PRIME

Prevention and Remediation of Insulin Multimorbidity in Europe

H2020 – 847879

D4.2. – Report on epigenetic analysis of insulinopathies

Dissemination level	Public
Contractual date of delivery	31. December 2022
Actual date of delivery	31. December 2022
Type	Report
Version	1
Filename	PRIME_Deliverable Report_D4.2.
Workpackage	4
Workpackage leader	Janita Bralten (RadboudUmc)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847879.

Author list

Organisation	Name	Contact information
Semmeweis University	Abel Fothi	fothi.abel@gmail.com
Semmeweis University	Andrea Vereczkei	vereczkei.andrea@med.semmelweis-univ.hu
Radboud university	Janita Bralten	Janita.Bralten@radboudumc.nl
Radboud university	Giuseppe Fanelli	Giuseppe.Fanelli@radboudumc.nl
Semmeweis University	Csaba Barta	barta.csaba@med.semmelweis-univ.hu

Abbreviations

ADHD	Attention-Deficit Hyperactivity Disorder
ALZ	Alzheimer's disease
ASD	Autism spectrum disorder
BIP	Bipolar disorder
CGI	CpG islands
GWAS	genome-wide association study
HAT	histone acetyl-transferase
HDAC	histone deacetylase
H-MAGMA	Hi-C-coupled MAGMA
MDD	Major depressive disorder
MetS	Metabolic syndrome
OCD	Obsessive-compulsive disorder
SCZ	Schizophrenia
T2DM	Type 2 diabetes mellitus
GO	Gene Ontology

(Please, see abbreviations of specific gene names after Table 2. at the description of their functions.)

Table of Contents

1. Executive Summary

In PRIME we aim to explore the multimorbidity between somatic and brain-based disorders. Earlier in the PRIME project it was shown that shared genetic factors exists between somatic (type 2 diabetes mellitus (T2DM), obesity and metabolic syndrome (MetS)) and brain-based disorders (Attention-Deficit/Hyperactivity Disorder (ADHD), Alzheimer's disease (ALZ), autism spectrum disorder (ASD), Bipolar disorder (BIP), obsessive-compulsive disorder (OCD), major depressive disorder (MDD) and schizophrenia (SCZ)). Next to genetic factors, environmental factors can also play a role. Cells can react to environmental changes by altering their gene expression state, this may be mediated by epigenetic processes. In WP4 we investigated epigenetic factors influencing the disorders studied within the PRIME project by performing gene-based analysis methods that incorporate epigenetic information, by using chromatin interaction profiles from human brain tissues. We performed these H-MAGMA analyses on the cross-disorder summary statistics of on the one hand somatic disorders (T2DM, obesity and MetS) and on the other hand brain-based disorders (ADHD, ALZ, ASD, BIP, OCD, MDD and SCZ). Significant gene-based associations were found for all pairs. Several of the overlapping genes between these analyses have relevant neurobiological functions and a number of them have been found to be associated with the studied somatic or psychiatric disorders in earlier studies. Subsequently, we performed a Gene Ontology analysis on a subset of our overlapping genes to assess the possible enrichment in certain biological processes. Our findings highlight that epigenetic, immune system, and mitochondrial organization processes related genes might play an important role in the multimorbidity between somatic and brain-based disorders. These results are in good concert with earlier findings within WP4 of the PRIME project.

2. Deliverable report

We performed epigenetics analyses on 21 pairs of 3 insulin-related somatic diseases and 7 neuropsychiatric disorders, using the H-MAGMA software (Sey et al., 2020) The analyses and results will be incorporated in a scientific manuscript together with the cross-disorder meta-analyses from Task 2 of WP4. This manuscript is in preparation. The results of the H-MAGMA analyses are described below.

Input data sets

Details of the analyzed genome-wide association studies (GWAS) are provided in the table below, including sample size (N), number of cases and controls and the derived effective sample size (Neff).

Table 1: Pairwise cross disorder H-MAGMA analyses between somatic diseases and brain-based disorders

Disorder	Author	Year	PMID	Consortium	N	Cases	Controls	Neff
MetS	Lind	2019	31589552		291107	59677	231430	189773
Obesity	Watanabe et al.	2019	31427789		244890	9805	235085	37650
T2DM	Mahajan et al.	2018	30297969	DIAGRAM	898130	74124	824006	272026
ADHD	Demontis et al.	2019	30478444	PGC	53293	19099	34194	49017
ALZ	Wightman et al.	2021	34493870	PGC	762917	86531	676386	306866
ASD	Grove et al.	2019	30804558	PGC	46350	18381	27969	44367
BIP	Mullins et al.	2021	34002096	PGC	413466	41917	371549	150670
OCD	OCGAS/IOCDF-GC	2018	28761083	OCGAS/IOCDF-GC	9725	2688	7037	7780
MDD	Wray et al./Howard et al.	2019	29700475/29662059	PGC	500199	170756	329443	449856
SCZ	Trubetskoy et al.	2022	35396580	PGC	130644	53386	77258	126282

We used a computational tool (H-MAGMA) to extract biological insights from the input GWAS summary statistics (Sey et al., 2020) Hi-C-coupled MAGMA (H-MAGMA), relies on epigenetic information. It can incorporate chromatin interaction profiles from human brain tissues. We used the following adult brain Hi-C data:

Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. Science 362, eaat8464 (2018).

During the analyses we performed cross-disorder gene-based analyses for all 21 pairs, filtering on:

(1) Genome-wide significance in the meta-analysis.

(2) Genes with at least nominal significance in both the input single trait analyses.

(3) Genes with at least one order of magnitude more significance in the meta-analysis than in both the single trait analyses.

57317 Ensembl genes (including protein coding and non-coding genes) were included in our analysis and we performed Bonferroni correction for the number of tests equaling the genes in the given meta-analysis. We used a computational tool (H-MAGMA) to extract biological insights from the input GWAS summary statistics (Sey et al., 2020) Hi-C-coupled MAGMA (H-MAGMA), relies on epigenetic information. It can incorporate chromatin interaction profiles from human brain tissues. We used the following adult brain Hi-C data: Wang, D. et al. Comprehensive functional genomic resource and integrative model for the human brain. Science 362, eaat8464 (2018). During the analyses we performed cross-disorder gene-based analyses for all 21 pairs, filtering on:

(1) Genome-wide significance in the meta-analysis.

(2) Genes with at least nominal significance in both the input single trait analyses.

(3) Genes with at least one order of magnitude more significance in the meta-analysis than in both the single trait analyses. 57317 Ensemble genes (including protein coding and non-coding genes) were included in our analysis and we performed Bonferroni correction for the number of tests equaling the genes in the given meta-analysis. The number of genes in a meta-analysis depends on how many genes could be paired with the SNPs from the given summary statistics.

The number of genes in a meta-analysis depends on how many genes could be paired with the SNPs from the given summary statistics.

Results

Significant gene-wide associations were found for all of the 21 cross-disorder pairs. The number of significant genes differs for different pairs of disorders however, several genes are significant in multiple pairs (see Table 1). Please find the details of part of the significant genes in Tables 3-17.

The most frequently overlapping genes

Genes are listed by the number of their occurrence in the somatic-neuropsychiatric disease analysis results. Their p-values and if they are linked to insulin (based on the Molecular signatures database (v.7.5.1.) (Liberzon et al., 2011) are also provide in the table below:

Table 2: Significant genes

Gene name	No of overlaps	METS_ ADHD	Obes_A DHD	T2DM_ ADHD	METS_ ALZ	Obes_ ALZ	T2DM_ ALZ	METS_ ASD	Obes_ ASD	T2DM_ ASD	METS _BIP	Obes_ BIP	T2DM_ BIP	METS _MDD	Obes_ MDD	T2DM_ MDD	METS _OCD	Obes_ OCD	T2DM_ OCD	METS _SCZ	Obes_ SCZ	T2DM_ SCZ	MSig Insulin
RHOA	14	2E-07	1E-08	2E-12				1E-09	5E-07	5E-11	1E-08		4E-09	8E-09	2E-09	1E-09				2E-13	1E-07	7E-23	X
IHO1	13	3E-12	8E-07	2E-12				3E-09	2E-07	1E-09			3E-07	4E-07	7E-09	2E-08				2E-08	2E-12	5E-12	
NDUFAF3	13	6E-17	3E-07	2E-07				9E-12	1E-07	7E-08			4E-10	2E-10	5E-08	1E-12				2E-12	9E-19	7E-13	
PKHD1	13	1E-15	8E-09	3E-08				4E-07	3E-08	7E-11	9E-09	3E-09	2E-11	4E-09						6E-20	9E-08	4E-28	
CCDC71	12	5E-19	1E-09	8E-07				3E-07		6E-13			1E-13	1E-11	1E-10	4E-17				1E-10	3E-19	2E-20	X
TMEM219	12				6E-15	1E-10	2E-31	5E-08		4E-09	3E-10	9E-08	1E-07					1E-07		8E-09	2E-07	2E-13	
AMT	11	2E-14		3E-11				2E-09		1E-08	7E-09		8E-08	5E-08		6E-09				2E-07	1E-17	1E-09	
ARIH2	11	3E-12	3E-08	3E-17							9E-14		6E-07	5E-07	3E-08	9E-07				9E-11	8E-08	3E-16	
NPIPB11	11				7E-13	5E-07	3E-20				1E-10	5E-08	1E-12				2E-14	5E-07		1E-10	4E-09	7E-09	
QRICH1	11	6E-12	5E-08	4E-08				9E-09		3E-08				9E-08	2E-10	3E-07				1E-10	7E-13	3E-11	
TAOK2	11				2E-16	5E-07	2E-19	1E-07		1E-08	4E-16	2E-09	9E-11							3E-12	1E-08	2E-33	X
BANK1	10	1E-07	3E-08	2E-12				5E-09		4E-08				8E-07		1E-09				2E-10	1E-08	8E-07	
FIBP	10				4E-35		3E-13				3E-19	5E-08	1E-11	5E-07	1E-09	4E-07				6E-07		7E-39	
RNF123	10	3E-18	7E-08	7E-15				1E-10		2E-12				3E-23		1E-09				1E-09	5E-12	1E-07	
TMEM161B	10	1E-08		6E-07				6E-07		2E-07	6E-14		8E-15	2E-07		2E-10				9E-10		1E-14	
CDIPT	9				1E-14	4E-07	1E-10				3E-09	8E-07	9E-21							7E-11	3E-10	3E-14	

GRM4	9	2E-11		1E-09			9E-16						1E-07		5E-09	5E-09				9E-10	7E-17	6E-08	
KCTD13	9	2E-11		3E-14	2E-10		1E-14				2E-08	1E-07	7E-08							9E-08	6E-11		
KCTD13-DT	9				8E-07		5E-14				8E-15	2E-09	2E-09					4E-07		2E-08	4E-11	1E-07	
KIF22	9				2E-12	1E-09	2E-13				6E-10	3E-07	3E-13							5E-13	3E-10	5E-18	
LAMB2	9	4E-07	5E-08	7E-07										5E-08	3E-07	4E-11				5E-10	3E-17	3E-09	
USP4	9	1E-07		2E-10							4E-07		9E-19	1E-10		7E-11				9E-12	4E-10	5E-07	
DOC2A	8				5E-15	7E-36	3E-39				4E-18		3E-08							1E-10	1E-09	4E-09	
IMPDH2	8	1E-17		2E-09				7E-11		6E-09				2E-07		7E-08				3E-19		2E-07	X
IP6K1	8	3E-09		2E-07							1E-08		9E-08	1E-09		3E-11				2E-12		2E-13	
ITPR3	8	2E-08		2E-08							2E-07		6E-08				7E-11			6E-07	2E-12	3E-15	X
LAMB2P1	8	1E-19		7E-11				4E-10		2E-15				7E-07		6E-08				1E-14		1E-20	
NUP160	8	7E-11		6E-11	4E-13		2E-08						1E-09			2E-09				3E-07		2E-10	
OR4C2P	8	4E-09		1E-07	3E-12		3E-11				2E-08		2E-07	1E-07		1E-10							
RN7SL309P	8										4E-10	2E-10	2E-07	3E-08		4E-09				2E-15	5E-08	3E-27	
SLC9B1	8	8E-07	4E-07	5E-08				2E-07		1E-14										2E-08	1E-11	5E-29	
TNXB	8				2E-08		6E-14				2E-08		1E-07	1E-07		6E-08			4E-22			2E-10	

The function and psychiatric/neurodevelopmental relevance of the most frequently overlapping genes

***RHOA* (Ras Homolog Family Member A)**

RHOA promotes reorganization of the actin cytoskeleton and regulate cell shape, attachment, and motility. Overexpression of this gene is associated with tumor cell proliferation and metastasis. RHOA along with Rho-kinase is indicated to have a functional role in mediating depression-like behaviors via dendritic remodeling of Nucleus Accumbens D1- medium spiny neurons and may prove a useful target for new depression therapeutics (Fox et al., 2020).

***IHO1* (Interactor Of HORMAD1)**

Predicted to be involved in gamete generation; meiosis I cell cycle process; and regulation of homologous chromosome segregation.

***NDUFAF3* (NADH:Ubiquinone Oxidoreductase Complex Assembly Factor 3)**

NDUFAF3 encodes a mitochondrial complex I assembly protein. Diseases associated with *NDUFAF3* include Mitochondrial Complex I Deficiency, Nuclear Type 18 and Leigh Syndrome with Cardiomyopathy.

***PKHD1* (PKHD1 Ciliary IPT Domain Containing Fibrocystin/Polyductin)**

Gene Ontology annotations related to this gene include signaling receptor activity. The protein encoded by this gene is associated with Polycystic Kidney Disease 4 with or without Polycystic Liver Disease and Polycystic Kidney Disease 4. Although primarily expressed in the kidney, the expression of PKHD1 in the liver and pancreas indicates a potential role in fat and glucose metabolism. A genome-wide association study for olanzapine treatment emergent weight gain indicated *PKHD1* among the most promising candidate genes (Muller & Kennedy, 2006).

***CCDC71* (Coiled-Coil Domain Containing 71)**

Predicted to be involved in cellular lipid metabolic process and positive regulation of fat cell differentiation. Diseases associated with *CCDC71* include Acute Endometritis.

***TMEM219* (Transmembrane Protein 219)**

Predicted to be involved in the apoptotic process. Using large genotyped cohorts, *TMEM219* within 16p11.2 was found to be associated with BMI (Vysotskiy et al., 2021).

***AMT* (Aminomethyltransferase)**

This gene encodes one of four critical components of the glycine cleavage system.

***ARIH2* (Ariadne RBR E3 Ubiquitin Protein Ligase 2)**

The protein encoded by this gene is an E3 ubiquitin-protein ligase that polyubiquitinates some proteins, tagging them for degradation. Among its related pathways are Class I MHC mediated antigen processing and presentation and Innate Immune System. SNP variant (c.338-6C>T) of *ARIH2* showed suggestive association with increased risk for inflammatory bowel disease (Prescott et al., 2015).

***NPIPB11* (Nuclear Pore Complex Interacting Protein Family Member B11)**

Predicted to act upstream of or within prevention of polyspermy. In a phenome-wide association study the *NPIPB11*-psychosis gene-trait pair was found to be significant (Vysotskiy et al., 2021).

***QRICH1* (Glutamine Rich 1)**

Enables DNA binding activity. Involved in several processes, including PERK-mediated unfolded protein response; intrinsic apoptotic signaling pathway in response to endoplasmic reticulum stress; and positive regulation of transcription.

***TAOK2* (TAO Kinase 2)**

Encodes a serine/threonine protein kinase that is involved in many different processes, including, cell signaling, microtubule organization and stability, and apoptosis. A critical role of TAOK2 in cortical development and its contribution to neurodevelopmental disorders, including ASD has been delineated (Scharrenberg et al., 2022). *TAOK2* within 16p11.2 was found to be associated with BMI (Vysotskiy et al., 2021). A noncoding single nucleotide polymorphism, rs4420550 ($p = 2.36 \times 10\text{-}9$ in an schizophrenia GWAS) in the 16p11.2 Schizophrenia-Associated Locus physically interacted with the *MAPK3* promoter and *TAOK2* promoter (Chang et al., 2021). Whole-genome and -exome sequencing of ASD families identified three de novo mutations in *TAOK2* and functional analysis in mice and human cells revealed that all the mutations impair protein stability. Data also provided evidence that *TAOK2* is a neurodevelopmental disorder risk gene and identify RhoA signaling as a mediator of TAOK2-dependent synaptic development (Richter et al., 2019).

***BANK1* (B Cell Scaffold Protein With Ankyrin Repeats 1)**

The protein encoded by this gene is a B-cell-specific scaffold protein that functions in B-cell receptor-induced calcium mobilization from intracellular stores. In a study aiming to evaluate the role of genetic variants associated with the gut microbiome in the susceptibility of individuals to four psychiatric disorders, *BANK1* was associated with SCZ in the gene-based association analysis (Martins-Silva et al., 2021).

***FIBP* (FGF1 Intracellular Binding Protein)**

The FIBP protein is an intracellular protein that binds selectively to acidic fibroblast growth factor. Diseases associated with *FIBP* include learning disability. In a study performing transcriptome and chromatin accessibility profiling in primary human microglia identified genetically driven variation and cell-specific enhancer-promoter interactions, putative regulatory mechanisms for 21 Alzheimer's disease risk loci were identified, of which 18 were refined to a single gene, including *FIBP* as a new candidate risk gene (Kosoy et al., 2022). A study utilizing an OCD-specific gene signature, which was identified using blood gene expression analysis to construct a predictive model of OCD found a six-gene panel including *FIBP*, which discriminated patients with OCD from healthy controls, MDD, and schizophrenia in a training set (Wang et al., 2018).

***RNF123* (Ring Finger Protein 123)**

The protein encoded by this gene contains a C-terminal RING finger domain, a motif present in a variety of functionally distinct proteins and known to be involved in protein-protein and protein-DNA interactions. A Transcriptome-Wide Association Study found RNF123 expression among the most significant genes in their anxiety study (Su et al., 2021). The expression level of the *RNF123* gene in blood cells has been identified as a disease risk marker, more than ten years before the diagnosis of depression (Glahn et al., 2012).

***TMEM161B* (Transmembrane Protein 161B)**

Predicted to enable nucleic acid binding activity. A gene-gene interaction study on MDD found *TMEM161B* (rs768705) positively associated with new onset MDD (Yue et al., 2022).

***IMPDH2* (Inosine Monophosphate Dehydrogenase 2)**

This gene encodes the rate-limiting enzyme in the de novo guanine nucleotide biosynthesis. Among its related pathways are Metabolism of nucleotides and Innate Immune System. A genome-wide Mendelian randomization study identified *IMPDH2* as an actionable novel drug target for depression (Liu et al., 2022).

***ITPR3* (Inositol 1,4,5-Trisphosphate Receptor Type 3)**

This gene encodes a receptor for inositol 1,4,5-trisphosphate, a second messenger that mediates the release of intracellular calcium. In a study on differential and spatial expression meta-analysis of genes identified in genome-wide association studies of depression, *ITPR3* was among the genes with the most consistent differential expression (Wu, Howard, Sibille, & French, 2021).

Somatic x Neuropsychiatric disorders

Table 3: T2DMxOCD Top results (of a total of 24 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
NKX6-3	NK6 homeobox 3	8	1,88E-29	
CRMP1	collapsin response mediator protein 1	4	9,36E-29	
TNXB	tenascin XB	6	2,50E-22	
FKBPL	FKBP prolyl isomerase like	6	3,80E-22	
BTNL2	butyrophilin like 2	6	2,43E-21	
MTNR1B	melatonin receptor 1B	11	3,59E-20	X
PABPC4	poly(A) binding protein cytoplasmic 4	1	9,29E-20	X
HLA-DRA	major histocompatibility complex, class II, DR alpha	6	9,95E-20	X
LY6G6C	lymphocyte antigen 6 family member G6C	6	3,58E-18	
TSBP1	testis expressed basic protein 1	6	1,98E-16	

Table 4: MetSxOCD Top results (of a total of 24 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
SLC12A3	solute carrier family 12 member 3	16	1,35E-22	X
HERPUD1	homocysteine inducible ER protein with ubiquitin like domain 1	16	4,53E-18	X
YPEL5P2	YPEL5 pseudogene 2	11	9,34E-16	
FNBP4	formin binding protein 4	11	1,82E-15	
PABPC4	poly(A) binding protein cytoplasmic 4	1	1,08E-14	X
TTBK2	tau tubulin kinase 2	15	1,52E-14	
FAM180B	family with sequence similarity 180 member B	11	1,94E-14	
C1QTNF4	C1q and TNF related 4	11	1,98E-14	

NDUFS3	NADH:ubiquinone oxidoreductase core subunit S3	11	8,30E-14	
CDAN1	codanin 1	15	1,43E-13	

Table 5: ObesxOCD Top 3

gene_name	Description	CHR	P_MULTI	MSig insulin
NPIPB11	nuclear pore complex interacting protein family member B11	16	1,19E-07	
TMEM219	transmembrane protein 219	16	3,94E-07	
KCTD13-DT	KCTD13 divergent transcript	16	4,76E-07	

Table 6: T2DMxADHD Top results (of a total of 284 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
RPS12P20	ribosomal protein S12 pseudogene 20	11	2,24E-21	
KRT18P60	keratin 18 pseudogene 60	12	6,74E-21	
SRP9P1	signal recognition particle 9 pseudogene 1	10	1,08E-20	
HMGA2	high mobility group AT-hook 2	12	6,44E-19	
MARCHF5	membrane associated ring-CH-type finger 5	10	7,50E-19	
HMG20A	high mobility group 20A	15	1,52E-18	
RNY3P12	RNY3 pseudogene 12	10	8,46E-18	
PKHD1	PKHD1 ciliary IPT domain containing fibrocystin/polyductin	6	2,89E-17	
SH3YL1	SH3 and SYLF domain containing 1	2	3,13E-17	
MARK2P9	microtubule affinity regulating kinase 2 pseudogene 9	10	3,86E-17	

Table 7: MetSxADHD Top results (of a total of 143 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
LEMD2	LEM domain nuclear envelope protein 2	6	1,20E-19	X
GRM4	glutamate metabotropic receptor 4	6	4,62E-19	
SH3YL1	SH3 and SYLF domain containing 1	2	2,01E-18	
RPL35P2	ribosomal protein L35 pseudogene 2	6	3,09E-18	
ARFGAP2	ADP ribosylation factor GTPase activating protein 2	11	1,17E-17	
NCK1-DT	NCK1 divergent transcript	3	2,43E-17	
PACSIN1	protein kinase C and casein kinase substrate in neurons 1	6	2,50E-17	
ATG13	autophagy related 13	11	5,65E-17	
TYW1B	tRNA-yW synthesizing protein 1 homolog B	7	8,85E-17	
AMBRA1	autophagy and beclin 1 regulator 1	11	1,36E-16	X

Table 8: ObesxADHD Top results (of a total of 19 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
RHOA	ras homolog family member A	3	1,32E-09	X
PKHD1	PKHD1 ciliary IPT domain containing fibrocystin/polyductin	6	8,47E-09	
IHO1	interactor of HORMAD1 1	3	1,10E-08	
SEMA3F	semaphorin 3F	3	2,53E-08	
SH3YL1	SH3 and SYLF domain containing 1	2	3,18E-08	
NAA80	N-alpha-acetyltransferase 80, NatH catalytic subunit	3	3,27E-08	
SLC9B1	solute carrier family 9 member B1	4	4,54E-08	
LAMB2	laminin subunit beta 2	3	4,88E-08	

HYAL3	hyaluronidase 3	3	5,93E-08	
CCDC71	coiled-coil domain containing 71	3	7,10E-08	X

Table 9: T2DMxASD Top results (of a total of 250 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
GPSM1	G protein signaling modulator 1	9	3,99E-21	
XKR6	XK related 6	8	1,17E-19	
CARD9	caspase recruitment domain family member 9	9	1,58E-19	
MSRA	methionine sulfoxide reductase A	8	1,57E-18	
MTMR9	myotubularin related protein 9	8	2,14E-18	
OPTN	Optineurin	10	3,69E-18	X
COL13A1	collagen type XIII alpha 1 chain	10	8,84E-18	
FAM167A-AS1	FAM167A antisense RNA 1	8	1,73E-17	
LINC02636	long intergenic non-protein coding RNA 2636	10	3,93E-17	
BPTF	bromodomain PHD finger transcription factor	17	5,54E-17	X

Table 10: MetSxASD Top results (of a total of 121 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
SPDYE7P	speedy/RINGO cell cycle regulator family member E7, pseudogene	7	1,54E-21	
JMJD1C	jumonji domain containing 1C	10	1,68E-18	
TYW1B	tRNA-yW synthesizing protein 1 homolog B	7	1,47E-16	
CDAN1	codanin 1	15	1,38E-13	
ERI1	exoribonuclease 1	8	3,62E-13	

gene_name	Description	CHR	P_MULTI	MSig insulin
MAP1A	microtubule associated protein 1A	15	4,78E-13	
BAK1	BCL2 antagonist/killer 1	6	5,62E-13	
BPTF	bromodomain PHD finger transcription factor	17	5,84E-13	X
TP53BP1	tumor protein p53 binding protein 1	15	6,23E-13	
NCOR2	nuclear receptor corepressor 2	12	3,74E-12	

Table 11: ObesxASD Top results (of a total of 25 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
SPPL2C	signal peptide peptidase like 2C	17	6,40E-09	
FMNL1	formin like 1	17	7,90E-09	
PLEKHM1	pleckstrin homology and RUN domain containing M1	17	1,09E-08	
LINC02210	long intergenic non-protein coding RNA 2210	17	3,41E-08	
ARL17A	ADP ribosylation factor like GTPase 17A	17	3,76E-08	
RPS26P8	ribosomal protein S26 pseudogene 8	17	4,23E-08	
LRRC37A4P	leucine rich repeat containing 37 member A4, pseudogene	17	5,05E-08	
ARHGAP27	Rho GTPase activating protein 27	17	5,47E-08	
ARL17B	ADP ribosylation factor like GTPase 17B	17	6,98E-08	
STH	Saitohin	17	8,82E-08	

Table 12: T2DMxALZ Top results (of a total of 311 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
APOE	apolipoprotein E	19	1,96E-48	
APOC1P1	apolipoprotein C1 pseudogene 1	19	4,65E-46	
NECTIN2	nectin cell adhesion molecule 2	19	2,17E-44	
TOMM40	translocase of outer mitochondrial membrane 40	19	1,16E-39	
EML2	EMAP like 2	19	2,55E-39	
SKIV2L	Ski2 like RNA helicase	6	4,03E-37	
STK19	serine/threonine kinase 19	6	7,84E-37	
CEACAM19	CEA cell adhesion molecule 19	19	1,88E-36	
MSH5	mutS homolog 5	6	1,22E-31	
TSBP1-AS1	TSBP1 and BTNL2 antisense RNA 1	6	1,32E-31	

Table 13: MetSxALZ Top results (of a total of 247 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
APOE	apolipoprotein E	19	5,84E-49	
APOC1P1	apolipoprotein C1 pseudogene 1	19	2,20E-46	
RSPH6A	radial spoke head 6 homolog A	19	3,65E-45	
TOMM40	translocase of outer mitochondrial membrane 40	19	4,82E-45	
NECTIN2	nectin cell adhesion molecule 2	19	4,49E-41	
CEACAM19	CEA cell adhesion molecule 19	19	9,61E-36	
APOC1	apolipoprotein C1	19	4,39E-35	

gene_name	Description	CHR	P_MULTI	MSig insulin
SPI1	Spi-1 proto-oncogene	11	7,22E-27	
DGKZ	diacylglycerol kinase zeta	11	2,82E-25	X
EML2	EMAP like 2	19	4,73E-25	

Table 14: ObesxALZ Top results (of a total of 72 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
APOE	apolipoprotein E	19	1,99E-38	
NECTIN2	nectin cell adhesion molecule 2	19	6,66E-36	
TOMM40	translocase of outer mitochondrial membrane 40	19	1,16E-30	
CEACAM19	CEA cell adhesion molecule 19	19	1,30E-21	
TRAPPC6A	trafficking protein particle complex subunit 6A	19	5,92E-19	
MARK4	microtubule affinity regulating kinase 4	19	2,38E-12	
ZNF652-AS1	ZNF652 antisense RNA 1	17	1,15E-10	
NPIPB11	nuclear pore complex interacting protein family member B	16	2,73E-10	
SCIMP	SLP adaptor and CSK interacting membrane protein	17	3,84E-10	
ZNF652	zinc finger protein 652	17	6,46E-10	

Table 15: T2DMxBIP Top results (of a total of 791 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
RBM26-AS1	RBM26 antisense RNA 1	13	1,90E-30	
NOTCH4	notch receptor 4	6	1,16E-27	

DHX16	DEAH-box helicase 16	6	2,97E-27	
ATAT1	alpha tubulin acetyltransferase 1	6	3,18E-27	X
MSRA	methionine sulfoxide reductase A	8	6,71E-26	
LINC01149	long intergenic non-protein coding RNA 1149	6	1,34E-25	
HAAO	3-hydroxyanthranilate 3,4-dioxygenase	2	9,54E-25	
NFKBIL1	NFKB inhibitor like 1	6	2,50E-24	
MDC1	mediator of DNA damage checkpoint 1	6	4,63E-24	
TNXB	tenascin XB	6	5,72E-24	

Table 16: T2DMxMDD Top results (of a total of 465 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
TCF4	transcription factor 4	18	1,83E-28	
ATAT1	alpha tubulin acetyltransferase 1	6	9,83E-27	X
TCF19	transcription factor 19	6	1,89E-26	
NFKBIL1	NFKB inhibitor like 1	6	3,17E-23	
TNXB	tenascin XB	6	1,10E-22	
LINC01149	long intergenic non-protein coding RNA 1149	6	1,31E-22	
HLA-B	major histocompatibility complex, class I, B	6	4,22E-22	X
RPL31P12	ribosomal protein L31 pseudogene 12	1	4,86E-21	
LINC02323	long intergenic non-protein coding RNA 2323	14	6,42E-21	
HCG27	HLA complex group 27	6	1,23E-20	

Table 17: T2DMxSCZ Top results (of a total of 1295 gene-wide significant genes)

gene_name	Description	CHR	P_MULTI	MSig insulin
NOTCH4	notch receptor 4	6	3,14E-44	
TSBP1-AS1	TSBP1 and BTNL2 antisense RNA 1	6	4,98E-42	
ATAT1	alpha tubulin acetyltransferase 1	6	2,39E-41	X
STK19	serine/threonine kinase 19	6	1,42E-40	
LINC01149	long intergenic non-protein coding RNA 1149	6	4,04E-40	
MSH5	mutS homolog 5	6	4,99E-40	
SKIV2L	Ski2 like RNA helicase	6	5,93E-39	
MDC1	mediator of DNA damage checkpoint 1	6	6,91E-39	
RBM26-AS1	RBM26 antisense RNA 1	13	1,23E-38	
TNXB	tenascin XB	6	1,72E-38	

Subsequently a Gene Ontology (GO) analysis was run on genes that were significant in at least 3 disorder pairs in our previous cross-disorder analyses (n=420 genes). As a background, we used all protein-coding genes in the human genome (N=19947) listed by the database. GO analysis was performed by GOrilla (http://cbl-gorilla.cs.technion.ac.il/).

The results oft he GO analysis showed that gene sets were enriched in 3 major domains of biological processes (epigenetics, immune system, and mitochondrial organization) and a few minor ones (e.g. presynaptic active zone assembly/presynaptic signal transduction, negative regulation of transcription by RNA polymerase II), see Table 16-18.

Within the co called „epigenetics" domain we found genes related to Covalent chromatin modifications, Histone modifications and Lysine acetylation, as well as miscellaneous Protein modifications.

Table 18: Epigenetic GO processes

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b

GO:0036211	protein modification process	0.000000963	0.00373	1.59	19947	3025	420	101
GO:0006325	chromatin organization	0.000236	0.141	2.02	19947	707	420	30
GO:0016570	histone modification	0.0000475	0.0409	2.75	19947	345	420	20

Table 19: Immune system GO processes

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0002682	regulation of immune system process	0.00000505	0.00869	1.78	19947	1681	420	63
GO:0050852	T cell receptor signaling pathway	0.0000153	0.0159	3.91	19947	170	420	14
GO:0060333	interferon-gamma-mediated signaling pathway	0.000668	0.216	4.75	19947	70	420	7

Table 20: Mitochondrial GO processes

GO Term	Description	P-value	FDR q-value	Enrichment	N	B	n	b
GO:0010821	regulation of mitochondrion organization	0.000165	0.107	3.32	19947	186	420	13

N: the total number of protein-coding genes in the human genome (N=19947) listed by the database, n: number of genes in our previous cross-disorder analysis gene-set that were significant in at least 3 disorder pairs (n=420 genes), B: total number of genes in the genome under a GO term, b: number of genes in our gene-set under a GO term.

3. Conclusion

We performed H-MAGMA analyses on publicly available genetic datasets to reveal genetic and epigenetic interactions within the studied phenotypes. We have included data on somatic insulinopathies and psychiatric and neurological disorders and performed 21 cross-disorder analyses at gene level. Then we collated the gene-lists and found that there is a great overlap between the top genes that popped up in the separate cross-disorder analyses. In other words, the results show a considerable overlap between the studied conditions. Due to the hypothesis in PRIME that insulin might be a key mediator in the multimorbidity of somatic and brain-based disorders, we also analyzed, which of the genes within our sets are insulin related, but an enrichment was not particularly evident. To investigate if there were other patterns to be found in our genetic data, we performed subsequent Gene Ontology analysis on the gene-set that was significant in at least 3 disorder pairs in the cross-disorder study. We found that genes were enriched in 3 major domains of biological processes i.e. epigenetics, immune system, and mitochondrial organization. This is in good concert with previous data in the literature, as well as some of the current findings of other groups within WP4 of the PRIME project. Therefore, common analysis of our genetic and epigenetic findings is planned within WP4 and dissemination of our results in the form of a publication is foreseen.

References:

Chang, H., Cai, X., Li, H. J., Liu, W. P., Zhao, L. J., Zhang, C. Y., . . . Xiao, X. (2021). Functional Genomics Identify a Regulatory Risk Variation rs4420550 in the 16p11.2 Schizophrenia-Associated Locus. *Biol Psychiatry, 89*(3), 246-255. doi: 10.1016/j.biopsych.2020.09.016

Fox, M. E., Chandra, R., Menken, M. S., Larkin, E. J., Nam, H., Engeln, M., . . . Lobo, M. K. (2020). Dendritic remodeling of D1 neurons by RhoA/Rho-kinase mediates depression-like behavior. *Mol Psychiatry, 25*(5), 1022-1034. doi: 10.1038/s41380-018-0211-5

Girdhar, K., Rahman, S., Dong, P. F., Fullard, J. F., & Roussos, P. (2022). The Neuroepigenome: Implications of Chemical and Physical Modifications of Genomic DNA in Schizophrenia. *Biol Psychiatry, 92*(6), 443-449. doi: 10.1016/j.biopsych.2022.04.018

Glahn, D. C., Curran, J. E., Winkler, A. M., Carless, M. A., Kent, J. W., Charlesworth, J. C., . . . Blangero, J. (2012). High Dimensional Endophenotype Ranking in the Search for Major Depression Risk Genes. *Biol Psychiatry, 71*(1), 6-14. doi: 10.1016/j.biopsych.2011.08.022

Kosoy, R., Fullard, J. F., Zeng, B., Bendl, J., Dong, P. F., Rahman, S., . . . Roussos, P. (2022). Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nature Genetics, 54*(8), 1145-+. doi: 10.1038/s41588-022-01149-1

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics, 27(12), 1739-1740.

Liu, J. W., Cheng, Y. Q., Li, M., Zhang, Z. J., Li, T., & Luo, X. J. (2022). Genome-wide Mendelian randomization identifies actionable novel drug targets for psychiatric disorders. *Neuropsychopharmacology*. doi: 10.1038/s41386-022-01456-5

Martins-Silva, T., Salatino-Oliveira, A., Genro, J. P., Meyer, F. D. T., Li, Y., Rohde, L. A., . . . Tovo-Rodrigues, L. (2021). Host genetics influences the relationship between the gut microbiome and psychiatric disorders. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 106*. doi: Artn 11015310.1016/J.Pnpbp.2020.110153

Muller, D. J., & Kennedy, J. L. (2006). Genetics of antipsychotic treatment emergent weight gain in schizophrenia. *Pharmacogenomics, 7*(6), 863-887. doi: 10.2217/14622416.7.6.863

Pagliaroli L., Veto B., Aranyi T., & Barta C. (2016). From genetics to epigenetics: new perspectives in Tourette Syndrome research*. Front. Neurosci. - Child and Adolescent Psychiatry*, doi: 10.3389/fnins.2016.00277

Prescott, N. J., Lehne, B., Stone, K., Lee, J. C., Taylor, K., Knight, J., . . . Consortium, U. I. G. (2015). Pooled sequencing of 531 genes in inflammatory bowel disease identifies an associated rare variant in BTNL2 and implicates other immune related genes. *PLoS Genet, 11*(2), e1004955. doi: 10.1371/journal.pgen.1004955

Richter, M., Murtaza, N., Scharrenberg, R., White, S. H., Johanns, O., Walker, S., . . . de Anda, F. C. (2019). Altered TAOK2 activity causes autism-related neurodevelopmental and cognitive abnormalities through RhoA signaling. *Mol Psychiatry, 24*(9), 1329-1350. doi: 10.1038/s41380-018-0025-5

Scharrenberg, R., Richter, M., Johanns, O., Meka, D. P., Rucker, T., Murtaza, N., . . . de Anda, F. C. (2022). TAOK2 rescues autism-linked developmental deficits in a 16p11.2 microdeletion mouse model. *Mol Psychiatry, 27*(11), 4707-4721. doi: 10.1038/s41380-022-01785-3

Sey, N. Y., Hu, B., Mah, W., Fauni, H., McAfee, J. C., Rajarajan, P., ... & Won, H. (2020). A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. Nature Neuroscience, 23(4), 583-593.

Su, X., Li, W. Q., Lv, L. X., Li, X. Y., Yang, J. F., Luo, X. J., & Liu, J. W. (2021). Transcriptome-Wide Association Study Provides Insights Into the Genetic Component of Gene Expression in Anxiety. *Frontiers in Genetics, 12*. doi: Artn 74013410.3389/Fgene.2021.740134

Vysotskiy, M., Zhong, X., Miller-Fleming, T. W., Zhou, D., Cox, N. J., Weiss, L. A., & Consortium, P. G. (2021). Integration of genetic, transcriptomic, and clinical data provides insight into 16p11.2 and 22q11.2 CNV genes. *Genome Medicine, 13*(1). doi: ARTN 17210.1186/s13073-021-00972-1

Wang, Y., Cheng, C. M., Zhang, Z. F., Wang, J. Y., Wang, Y., Li, X. P., . . . Liew, C. C. (2018). Blood-based dynamic genomic signature for obsessive-compulsive disorder. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics, 177*(8), 709-716. doi: 10.1002/ajmg.b.32675

Wu, W. N., Howard, D., Sibille, E., & French, L. (2021). Differential and spatial expression meta-analysis of genes identified in genome-wide association studies of depression. *Translational Psychiatry, 11*(1). doi: ARTN 810.1038/s41398-020-01127-3

Yue, S., Luo, L. L., Feng, Y. T., Liu, D. B., Wang, F. T., Che, R. B., . . . Liu, Y. (2022). Gene-gene interaction and new onset of major depressive disorder: Findings from a Chinese freshmen nested case-control study. *Journal of Affective Disorders, 300*, 505-510. doi: 10.1016/j.jad.2021.12.138