*(1), 1–7. https://doi.org/10.1186/s12916-017-0849-x

Fernandez-Pujals, A. M., Adams, M. J., Thomson, P., McKechanie, A. G., Blackwood, D. H., Smith, B. H., Dominiczak, A. F., Morris, A. D., Matthews, K., Campbell, A., Linksted, P., Haley, C. S., Deary, I. J., Porteous, D. J., MacIntyre, D. J., & McIntosh, A. M. (2015). Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in generation Scotland: Scottish family health study (GS: SFHS). *PLoS ONE*, *10*(11), 1–18. https://doi.org/10.1371/journal.pone.0142197

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., Vos, T., & Whiteford, H. A. (2013). Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. *PLoS Medicine*, *10*(11). https://doi.org/10.1371/journal.pmed.1001547

Fiedorowicz, J. G., Palagummi, N. M., Forman-Hoffman, V. L., Miller, D. D., & Haynes, W. G. (2008). Elevated prevalence of obesity, metabolic syndrome, and cardiovascular risk factors in bipolar disorder. *Annals of Clinical Psychiatry*, *20*(3), 131–137. https://doi.org/10.1080/10401230802177722

Fisher, R. A. (1934). *Statistical methods for research workers; 5th ed.* Oliver; Boyd.

Fluharty, M., Taylor, A. E., Grabski, M., & Munafò, M. R. (2016). The association of cigarette smoking with depression and anxiety: A systematic review. *Nicotine &amp Tobacco Research*, *19*(1), 3–13. https://doi.org/10.1093/ntr/ntw140

Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., et al. (2018). Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage*, *167*, 104–120.

Frach, L., Tierling, S., Schwaiger, M., Moser, D., Heinrichs, M., Hengstler, J. G., Walter, J., & Kumsta, R. (2020). The mediating role of KITLG DNA methylation in the association between childhood adversity and cortisol stress reactivity does not replicate in monocytes. *Psychoneuroendocrinology*, *116*, 104653. https://doi.org/10.1016/j.psyneuen.2020.104653

Franco, E. F., Rana, P., Cruz, A., Calderón, V. V., Azevedo, V., Ramos, R. T., & Ghosh, P. (2021). Performance comparison of deep learning autoencoders for cancer subtype detection using multi-omics data. *Cancers*, *13*(9), 2013.

Friborg, O., Martinsen, E. W., Martinussen, M., Kaiser, S., Overgård, K. T., & Rosenvinge, J. H. (2014). Comorbidity of personality disorders in mood disorders: A meta-analytic review of 122 studies from 1988 to 2010. *Journal of Affective Disorders*, *152-154*(1), 1–11. https://doi.org/10.1016/j.jad.2013.08.023

Fries, G. R., Bauer, I. E., Scaini, G., Wu, M.-J., Kazimi, I. F., Valvassori, S. S., Zunta-Soares, G., Walss-Bass, C., Soares, J. C., & Quevedo, J. (2017). Accelerated epigenetic aging and mitochondrial DNA copy number in bipolar disorder. *Translational psychiatry*, *7*(12), 1283. https://doi.org/10.1038/s41398-017-0048-8

Fries, G. R., Carvalho, A. F., & Quevedo, J. (2018). The miRNome of bipolar disorder. *Journal of Affective Disorders*, *233*, 110–116. https://doi.org/10.1016/j.jad.2017.09.025

Fries, G. R., Zamzow, M. J., Andrews, T., Pink, O., Scaini, G., & Quevedo, J. (2020). Accelerated aging in bipolar disorder: A comprehensive review of molecular findings and their clinical implications. *Neuroscience and Biobehavioral Reviews*, *112*(October 2019), 107–116. https://doi.org/10.1016/j.neubiorev.2020.01.035

Frokjaer, V. G., Pinborg, A., Holst, K. K., Overgaard, A., Henningsson, S., Heede, M., Larsen, E. C., Jensen, P. S., Agn, M., Nielsen, A. P., Stenbæk, D. S., da Cunha-Bang, S., Lehel, S., Siebner, H. R., Mikkelsen, J. D., Svarer, C., & Knudsen, G. M. (2015). Role of serotonin transporter changes in depressive responses to sex-steroid hormone manipulation: A positron emission tomography study. *Biological Psychiatry*, *78*(8), 534–543. https://doi.org/10.1016/j.biopsych.2015.04.015

Gaine, M. E., Seifuddin, F., Sabunciyan, S., Lee, R. S., Benke, K. S., Monson, E. T., Zandi, P. P., Potash, J. B., & Willour, V. L. (2019). Differentially methylated regions in bipolar disorder and suicide. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, *180*(7), 496–507. https://doi.org/10.1002/ajmg.b.32754

Gallo, E. A. G., Munhoz, T. N., de Mola, C. L., & Murray, J. (2018). Gender differences in the effects of childhood maltreatment on adult depression and anxiety: A systematic review and meta-analysis. *Child Abuse &amp Neglect*, *79*, 107–114. https://doi.org/10.1016/j.chiabu.2018.01.003

Gamazon, E. R., Zwinderman, A. H., Cox, N. J., Denys, D., & Derks, E. M. (2019). Multi-tissue transcriptome analyses identify genetic mechanisms underlying neuropsychiatric traits (2019/05/13). *Nature genetics*, *51*(6), 933–940. https://doi.org/10.1038/s41588-019-0409-8

Gandal, M. J., Leppa, V., Won, H., Parikshak, N. N., & Geschwind, D. H. (2016). The road to precision psychiatry: Translating genetics into disease mechanisms. *Nature Neuroscience*, *19*(11), 1397–1407. https://doi.org/10.1038/nn.4409

Gao, X., Jia, M., Zhang, Y., Breitling, L. P., & Brenner, H. (2015). DNA methylation changes of whole blood cells in response to active smoking exposure in adults: A systematic review of DNA methylation studies. *Clinical Epigenetics*, *7*(1). https://doi.org/10.1186/s13148-015-0148-3

Garcia, M., Montalvo, I., Creus, M., Cabezas, Á., Solé, M., Algora, M. J., Moreno, I., Gutiérrez-Zotes, A., & Labad, J. (2016). Sex differences in the effect of childhood trauma on the clinical expression of early psychosis. *Comprehensive Psychiatry*, *68*, 86–96. https://doi.org/10.1016/j.comppsych.2016.04.004

Garno, J. L., Goldberg, J. F., Ramirez, P. M., & Ritzler, B. A. (2005). Impact of childhood abuse on the clinical course of bipolar disorder. *The British Journal of Psychiatry*, *186*(2), 121–125.

Giannakopoulou, O., Lin, K., Meng, X., Su, M.-H., Kuo, P.-H., Peterson, R. E., Awasthi, S., Moscati, A., Coleman, J. R. I., Bass, N., Millwood, I. Y., Chen, Y., Chen, Z., Chen, H.-C., Lu, M.-L., Huang, M.-C., Chen, C.-H., Stahl, E. A., Loos, R. J. F., … and, M. Z. (2021). The genetic architecture of depression in individuals of east asian ancestry. *JAMA Psychiatry*, *78*(11), 1258. https://doi.org/10.1001/jamapsychiatry.2021.2099

Gibbons, A., Sundram, S., & Dean, B. (2020). Changes in non-coding RNA in depression and bipolar disorder: Can they be used as diagnostic or theranostic biomarkers? *Non-coding RNA*, *6*(3). https://doi.org/10.3390/ncrna6030033

Gordovez, F. J. A., & McMahon, F. J. (2020). The genetics of bipolar disorder. *Molecular Psychiatry*, *25*(3), 544–559. https://doi.org/10.1038/s41380-019-0634-7

Green, E. K., Rees, E., Walters, J. T., Smith, K. G., Forty, L., Grozeva, D., Moran, J. L., Sklar, P., Ripke, S., Chambert, K. D., Genovese, G., McCarroll, S. A., Jones, I., Jones, L., Owen, M. J., O'Donovan, M. C., Craddock, N., & Kirov, G. (2016). Copy number variation in bipolar disorder. *Molecular Psychiatry*, *21*(1), 89–93. https://doi.org/10.1038/mp.2014.174

Green, J. G., McLaughlin, K. A., Berglund, P. A., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2010). Childhood adversities and adult psychiatric disorders in the national comorbidity survey replication I: Associations with first onset of DSM-IV disorders. *Archives of General Psychiatry*, *67*(2), 113–123. https://doi.org/10.1001/archgenpsychiatry.2009.186

Greenberg, M. V., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology*, *20*(10), 590–607. https://doi.org/10.1038/s41580-019-0159-6

Guan, J., Cai, J. J., Ji, G., & Sham, P. C. (2019). Commonality in dysregulated expression of gene sets in cortical brains of individuals with autism, schizophrenia, and bipolar disorder. *Translational psychiatry*, *9*(1), 152. https://doi.org/10.1038/s41398-019-0488-4

Gururajan, A., Naughton, M. E., Scott, K. A., O'Connor, R. M., Moloney, G., Clarke, G., Dowling, J., Walsh, A., Ismail, M. F., Shorten, G., Scott, L., McLoughlin, D. M., Cryan, J. F., & Dinan, T. G. (2016). MicroRNAs as biomarkers for major depression: A role for let-7b and let-7c. *Translational Psychiatry*, *6*(8), e862–9. https://doi.org/10.1038/tp.2016.131

Gutiérrez-Rojas, L., Porras-Segovia, A., Dunne, H., Andrade-González, N., & Cervilla, J. A. (2020). Prevalence and correlates of major depressive disorder: A systematic review. *Brazilian Journal of Psychiatry*, *42*(6), 657–672. https://doi.org/10.1590/1516-4446-2020-0650

Ha, M. J., Baladandayuthapani, V., & Do, K.-A. (2015). DINGO: Differential network analysis in genomics. *Bioinformatics*, *31*(21), 3413–3420. https://doi.org/10.1093/bioinformatics/btv406

HAMILTON, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology*, *6*(4), 278–296. https://doi.org/10.1111/j.2044-8260.1967.tb00530.x

Han, L. K. M., Aghajani, M., Clark, S. L., Chan, R. F., Hattab, M. W., Shabalin, A. A., Zhao, M., Kumar, G., Xie, L. Y., Jansen, R., Milaneschi, Y., Dean, B., Aberg, K. A., van den Oord, E. J. C. G., & Penninx, B. W. J. H. (2018). Epigenetic Aging in Major Depressive Disorder (2018/04/16). *The American journal of psychiatry*, *175*(8), 774–782. https://doi.org/10.1176/appi.ajp.2018.17060595

Hankin, B. L., & Abramson, L. Y. (1999). Development of gender differences in depression: Description and possible explanations. *Annals of Medicine*, *31*(6), 372–379. https://doi.org/10.3109/07853899908998794

Hannon, E., Lunnon, K., Schalkwyk, L., & Mill, J. (2015). Interindividual methylomic variation across blood, cortex, and cerebellum: Implications for epigenetic studies of neu-

rological and neuropsychiatric phenotypes. *Epigenetics*, *10*(11), 1024–1032. https://doi.org/10.1080/15592294.2015.1100786

Hasin, D. S., Sarvet, A. L., Meyers, J. L., Saha, T. D., Ruan, W. J., Stohl, M., & Grant, B. F. (2018). Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the united states. *JAMA Psychiatry*, *75*(4), 336. https://doi.org/10.1001/jamapsychiatry.2017.4602

He, H., Hu, C., Ren, Z., Bai, L., Gao, F., & Lyu, J. (2020). Trends in the incidence and DALYs of bipolar disorder at global, regional, and national levels: Results from the global burden of Disease Study 2017. *Journal of Psychiatric Research*, *125*(December 2019), 96–105. https://doi.org/10.1016/j.jpsychires.2020.03.015

He, Y., & Ecker, J. R. (2015). Non-CG methylation in the human genome. *Annual Review of Genomics and Human Genetics*, *16*(1), 55–77. https://doi.org/10.1146/annurev-genom-090413-025437

Heard, N. A., & Rubin-Delanchy, P. (2018). Choosing between methods of combining p -values. *Biometrika*, *105*(1), 239–246. https://doi.org/10.1093/biomet/asx076

Heim, C., & Nemeroff, C. B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: Preclinical and clinical studies. *Biological psychiatry*, *49*(12), 1023–1039.

Herman, J. P., McKlveen, J. M., Ghosal, S., Kopp, B., Wulsin, A., Makinson, R., Scheimann, J., & Myers, B. (2016, March). Regulation of the hypothalamic-pituitary-adrenocortical stress response. https://doi.org/10.1002/cphy.c150015

Hess, J. L., Tylee, D. S., Barve, R., de Jong, S., Ophoff, R. A., Kumarasinghe, N., Tooney, P., Schall, U., Gardiner, E., Beveridge, N. J., Scott, R. J., Yasawardene, S., Perera, A., Mendis, J., Carr, V., Kelly, B., Cairns, M., Tsuang, M. T., Glatt, S. J., … Glatt, S. J. (2019). Transcriptomic abnormalities in peripheral blood in bipolar disorder, and discrimination of the major psychoses. *Schizophrenia research*, *217*, S0920–9964(19)30316–0. https://doi.org/10.1016/j.schres.2019.07.036

Hett, D., Etain, B., & Marwaha, S. (2022). Childhood trauma in bipolar disorder : new targets for future interventions, 1–3. https://doi.org/10.1192/bjo.2022.529

Hillegers, M. H., Burger, H., Wals, M., Reichart, C. G., Verhulst, F. C., Nolen, W. A., & Ormel, J. (2004). Impact of stressful life events, familial loading and their interaction on the onset of mood disorders: Study in a high-risk cohort of adolescent offspring of parents with bipolar disorder. *British Journal of Psychiatry*, *185*(AUG.), 97–101. https://doi.org/10.1192/bjp.185.2.97

Hiller-Sturmhöfel, S, & Bartke, A. (1998). The endocrine system: An overview. *Alcohol Health Res. World*, *22*(3), 153–164.

Hirschfeld, R. M. (2001). The comorbidity of major depression and anxiety disorders: Recognition and management in primary care. *Primary Care Companion to the Journal of Clinical Psychiatry*, *3*(6), 244–254. https://doi.org/10.4088/pcc.v03n0609

Ho, A. M.-C., Winham, S. J., Armasu, S. M., Blacker, C. J., Millischer, V., Lavebratt, C., Overholser, J. C., Jurjus, G. J., Dieter, L., Mahajan, G., Rajkowska, G., Vallender, E. J., Stockmeier, C. A., Robertson, K. D., Frye, M. A., Choi, D.-S., & Veldic, M. (2019). Genome-wide DNA methylomic differences between dorsolateral prefrontal and temporal pole cortices of bipolar disorder (2019/06/26). *Journal of psychiatric research*, *117*, 45–54. https://doi.org/10.1016/j.jpsychires.2019.05.030

Hoffman, G. E., & Schadt, E. E. (2016). variancePartition: Interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*, *17*(1). https://doi.org/10.1186/s12859-016-1323-z

Hollins, S. L., & Cairns, M. J. (2016). MicroRNA: Small RNA mediators of the brains genomic response to environmental stress. *Progress in Neurobiology*, *143*, 61–81. https://doi.org/10.1016/j.pneurobio.2016.06.005

Hori, H., Nakamura, S., Yoshida, F., Teraishi, T., Sasayama, D., Ota, M., Hattori, K., Kim, Y., Higuchi, T., & Kunugi, H. (2018). Integrated profiling of phenotype and blood transcriptome for stress vulnerability and depression (2018/08/06). *Journal of psychiatric research*, *104*(August), 202–210. https://doi.org/10.1016/j.jpsychires.2018.08.010

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Horvath, S. (2015). Erratum to: DNA methylation age of human tissues and cell types. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0649-6

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., Wiencke, J. K., & Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*(1). https://doi.org/10.1186/1471-2105-13-86

Hu, B., Won, H., Mah, W., Park, R. B., Kassim, B., Spiess, K., Kozlenkov, A., Crowley, C. A., Pochareddy, S., Ashley-Koch, A. E., Crawford, G. E., Garrett, M. E., Song, L., Safi, A., Johnson, G. D., Wray, G. A., Reddy, T. E., Goes, F. S., Zandi, P., … and, D. H. G. (2021). Neuronal and glial 3d chromatin architecture informs the cellular etiology of brain disorders. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-24243-0

Hüls, A., Robins, C., Conneely, K. N., De Jager, P. L., Bennett, D. A., Epstein, M. P., Wingo, T. S., & Wingo, A. P. (2020). Association between DNA methylation levels in brain tissue and late-life depression in community-based participants. *Translational psychiatry*, *10*(1), 262. https://doi.org/10.1038/s41398-020-00948-6

Hussaini, S. M. Q., Choi, C.-I., Cho, C. H., Kim, H. J., Jun, H., & Jang, M.-H. (2014). Wnt signaling in neuropsychiatric disorders: Ties with adult hippocampal neurogenesis and behavior. *Neuroscience &amp Biobehavioral Reviews*, *47*, 369–383. https://doi.org/10.1016/j.neubiorev.2014.09.005

Jaffe, D. H., Rive, B., & Denee, T. R. (2019). The humanistic and economic burden of treatment-resistant depression in europe: A cross-sectional study. *BMC Psychiatry*, *19*(1). https://doi.org/10.1186/s12888-019-2222-4

Janbain, A., Reynès, C., Assaghir, Z., Zeineddine, H., Sabatier, R., & Journot, L. (2021). TopoFun: A machine learning method to improve the functional similarity of gene co-expression modules. *NAR Genomics and Bioinformatics*, *3*(4). https://doi.org/10.1093/nargab/lqab103

Jansen, R, Penninx, B, Madar, V, Xia, K, & ... (2016). Gene expression in major depressive disorder. *Molecular …* https://www.nature.com/articles/mp201557

Jeremian, R., Chen, Y.-A., De Luca, V., Vincent, J. B., Kennedy, J. L., Zai, C. C., & Strauss, J. (2017). Investigation of correlations between DNA methylation, suicidal behavior and aging. *Bipolar disorders*, *19*(1), 32–40. https://doi.org/10.1111/bdi.12466

Johansson, V., Kuja-Halkola, R., Cannon, T. D., Hultman, C. M., & Hedman, A. M. (2019). A population-based heritability estimate of bipolar disorder – In a Swedish twin sample. *Psychiatry Research*, *278*(June), 180–187. https://doi.org/10.1016/j.psychres.2019.06.010

Johnson, W. E., Li, C., & Rabinovic, A. (2006). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*, *8*(1), 118–127.

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, *13*(7), 484–492. https://doi.org/10.1038/nrg3230

Karg, K., Burmeister, M., Shedden, K., & Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited. *Archives of General Psychiatry*, *68*(5), 444. https://doi.org/10.1001/archgenpsychiatry.2010.189

Karos, K., Niederstrasser, N., Abidi, L., Bernstein, D. P., & Bader, K. (2014). Factor structure, reliability, and known groups validity of the German version of the Childhood Trauma Questionnaire (short-form) in Swiss patients and nonpatients. *Journal of Child Sexual Abuse*, *23*(4), 418–430. https://doi.org/10.1080/10538712.2014.896840

Kehrer-Sawatzki, H. (2007). Copy Number Variation in the Human Genome. *eLS*, 1–5. https://doi.org/10.1002/9780470015902.a0020824

Kendall, K. M., Rees, E., Escott-Price, V., Einon, M., Thomas, R., Hewitt, J., O'Donovan, M. C., Owen, M. J., Walters, J. T., & Kirov, G. (2017). Cognitive Performance Among Carri-

ers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biological Psychiatry*, *82*(2), 103–110. https://doi.org/10.1016/j.biopsych.2016.08.014

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Causal relationship between stressful life events and the onset of major depression. *American Journal of Psychiatry*, *156*(6), 837–841. https://doi.org/10.1176/ajp.156.6.837

Kennedy, N., Psych, M. R. C., Boydell, J., Kalidindi, S., Fearon, P., Jones, P. B., Ph, D, Os, J. V., Murray, R. M., Sc, D, & Psych, F. R. (2005). Gender Differences in Incidence and Age at Onset of Mania and Bipolar Disorder Over a 35-Year Period. *Am J Psychiatry*, *162*(2), 257–262.

Kim, H. K., Tyryshkin, K., Elmi, N., Dharsee, M., Evans, K. R., Good, J., Javadi, M., McCormack, S., Vaccarino, A. L., Zhang, X., Andreazza, A. C., & Feilotter, H. (2019). Plasma microRNA expression levels and their targeted pathways in patients with major depressive disorder who are responsive to duloxetine treatment. *Journal of Psychiatric Research*, *110*(December 2018), 38–44. https://doi.org/10.1016/j.jpsychires.2018.12.007

Kim, S., Hwang, Y., Webster, M. J., & Lee, D. (2016). Differential activation of immune/inflammatory response-related co-expression modules in the hippocampus across the major psychiatric disorders. *Molecular Psychiatry*, *21*(3), 376–385. https://doi.org/10.1038/mp.2015.79

Kirov, G. (2015). CNVs in neuropsychiatric disorders. https://doi.org/10.1093/hmg/ddv253

Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J. Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., & Hoh, J. (2005). Complement factor H polymorphism in age-related macular degeneration. *Science*, *308*(5720), 385–389. https://doi.org/10.1126/science.1109557

Klein, S. L., & Flanagan, K. L. (2016). Sex differences in immune responses. *Nature Reviews Immunology*, *16*(10), 626–638. https://doi.org/10.1038/nri.2016.90

Kokras, N., Hodes, G. E., Bangasser, D. A., & Dalla, C. (2019). Sex differences in the hypothalamic–pituitary–adrenal axis: An obstacle to antidepressant drug development? *British Journal of Pharmacology*, *176*(21), 4090–4106. https://doi.org/10.1111/bph.14710

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2016). Fast gene set enrichment analysis. https://doi.org/10.1101/060012

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2018). miRBase: From microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155–D162. https://doi.org/10.1093/nar/gky1141

Krassowski, M., Das, V., Sahu, S. K., & Misra, B. B. (2020). State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Frontiers in Genetics*, *11*(December), 1–17. https://doi.org/10.3389/fgene.2020.610798

Krebs, C. E., Ori, A. P. S., Vreeker, A., Wu, T., Cantor, R. M., Boks, M. P. M., Kahn, R. S., Loohuis, L. M. O., Ophoff, R. A., Olde Loohuis, L. M., Ophoff, R. A., Loohuis, L. M. O., & Ophoff, R. A. (2020). Whole blood transcriptome analysis in bipolar disorder reveals strong lithium effect. *Psychological Medicine*, *50*(15), 2575–2586. https://doi.org/10.1017/S0033291719002745

Kubota, T., Miyake, K., & Hirasawa, T. (2014). Epigenetic understanding of gene-environment interactions in psychiatric disorders: A new concept of clinical genetics. *Specific Gene Expression and Epigenetics: The Interplay Between the Genome and its Environment*, 241–258. https://doi.org/10.1201/b16680

Kupfer, D. J. (2005). The increasing medical burden in bipolar disorder. *JAMA*, *293*(20), 2528. https://doi.org/10.1001/jama.293.20.2528

Labonté, B., Engmann, O., Purushothaman, I., Menard, C., Wang, J., Tan, C., Scarpa, J. R., Moy, G., Loh, Y.-H. E., Cahill, M., Lorsch, Z. S., Hamilton, P. J., Calipari, E. S., Hodes, G. E., Issler, O., Kronman, H., Pfau, M., Obradovic, A. L. J., Dong, Y., … Nestler, E. J. (2017a). Sex-specific transcriptional signatures in human depression. *Nature Medicine*, *23*(9), 1102–1111. https://doi.org/10.1038/nm.4386

Labonté, B., Engmann, O., Purushothaman, I., Menard, C., Wang, J., Tan, C., Scarpa, J. R., Moy, G., Loh, Y. H. E., Cahill, M., Lorsch, Z. S., Hamilton, P. J., Calipari, E. S., Hodes, G. E., Issler, O., Kronman, H., Pfau, M., Obradovic, A. L., Dong, Y., … ... (2017b). Sex-specific transcriptional signatures in human depression. *Nature Medicine*, *23*(9), 1102–1111. https://doi.org/10.1038/nm.4386

Lam, M., Chen, C.-Y., Li, Z., Martin, A. R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B. C., Liu, R., Zhou, W., Guan, L., Kamatani, Y., Kim, S.-W., Kubo, M., Kusumawardhani, A. A. A. A., Liu, C.-M., … and. (2019). Comparative genetic architectures of schizophrenia in east asian and european populations. *Nature Genetics*, *51*(12), 1670–1678. https://doi.org/10.1038/s41588-019-0512-x

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K. N., et al. (2006). The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *science*, *313*(5795), 1929–1935.

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*. https://doi.org/10.1186/1471-2105-9-559

Langfelder, P., & Horvath, S. (2012). Fast r functions for robust correlations and hierarchical clustering. *Journal of statistical software*, *46*(11).

Le, T. T., Savitz, J., Suzuki, H., Misaki, M., Teague, T. K., White, B. C., Marino, J. H., Wiley, G., Gaffney, P. M., Drevets, W. C., McKinney, B. A., & Bodurka, J. (2018). Identification

and replication of RNA-Seq gene network modules associated with depression severity. *Translational Psychiatry*, *8*(1). https://doi.org/10.1038/s41398-018-0234-3

Leday, G. G. R. G., Vértes, P. E., Richardson, S., Greene, J. R., Regan, T., Khan, S., Henderson, R., Freeman, T. C., Pariante, C. M., Harrison, N. A., Vertes, P. E., Cardinal, R., Richardson, S., Leday, G. G. R. G., Freeman, T. C., Regan, T., Hume, D., Wu, Z., Pariante, C. M., … Bullmore, E. T. (2018). Replicable and Coupled Changes in Innate and Adaptive Immune Gene Expression in Two Case-Control Studies of Blood Microarrays in Major Depressive Disorder. *Biological Psychiatry*, *83*(1), 70–80. https://doi.org/10.1016/j.biopsych.2017.01.021

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*(6755), 788–791. https://doi.org/10.1038/44565

Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., Perlis, R. H., Mowry, B. J., Thapar, A., Goddard, M. E., Witte, J. S., Absher, D., Agartz, I., Akil, H., Amin, F., Andreassen, O. A., Anjorin, A., Anney, R., Anttila, V., Arking, D. E., … Wray, N. R. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, *45*(9), 984–994. https://doi.org/10.1038/ng.2711

Lee, S.-Y., Lu, R.-B., Wang, L.-J., Chang, C.-H., Lu, T., Wang, T.-Y., & Tsai, K.-W. (2020). Serum miRNA as a possible biomarker in the diagnosis of bipolar II disorder. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-58195-0

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, *28*(6), 882–883. https://doi.org/10.1093/bioinformatics/bts034

Lemsara, A., Ouadfel, S., & Fröhlich, H. (2020). PathME: Pathway based multi-modal sparse autoencoders for clustering of patient-level multi-omics data. *BMC Bioinformatics*, *21*(1). https://doi.org/10.1186/s12859-020-3465-2

Levine, A. B., Schlosser, C., Grewal, J., Coope, R., Jones, S. J., & Yip, S. (2019). Rise of the Machines: Advances in Deep Learning for Cancer Diagnosis. *Trends in Cancer*, *5*(3), 157–169. https://doi.org/10.1016/j.trecan.2019.02.002

Li, A., & Horvath, S. (2007). Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics*, *23*(2), 222–231. https://doi.org/10.1093/bioinformatics/btl581

Li, C.-X., Wheelock, C. E., Sköld, C. M., & Wheelock, Å. M. (2018). Integration of multi-omics datasets enables molecular classification of COPD. *European Respiratory Journal*, *51*(5), 1701930. https://doi.org/10.1183/13993003.01930-2017

Li, M., Fu, X., Xie, W., Guo, W., Li, B., Cui, R., & Yang, W. (2020). Effect of Early Life Stress on the Epigenetic Profiles in Depression. *Frontiers in Cell and Developmental Biology*, *8*(October), 1–13. https://doi.org/10.3389/fcell.2020.00867

Li, M., D'Arcy, C., Li, X., Zhang, T., Joober, R., & Meng, X. (2019). What do DNA methylation studies tell us about depression? A systematic review. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-019-0412-y

Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., & Zhang, B. (2019). WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Research*, *47*(W1), W199–W205. https://doi.org/10.1093/nar/gkz401

Lim, C. H., Zainal, N. Z., Kanagasundram, S., Zain, S. M., & Mohamed, Z. (2016). Preliminary examination of microRNA expression profiling in bipolar disorder I patients during antipsychotic treatment (2016/05/13). *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, *171*(6), 867–874. https://doi.org/10.1002/ajmg.b.32457

Lisoway, A. J., Zai, C. C., Tiwari, A. K., & Kennedy, J. L. (2018). DNA methylation and clinical response to antidepressant medication in major depressive disorder: A review and recommendations. *Neuroscience Letters*, *669*, 14–23. https://doi.org/10.1016/j.neulet.2016.12.071

Lock, E. F., Hoadley, K. A., Marron, J. S., & Nobel, A. B. (2013). Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *Annals of Applied Statistics*, *7*(1), 523–542. https://doi.org/10.1214/12-AOAS597

Løkhammer, S., Stavrum, A.-K., Polushina, T., Aas, M., Ottesen, A. A., Andreassen, O. A., Melle, I., & Hellard, S. L. (2022). An epigenetic association analysis of childhood trauma in psychosis reveals possible overlap with methylation changes associated with PTSD. *Translational Psychiatry*, *12*(1). https://doi.org/10.1038/s41398-022-01936-8

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., … Moore, H. F. (2013). The genotype-tissue expression (GTEx) project. *Nature Genetics*, *45*(6), 580–585. https://doi.org/10.1038/ng.2653

Loos, R. J. (2020). 15 Years of Genome-Wide Association Studies and No Signs of Slowing Down. *Nature Communications*, *11*(1), 10–12. https://doi.org/10.1038/s41467-020-19653-5

Lopez, J. P., Fiori, L. M., Cruceanu, C., Lin, R., Labonte, B., Cates, H. M., Heller, E. A., Vialou, V., Ku, S. M., Gerald, C., Han, M.-H., Foster, J., Frey, B. N., Soares, C. N., Müller, D. J., Farzan, F., Leri, F., MacQueen, G. M., Feilotter, H., … Turecki, G. (2017). MicroRNAs 146a/b-5 and 425-3p and 24-3p are markers of antidepressant response and regulate MAPK/wnt-system genes. *Nature Communications*, *8*(1). https://doi.org/10.1038/ncomms15497

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8

Low, N., Dugas, E., Rodriguez, D., Contreras, G., O'Loughlin, E., Chaiton, M., & O'Loughlin, J. (2011). Stressful life events and relationships are associated with mental health symptoms and substance use in young adolescents. *Comprehensive Psychiatry*, *52*(6), e10. https://doi.org/10.1016/j.comppsych.2011.04.034

Lucero, M. M., Satz, S., Miceli, R., Swartz, H. A., & Manelis, A. (2022). The effects of mood disorders and childhood trauma on fear of positive and negative evaluation. *Acta Psychologica*, *227*(May), 103603. https://doi.org/10.1016/j.actpsy.2022.103603

Lutz, P. E., Chay, M. A., Pacis, A., Chen, G. G., Aouabed, Z., Maffioletti, E., Théroux, J. F., Grenier, J. C., Yang, J., Aguirre, M., Ernst, C., Redensek, A., van Kempen, L. C., Yalcin, I., Kwan, T., Mechawar, N., Pastinen, T., & Turecki, G. (2021). Non-CG methylation and multiple histone profiles associate child abuse with immune and small GTPase dysregulation. *Nature Communications*, *12*(1), 1–16. https://doi.org/10.1038/s41467-021-21365-3

Lutz, P. E., Tanti, A., Gasecka, A., Barnett-Burns, S., Kim, J. J., Zhou, Y., Chen, G. G., Wakid, M., Shaw, M., Almeida, D., Chay, M. A., Yang, J., Larivière, V., M'Boutchou, M. N., Van Kempen, L. C., Yerko, V., Prud'Homme, J., Davoli, M. A., Vaillancourt, K., … Turecki, G. (2017). Association of a history of child abuse with impaired myelination in the anterior cingulate cortex: Convergent epigenetic, transcriptional, and morphological evidence. *American Journal of Psychiatry*, *174*(12), 1185–1194. https://doi.org/10.1176/appi.ajp.2017.16111286

Mackenzie, C. S., Reynolds, K., Cairney, J., Streiner, D. L., & Sareen, J. (2011). Disorder-specific mental health service use for mood and anxiety disorders: Associations with age, sex, and psychiatric comorbidity. *Depression and Anxiety*, *29*(3), 234–242. https://doi.org/10.1002/da.20911

Maffioletti, E., Cattaneo, A., Rosso, G., Maina, G., Maj, C., Gennarelli, M., Tardito, D., & Bocchio-Chiavetto, L. (2016). Peripheral whole blood microRNA alterations in major depression and bipolar disorder. *Journal of Affective Disorders*, *200*, 250–258. https://doi.org/10.1016/j.jad.2016.04.021

Mahajan, G. J., Vallender, E. J., Garrett, M. R., Challagundla, L., Overholser, J. C., Jurjus, G., Dieter, L., Syed, M., Romero, D. G., Benghuzzi, H., & Stockmeier, C. A. (2018). Altered neuro-inflammatory gene expression in hippocampus in major depressive disorder. *Progress in neuro-psychopharmacology & biological psychiatry*, *82*, 177–186. https://doi.org/10.1016/j.pnpbp.2017.11.017

Malki, K, Koritskaya, E, Harris, F, Bryson, K, Herbster, M, & Tosto, M. G. (2016). Epigenetic differences in monozygotic twins discordant for major depressive disorder. *Translational psychiatry*, *6*(6), e839. https://doi.org/10.1038/tp.2016.101

Männik, K., Mägi, R., Macé, A., Cole, B., Guyatt, A. L., Shihab, H. A., Maillard, A. M., Alavere, H., Kolk, A., Reigo, A., Mihailov, E., Leitsalu, L., Ferreira, A. M., Nõukas, M., Teumer, A., Salvi, E., Cusi, D., McGue, M., Iacono, W. G., … Reymond, A. (2015). Copy number variations and cognitive phenotypes in unselected populations. *JAMA - Journal of the American Medical Association*, *313*(20), 2044–2054. https://doi.org/10.1001/jama.2015.4845

Marchese, S., Cancelmo, L., Diab, O., Cahn, L., Aaronson, C., Daskalakis, N. P., Schaffer, J., Horn, S. R., Johnson, J. S., Schechter, C., Desarnaud, F., Bierer, L. M., Makotkine, I., Flory, J. D., Crane, M., Moline, J. M., Udasin, I. G., Harrison, D. J., Roussos, P., … Feder, A. (2022). Altered gene expression and PTSD symptom dimensions in world trade center responders. *Molecular Psychiatry*, *27*(4), 2225–2246. https://doi.org/10.1038/s41380-022-01457-2

Mariani, N., Cattane, N., Pariante, C., & Cattaneo, A. (2021). Gene expression studies in depression development and treatment: An overview of the underlying molecular mechanisms and biological processes to identify biomarkers. *Translational Psychiatry*, *11*(1). https://doi.org/10.1038/s41398-021-01469-6

Marie-Claire, C., Lejeune, F. X., Mundwiller, E., Ulveling, D., Moszer, I., Bellivier, F., & Etain, B. (2020). A DNA methylation signature discriminates between excellent and non-response to lithium in patients with bipolar disorder type 1. *Scientific Reports*, *10*(1), 1–9. https://doi.org/10.1038/s41598-020-69073-0

McCabe, S. D., Lin, D. Y., & Love, M. I. (2019). Consistency and overfitting of multi-omics methods on experimental data. *Briefings in Bioinformatics*, *21*(4), 1277–1284. https://doi.org/10.1093/bib/bbz070

McEwen, B. S. (2008). Central effects of stress hormones in health and disease: Understanding the protective and damaging effects of stress and stress mediators. *European Journal of Pharmacology*, *583*(2-3), 174–185. https://doi.org/10.1016/j.ejphar.2007.11.071

McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R., & Cardno, A. (2003). The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Archives of General Psychiatry*, *60*(5), 497–502. https://doi.org/10.1001/archpsyc.60.5.497

McIntyre, R. S., Zimmerman, M., Goldberg, J. F., & First, M. B. (2019). Differential diagnosis of major depressive disorder versus bipolar disorder: Current status and best clinical practices. *Journal of Clinical Psychiatry*, *80*(3), 15–24. https://doi.org/10.4088/JCP.OT18043AH2

McKay, M. T., Cannon, M., Chambers, D., Conroy, R. M., Coughlan, H., Dodd, P., Healy, C., O'Donnell, L., & Clarke, M. C. (2021). Childhood trauma and adult mental disorder: A systematic review and meta-analysis of longitudinal cohort studies. *Acta Psychiatrica Scandinavica*, *143*(3), 189–205. https://doi.org/10.1111/acps.13268

McKenzie, A. T., Katsyv, I., Song, W.-M., Wang, M., & Zhang, B. (2016). DGCA: A comprehensive r package for differential gene correlation analysis. *BMC Systems Biology*, *10*(1). https://doi.org/10.1186/s12918-016-0349-1

Mechawar, N, & Savitz, J. (2016). Neuropathology of mood disorders: Do we see the stigmata of inflammation? *Translational Psychiatry*, *6*(11), e946–e946. https://doi.org/10.1038/tp.2016.212

Mehta, A., & Baltimore, D. (2016). MicroRNAs as regulatory elements in immune system logic. *Nature Reviews Immunology*, *16*(5), 279–294. https://doi.org/10.1038/nri.2016.40

Mehta, D., Klengel, T., Conneely, K. N., Smith, A. K., Altmann, A., Pace, T. W., Rex-Haffner, M., Loeschner, A., Gonik, M., Mercer, K. B., Bradley, B., Müller-Myhsok, B., Ressler, K. J., & Binder, E. B. (2013). Childhood maltreatment is associated with distinct genomic and epigenetic profiles in posttraumatic stress disorder. *Proceedings of the National Academy of Sciences*, *110*(20), 8302–8307. https://doi.org/10.1073/pnas.1217750110

Mellon, S. H., Wolkowitz, O. M., Schonemann, M. D., Epel, E. S., Rosser, R, Burke, H. B., Mahan, L, Reus, V. I., Stamatiou, D, Liew, C.-C., & Cole, S. W. (2016). Alterations in leukocyte transcriptional control pathway activity associated with major depressive disorder and antidepressant treatment. *Translational psychiatry*, *6*, e821. https://doi.org/10.1038/tp.2016.79

Mendes-Silva, A. P., Fujimura, P. T., Silva, J. R. d. C., Teixeira, A. L., Vieira, E. M., Guedes, P. H. G., Barroso, L. S. S., Nicolau, M. d. S., Ferreira, J. D. R., Bertola, L., Nicolau, E. d. S., Tolentino-Araújo, G. T., Berlezzi, C. M. S. F., Rodrigues, T. S., Borges, L. D. F., Gomes, M. d. S., do Amaral, L. R., Bonetti, A. M., Ueira-Vieira, C., & Diniz, B. S. (2019). Brain-enriched MicroRNA-184 is downregulated in older adults with major depressive disorder: A translational study. *Journal of Psychiatric Research*, *111*, 110–120. https://doi.org/10.1016/j.jpsychires.2019.01.019

Meng, C., Kuster, B., Culhane, A. C., & Gholami, A. M. (2014). A multivariate approach to the integration of multi-omics datasets. *BMC Bioinformatics*, *15*(1), 1–13. https://doi.org/10.1186/1471-2105-15-162

Meng, C., Zeleznik, O. A., Thallinger, G. G., Kuster, B., Gholami, A. M., & Culhane, A. C. (2016). Dimension reduction techniques for the integrative analysis of multi-omics data. *Briefings in Bioinformatics*, *17*(4), 628–641. https://doi.org/10.1093/bib/bbv108

Mental, G. B. D., & Collaborators, D. (2022). Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for

the Global Burden of Disease Study 2019. *The Lancet Psychiatry*, *9*(2), 137–150. https://doi.org/10.1016/S2215-0366(21)00395-3

Menyhárt, O., & Győrffy, B. (2021). Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. *Computational and Structural Biotechnology Journal*, *19*, 949–960. https://doi.org/10.1016/j.csbj.2021.01.009

Min, E. J., & Long, Q. (2020). Sparse multiple co-inertia analysis with application to integrative analysis of multi -omics data. *BMC Bioinformatics*, *21*(1). https://doi.org/10.1186/s12859-020-3455-4

min Fan, H., Sun, X. Y., Guo, W., Zhong, A. F., Niu, W., Zhao, L., Dai, Y. H., Guo, Z. M., Zhang, L. Y., & Lu, J. (2014). Differential expression of microRNA in peripheral blood mononuclear cells as specific biomarker for major depressive disorder patients. *Journal of Psychiatric Research*, *59*, 45–52. https://doi.org/10.1016/j.jpsychires.2014.08.007

Mineka, S., Watson, D., & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology*, *49*, 377–412. https://doi.org/10.1146/annurev.psych.49.1.377

Mistry, S., Harrison, J. R., Smith, D. J., Escott-Price, V., & Zammit, S. (2018). The use of polygenic risk scores to identify phenotypes associated with genetic risk of bipolar disorder and depression: A systematic review. *Journal of Affective Disorders*, *234*, 148–155. https://doi.org/10.1016/j.jad.2018.02.005

Mokhtari, A., Porte, B., Belzeaux, R., Etain, B., Ibrahim, E. C., Marie-Claire, C., Lutz, P. E., & Delahaye-Duriez, A. (2022). The molecular pathophysiology of mood disorders: From the analysis of single molecular layers to multi-omic integration. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *116*(October 2021). https://doi.org/10.1016/j.pnpbp.2022.110520

Mondelli, V, & Vernon, A. C. (2019). From early adversities to immune activation in psychiatric disorders: The role of the sympathetic nervous system. *Clinical and Experimental Immunology*, *197*(3), 319–328. https://doi.org/10.1111/cei.13351

Moraes, J. B., Maes, M., Roomruangwong, C., Bonifacio, K. L., Barbosa, D. S., Vargas, H. O., Anderson, G., Kubera, M., Carvalho, A. F., & Nunes, S. O. V. (2018). In major affective disorders, early life trauma predict increased nitro-oxidative stress, lipid peroxidation and protein oxidation and recurrence of major affective disorders, suicidal behaviors and a lowered quality of life. *Metabolic Brain Disease*, *33*(4), 1081–1096. https://doi.org/10.1007/s11011-018-0209-3

Moran, S., Arribas, C., & Esteller, M. (2016). Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*, *8*(3), 389–399. https://doi.org/10.2217/epi.15.114

Mostafavi, S., Battle, A., Zhu, X., Potash, J. B., ..., Weissman, M. M., Shi, J., Beckman, K., Haudenschild, C., Mccormick, C., Mei, R., Gameroff, M. J., Gindes, H., Adams, P., Goes, F. S., Mondimore, F. M., Mackinnon, D. F., Notes, L., Schweizer, B., … Levinson, D. F. (2014). Type I interferon signaling genes in recurrent major depression: increased expression detected by whole-blood RNA sequencing. *Molecular …*, *19*(12), 1267–1274. https://doi.org/10.1038/mp.2013.161

Nagy, C., Maitra, M., Tanti, A., Suderman, M., Théroux, J. F., Davoli, M. A., Perlman, K., Yerko, V., Wang, Y. C., Tripathy, S. J., Pavlidis, P., Mechawar, N., Ragoussis, J., & Turecki, G. (2020). Single-nucleus transcriptomics of the prefrontal cortex in major depressive disorder implicates oligodendrocyte precursor cells and excitatory neurons. *Nature Neuroscience*, *23*(6), 771–781. https://doi.org/10.1038/s41593-020-0621-y

Nestler, E. J., Peña, C. J., Kundakovic, M., Mitchell, A., & Akbarian, S. (2016). *Epigenetic Basis of Mental Illness* (Vol. 22). https://doi.org/10.1177/1073858415608147

Nie, P.-Y., Tong, L., Li, M.-D., Fu, C.-H., Peng, J.-B., & Ji, L.-L. (2021). miR-142 downregulation alleviates rat PTSD-like behaviors, reduces the level of inflammatory cytokine expression and apoptosis in hippocampus, and upregulates the expression of fragile x mental retardation protein. *Journal of Neuroinflammation*, *18*(1). https://doi.org/10.1186/s12974-020-02064-0

Niles, A. N., Smirnova, M., Lin, J., & O'Donovan, A. (2018). Gender differences in longitudinal relationships between depression and anxiety symptoms and inflammation in the health and retirement study. *Psychoneuroendocrinology*, *95*, 149–157. https://doi.org/10.1016/j.psyneuen.2018.05.035

Nygaard, V., Rødland, E. A., & Hovig, E. (2015). Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics*, *17*(1), 29–39. https://doi.org/10.1093/biostatistics/kxv027

O'Connell, K. S., & Coombes, B. J. (2021). Genetic contributions to bipolar disorder: Current status and future directions. *Psychological Medicine*, *51*(13), 2156–2167. https://doi.org/10.1017/S0033291721001252

O'Connor, D. B., Ferguson, E., Green, J. A., O'Carroll, R. E., & O'Connor, R. C. (2016). Cortisol levels and suicidal behavior: A meta-analysis. *Psychoneuroendocrinology*, *63*, 370–379. https://doi.org/10.1016/j.psyneuen.2015.10.011

Okazaki, S., Numata, S., Otsuka, I., Horai, T., Kinoshita, M., Sora, I., Ohmori, T., & Hishimoto, A. (2020). Decelerated epigenetic aging associated with mood stabilizers in the blood of patients with bipolar disorder. *Translational psychiatry*, *10*(1), 129. https://doi.org/10.1038/s41398-020-0813-y

Oliveira, J., Etain, B., Lajnef, M., Hamdani, N., Bennabi, M., Bengoufa, D., Sundaresh, A., Chaabane, A. B., Bellivier, F., Henry, C., Kahn, J.-P., Charron, D., Krishnamoorthy, R.,

Leboyer, M., & Tamouza, R. (2015). Combined effect of TLR2 gene polymorphism and early life stress on the age at onset of bipolar disorders (E. A. Eugenin, Ed.). *PLOS ONE*, *10*(3), e0119702. https://doi.org/10.1371/journal.pone.0119702

Oliveira, J., Hamdani, N., Busson, M., Etain, B., Bennabi, M., Amokrane, K., Boukouaci, W., Fortier, C., Marzais, F., Bengoufa, D., Bellivier, F., Henry, C., Kahn, J.-P., Charron, D., Krishnamoorthy, R., Corvoisier, P. L., Leboyer, M., & Tamouza, R. (2014). Association between toll-like receptor 2 gene diversity and early-onset bipolar disorder. *Journal of Affective Disorders*, *165*, 135–141. https://doi.org/10.1016/j.jad.2014.04.059

Ormel, J., Hartman, C. A., & Snieder, H. (2019). The genetics of depression: successful genome-wide association studies introduce new challenges. *Translational Psychiatry*, *9*(1). https://doi.org/10.1038/s41398-019-0450-5

Owen, M. J., & Williams, N. M. (2021). Explaining the missing heritability of psychiatric disorders. *World Psychiatry*, *20*(2), 294–295. https://doi.org/10.1002/wps.20870

Pace, T. W., Hu, F., & Miller, A. H. (2007). Cytokine-effects on glucocorticoid receptor function: Relevance to glucocorticoid resistance and the pathophysiology and treatment of major depression. *Brain, Behavior, and Immunity*, *21*(1), 9–19. https://doi.org/10.1016/j.bbi.2006.08.009

Pacifico, R., & Davis, R. L. (2017). Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder (2016/06/28). *Molecular Psychiatry*, *22*(3), 441–449. https://doi.org/10.1038/mp.2016.94

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Palmier-Claus, J. E., Berry, K., Bucci, S., Mansell, W., & Varese, F. (2016). Relationship between childhood adversity and bipolar affective disorder: Systematic review and meta-analysis. *British Journal of Psychiatry*, *209*(6), 454–459. https://doi.org/10.1192/bjp.bp.115.179655

Pandey, G. N. (2013). Biological basis of suicide and suicidal behavior. *Bipolar Disorders*, *15*(5), 524–541. https://doi.org/10.1111/bdi.12089

Parade, S. H., Huffhines, L., Daniels, T. E., Stroud, L. R., Nugent, N. R., & Tyrka, A. R. (2021). A systematic review of childhood maltreatment and dna methylation: Candidate gene and epigenome-wide approaches. *Translational psychiatry*, *11*(1), 1–33.

Parel, S. T., & Peña, C. J. (2022). Genome-wide signatures of early-life stress: Influence of sex. *Biological Psychiatry*, *91*(1), 36–42. https://doi.org/10.1016/j.biopsych.2020.12.010

Parker, K. J., Schatzberg, A. F., & Lyons, D. M. (2003). Neuroendocrine aspects of hypercortisolism in major depression. *Hormones and Behavior*, *43*(1), 60–66. https://doi.org/10.1016/s0018-506x(02)00016-8

Pearson, K. (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, *25*(3/4), 379. https://doi.org/10.2307/2332290

Peng, H., Zhu, Y., Strachan, E., Fowler, E., Bacus, T., Roy-Byrne, P., Goldberg, J., Vaccarino, V., & Zhao, J. (2018). Childhood Trauma, DNA Methylation of Stress-Related Genes, and Depression: Findings from Two Monozygotic Twin Studies. *Psychosomatic Medicine*, *80*(7), 599–608. https://doi.org/10.1097/PSY.0000000000000604

Penner-Goeke, S., & Binder, E. B. (2019). Epigenetics and depression. *Dialogues in Clinical Neuroscience*, *21*(4), 397–405. https://doi.org/10.31887/dcns.2019.21.4/ebinder

Perlick, D. A., Rosenheck, R. A., Miklowitz, D. J., Chessick, C., Wolff, N., Kaczynski, R., Ostacher, M., Patel, J., & and, R. D. (2007). Prevalence and correlates of burden among caregivers of patients with bipolar disorder enrolled in the systematic treatment enhancement program for bipolar disorder. *Bipolar Disorders*, *9*(3), 262–273. https://doi.org/10.1111/j.1399-5618.2007.00365.x

Pezzoli, P., Antfolk, J., Hatoum, A. S., & Santtila, P. (2019). Genetic vulnerability to experiencing child maltreatment. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00852

PGC-Consortium. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *The Lancet*, *381*(9875), 1371–1379. https://doi.org/10.1016/s0140-6736(12)62129-1

Picard, M., Scott-Boyer, M. P., Bodein, A., Périn, O., & Droit, A. (2021). Integration strategies of multi-omics data for machine learning analysis. *Computational and Structural Biotechnology Journal*, *19*, 3735–3746. https://doi.org/10.1016/j.csbj.2021.06.030

Pini, S., de Queiroz, V, Pagnin, D., Pezawas, L., Angst, J., Cassano, G. B., & Wittchen, H. U. (2005). Prevalence and burden of bipolar disorders in European countries. (Special issue: Size & Burden of Mental Disorders in Europe). *European Neuropsychopharmacology*, *15*, 425–434. http://www.qucosa.de/fileadmin/data/qucosa/documents/11019/796{_}PP.pdf{\%}0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS{\&}CSC=Y{\&}NEWS=N{\&}PAGE=fulltext{\&}D=cagh{\&}AN=20073004773http://lshtmsfx.hosted.exlibrisgroup.com/lshtm?sid=OVID:caghdb{\&}id=pmid:{\&}id=doi:10.1016{\%}2Fj.euron

Poirion, O. B., Jing, Z., Chaudhary, K., Huang, S., & Garmire, L. X. (2021). DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using

multi-omics data. *Genome Medicine*, *13*(1), 1–15. https://doi.org/10.1186/s13073-021-00930-x

Pompili, M., Gonda, X., Serafini, G., Innamorati, M., Sher, L., Amore, M., Rihmer, Z., & Girardi, P. (2013). Epidemiology of suicide in bipolar disorders: A systematic review of the literature. *Bipolar Disorders*, *15*(5), 457–490. https://doi.org/10.1111/bdi.12087

Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B., & Szemes, T. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical Journal*, *44*(5), 548–559. https://doi.org/10.1016/j.bj.2021.02.003

Price, E. M., & Robinson, W. P. (2018). Adjusting for batch effects in DNA methylation microarray data, a lesson learned. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/fgene.2018.00083

Prodromidou, K., & Matsas, R. (2019). Species-Specific miRNAs in Human Brain Development and Disease. *Frontiers in Cellular Neuroscience*, *13*(December), 1–12. https://doi.org/10.3389/fncel.2019.00559

Pruessner, M., King, S., Vracotas, N., Abadi, S., Iyer, S., Malla, A. K., Shah, J., & Joober, R. (2019). Gender differences in childhood trauma in first episode psychosis: Association with symptom severity over two years. *Schizophrenia Research*, *205*, 30–37. https://doi.org/10.1016/j.schres.2018.06.043

Pucher, B. M., Zeleznik, O. A., & Thallinger, G. G. (2019). Comparison and evaluation of integrative methods for the analysis of multilevel omics data: A study based on simulated and experimental cancer data. *Briefings in Bioinformatics*, *20*(2), 671–681. https://doi.org/10.1093/bib/bby027

Qin, Z., Wang, P.-Y., Su, D.-F., & Liu, X. (2016). miRNA-124 in immune system and immune disorders. *Frontiers in Immunology*, *7*. https://doi.org/10.3389/fimmu.2016.00406

Quidé, Y., Tozzi, L., Corcoran, M., Cannon, D. M., & Dauvermann, M. R. (2020). The impact of childhood trauma on developing bipolar disorder: Current understanding and ensuring continued progress. *Neuropsychiatric Disease and Treatment*, *16*, 3095–3115. https://doi.org/10.2147/NDT.S285540

Ramachandram, D., & Taylor, G. W. (2017). Deep Learning for VisuaL unDerstanDing Deep Multimodal Learning. *IEEE SIgnal ProcESSIng MagazInE*, (November).

Ramaker, R. C., Bowling, K. M., Lasseigne, B. N., Hagenauer, M. H., Hardigan, A. A., Davis, N. S., Gertz, J., Cartagena, P. M., Walsh, D. M., Vawter, M. P., Jones, E. G., Schatzberg, A. F., Barchas, J. D., Watson, S. J., Bunney, B. G., Akil, H., Bunney, W. E., Li, J. Z., Cooper, S. J., & Myers, R. M. (2017). Post-mortem molecular profiling of three psychiatric disorders. *Genome medicine*, *9*(1), 72. https://doi.org/10.1186/s13073-017-0458-5

Ramsey, J. M., Cooper, J. D., Bot, M., Guest, P. C., Lamers, F., Weickert, C. S., Penninx, B. W. J. H., & Bahn, S. (2016). Sex differences in serum markers of major depressive disorder in the netherlands study of depression and anxiety (NESDA) (K. Ebmeier, Ed.). *PLOS ONE*, *11*(5), e0156624. https://doi.org/10.1371/journal.pone.0156624

Rappoport, N., & Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Research*, *46*(20), 10546–10562. https://doi.org/10.1093/nar/gky889

Ray, S., Hossain, S. M. M., Khatun, L., & Mukhopadhyay, A. (2017). A comprehensive analysis on preservation patterns of gene co-expression networks during alzheimer's disease progression. *BMC Bioinformatics*, *18*(1). https://doi.org/10.1186/s12859-017-1946-8

Ressler, K. J., & Williams, L. M. (2021). Big data in psychiatry: multiomics, neuroimaging, computational modeling, and digital phenotyping. *Neuropsychopharmacology*, *46*(1), 1–2. https://doi.org/10.1038/s41386-020-00862-x

Richardson, S., Tseng, G. C., & Sun, W. (2016). Statistical Methods in Integrative Genomics. *Annual Review of Statistics and Its Application*, *3*(April), 181–209. https://doi.org/10.1146/annurev-statistics-041715-033506

Ripke, S., Neale, B. M., Corvin, A., Walters, J. T., Farh, K. H., Holmans, P. A., Lee, P., Bulik-Sullivan, B., Collier, D. A., Huang, H., Pers, T. H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S. A., Begemann, M., Belliveau, R. A., … O'Donovan, M. C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., Lin, D. Y., Duan, J., Ophoff, R. A., Andreassen, O. A., Scolnick, E., Cichon, S., St. Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D. H., Pato, C. N., … Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*(10), 969–978. https://doi.org/10.1038/ng.940

Roberson-Nay, R., Lapato, D. M., Wolen, A. R., Lancaster, E. E., Webb, B. T., Verhulst, B., Hettema, J. M., & York, T. P. (2020). An epigenome-wide association study of early-onset major depression in monozygotic twins. *Translational Psychiatry*, *10*(1). https://doi.org/10.1038/s41398-020-00984-2

Ronen, J., Hayat, S., & Akalin, A. (2019). Evaluation of colorectal cancer subtypes and cell lines using deep learning. *Life Science Alliance*, *2*(6), 1–16. https://doi.org/10.26508/lsa.201900517

Rowland, T. A., & Marwaha, S. (2018). Epidemiology and risk factors for bipolar disorder. *Therapeutic Advances in Psychopharmacology*, *8*(9), 251–269. https://doi.org/10.1177/2045125318769235

Roy, B., Dunbar, M., Shelton, R. C., & Dwivedi, Y. (2016). Identification of MicroRNA-124-3p as a putative epigenetic signature of major depressive disorder. *Neuropsychopharmacology*, *42*(4), 864–875. https://doi.org/10.1038/npp.2016.175

Roy, B., Shelton, R. C., & Dwivedi, Y. (2017). DNA methylation and expression of stress related genes in PBMC of MDD patients with and without serious suicidal ideation. *Journal of Psychiatric Research*, *89*, 115–124. https://doi.org/10.1016/j.jpsychires.2017.02.005

Roy, B., Yoshino, Y., Allen, L., Prall, K., Schell, G., & Dwivedi, Y. (2020). Exploiting Circulating MicroRNAs as Biomarkers in Psychiatric Disorders. *Molecular diagnosis & therapy*, *24*(3), 279–298. https://doi.org/10.1007/s40291-020-00464-9

Ruzicka, W. B., Subburaju, S., Coyle, J. T., & Benes, F. M. (2018). Location matters: distinct DNA methylation patterns in GABAergic interneuronal populations from separate microcircuits within the human hippocampus. *Human molecular genetics*, *27*(2), 254–265. https://doi.org/10.1093/hmg/ddx395

Salk, R. H., Hyde, J. S., & Abramson, L. Y. (2017). Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms. *Psychological Bulletin*, *143*(8), 783–822. https://doi.org/10.1037/bul0000102

Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., Abbafati, C., Adolph, C., Amlag, J. O., Aravkin, A. Y., Bang-Jensen, B. L., Bertolacci, G. J., Bloom, S. S., Castellano, R., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., … Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, *398*(10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7

Schaffer, A., McIntosh, D., Goldstein, B. I., Rector, N. A., McIntyre, R. S., Beaulieu, S., Swinson, R., & Yatham, L. N. (2012). The Canadian Network for Mood and Anxiety Treatments (CANMAT) task force recommendations for the management of patients with mood disorders and comorbid anxiety disorders. *Annals of Clinical Psychiatry*, *24*(1), 6–22.

Schiweck, C., Claes, S., Oudenhove, L. V., Lafit, G., Vaessen, T., de Beeck, G. O., Berghmans, R., Wijkhuijs, A., Müller, N., Arolt, V., Drexhage, H., & Vrieze, E. (2020a). Childhood trauma, suicide risk and inflammatory phenotypes of depression: Insights from monocyte gene expression. *Translational Psychiatry*, *10*(1). https://doi.org/10.1038/s41398-020-00979-z

Schiweck, C., Claes, S., Oudenhove, L. V., Lafit, G., Vaessen, T., de Beeck, G. O., Berghmans, R., Wijkhuijs, A., Müller, N., Arolt, V., Drexhage, H., & Vrieze, E. (2020b). Childhood trauma, suicide risk and inflammatory phenotypes of depression: Insights from mono-

cyte gene expression. *Translational Psychiatry*, *10*(1). https://doi.org/10.1038/s41398-020-00979-z

Seedat, S., Scott, K. M., Angermeyer, M. C., Bromet, E. J., Ph, D, Brugha, T. S., Girolamo, G. D., & Haro, J. M. (2009). Disorders in the WHO World Mental Health Surveys. *Archives of general psychiatry*, *66*(7), 785–795. https://doi.org/10.1001/archgenpsychiatry.2009.36.Cross-national

Seney, M. L., Glausier, J., & Sibille, E. (2022). Large-scale transcriptomics studies provide insight into sex differences in depression. *Biological Psychiatry*, *91*(1), 14–24. https://doi.org/10.1016/j.biopsych.2020.12.025

Seney, M. L., Huo, Z., Cahill, K., French, L., Puralewski, R., Zhang, J., Logan, R. W., Tseng, G., Lewis, D. A., & Sibille, E. (2018). Opposite Molecular Signatures of Depression in Men and Women. *Biological Psychiatry*, *84*(1), 18–27. https://doi.org/10.1016/j.biopsych.2018.01.017

Seney, M. L., & Sibille, E. (2014). Sex differences in mood disorders: Perspectives from humans and rodent models. *Biology of Sex Differences*, *5*(1), 1–10. https://doi.org/10.1186/s13293-014-0017-3

Shen, X., Caramaschi, D., Adams, M. J., Walker, R. M., Min, J. L., Kwong, A., Hemani, G., Barbu, M. C., Whalley, H. C., Harris, S. E., Deary, I. J., Morris, S. W., Cox, S. R., Relton, C. L., Marioni, R. E., Evans, K. L., & and, A. M. M. (2022). DNA methylome-wide association study of genetic risk for depression implicates antigen processing and immune responses. *Genome Medicine*, *14*(1). https://doi.org/10.1186/s13073-022-01039-5

Shimada, M., Otowa, T., Miyagawa, T., Umekage, T., Kawamura, Y., Bundo, M., Iwamoto, K., Ikegame, T., Tochigi, M., Kasai, K., Kaiya, H., Tanii, H., Okazaki, Y., Tokunaga, K., & Sasaki, T. (2018). An epigenome-wide methylation study of healthy individuals with or without depressive symptoms. *Journal of human genetics*, *63*(3), 319–326. https://doi.org/10.1038/s10038-017-0382-y

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., Strawbridge, R. J., Pers, T. H., Fischer, K., Justice, A. E., Workalemahu, T., Wu, J. M. W., Buchkovich, M. L., Heard-Costa, N. L., Roman, T. S., Drong, A. W., Song, C., Gustafsson, S., … and. (2015). New genetic loci link adipose and insulin biology to body fat distribution. *Nature*, *518*(7538), 187–196. https://doi.org/10.1038/nature14132

Simidjievski, N., Bodnar, C., Tariq, I., Scherer, P., Andres Terre, H., Shams, Z., Jamnik, M., & Liò, P. (2019). Variational Autoencoders for Cancer Data Integration: Design Principles and Computational Practice. *Frontiers in Genetics*, *10*(December), 1–14. https://doi.org/10.3389/fgene.2019.01205

Slavich, G. M., & Irwin, M. R. (2014). From stress to inflammation and major depressive disorder: A social signal transduction theory of depression. *Psychological Bulletin*, *140*(3), 774–815. https://doi.org/10.1037/a0035302

Smith, A. K., Kilaru, V., Klengel, T., Mercer, K. B., Bradley, B., Conneely, K. N., Ressler, K. J., & Binder, E. B. (2014). DNA extracted from saliva for methylation studies of psychiatric traits: Evidence tissue specificity and relatedness to brain. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *168*(1), 36–44. https://doi.org/10.1002/ajmg.b.32278

Song, M., Greenbaum, J., Luttrell, J., Zhou, W., Wu, C., Shen, H., Gong, P., Zhang, C., & Deng, H. W. (2020). A Review of Integrative Imputation for Multi-Omics Datasets. *Frontiers in Genetics*, *11*(October), 1–15. https://doi.org/10.3389/fgene.2020.570255

Sonino, N., & Fava, G. A. (2001). Psychiatric disorders associated with cushing??s syndrome. *CNS Drugs*, *15*(5), 361–373. https://doi.org/10.2165/00023210-200115050-00003

Squassina, A., Niola, P., Lopez, J. P., Cruceanu, C., Pisanu, C., Congiu, D., Severino, G., Ardau, R., Chillotti, C., Alda, M., Turecki, G., & Del Zompo, M. (2020). MicroRNA expression profiling of lymphoblasts from bipolar disorder patients who died by suicide, pathway analysis and integration with postmortem brain findings. *European Neuropsychopharmacology*, *34*, 39–49. https://doi.org/10.1016/j.euroneuro.2020.03.005

Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J. R. I., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J. M. W., Trzaskowski, M., Byrne, E. M., Pers, T. H., Holmans, P. A., Richards, A. L., … and. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics*, *51*(5), 793–803. https://doi.org/10.1038/s41588-019-0397-8

Stordal, E., Bjartveit Krüger, M., Dahl, N. H., Krüger, O., Mykletun, A., & Dahl, A. A. (2001). Depression in relation to age and gender in the general population: The Nord-Trøndelag health study (HUNT). *Acta Psychiatrica Scandinavica*, *104*(3), 210–216. https://doi.org/10.1034/j.1600-0447.2001.00130.x

Stouffer, S. (1949). *The american soldier: Combat and its aftermath*. John Wiley & Sons. https://books.google.fr/books?id=hUMjzQEACAAJ

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Sullivan, P. F. (2018). Psychiatric genomics: An update and an Agenda. *American Journal of Psychiatry*, *175*(1), 15–27. https://doi.org/10.1176/appi.ajp.2017.17030283

Sullivan, P. F., & Geschwind, D. H. (2019). Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*, *177*(1), 162–183. https://doi.org/10.1016/j.cell.2019.01.015

Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry*, *157*(10), 1552–1562. https://doi.org/10.1176/appi.ajp.157.10.1552

Sun, H., Kennedy, P. J., & Nestler, E. J. (2013). Epigenetics of the depressed brain: Role of histone acetylation and methylation. *Neuropsychopharmacology*, *38*(1), 124–137. https://doi.org/10.1038/npp.2012.73

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2020). The STRING database in 2021: Customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074

Tabano, S., Caldiroli, A., Terrasi, A., Colapietro, P., Grassi, S., Carnevali, G. S., Fontana, L., Serati, M., Vaira, V., Altamura, A. C., Miozzo, M., & Buoli, M. (2019). A miRNome analysis of drug-free manic psychotic bipolar patients versus healthy controls. *European archives of psychiatry and clinical neuroscience*, 10.1007/s00406–019–01057–2. https://doi.org/10.1007/s00406-019-01057-2

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*, *6*(5), 377–382. https://doi.org/10.1038/nmeth.1315

Tang, R., Howe, L. D., Suderman, M., Relton, C. L., Crawford, A. A., & Houtepen, L. C. (2020). Adverse childhood experiences, DNA methylation age acceleration, and cortisol in UK children: A prospective population-based cohort study. *Clinical Epigenetics*, *12*(1), 1–9. https://doi.org/10.1186/s13148-020-00844-2

Tenenhaus, A., Philippe, C., Guillemot, V., Le Cao, K. A., Grill, J., & Frouin, V. (2014). Variable selection for generalized canonical correlation analysis. *Biostatistics*, *15*(3), 569–583. https://doi.org/10.1093/biostatistics/kxu001

Tenenhaus, A., & Tenenhaus, M. (2011). Regularized Generalized Canonical Correlation Analysis. *Psychometrika*, *76*(2), 257–284. https://doi.org/10.1007/s11336-011-9206-8

Tenenhaus, M., Tenenhaus, A., & Groenen, P. J. (2017). Regularized Generalized Canonical Correlation Analysis: A Framework for Sequential Multiblock Component Methods. *Psychometrika*, *82*(3), 737–777. https://doi.org/10.1007/s11336-017-9573-x

Teroganova, N., Girshkin, L., Suter, C. M., & Green, M. J. (2016). DNA methylation in peripheral tissue of schizophrenia and bipolar disorder: a systematic review. *BMC genetics*, *17*, 27. https://doi.org/10.1186/s12863-016-0332-2

Tesson, B. M., Breitling, R., & Jansen, R. C. (2010). DiffCoEx: A simple and sensitive method to find differentially coexpressed gene modules. *BMC Bioinformatics*, *11*(1). https://doi.org/10.1186/1471-2105-11-497

Thombs, B. D., Bernstein, D. P., Lobbestael, J., & Arntz, A. (2009). A validation study of the Dutch Childhood Trauma Questionnaire-Short Form: Factor structure, reliability, and known-groups validity. *Child Abuse and Neglect*, *33*(8), 518–523. https://doi.org/10.1016/j.chiabu.2009.03.001

Thumfart, K. M., Jawaid, A., Bright, K., Flachsmann, M., & Mansuy, I. M. (2022). Epigenetics of childhood trauma: Long term sequelae and potential for treatment. *Neuroscience &amp Biobehavioral Reviews*, *132*, 1049–1066. https://doi.org/10.1016/j.neubiorev.2021.10.042

Thygesen, J. H., Presman, A., Harju-Seppänen, J., Irizar, H., Jones, R., Kuchenbaecker, K., Lin, K., Alizadeh, B. Z., Austin-Zimmerman, I., Bartels-Velthuis, A., Bhat, A., Bruggeman, R., Cahn, W., Calafato, S., Crespo-Facorro, B., de Haan, L., de Zwarte, S. M., Di Forti, M., Díez-Revuelta, Á., … Bramon, E. (2021). Genetic copy number variants, cognition and psychosis: a meta-analysis and a family study. *Molecular Psychiatry*, *26*(9), 5307–5319. https://doi.org/10.1038/s41380-020-0820-7

Tian, Y., Morris, T. J., Webster, A. P., Yang, Z., Beck, S., Feber, A., & Teschendorff, A. E. (2017). ChAMP: Updated methylation analysis pipeline for illumina BeadChips (A. Valencia, Ed.). *Bioinformatics*, *33*(24), 3982–3984. https://doi.org/10.1093/bioinformatics/btx513

Tini, G., Marchetti, L., Priami, C., & Scott-Boyer, M. P. (2018). Multi-omics integration-A comparison of unsupervised clustering methodologies. *Briefings in Bioinformatics*, *20*(4), 1269–1279. https://doi.org/10.1093/bib/bbx167

Tiwari, A., & Gonzalez, A. (2018). Biological alterations affecting risk of adult psychopathology following childhood trauma: A review of sex differences. *Clinical Psychology Review*, *66*, 69–79. https://doi.org/10.1016/j.cpr.2018.01.006

Torshizi, A. D., & Wang, K. (2017). Deconvolution of transcriptional networks in post-traumatic stress disorder uncovers master regulators driving innate immune system function. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-15221-y

Trerotola, M., Relli, V., Simeone, P., & Alberti, S. (2015). Epigenetic inheritance and the missing heritability. *Human Genomics*, *9*(1), 1–12. https://doi.org/10.1186/s40246-015-0041-3

Tylee, D. S., Kawaguchi, D. M., & Glatt, S. J. (2013). On the outside, looking in: A review and evaluation of the comparability of blood and brain "-omes". *American Journal of*

Medical Genetics Part B: Neuropsychiatric Genetics, *162*(7), 595–603. https://doi.org/10.1002/ajmg.b.32150

Tyler, S. R., Chun, Y., Ribeiro, V. M., Grishina, G., Grishin, A., Hoffman, G. E., Do, A. N., & Bunyavanich, S. (2021). Merged Affinity Network Association Clustering: Joint multi-omic/clinical clustering to identify disease endotypes. *Cell Reports*, *35*(2), 108975. https://doi.org/10.1016/j.celrep.2021.108975

Uchida, S., Yamagata, H., Seki, T., & Watanabe, Y. (2018). Epigenetic mechanisms of major depression: Targeting neuronal plasticity (2017/12/26). *Psychiatry and clinical neurosciences*, *72*(4), 212–227. https://doi.org/10.1111/pcn.12621

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., Martin, H. C., Lappalainen, T., & Posthuma, D. (2021). Genome-wide association studies. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-021-00056-9

Vaisvila, R., Ponnaluri, V. K., Sun, Z., Langhorst, B. W., Saleh, L., Guan, S., Dai, N., Campbell, M. A., Sexton, B. S., Marks, K., Samaranayake, M., Samuelson, J. C., Church, H. E., Tamanaha, E., Corrêa, I. R., Pradhan, S., Dimalanta, E. T., Evans, T. C., Williams, L., & Davis, T. B. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research*, *31*(7), 1280–1289. https://doi.org/10.1101/gr.266551.120

van den Berg, M., Krauskopf, J., Ramaekers, J., Kleinjans, J., Prickaerts, J., & Briedé, J. (2020). Circulating microRNAs as potential biomarkers for psychiatric and neurodegenerative disorders. *Progress in Neurobiology*, *185*, 101732. https://doi.org/10.1016/j.pneurobio.2019.101732

Van Dongen, J., Nivard, M. G., Willemsen, G., Hottenga, J. J., Helmer, Q., Dolan, C. V., Ehli, E. A., Davies, G. E., Van Iterson, M., Breeze, C. E., Beck, S., Suchiman, H. E., Jansen, R., Van Meurs, J. B., Heijmans, B. T., Slagboom, P. E., & Boomsma, D. I. (2016). Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nature Communications*, *7*, 1–13. https://doi.org/10.1038/ncomms11115

Walker, D. M., Cates, H. M., Heller, E. A., & Nestler, E. J. (2015). Regulation of chromatin states by drugs of abuse. *Current Opinion in Neurobiology*, *30*, 112–121. https://doi.org/10.1016/j.conb.2014.11.002

Walton, E., Hass, J., Liu, J., Roffman, J. L., Bernardoni, F., Roessner, V., Kirsch, M., Schackert, G., Calhoun, V., & Ehrlich, S. (2015). Correspondence of DNA methylation between blood and brain tissue and its application to schizophrenia research. *Schizophrenia Bulletin*, *42*(2), 406–414. https://doi.org/10.1093/schbul/sbv074

Wang, B., Mezlini, A. M., Demir, F., Fiume, M., Tu, Z., Brudno, M., Haibe-Kains, B., & Goldenberg, A. (2014). Similarity network fusion for aggregating data types on a genomic scale. *Nature Methods*, *11*(3), 333–337. https://doi.org/10.1038/nmeth.2810

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., Clarke, D., Gu, M., Emani, P., Yang, Y. T., Xu, M., Gandal, M. J., Lou, S., Zhang, J., Park, J. J., Yan, C., Rhie, S. K., Manakongtreecheep, K., Zhou, H., … and, A. A. (2018). Comprehensive functional genomic resource and integrative model for the human brain. *Science*, *362*(6420). https://doi.org/10.1126/science.aat8464

Wani, N., & Raza, K. (2019). Integrative approaches to reconstruct regulatory networks from multi-omics data: A review of state-of-the-art methods. *Computational Biology and Chemistry*, *83*(April 2018), 1–20. https://doi.org/10.1016/j.compbiolchem.2019.107120

Warrier, V., Kwong, A. S. F., Luo, M., Dalvie, S., Croft, J., Sallis, H. M., Baldwin, J., Munafò, M. R., Nievergelt, C. M., Grant, A. J., Burgess, S., Moore, T. M., Barzilay, R., McIntosh, A., van IJzendoorn, M. H., & Cecil, C. A. M. (2021). Gene–environment correlations and causal effects of childhood maltreatment on physical and mental health: A genetically informed approach. *The Lancet Psychiatry*, *8*(5), 373–386. https://doi.org/10.1016/s2215-0366(20)30569-1

Waterland, R. A., & Michels, K. B. (2007). Epigenetic epidemiology of the developmental origins hypothesis. *Annual Review of Nutrition*, *27*, 363–388. https://doi.org/10.1146/annurev.nutr.27.061406.093705

Watson, M. (2006). CoXpress: Differential co-expression in gene expression data. *BMC Bioinformatics*, *7*(1). https://doi.org/10.1186/1471-2105-7-509

Webb, L. M., Phillips, K. E., Ho, M. C., Veldic, M., & Blacker, C. J. (2020). The relationship between DNA methylation and antidepressant medications: A systematic review. *International Journal of Molecular Sciences*, *21*(3), 826. https://doi.org/10.3390/ijms21030826

Wissel, D., Rowson, D., & Boeva, V. (2022). Hierarchical autoencoder-based integration improves performance in multi-omics cancer survival models through soft modality selection.

Wittenberg, G. M., Greene, J., Vértes, P. E., Drevets, W. C., & Bullmore, E. T. (2020). Major Depressive Disorder Is Associated With Differential Expression of Innate Immune and Neutrophil-Related Gene Networks in Peripheral Blood: A Quantitative Review of Whole-Genome Transcriptional Data From Case-Control Studies. *Biological Psychiatry*, *88*(8), 625–637. https://doi.org/10.1016/j.biopsych.2020.05.006

Wörheide, M. A., Krumsiek, J., Kastenmüller, G., & Arnold, M. (2021). Multi-omics integration in biomedical research – A metabolomics-centric review. *Analytica Chimica Acta*, *1141*, 144–162. https://doi.org/10.1016/j.aca.2020.10.038

World Health Organization(WHO). (1993, November). *The ICD-10 classification of mental and behavioural disorders*. World Health Organization.

Wray, N. R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E. M., Abdellaoui, A., Adams, M. J., Agerbo, E., Air, T. M., Andlauer, T. M. F., Bacanu, S.-A., Bækvad-Hansen, M., Beek-

man, A. F. T., Bigdeli, T. B., Binder, E. B., Blackwood, D. R. H., Bryois, J., Buttenschøn, H. N., … and. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics*, *50*(5), 668–681. https://doi.org/10.1038/s41588-018-0090-3

Xing, X., Zhang, B., Li, D., & Wang, T. (2018). *Comprehensive whole DNA methylome analysis by integrating MeDIP-seq and MRE-seq* (Vol. 1708). https://doi.org/10.1007/978-1-4939-7481-8_12

Yan, J., Risacher, S. L., Shen, L., & Saykin, A. J. (2017). Network approaches to systems biology analysis of complex disease: Integrative methods for multi-omics data. *Briefings in Bioinformatics*, *19*(6), 1370–1381. https://doi.org/10.1093/bib/bbx066

Yeh, H., Duncan, B., Chmidt, M., Wang, N., & FL., B. (2010). HHS Public Access. *Ann Intern Med.*, *152*(1), 10–17. https://doi.org/10.1038/nature08822.The

Yoo, W., Mayberry, R., Bae, S., Singh, K., He, Q. P., & Lillard Jr, J. W. (2014). A study of effects of multicollinearity in the multivariable analysis. *International journal of applied science and technology*, *4*(5), 9.

Yoshino, Y., Roy, B., & Dwivedi, Y. (2020a). Altered miRNA landscape of the anterior cingulate cortex is associated with potential loss of key neuronal functions in depressed brain. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, *40*, 70–84. https://doi.org/10.1016/j.euroneuro.2020.06.004

Yoshino, Y., Roy, B., & Dwivedi, Y. (2020b). Differential and unique patterns of synaptic miRNA expression in dorsolateral prefrontal cortex of depressed subjects. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. https://doi.org/10.1038/s41386-020-00861-y

Yrondi, A., Fiori, L. M., Frey, B. N., Lam, R. W., MacQueen, G. M., Milev, R., Müller, D. J., Foster, J. A., Kennedy, S. H., & Turecki, G. (2020). Association Between Side Effects and Blood microRNA Expression Levels and Their Targeted Pathways in Patients With Major Depressive Disorder Treated by a Selective Serotonin Reuptake Inhibitor, Escitalopram: A CAN-BIND-1 Report. *The international journal of neuropsychopharmacology*, *23*(2), 88–95. https://doi.org/10.1093/ijnp/pyz066

Yuan, H., Mischoulon, D., Fava, M., & Otto, M. W. (2018). Circulating microRNAs as biomarkers for depression: Many candidates, few finalists. *Journal of Affective Disorders*, *233*, 68–78. https://doi.org/10.1016/j.jad.2017.06.058

Zeggini, E. (2009). Meta-analysis in genome-wide association studies R eview. *10*, 191–201.

Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*, *4*(1). https://doi.org/10.2202/1544-6115.1128

Zhang, Y., Qi, G., Park, J. H., & Chatterjee, N. (2018). Estimation of complex effect-size distributions using summary-level statistics from genome-wide association studies across 32 complex traits. *Nature Genetics*, *50*(9), 1318–1326. https://doi.org/10.1038/s41588-018-0193-x

Zhao, L., Yang, X., Cui, L., Wei, J., Ni, P., Li, M., Wang, Y., He, Y., Li, X., Liang, S., Tian, Y., Wang, Q., Cui, W., Lin, D., Ma, X., & Li, T. (2019). Increased expression of a novel miRNA in peripheral blood is negatively correlated with hippocampal volume in patients with major depressive disorder. *Journal of Affective Disorders*, *245*(1), 205–212. https://doi.org/10.1016/j.jad.2018.10.363

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, *9*(1). https://doi.org/10.1371/journal.pone.0078644

Zhao, Y., & Garcia, B. A. (2015). Comprehensive catalog of currently documented histone modifications. *Cold Spring Harbor Perspectives in Biology*, *7*(9), 1–20. https://doi.org/10.1101/cshperspect.a025064

Zhou, W., Laird, P. W., & Shen, H. (2016). Comprehensive characterization, annotation and innovative use of infinium DNA methylation BeadChip probes. *Nucleic Acids Research*, gkw967. https://doi.org/10.1093/nar/gkw967

Zindler, T., Frieling, H., Neyazi, A., Bleich, S., & Friedel, E. (2020). Simulating ComBat: How batch correction can lead to the systematic introduction of false positive results in DNA methylation microarray studies. *BMC Bioinformatics*, *21*(1). https://doi.org/10.1186/s12859-020-03559-6

Žitnik, M., & Zupan, B. (2015a). *Learning by fusing heterogeneous data: A dissertation*. M. Žitnik. https://books.google.fr/books?id=WSfLjwEACAAJ

Zitnik, M., Nguyen, F., Wang, B., Leskovec, J., Goldenberg, A., & Hoffman, M. M. (2019). Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities. *Information Fusion*, *50*, 71–91. https://doi.org/10.1016/j.inffus.2018.09.012

Žitnik, M., & Zupan, B. (2015b). Data fusion by matrix factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(1), 41–53. https://doi.org/10.1109/TPAMI.2014.2343973

Żurawek, D., & Turecki, G. (2021a). The miRNome of depression. *International Journal of Molecular Sciences*, *22*(21), 11312. https://doi.org/10.3390/ijms222111312

Żurawek, D., & Turecki, G. (2021b). The miRNome of depression. *International Journal of Molecular Sciences*, *22*(21), 11312. https://doi.org/10.3390/ijms222111312

Appendix

Supplementary table 01: Mood disorders associated genome wide association studies.

Study	Cohort	Disease	Ethnicity	Number of significant genome-wide identified risk loci	Nearest candidate genes to the GWAS SNP*
Charney et al. 2017	13 903 BD, 19 279 HC	BD	EU	8	*ADD3, ANK3, CACNA1C, ITIH1, LMAN2L, LOC100505933, RHEBL1/DHH, SYNE1, TRANK1, XPNPEP1*
Chen et al. 2013	5568 BD EUR, 7187 HC EU; 2000 BD EAS 1000, 1000 HC EAS	BD	EU and EA	5	*ANK3, LMAN2, PRM1M, PTGFR/IFI44L/ADGRL4L, TRANK1*
CONVERGE consortium. 2015	5303 MDD, 5337 HC	MDD	EA ♀	2	*LHPP, SIRT1*
Direk et al. 2016	9240 MDD, 9519 HC	MDD	EU	1	*FHIT*
Ferreira et al. 2008	4387BD, 6209 HC	BD	EU	1	*ANK3*
Hall et al. 2018	3852 MDD, 16 034 HC	MDD	EU	1	*CRTAP*
Hou et al. 2016	7647 BD, 27 303 HC	BD	EU	6	*DDN, ERBB2, MAD1L1, MIR2113, POU3F2, TRANK1*
Howard et al. 2018	8276 MDD, 209 308 HC	MDD	EU	1	*MAD1L1*
Howard et al. 2019	246 363 MDD, 561 190 HC	MDD	EU	101	*AREL1, ASTN2, ATP2A2, B3GLCT, BAG5, BEND4, C3orf84, CACNA1E, CELF4, CTTNBP2, CYP7B1, DAGLA, DCC, DCDC1, DENND1A, DENND1B, ELAVL2, ELAVL4, EMILIN3, EP300, ESR2, FAM172A, FHIT, GRM5, HIST1H1B, HIST1H2AL, HIST1H3I, HTT, KIAA1109, KLHDC8B, LINC01876, LOC102724552, LRFN5, LTBP3, MAD1L1, METTL9, MIR1281, NRDC, PAX5, PCDH9, PCLO, POU5F2, RBFOX1, RERE, RFWD2, RNF103/CHMP3, RSRC1, RTN1, SEMA6D, SGIP1, SHANK2, SHISA9, SLC12A5, SORCS3, SOX5, SPPL3, STK24, SYNE2, TCF4, TLR4, TMEM106B, TMEM161B, ZNF536*
Hyde et al. 2016	130 620 MDD, 347 620 HC	MDD	EU	15	*HACE1/LIN28B, KIAA0020/RFX3, L3MBTL2, MEIS2/TMCO5A, NEGR, NEGR1, OLFM4, PAX5, RERE, RSRC1/MLF1, SLC6A15, SORCS3, TMEM161/MEF2C, TMEM161B/MEF2C, VRK2*
Ikeda et al. 2018	2964 EAS BD, 61 887 EAS HC; 7481 EU BD, 9250 EU HC	BD	EA and EU	6	*C11orf10, C11orf9, DAGLA, FADS1, FADS2, FEN1, MAD1L1, NFIX, ODZ4, TRANK1*
Li et al. 2018	90 150 MDD, 246 603 HC	MDD	EU and EA ♀	10	*LINC00461, NEGR1, PCDH8P1, RSRC1, SPPL3*
Mühleisen et al. 2014	9747 BD, 14 278 HC	BD	EU	5	*ADCY2, ANK3, MIR2113, ODZ4, POU3F2, TRANK1*
Okbay et al. 2016	16 471 MDD,58 835 HC	MDD	EU	2	*DCC, KSR2*
PGC BD Group. 2011	7481 BD, 9250 HC	BD	EU	2	*CANCNA1C, ODZ4*
Power et al. 2017	4000 MDD,9519 HC	MDD	EU	1	*EHHADH-AS1*
Stahl et al. 2019	20 352 EU BD, 31 358 HC	BD	EU	31	*ACOT12, ADCY2, ADD3, ALPK3, ANK3, ANP32E, APH1A, ASB16, ASB16-AS1, ATXN7L3, BANF1, BRAF, BRMS1, C11orf80, C17orf53, C1orf51, C1orf54, CA14, CACNA1C, CATSPER1, CD248, CD47, CNIH2, CSRNP3, CST6, DNM1P41, EIF1AD, FADS1, FSTL5, G6PC3, GAL3ST3, GOLGA6L4, GOLGA6L5, GRIN2A, HAUS2, HDAC5, IFT57, ITIH1, ITIH3, KCNS1, KCNS1, KLC2, LMAN2L, LOC100506874, LOC388152, LOC440300, LOC642423, LRFN4, LRRC57, LSM12, MRPS33, NCAN, NMB, OTUD7B, PABPC1L, PACS1, PC, PI3, PLEKHO1, POU3F2, PUS7, RAB1B, RCE1, RIN1, RNU5D-1, RNU5E-1, RPS6KA2, SART1, SCAND2P, SCN2A, SEC11A, SF3B2, SHANK2, SHANK2-AS1, SLC28A1, SLC4A1, SNAP23, SPTBN2, SRPK2, SSBP2, STARD9, STK4, STK4-AS1, THSD7A, TMEM151A, TMEM178B, TMUB2, TOMM34, TRANK1, UBE2Q2P1, UBTF, VPS45, WDR73, WFDC12, WFDC5, YIF1A, ZCCHC2, ZNF592, ZSCAN2*
Wray et al. 2018	135 458 MDD, 344 901 HC	MDD	EU	44	*APOPT1, ASTN2, BAG5, C6orf168, CCDC122, CCDC68, CHADL, CPPED1, CRYBA1, DCAF4L1, DCC, DENND1A, DHX38, DKFZp686K1684, DLST, ELP4, ENOX1, EP300-AS1, ESR2, FBXL4, GPD2, L3MBTL2, LACC1, LHX2, LINC00461, LINC00682, LINC01065, LINC01231, LINC01360, LINC01876, LOC101927421, LRFN5, MEF2C, MIR4528, MIR4529, MIR548H1, MIR924HG, MYO18A, NEGR1, NR4A2, NUFIP2, OLFM4, PAUPAR, PAX6, PMFBP1, PROX2, PUM3, RAB27B, RBFOX1, RERE, RPS6KL1, SHISA9, SLC30A9, SLC45A1, SORCS3, SOX5, SYNE2, TCAIM, TCF4, TENM2, TOPAZ1, VRK2, VWDE, ZNF445*
Coleman et al. 2020	185,285 MDD and BD, 439,741 HC	MDD and BP	EU	73	*AL513166.1, LOC105378797, COP1, AL136114.1, LINC01645, CNNM2, SORCS3, SOX6, PAUPAR, AL035078.4, PAX6-AS1, CCDC73, TMEM258, MYRF, FEN1, MIR611, FADS2, MARK2, GRM5, CNTN5, CACNA1C, SOX5, AC069234.1, KDM2B, ENOX1, OLFM4, LRFN5, AL121821.2, LOC105370469, SYNE2, ESR2, AREL1, AC007956.2, BAG5, RNU7-160P, AL139300.2, AC021351.1, RBFOX1, AC005774.2, SHISA9, LOC105371092, PMFBP1, PIPOX, LOC107985157, MIR924HG, SETBP1, SETBP1-DT, LOC101927943, DCC, RAB27B, TCF4, ZNF536, AC096570.1, ACTG1P22, LINC01876, PLCL1, ZMYND8, L3MBTL2, CHADL, L3MBTL2-AS1, NFAM1, HDAC10, TUBGCP6, MAPK12, MYH15, RSRC1, SLC30A9, SCAMP1, MEF2C-AS2, FAM172A, AC106818.2, AC099520.1, LINC02240, AC109466.1, TENM2, AL133338.2, LOC107984041, LIN28B-AS1, TMEM106B, AC003093.1, LOC107986779, PCLO, AL355672.1, LOC105375974, ZCCHC7, AL512604.1, ASTN2, LOC107987122, DENND1A, AL513166.2*
PGC Cross disorder Group , 2013	33,332 cases (among which 6990 BP and 9227 MDD) 27,842 HC	ASD, ADHD, BP, MDD, SCZ	EU	4	*ITIH3, CACNB2, AS3MT, CACNA1C*

Legends: BD, bipolar disorder; EA, east asia; EU, europe; HC, healthy controls; MDD, major depressive disorder; ♀, female individuals only; * SNPs to gene assignement was based on the nearest gene if available retrived from Ensembl Variant Effect Predictor Tool (VEP: http://www.ensembl.org/Multi/Tools/VEP?db=core)

Supplementary 6.1: Mood disorders associated genome wide association studies.

Studies	Total in common	miRNAs*
Amoah et al. 2019; Fan et al. 2014 ; Maffioletti et al. 2016	2	miR-1291 miR-4521
Amoah et al. 2019; Lim et al. 2016 ; Yrondi et al. 2020	2	miR-17-5p miR-106a-5p
Amoah et al. 2019 ; Tabano et al. 2019 ; Yrondi et al. 2020	1	miR-15b-5p
Amoah et al. 2019; Yoshino et al. a 2020 ; Yoshino et al. b 2020	1	miR-487a-3p
Amoah et al. 2019; Gururajan et al. 2016; Kim et al. 2019	1	let-7c-5p
Fan et al. 2014 ; Lim et al. 2016 ; Maffioletti et al. 2016	1	miR-146b-5p
Fan et al. 2014; Maffioletti et al. 2016; Yoshino et al. a 2020	1	miR-4498
Tabano et al. 2019; Yoshino et al. a 2020; Yoshino et al. b 2020	1	miR-376a-3p
Amoah et al. 2019; Tabano et al. 2019	8	miR-23a-3p miR-199a-3p miR-196a-5p let-7d-5p miR-16-5p miR-199b-3p miR-1290 miR-498
Amoah et al. 2019; Yoshino et al. a 2020	4	miR-671-3p miR-548ar-3p miR-543 miR-154-5p
Amoah et al. 2019; Yoshino et al. b 2020	2	miR-193a-5p miR-223-3p
Amoah et al. 2019; Gururajan et al. 2016	1	let-7b-5p
Fan et al. 2014; Maffioletti et al. 2016	21	miR-4485 miR-4640-5p miR-26b miR-3679-5p miR-3147 miR-4298 miR-143 miR-29b miR-92b miR-874 miR-933 miR-4743 miR-1972 miR-1244 miR-4448 miR-4669 miR-4443 miR-210 miR-338-5p miR-4327 miR-4776-5p
Lim et al. 2016; Yoshino et al. a 2020	1	miR-92b-5p
Lim et al. 2016; Yrondi et al. 2020	1	miR-20a-5p
Lim et al. 2016; Zhao et al. 2019	1	miR-148b-3p
Tabano et al. 2019; Yoshino et al. a 2020	2	miR-146a-5p miR-520f-3p
Kim et al. 2019; Tabano et al. 2019	1	miR-21-5p
Yoshino et al. a 2020; Yoshino et al. b 2020	1	miR-376c-3p
Amoah et al. 2019	58	miR-378b miR-361-5p miR-1205 miR-4454 miR-211-5p miR-892a miR-483-5p miR-374c-5p miR-411-5p miR-28-5p miR-155-5p miR-198 miR-3161 miR-532-5p miR-3605-3p miR-3136-5p miR-199a-5p miR-1268a miR-204-5p miR-874-3p miR-513c-5p miR-630 miR-152-3p miR-608 miR-132-3p miR-365b-5p miR-143-3p miR-374b-5p miR-592 miR-28-3p miR-936 miR-454-3p miR-548b-3p miR-1185-1-3p miR-203a-3p miR-361-3p miR-924 miR-3613-5p miR-193b-3p miR-10a-5p miR-4755-5p miR-493-3p miR-125a-5p miR-4461 miR-320a miR-7975 miR-624-3p miR-627-5p miR-580-3p miR-326 miR-522-3p miR-1908-5p miR-193b-5p miR-874-5p miR-1245b-3p miR-4284 miR-561-5p miR-181b-2-3p
Lim et al. 2016	13	miR-1343-5p miR-210-3p miR-664b-5p miR-19b-3p miR-106b-5p miR-145-5p miR-339-5p miR-17-3p miR-18a-5p miR-27a-3p miR-15a-5p miR-6778-5p miR-30b-5p
Mendes-Silva et al. 2019	2	miR-1-3p miR-184
Squassina et al. 2020	4	miR-423-5p miR-186-5p miR-4286 miR-337-3p
Tabano et al. 2019	22	miR-548ah-5p let-7f-5p miR-1297 miR-640 let-7a-5p miR-549a miR-644a miR-379-5p miR-4536-5p miR-107 miR-128-1-5p miR-520h miR-587 let-7g-5p miR-212-3p miR-98-5p miR-553 miR-1253 miR-548q miR-219b-3p miR-26b-5p
Yoshino et al. a 2020	164	mir-3648-1 miR-3155a miR-1234-3p miR-4318 miR-6737-5p miR-153-5p miR-3689a-5p miR-6808-3p miR-521-1 miR-7-1-3p miR-342-5p miR-4638-5p miR-497-5p miR-6753-5p miR-518e-5p miR-6814-3p miR-376a-5p miR-4632-3p miR-135b-5p miR-22-3p let-7c-3p miR-516a-3p miR-30d-5p miR-136-5p miR-2861 miR-4466 miR-4736 miR-4695-5p miR-582-3p miR-6810-3p let-7e-5p miR-6805-3p miR-3648 miR-125b-2-3p miR-6821-5p miR-5582-3p miR-4430 miR-335-5p miR-4512 miR-486-3p miR-6826-3p miR-2277-5p miR-4538 miR-519b-3p miR-6850-5p miR-4281 miR-4796-3p miR-6857-5p miR-6869-5p miR-6851-5p miR-6802-5p miR-496 miR-1227-5p miR-670-3p miR-4497 miR-548ac miR-6853-3p miR-572 miR-544a miR-636 miR-484 miR-4258 miR-337-5p miR-4746-3p miR-4720-3p miR-20b-3p miR-1180-5p miR-328-3p miR-6075 miR-6082 miR-488-3p miR-4685-3p miR-302c-5p mir-3648-2 miR-6510-5p miR-4522 miR-4510 miR-7156-3p miR-29b-2-5p miR-495-5p miR-6511b-5p miR-6892-5p miR-4758-5p miR-4477b miR-551b-3p miR-519d-3p miR-367-3p miR-3189-3p miR-4705 miR-760 miR-4505 miR-153-3p miR-3609 miR-6844 miR-4728-3p miR-4485-5p miR-6769a-5p miR-4530 miR-378c miR-9-3p miR-29b-3p miR-4783-5p miR-1197 miR-595 miR-1193 miR-3928-5p miR-3591-5p miR-6131 miR-548a-5p miR-6800-3p miR-6077 miR-4426 miR-6851-3p miR-550b-2-5p miR-668-5p miR-710B-3p miR-1251-5p miR-6890-3p miR-6769b-5p miR-3621 miR-92b-3p miR-4761-3p miR-5189-5p miR-377-3p miR-523-5p miR-5572 miR-1343-3p miR-521-2 miR-758-3p miR-5590-5p miR-3689f miR-494-3p miR-376c-5p miR-521 miR-6789-3p miR-4433a-3p miR-6889-3p miR-6794-5p miR-6879-3p miR-138-5p miR-663b miR-1273h-3p miR-7112-5p miR-6084 miR-519e-5p miR-3182 miR-1470 miR-622 miR-495-3p miR-4663 miR-128-2-5p miR-548a-3p miR-22-5p miR-1-5p miR-1247-5p miR-648 miR-1307-5p miR-6781-5p miR-5580-5p miR-5587-5p miR-6815-3p miR-3651 miR-501-3p miR-4701-3p
Yoshino et al. b 2020	13	miR-32-5p miR-205-5p miR-431-5p miR-455-3p miR-372-3p miR-629-5p miR-324-5p miR-181c-5p miR-33a-5p miR-374a-3p miR-136-3p miR-217-5p miR-30e-5p
Yrondi et al. 2020	8	miR1301-3p miR-20b-5p miR-103b miR-93-5p miR-103a-3p miR-185-5p miR-660-5p miR-182-5p
Zhao et al. 2019	9	miR-1275 miR-3158-3p pmiR-chr11 miR-3611 miR-4532 miR-433 miR-4516 miR-30e-3p miR-3944-5p
Kim et al. 2019	1	miR-144-3p

(*) miRNAs, micro RNAs identified as differentially expressed or altered were overlapped across the miRNA publication corpus.

Supplementary 6.2: Common dysregulated miRNAs across the miRNA studeis

CHILD TRAUMA QUESTIONAIRE (CTQ) – SHORT FORM

These questions ask about some of your experiences growing up as a child and a teenager. For each question, circle (or select in any other way if completing online) the number that best describes how you feel. Although some of these questions are of a personal nature, please try to answer as honestly as you can. Your answers will be kept confidential.

Q	QUESTION	NEVER TRUE	RARELY TRUE	SOMETIMES TRUE	OFTEN TRUE	VERY OFTEN TRUE
When I was growing up						
1	I didn't have enough to eat.	1	2	3	4	5
2	I knew that there was someone to take care of me and protect me.	1	2	3	4	5
3	People in my family called me things like "stupid", "lazy", or "ugly".	1	2	3	4	5
4	My parents were too drunk or high to take care of the family.	1	2	3	4	5
5	There was someone in my family who helped me feel important or special	1	2	3	4	5
When I was growing up						
6	I had to wear dirty clothes	1	2	3	4	5
7	I felt loved.	1	2	3	4	5
8	I thought that my parents wished I had never been born	1	2	3	4	5
9	I got hit so hard by someone in my family that I had to see a doctor or go to the hospital.	1	2	3	4	5
10	There was nothing I wanted to change about my family.	1	2	3	4	5
When I was growing up						
11	People in my family hit me so hard that it left me with bruises or marks.	1	2	3	4	5
12	I was punished with a belt, a board, a cord (or some other hard object).	1	2	3	4	5
13	People in my family looked out for each other.	1	2	3	4	5
14	People in my family said hurtful or insulting things to me.	1	2	3	4	5
15	I believe that I was physically abused.	1	2	3	4	5
When I was growing up						
16	I had the perfect childhood.	1	2	3	4	5
17	I got hit or beaten so badly that it was noticed by someone like a teacher, neighbour, or doctor.	1	2	3	4	5
18	Someone in my family hated me.	1	2	3	4	5
19	People in my family felt close to each other.	1	2	3	4	5
20	Someone tried to touch me in a sexual way or tried to make me touch them.	1	2	3	4	5
When I was growing up						
21	Someone threatened to hurt me or tell lies about me unless I did something sexual with them.	1	2	3	4	5
22	I had the best family in the world.	1	2	3	4	5
23	Someone tried to make me do sexual things or watch sexual things.	1	2	3	4	5
24	Someone molested me (took advantage of me sexually).	1	2	3	4	5
25	I believe that I was emotionally abused.	1	2	3	4	5
When I was growing up						
26	There was someone to take me to the doctor if I needed it	1	2	3	4	5
27	I believe that I was sexually abused.	1	2	3	4	5
28	My family was a source of strength and support.	1	2	3	4	5

Copyright (1995) David P. Bernstein (Berstein et al., 2003)

Re-designed retaining exact same contact as original by Tor Alexander Bruce (2019) for use in academic studies

Supplementary 6.3: Short form of the Childhood trauma questionnaire (CTQ).

Hamilton Depression Rating Scale (HDRS)

Reference: **Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960; 23:56–62**

Rating Clinician-rated

Administration time 20–30 minutes

Main purpose To assess severity of, and change in, depressive symptoms

Population Adults

Commentary

The HDRS (also known as the Ham-D) is the most widely used clinician-administered depression assessment scale. The original version contains 17 items ($HDRS_{17}$) pertaining to symptoms of depression experienced over the past week. Although the scale was designed for completion after an unstructured clinical interview, there are now semi-structured interview guides available. The HDRS was originally developed for hospital inpatients, thus the emphasis on melancholic and physical symptoms of depression. A later 21-item version ($HDRS_{21}$) included 4 items intended to subtype the depression, but which are sometimes, incorrectly, used to rate severity. A limitation of the HDRS is that atypical symptoms of depression (e.g., hypersomnia, hyperphagia) are not assessed (see SIGH-SAD, page 55).

Scoring

Method for scoring varies by version. For the $HDRS_{17}$, a score of 0–7 is generally accepted to be within the normal range (or in clinical remission), while a score of 20 or higher (indicating at least moderate severity) is usually required for entry into a clinical trial.

Versions

The scale has been translated into a number of languages including French, German, Italian, Thai, and Turkish. As well, there is an Interactive Voice Response version (IVR), a Seasonal Affective Disorder version (SIGH-SAD, see page 55), and a Structured Interview Version (HDS-SIV). Numerous versions with varying lengths include the HDRS17, HDRS21, HDRS29, HDRS8, HDRS6, HDRS24, and HDRS7 (see page 30).

Additional references

Hamilton M. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol 1967; 6(4):278–96.

Williams JB. A structured interview guide for the Hamilton Depression Rating Scale. Arch Gen Psychiatry 1988; 45(8):742–7.

Address for correspondence

The HDRS is in the public domain.

Hamilton Depression Rating Scale (HDRS)

PLEASE COMPLETE THE SCALE BASED ON A STRUCTURED INTERVIEW

Instructions: for each item select the one "cue" which best characterizes the patient. Be sure to record the answers in the appropriate spaces (positions 0 through 4).

1 DEPRESSED MOOD *(sadness, hopeless, helpless, worthless)*
- 0 |__| Absent.
- 1 |__| These feeling states indicated only on questioning.
- 2 |__| These feeling states spontaneously reported verbally.
- 3 |__| Communicates feeling states non-verbally, i.e. through facial expression, posture, voice and tendency to weep.
- 4 |__| Patient reports virtually only these feeling states in his/her spontaneous verbal and non-verbal communication.

2 FEELINGS OF GUILT
- 0 |__| Absent.
- 1 |__| Self reproach, feels he/she has let people down.
- 2 |__| Ideas of guilt or rumination over past errors or sinful deeds.
- 3 |__| Present illness is a punishment. Delusions of guilt.
- 4 |__| Hears accusatory or denunciatory voices and/or experiences threatening visual hallucinations.

28

Supplementary 6.4: Hamilton Depression Rating Scale (HDRS) Form

3 SUICIDE
0 |__| Absent.
1 |__| Feels life is not worth living.
2 |__| Wishes he/she were dead or any thoughts of possible death to self.
3 |__| Ideas or gestures of suicide.
4 |__| Attempts at suicide (any serious attempt rate 4).

4 INSOMNIA: EARLY IN THE NIGHT
0 |__| No difficulty falling asleep.
1 |__| Complains of occasional difficulty falling asleep, i.e. more than ½ hour.
2 |__| Complains of nightly difficulty falling asleep.

5 INSOMNIA: MIDDLE OF THE NIGHT
0 |__| No difficulty.
1 |__| Patient complains of being restless and disturbed during the night.
2 |__| Waking during the night – any getting out of bed rates 2 (except for purposes of voiding).

6 INSOMNIA: EARLY HOURS OF THE MORNING
0 |__| No difficulty.
1 |__| Waking in early hours of the morning but goes back to sleep.
2 |__| Unable to fall asleep again if he/she gets out of bed.

7 WORK AND ACTIVITIES
0 |__| No difficulty.
1 |__| Thoughts and feelings of incapacity, fatigue or weakness related to activities, work or hobbies.
2 |__| Loss of interest in activity, hobbies or work – either directly reported by the patient or indirect in listlessness, indecision and vacillation (feels he/she has to push self to work or activities).
3 |__| Decrease in actual time spent in activities or decrease in productivity. Rate 3 if the patient does not spend at least three hours a day in activities (job or hobbies) excluding routine chores.
4 |__| Stopped working because of present illness. Rate 4 if patient engages in no activities except routine chores, or if patient fails to perform routine chores unassisted.

8 RETARDATION (slowness of thought and speech, impaired ability to concentrate, decreased motor activity)
0 |__| Normal speech and thought.
1 |__| Slight retardation during the interview.
2 |__| Obvious retardation during the interview.
3 |__| Interview difficult.
4 |__| Complete stupor.

9 AGITATION
0 |__| None.
1 |__| Fidgetiness.
2 |__| Playing with hands, hair, etc.
3 |__| Moving about, can't sit still.
4 |__| Hand wringing, nail biting, hair-pulling, biting of lips.

10 ANXIETY PSYCHIC
0 |__| No difficulty.
1 |__| Subjective tension and irritability.
2 |__| Worrying about minor matters.
3 |__| Apprehensive attitude apparent in face or speech.
4 |__| Fears expressed without questioning.

11 ANXIETY SOMATIC (physiological concomitants of anxiety) such as:
gastro-intestinal – dry mouth, wind, indigestion, diarrhea, cramps, belching
cardio-vascular – palpitations, headaches
respiratory – hyperventilation, sighing
urinary frequency
sweating
0 |__| Absent.
1 |__| Mild.
2 |__| Moderate.
3 |__| Severe.
4 |__| Incapacitating.

12 SOMATIC SYMPTOMS GASTRO-INTESTINAL
0 |__| None.
1 |__| Loss of appetite but eating without staff encouragement. Heavy feelings in abdomen.
2 |__| Difficulty eating without staff urging. Requests or requires laxatives or medication for bowels or medication for gastro-intestinal symptoms.

13 GENERAL SOMATIC SYMPTOMS
0 |__| None.
1 |__| Heaviness in limbs, back or head. Backaches, headaches, muscle aches. Loss of energy and fatigability.
2 |__| Any clear-cut symptom rates 2.

14 GENITAL SYMPTOMS (symptoms such as loss of libido, menstrual disturbances)
0 |__| Absent.
1 |__| Mild.
2 |__| Severe.

15 HYPOCHONDRIASIS
0 |__| Not present.
1 |__| Self-absorption (bodily).
2 |__| Preoccupation with health.
3 |__| Frequent complaints, requests for help, etc.
4 |__| Hypochondriacal delusions.

16 LOSS OF WEIGHT ***(RATE EITHER a OR b)***

a) According to the patient:	**b) According to weekly measurements:**
0 |__| No weight loss.	0 |__| Less than 1 lb weight loss in week.
1 |__| Probable weight loss associated with present illness.	1 |__| Greater than 1 lb weight loss in week.
2 |__| Definite (according to patient) weight loss.	2 |__| Greater than 2 lb weight loss in week.
3 |__| Not assessed.	3 |__| Not assessed.

17 INSIGHT
0 |__| Acknowledges being depressed and ill.
1 |__| Acknowledges illness but attributes cause to bad food, climate, overwork, virus, need for rest, etc.
2 |__| Denies being ill at all.

Total score: |__|__|

This scale is in the public domain.