

University of Helsinki
Dissertationes Universitatis Helsingiensis
221/2025

The effect of the environment on depression and underlying genetic influence in adolescents and young adults

Zhiyang Wang

Institute for Molecular Medicine Finland (FIMM)
Helsinki Institute of Life Science (HiLIFE)
University of Helsinki, Finland

Doctoral Programme in Population Health
University of Helsinki Doctoral School
University of Helsinki, Finland

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University
of Helsinki, for public examination in Hall 2, Biomedicum, on the 13th of June
2025, at 14 o'clock.
Helsinki 2025

Pre-examiners

Chief Researcher Jaana I. Halonen, Ph.D.
Department of Public Health,
Finnish Institute for Health and Welfare, Finland;
Department of Psychology,
Stockholm University, Sweden

Assistant Research Professor Léa Maitre, Ph.D.
Barcelona Institute for Global Health, Spain

Custos

Associate professor Andrea Ganna, Ph.D.
Institute for Molecular Medicine Finland (FIMM),
HiLIFE, University of Helsinki, Finland

Supervisor

Professor Jaakko Kaprio, M.D., Ph.D.
Institute for Molecular Medicine Finland (FIMM),
HiLIFE, University of Helsinki, Finland

Associate professor Andrea Ganna, Ph.D.
Institute for Molecular Medicine Finland (FIMM),
HiLIFE, University of Helsinki, Finland

Opponent

Professor Sylvain Sebert, Ph.D.
Research Unit of Population Health,
Faculty of Medicine, University of Oulu, Finland

Publisher: University of Helsinki
Series: Dissertationes Universitatis Helsingiensis 221/2025
Cover Image: Our Friday Goldfish Design Studio

ISBN 978-952-84-1041-6 (Paper back)
ISBN 978-952-84-1040-9 (PDF)
ISSN 2954-2898 (Print)
ISSN 2954-2952 (PDF)

PunaMusta, Joensuu 2025

莫愁前路无知己，天下谁人不识君

Abstract

Depression is a significant public health burden among youth, and it has a complex etiology with the interplay between the environment and gene. The exposome, the totality of environmental effects, offers a holistic framework to assess environmental contributions to depression. However, systematic evaluations integrating both the exposome and genetic perspectives remain limited. The thesis aimed to 1) investigate the association between the exposome and depression (depressive symptoms and major depressive disorder) in late adolescence and young adulthood (Studies I and II); 2) examine the relationship between the mid-adolescent exposome and proteomic biomarker of mental health (Study III); and 3) explore the genetic influences underlying these associations by twin design (Studies I, II, and III).

Data from the FinnTwin12 (Finland, used in Studies I, II, and III) and WALNUTs (Spain, used in Study III) cohorts were used. In FinnTwin12, depressive symptoms were assessed using the General Behavior Inventory (GBI) at age 17 and in young adulthood (Studies I and II). Major depressive disorder was diagnosed in young adulthood (Study II). Study I evaluated the effect of age at separation between cotwins as a proxy for the familial exposome (reflecting the duration of living with siblings and parents) on GBI scores and their trajectories through linear regression and mixed models for repeated measures (MMRM). Study II conducted exposome-wide association studies (ExWAS) based on up to 385 exposures from 12 domains to identify associations with GBI scores and incidence of major depressive disorder in young adulthood via repeated generalized linear regression models. In Study III, the mid-adolescent exposome included 60 exposures available in both cohorts, grouped into seven concept areas. Biomarkers included 26 plasma proteins (sampling at the mean age of 22) for FinnTwin12 and 31 (sampling at the mean age of 13) for WALNUTs. ExWASes were conducted for each proteomic biomarker of

mental health separately in both cohorts. Different twin analyses explored genetic influences across studies. In Study I, a bivariate cross-lagged path model assessed the relationship between cotwin differences in GBI scores and separation status. Study II used the factor analysis to generate exposome scores based on significant exposures from ExWASes, and then fitted it into bivariate Cholesky twin models with GBI scores. In Study III, within-pair fixed-effect regression models and cross-twin cross-trait (CTCT) correlations were performed in FinnTwin12 based on significant exposure—protein sets identified from ExWAS. Due to data availability, sample sizes varied across analyses in FinnTwin12, and 187 participants were included from the WALNUTs cohort.

Twins who separated before age 17 had significantly higher GBI scores at age 17 and in young adulthood compared to those separated later. ExWASes identified 29 out of 385 and 46 out of 286 exposures significantly associated with GBI scores in young adulthood and at age 17, respectively. Most of the significant exposures were from the domain of family and parents. Seven exposures related to lifestyle, indoor environmental quality, natural environment, and family social environment were associated with six proteins. Exposures related to greenspaces were linked to proteins in both cohorts, but the further effects on mental health were heterogeneous between the two cohorts. The cross-lagged path model found a small genetic effect that within-pair differences in GBI at age 17 linked to separation status before age 22. Bivariate twin models showed significant genetic contributions to the covariances between exposome scores and depressive symptoms, with sex differences. Within-pair regression models and CTCT correlations highlighted both genetic and environmental influences on exposure—protein associations, notably between videogame playing and serpin family G member 1.

This thesis provides evidence about the comprehensive impact of the exposome on depression in late adolescence and young adulthood. Familial exposures emerged as the most influential. The effects of greenspaces exerted opposite effects in Finland and Spain. Twin analyses highlighted the complex genetic-environmental interplay, that depression's etiology is not solely genetic or environmental. Despite limitations, these findings support the use of integrated exposome and genetic frameworks in mental health research and inform precise preventive strategies.

List of abbreviations

A:	additive genetic factor
C:	common environmental factor
CI:	confidence interval
CFA:	confirmatory factor analysis
CTCT:	cross-twin cross-trait
D:	dominant genetic factor
DZ:	dizygotic
E:	unique environmental factor
ExWAS:	exposome-wide association studies
GBI:	General Behavior Inventory
HDL:	high-density lipoprotein
MDD:	major depressive disorder
MMRM:	mixed model for repeated measures
MZ:	monozygotic
NDVI:	Normalized Difference Vegetation Index
PCA:	principal component analysis
SSAGA:	Semi-Structured Assessment for the Genetics of Alcoholism
SD:	standard deviation

Content

Abstract	v
List of abbreviations	vii
List of original publications	x
1 Introduction	1
2 Review of literature	3
2.1 Introduction to the exposome	3
2.1.1 Definition and characteristics of the exposome.....	3
2.1.2 Analysis of the exposome.....	6
2.1.3 Implication of the exposome research	8
2.1.4 Underlying heritable influence behind the exposome.....	10
2.2 Introduction to depression in adolescents and young adults.....	11
2.2.1 Definition of depression	11
2.2.2 Epidemiology and etiology of depression.....	12
2.3 The exposome and depression	15
2.3.1 The external exposome and depression	15
2.3.2 The proteomic biomarkers and depression.....	17
3 Aims of the study	19
4 Material and methods.....	20
4.1 Study participants	20
4.1.1 FinnTwin12.....	20
4.1.2 WALNUTs	20
4.2 Ethical approval.....	21
4.3 Study measurements	21
4.3.1 Exposome measurement.....	21
4.3.2 Depression measurement.....	24
4.3.3 Assay of proteome (biomarker of mental health)	25
4.3.4 Covariates	26
4.4 Hypothesis.....	26
4.5 Data analysis	27
4.5.1 Regression between the age at separation and depressive symptoms	
(Study I).....	28

4.5.2	Exposome-wide association studies (ExWAS) (Studies II-III)	28
4.5.3	Twin analysis (Studies I-III)	29
5	Results	32
5.1	Characteristic of participants included in each study	32
5.2	Association between age at separation and depressive symptoms (Study I results)	32
5.3	Association between the exposome and depressive symptoms (Study II results)	34
5.4	Association between the mid-adolescence exposome and proteomic markers of mental health (Study III results)	37
5.5	Outputs from twin analysis	40
5.5.1	Relationship between separation and within-pair similarity on depressive symptoms (Study I)	40
5.5.2	Bivariate twin modeling between the exposome score and depressive symptoms (Study II)	40
5.5.3	Twin analysis between significant exposure—protein sets (Study III)	41
6	Discussion	44
6.1	Main findings of the thesis	44
6.2	Interpretation of the results and comparison with previous literature	44
6.3	Complex reality and implication	49
6.4	Limitations	51
7	Conclusion	53
	Author contribution	54
	Acknowledgements	55
	Reference	57
	Appendices	81
	Appendix 1. General Behavior Inventory (GBI)	81

List of original publications

This thesis is based on the following articles, which are referred to the text by Roman numeral:

- I. **Z. Wang**, A. M. Whipp, M. Heinonen-Guzejev, J. Kaprio, Age at separation, residential mobility, and depressive symptoms among twins in late adolescence and young adulthood: a FinnTwin12 cohort study. *BMC Public Health* 24, 2239 (2024).
- II. **Z. Wang**, et al., The effect of environment on depressive symptoms in late adolescence and early adulthood: an exposome-wide association study and twin modeling. *Nat. Ment. Heal.* 1, 751–760 (2023).
- III. **Z. Wang**, et al., Associations between the mid-adolescent external exposome and proteomic biomarkers of mental health. *medRxiv*, 2024.12.18.24319208 (2024).

Studies I and II are licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). Study III was deposited on a preprint server under the Creative Commons Attribution 4.0 International License and is in a peer-review process of an academic journal.

1 Introduction

Depressive symptoms encompass a range of normative mood fluctuations, including sadness and anhedonia. If individuals persistently experience a depressed mood and/or a loss of interest or pleasure in daily activities, accompanied by other symptoms that impair functioning, they will be diagnosed with major depressive disorder (MDD) (1). Both depressive symptoms and MDD are referred to as depression, which poses a salient public health concern. The prevalence of depression among adolescents and young adults has risen in recent decades (2, 3). Adolescence and early adulthood are critical periods of physiological development, emotional regulation, and the acquisition of essential life skills (2, 3). During these stages, external deleterious stressors can increase the risk of developing depression. Family and twin studies have demonstrated that both environmental and genetic factors contribute to the etiology of depression (4).

The exposome concept provides a framework to incorporate the totality of the environment (5). It is classified into specific external (e.g., lifestyle), general external (e.g., infrastructure), and internal domains. Due to the highly spatiotemporal nature of environmental factors (exposures), exposome research is often conducted within large, longitudinal population cohorts to capture dynamics. Unlike traditional studies that investigate a single exposure or a limited set of factors, exposome research employs advanced technologies to assess a broad array of exposures, from parental relationships, urban facilities, to social support networks. This holistic approach offers an in-depth understanding of how these exposures interact and collectively affect depression.

Depression is typically assessed through self-reports or evaluations by close relatives and professionals, while it can also be reflected in molecular markers. Omics is the characterization and quantification of these molecules, and research has identified biomarkers from various omics layers (e.g., proteomics) that are

associated with depression (6). Internal exposures are biomarkers that have proven to be related to external exposures. By integrating these biomarkers, it is possible to map the biological networks that link exposures to psychophysiological changes.

Furthermore, the impact of environmental exposures can be modified or intensified by genetic predispositions. Twin studies compare traits between monozygotic (MZ) twins, who are genetically identical, and dizygotic (DZ) twins, who share approximately 50% of their segregating genes, to disentangle genetic and environmental contributions to the phenotypic variation, such as depression, in the population (7). By controlling genetic and common environmental confounding, twin studies improve causal inference for environmental health impact assessment toward the risk of depression, as well.

Although some exposome studies have linked depression, genes, and external environments together (8, 9), research in this area remains limited, particularly among younger populations. Expanding this body of knowledge and replicating findings are essential for developing precise and cost-effective implications.

2 Review of literature

2.1 Introduction to the exposome

2.1.1 Definition and characteristics of the exposome

The concept of the exposome was first raised by Dr. Christopher P. Wild in 2005 (10), which refers to the totality of non-genetic environmental exposures from various external and internal sources. In short, the exposome is everything besides the genome. It is generally classified into three categories: the general external (e.g., greenspaces), specific external (e.g., dietary), and internal (e.g., proteins) exposome. Initially, the exposome referred to the effects starting from conception, but given famine epigenetic studies (11, 12), the impact of the environment on people begins much earlier. Increasing evidence suggests that a single type of exposure cannot fully account for the environmental contributions to diseases and conditions. Recent reviews have identified a range of environmental factors, including noise, air pollution, and climate change, as risk factors for cardiovascular disease (13), mental health disorders (14), and diabetes (15). Many environmental exposures are sufficient factors. Unlike the conventional “one exposure” model, the exposome framework highlighted the need for a holistic view of the complex environmental effect on people living in actual contexts rather than in controlled laboratory conditions studying model organisms.

Every exposure constitutes the exposome, and none of them are isolated, which forms a huge and intricate network. Within the network, each exposure may act independently, synergistically, or antagonistically, and high-dimensional data make the situation more challenging. For example, three-way interactions indicated that a two-way interaction might vary depending on a third variable (16), while introducing a fourth variable would further sheer the complexity. For another instance, Gennings et al. has proposed another complex interaction effect with

chemical exposures, that there is a more-than-additive effect on thyroid function in rats when the total dose of the mixture of polyhalogenated aromatic hydrocarbons exceeded a specific interaction point (upper limit of the confidence interval on the dose-dependent interaction threshold) (17). Additionally, highly correlated exposures (multicollinearity) can reduce the precision of estimated effect sizes and inflate standard errors. Large-scale, high-dimensional datasets are very common in exposome research. Therefore, choosing appropriate analytical methods is essential, as discussed in Section 2.1.2.

Modern tools allow us to better profile exposures. For external exposures, questionnaires and interviews capture lifestyle, family influences, and perceptions of the living environment. Indoor and outdoor sensors monitor air pollution, noise, and chemical exposures, while smartphone applications track activities like consumer product use and live location data for pollution exposure assessment (18). For internal exposures, high-throughput omics techniques enable comprehensive analysis of biological molecules, including small metabolites, larger proteins, gene expression profiles, and so on (18). Toxicologists have introduced the eco-exposome, which assays all anthropogenic and natural chemicals, biotransformation products, and signaling molecules sensitive to external exposures (19). They seek to trace the systematic pathway from external exposures to adverse outcomes (19).

Gathering all available exposures is able to reduce concerns about confounding to some extent, that traditional covariates, such as gender/sex (20), become environmental indicators too. Of course, this does not mean that we ignore confounding, and covariates are often included in exposome analyses depending on the research questions and methodology. Our concern is whether we have overlooked any important exposures, like a selection bias in exposures that people only included those perceived to be more important. In fact, it is impossible to include every exposure, and new exposures will continue to emerge, as we would not worry about smartphones' impact on children's mental health 30 years ago. To address this concern, one approach is through the twin study, which estimates the overall environmental contribution to the variance of phenotypes (21). We can explore to what extent the currently available exposome set explains environmental

contribution and what may still be missing. Exposome studies should aim to thoroughly measure and include exposures and carefully review possible omissions.

The exposome is dynamic across time and space, with the presence, frequency, and intensity of exposures constantly shifting. Extensive monitoring demonstrated a spatiotemporal and location-dependent exposome (both internal and external) interaction network within the broader ecological system (22, 23). Given this temporospatial complexity, the exposome is often operationalized and conceptualized into practical sub-domains (e.g., chemical, physical, social exposures), spatial contexts (e.g., home, school), and critical life periods (e.g., early life or pregnancy). The ATHLETE (Advancing Tools for Human Early Lifecourse Exposome Research and Translation) project, from over 100 exposures, has identified that maternal tobacco smoking during pregnancy and exposures to air pollution and chemicals during childhood were associated with accelerated aging in children (24). More specifically, exposure during the first trimester of pregnancy, the critical period of organogenesis, is more likely to result in birth defects, whereas harmful exposures like maternal smoking in the second and third trimesters are more likely to cause growth problems (25, 26). The lifecourse exposome broadens the focus to the entire lifespan, assesses the longitudinal and dynamic effects of exposures, and explores optimal windows for intervention. Repeated exposures are included, and multiple risk models are considered, such as cumulative and trajectory. A review summarized that cumulative exposure to neighborhood disadvantage, including various indicators, is saliently associated with functional decline, more chronic health conditions, and increased mortality (27). The urban, school, or occupational exposome expresses diverse settings or contexts. The urban exposome describes that most people live in urban areas and experience the complex interplay between the built, social, chemical, food, and lifestyle factors. Researchers have collected exposures about urban characteristics to study their impact on a range of phenotypes (28–31). Similarly, the working-life exposome links daily life and the occupational environment, which tends to be more frequent and intense than usual urban exposures and is closely tied to lifestyle, behavior, and socioeconomic status (32).

2.1.2 Analysis of the exposome

Due to the complex, high-dimensional nature of exposome data, analyses utilize multiple advanced methods to uncover how a large number of environmental factors collectively affect health through various technical approaches. The selected approaches also reflect how researchers view the exposome.

The first involves testing exposures individually for association with health outcomes. One widely used method is exposome-wide association studies (ExWAS), a hypothesis-free approach that assesses associations between each exposure and the outcome based on regression models (41). ExWAS uses adjusted significance thresholds to select “more” significant exposures. An ExWAS study based on the Florida Vital Statistics Birth Records has found 12 exposures significantly associated with hypertensive disorders in pregnancy from 5784 exposures (42). ExWAS can also include interaction terms, adding depth to the analysis of complex relationships (43). Another method, correlation analysis with adjusted thresholds, maps critical networks between exposures and outcomes. In the HELIX (Human Early-Life Exposome) Child Panel study, Fabbri et al. employed pairwise partial correlations with a feature shrinkage step using Gaussian graphical models to reveal 23 direct relationships between endocrine-disrupting chemicals and omics molecules (33). This method further considered the influence from other exposure between the exposure of interest and health outcomes, in addition to ExWAS.

Feature selection refers to the selection of a subset of exposures from the original set in a multivariable context. Regression models with penalized terms can identify significant exposures by applying regularization, which shrinks the coefficients of weaker exposures. Methods such as the elastic net algorithm have been used to examine the impact of the urban exposome on type 2 diabetes (28) and acute myocardial infarction (45). Machine learning models, increasingly adopted for exposome research, offer advantages in handling high-dimensional data and improving predictive accuracy. This type of technique can guide variable selection by ranking the relative importance of exposures. For instance, in the longitudinal Doetinchem Cohort Study in the Netherlands, the random forest model identified nine key exposures across multiple domains that were most important for

predicting self-perceived health, because they were largely responsible for the model's performance, while 87 exposures contributed little to the performance (34).

Feature extraction is a process of transforming a large number of exposures into informative, compact indices. The first technique is calculating the summary measure based on exposure weights or importance, through either expert knowledge or modeling. Lam et al. developed a neighborhood obesogenic built environment index in the Netherlands using an evidence-based, expert-informed framework (35). In Kuopio, Finland, inspired by approaches like "unhealthy or low-risk lifestyle scores," exposome scores have also been created, which were linked to various metabolites among children (36). Pries et al. calculated supervised exposome scores for schizophrenia from methods such as logistic regression and the Ridge algorithm (37). The second technique, feature projection, reduces the data from a high-dimensional space to fewer dimensions. Principal component analysis (PCA), which calculates eigenvalues and eigenvectors of the exposure correlation matrix, has illustrated correlations between eco-exposome variables and the microbiome and resistome (all antibiotic resistance genes) in hospital and urban wastewater (38). Moreover, based on 348 exposures, a series of data-driven iterative factor analyses and bifactor modeling have produced six orthogonal subfactors, some of which were associated with psychosis (39). The third technique is to distinguish participants with different patterns of exposure profiles. Hierarchical clustering has been used to generate distinct patterns of environmental stressors, further linked to age-standardized all-cause mortality rates in Belgium (40).

Co-exposure assessment typically employs a single analytic platform that incorporates all available exposures to evaluate their combined effects on the outcome of interest. Various mixture models have been developed to analyze the health effects of multiple simultaneously occurring exposures, as well as their interactions, such as weighted quantile sum regression (41). This approach simplifies inference and also stresses the role of individual exposures (41). For instance, Bayesian kernel machine regression and quantile-based G-computation have been used to explore joint effects and effect modifications between air pollutants and neighborhood socioeconomic factors on cognitive decline among

adults over 50 in the US (42). Additionally, some extensions of these mixture models have been proposed to account for the temporal variation (43, 44). Machine learning models also support co-exposure assessment by identifying nonlinear and interactive relationships. Techniques like gradient boosting and random forests have been applied in a series of analyses to connect the exposome and health outcomes (28, 29, 34, 45, 46). However, one limitation of machine learning models is their interpretability. Researchers have developed several explainable artificial intelligence methods, such as Shapley Additive Explanations and Local Interpretable Model-agnostic Explanations, for better transparency and understanding of outputs. Furthermore, the deletion/substitution/addition algorithm constructs multi-exposure regression models to simultaneously examine the joint effects of exposures on outcomes. This method employs iterative steps to delete, substitute, or add variables, thereby refining predictive models, achieving feature selection and combined-effect evaluation together (47).

The abovementioned methods are meant to be more illustrative than exhaustive, and I intend to introduce relevant studies and findings across these various designs. The same method can also be used for different approaches based on the research question, and some studies also use multiple methods at the same time as multi-stage analysis (48, 49). The HELIX/ATHLETE projects have published a series of papers reviewing state-of-the-art exposome analysis methods across different scenarios, including applications, assumptions, strengths, and limitations (50–54).

2.1.3 Implication of the exposome research

A holistic perspective of environmental effects effectively enhances disease prevention. A deep and precise knowledge base of the complex relationships between exposures and health has been accumulating to inform evidence-based prevention strategies and policies. The STOP (Science and Technology in Childhood Obesity Policy) consortium suggests a few interventions to reduce childhood obesity, such as limiting the ultra-processed food intake for children and diversifying children's diets, through intermediate omic markers (the internal exposome) to connect the dietary habits and outcomes (55). The Equal-Life project aims to develop evidence-based guidance for policy, research, and interventions

that help anticipate changes in the distribution of mental health outcomes if elements of the exposome are modified (56). Through high-granularity information, agent-based modeling effectively captures the change in exposure-response to interventions, estimates behavior-environment interactions, robustly evaluates health impacts and social costs, and empowers the intervention and policy design (57).

The exposome advances precision medicine, in which the environment plays a critical role in many diseases and their progression. Adjusting treatments based on the environmental context can lead to improved outcomes. Drug-exposome interactions reveal that adverse effects can arise from environmental, dietary, and microbiome-derived xenobiotics on drug efficacy (58). Researchers now advocate for calibrating the implementation of polygenic risk scores on clinical use by ancestry (59), and environmental profiles have the same potential to enhance personalization based on abundant evidence of gene-environment interactions (60).

The exposome can also help us evaluate past policies. Often, certain policies focus on specific elements such as greenspace or walkability, and their purposes may not be related to health. After a policy has been implemented, there will be various collateral effects, such as bringing economic, social, and cultural benefits within the city's complex system (61, 62). For example, an urban riverside park regeneration project in Barcelona, Spain was estimated to attract over five thousand adult users daily to perform different types of physical activity (63). Beyond the project's basic objectives, an open-air museum will be built there, transforming social and built environments as well. In China, residents have expressed concerns that subway construction negatively impacted their daily lives and the city image, pointing out that more efforts from the government and contractors were warranted (64), while another study showed that rail transit can drive urban economic growth (65). The environmental impacts associated with different stages of policy implementation can vary significantly. By integrating the exposome framework with natural experiments, people can gain a comprehensive understanding of the benefits and potential harms in the health field associated with policies, facilitating refinement for future initiatives.

The exposome concept fosters collaboration across multiple disciplines, including epidemiology, toxicology, sociology, and geography, promoting a more integrated approach to understanding health impacts. Similar to how the Human Genome Project advanced sequencing technology, the development of analytical platforms for internal assays and general environmental monitoring infrastructure is expected to follow with economies of scale (66). The high accuracy and relevance of data will motivate both private and public sectors to develop and commercialize them (66). This emerging field holds potential for translational research, requiring collaborative efforts from the academic community.

2.1.4 Underlying heritable influence behind the exposome

Exposome researchers usually claim that the exposome complements the genome, and some even emphasize that exposures are non-genetic factors (67, 68). In fact, exposures do not necessarily imply a pure environmental effect. Both familial and genetic influences are considered ‘inheritable factors’ between generations. According to intergenerational transmission theory, factors affecting individuals can be passed down from parents to children through both shared genetic and environmental contexts. Aspects such as parental life satisfaction and experiences of family violence significantly influence the development of similar family dynamics in offspring (69, 70). A US study has shown that parental resource capital was related to children's academic achievement (71). Income, which is closely related to educational achievement, also affects living conditions later. Social factors, including social support, peer relationships, marital quality, and lifestyle factors, such as diet, smoking, physical activity, and alcohol consumption, also have a genetic component (72, 73). Additionally, parents are able to affect where their offspring to live. Increased support needs among parents and adult children often lead to co-residence and a higher likelihood of one generation moving into the other's home (74). The physical characteristics of living and working environments could then be inferred to have heritable influences. For instance, individuals who love sports may prioritize access to greenspaces and recreational facilities in selecting their residences.

2.2 Introduction to depression in adolescents and young adults

2.2.1 Definition of depression

In the thesis, depression is defined as an overarching concept that encompasses a broad spectrum of mood-related experiences and difficulties. At one end of the spectrum, depressive symptoms can refer to any mood state in the context of normative mood fluctuations, including sadness, loss of interest or pleasure, fatigue, changes in appetite or sleep patterns, difficulty concentrating, and feelings of worthlessness or excessive guilt (3). These symptoms reflect temporary or persistent emotional states. In contrast, MDD, another end of the spectrum, is a specific clinical disease. While anyone can experience depressive symptoms, not everyone meets the criteria for MDD. According to the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision, MDD is diagnosed when an individual experiences a depressed mood and/or a loss of interest or pleasure in daily activities for at least two weeks, accompanied by at least five additional symptoms that significantly impair social, occupational, or other important areas of functioning (1). The core criteria did not change in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (75). Despite the diagnostic criteria, MDD remains highly heterogeneous due to a variety of symptom combinations, clinical severity, age at onset, comorbidities, and progression over time (3). Therefore, in practical research, measuring depressive symptoms by considering subtypes, severity, and duration captures more information than a binary diagnosis.

Adolescence and young adulthood are vulnerable periods for susceptibility to developing depression due to biological and psychosocial changes and life transitions. Suboptimal hormonal fluctuations and brain development during these periods disrupt neural activity, mood regulation, and emotional stability, increasing the risk of depression (76, 77). Additionally, core life skills, such as planning, resilience, and self-control, are learned and practiced during this period. These skills enable adolescents and young adults to make right decisions, solve problems, think critically and creatively, communicate effectively, and build healthy

relationships, which ultimately help to protect them from depression and other mental health problems (78–80). Studies have shown that depression rises substantially throughout adolescence (3, 81). Depression in adolescence is associated with a high risk of recurrence and other comorbidity (3). Approximately 50% of adolescents with depression will have an adult episode of depression or anxiety disorder (82). Depressed adolescents and young adults are also more likely to develop anxiety disorders, bipolar disorder, functional somatic symptoms, and high-risk behaviors, such as suicide, substance use, and unsafe sexual behavior (3, 83–85). In addition, they face a greater risk of physical diseases such as cardiovascular disease and obesity and low socioeconomic status, including lower educational attainment, higher welfare dependence, and unemployment in adulthood (3, 85).

2.2.2 Epidemiology and etiology of depression

Depression is a leading contributor to the global burden of disease. According to the Global Burden of Disease Study 2019, the age-standardized prevalence of depressive disorders, which include MDD and dysthymia, was 3440.1 per 100000 people, while the number (3851.3 per 100000) was slightly higher in Western Europe (the setting of the thesis) (86). The prevalence was notably higher among females (4,158.4 per 100,000) compared to males (2,713.3 per 100,000) at the global level (86). Globally, depressive disorders contributed most to the proportion of disability-adjusted life years caused by mental disorders, and it ranked as the second top leading cause of years lived with disability (86). As vulnerable periods, the burden of depression is even more pronounced among adolescents and young adults. A systematic review by Shorey et al. estimated that the pooled prevalence from 2001 to 2020 was 8% for MDD and 4% for depressive dysthymia among adolescents (87). In the US, 21.0% of adults aged 18–29 experienced any depressive symptoms, which was the highest among all age groups (88).

In recent years, there has been an increasing trend in the occurrence of depression. Shorey et al. also revealed that the point prevalence of elevated depressive symptoms among adolescents increased from 24% (between 2001 and 2010) to 37%

between 2011 and 2020 globally (87). Based on the US National Survey on Drug Use and Health, the past-year prevalence of major depressive episodes in adolescents rose from 8.1% in 2009 to 15.8% in 2019 (89). In Germany, the prevalence of diagnosed depressive disorders among young people also increased significantly between 2009 and 2017 regardless of the severity (90). Among males aged 15–19 and 20–24, cases of mild diagnosed depressive disorders rose by 166% and 147%, respectively, and the numbers were 103% and 70% among females in these age groups, respectively (90).

Risk or protective factors for depression among adolescents and young adults (the focus of the thesis) are across areas and contexts. A review of over one hundred longitudinal studies by Cairns et al. has indicated that behavior modification on substance use (including alcohol, tobacco, and cannabis), dietary, coping strategies, and weight management could alter the development of depression during adolescence without professional intervention (91). Parent and family dynamics play a crucial role, in which factors like parental mental health history, cognitive styles, and the overall emotional climate within the family were strongly associated with the risk of adolescent depression (92). The efficacy of familial interventions in reducing depressive symptoms among adolescents was observed (93, 94). Unfavorable community factors, such as neighborhood unsafety and discrimination, have exacerbated the risk of depression among students (ages 4–18) (95). Some advantageous exposures from the built environment, such as low levels of pollution, greenspaces, and neighborhood aesthetics, were associated with less adolescent depression shown in numerous studies and reviews (96–98). These factors also apply to young adults, while young adults face many important life transitions and enter new environments. In in-depth interviews with 15 emerging adults with major or minor depression, concerns about self-status on depression, financial uncertainty, feelings of alienation from family and peers, and the perceived failure to meet developmental milestones appeared to interact and exacerbate depression (99). Suboptimal school experiences related to learning, achievement motivation, emotional well-being, and social relationships might also amplify challenges in mental health (100). Moreover, from the perspective of the internal environment and biomarkers, a 2021 systematic review has identified several internal risk factors, including inflammation, telomere length, and brain

abnormalities, associated with MDD in adolescents, especially in the context of early life adversity (101). This growing body of evidence underscores the importance of studying depression through a holistic view to integrate all external and internal environmental influences.

Depression is a complex phenotype with a considerable genetic component, particularly during adolescence and young adulthood. Across all age groups, the heritability of major depression has been estimated at approximately 37% (4). In adolescents, genetic influences account for around 40% of the variation in MDD and for 30 - 50% of the variation in depressive symptoms (102, 103). In South Korea, genetic contributions to depressive symptoms among adolescents and young adults were notably higher in females (41%) compared to males (11%) (104). The longitudinal Netherlands Twin Registry study has found a moderate genetic influence on co-symptoms of anxiety and depression from adolescence into young adulthood, as well (105). Multiple gene associations have been found in the largest meta-genome-wide association studies to date (106), and some genetic variants were identified to be associated with depressive symptoms in adolescents and young adults, such as the BDNF Val66Met genotype (107). Genetic factors can further influence depression in adolescents and young adults by modulating DNA methylation and microRNA levels (108).

The development and progression of depression result from the long-term interplay between genetic predispositions and environmental factors. Adolescents genetically at risk for depression may experience more social adversity, such as negative life events and maternal punitive discipline, due to gene-environment correlation (109). These individuals are also more susceptible to developing depressive symptoms in response to such environmental stressors, as gene-environment interaction (109). The interaction between the serotonin transporter-linked promoter region gene and family support has been observed to impact self-reported depression symptoms in adolescents and young adults in the US (110). Thus, understanding the etiology of depression requires considering both genetic and environmental factors.

2.3 The exposome and depression

2.3.1 The external exposome and depression

Here the exposome research is defined as including three or more different types of external exposure (or their biomarkers of effect). Some studies that only included many toxic or nutritional exposures were not considered in this scoping review section, even if they claim to use the concept of the exposome. Any solely multi-omics study was neither considered, since this thesis focused on the external environment.

Many studies were based on the UK Biobank. Choi et al. examined 106 modifiable factors across lifestyle, social (e.g., support, engagement), and environmental (e.g., green space, pollution) domains and identified 29 exposures associated with the incidence of depressive symptoms after full adjustment for covariates (9). Further Mendelian randomization analysis strengthened causal inferences for some exposures, such as confiding in others and television use (9). After stratifying ExWAS analyses by genetic risk (using polygenic risk scores), they found that several significant associations attenuated to null, indicating potential genetic interactions (9). Another study examined 128 urban exposures categorized into 53 groups (8). Through multi-stage data integration, they found a significant correlation between the urban living environment profile and affective symptom groups, including the frequency of depressed mood (8). Exposures like deprivation, air pollution, and urban infrastructure had high weights in the urban living environmental profile (8). Meng et al. collected 14 environmental risk factors for mental disorders in five categories: sleep behaviors, stress, childhood trauma, social support, and shift work (111). There were significant associations between sleep behaviors, childhood trauma, and social support with specific single nucleotide polymorphisms linked to depressive episodes (111).

In the Netherland Twin Registry, researchers created objective (from physical, cultural, and other aspects) and subjective (reflecting subjective evaluations on their relationship, stress, and so on) poly-environmental scores of wellbeing based on 69 objective exposures and 21 subjective socio-environmental indicators, respectively (112). The subjective poly-environmental score of life satisfaction was

associated with depressive problems regardless of before and during the COVID-19 pandemic (112). As a replication, researchers created a subjective poly-environmental score in the UK Biobank and found a significant correlation between the subjective poly-environmental score and polygenetic risk score of life satisfaction in the UK Biobank but not in the Netherlands Twin Registry (112). In Spain, researchers developed a general exposome risk score and three sub-scores (social and household, lifestyle and behavior, and wider environment and health) using data on 18 exposures and health status collected during the COVID-19 lockdown based on 9218 individuals (113). All three sub-scores were significantly associated with depressive symptoms (113). Additionally, an interaction between the general exposome score and the polygenic risk score for broad depression suggested a potential mitigating effect and highlighted the population susceptible to the pandemic (113). In Poland, a correlation network was constructed using an exposome score based on 14 exposures, ranging from childhood trauma to urban upbringing, along with psychopathological symptoms and demographic variables such as education level (114). This exposome score was not directly linked to depressive symptoms (114). Based on data from the Belgian Health Interview Surveys, researchers included 13 exposures related to air pollution, noise, greenness, and building morphology (115). In single-exposure models, ozone, black carbon, and PM₁₀ levels were associated with depressive disorders (115). However, in multi-exposure models after further adjustment of social exposures, only the effect of black carbon remained significant (115).

More than half of the referenced studies have utilized correlation or ExWAS to spot specific exposures. The coefficient estimation from these methods makes interpretations easier. In contrast, two studies employed feature extraction approaches to create exposome scores. van de Weijer et al. used supervised elastic net regression based on life satisfaction (112), while Farré et al. applied hierarchical clustering to group exposures, followed by simple addition to generate scores (113). These studies emphasized the holistic effects of exposome scores in their interpretations. Xu et al. used sparse canonical correlation analysis, allowing multiple exposures and psychiatric symptoms in the same model to present a comprehensive and complicated perspective (8). Many studies acknowledged the role of genetic influences, typically through molecular approaches. Nevertheless,

these studies often examined a limited number of exposures or overlooked vulnerable populations of adolescents and young adults, indicating a research gap.

There are a relatively low number of studies integrating the exposome and depression. One reason is the limited acceptance or awareness of the exposome framework. Many studies did not explicitly state that they employed the exposome framework, but they met the pre-specified criteria. Most epidemiological studies still tend to focus on one or a few types of exposures, as observed during the 2024 annual conference of the International Society of Environmental Epidemiology, the leading conference in the field. Additionally, limited infrastructure and funding disparities are reasons for few exposome studies. Developed regions are better equipped with large population cohorts, facilities for testing biological samples, and environmental monitoring networks, such as the European Union's Copernicus program.

2.3.2 The proteomic biomarkers and depression

There have been a lot of bioinformatics studies to find various types to biomarkers related to depression. In this scoping review section, the proteomic biomarker was the focus due to the research interest of this thesis.

In a 2019 book chapter, Silva-Costa et al. provided an overview of proteomic markers linked to depression (116), while new studies are emerging. There were two studies connecting proteomics and depressive symptoms in FinnTwin12. Afonin et al. identified 13 proteins significantly associated with the general psychopathology factor in young adulthood, of which depression is one dimension (117). Moreover, Whipp et al. found 47 proteins related to depressive symptoms or general psychopathology, assessed longitudinally from adolescence to young adulthood (118). In WALNUTs, eight proteins were significantly associated with the Strengths and Difficulties Questionnaire score, another measure encompassing depressive symptoms (119). These three studies form a part of the empirical foundation for this thesis. Expanding beyond these cohorts, a large-scale study of over 46 thousand participants in UK Biobank identified 157 of 2920 plasma proteins associated with incident depression, which these proteins correlated with brain regions implicated in depression (120). Further, a two-sample Mendelian randomization study using

data from the UK Biobank Pharma Proteomics Project and deCODE Genetics found that 17 proteins (via cis-acting pQTLs) and 3 proteins (via trans-acting pQTLs) showed evidence of causal associations with MDD (121). A case-control study conducted among North American participants identified 66 signaling proteins that were differentially expressed in individuals with late-life depression (122). Among these, 33 proteins were uniquely associated with depression in females and 6 in males, highlighting potential sex-specific biological pathways (122).

While the proteins identified in above studies may serve as potential biomarkers of depression, they can only be considered as internal exposures within the exposome framework if they are also further linked to external exposures. The concept of redox proteomics was raised to recognize the oxidative stress which was one of the most important mechanisms of action for pollutant-induced toxicity (123). For instance, pesticide-induced toxicity has been shown to involve oxidative changes, which reflected in accumulation of reactive oxygen species, lipid peroxidation, and so on (123). A 2022 systematic review highlighted how proteomic methods have been used to capture organism-level responses to persistent organic pollutants, such as brominated flame retardants and dioxins, providing an integrated understanding of toxicological impacts (124). Furthermore, in a study of over one thousand coronary heart disease patients, three proteins: insulin-like growth factor binding protein 2, plexin domain containing 1, and lymphocyte antigen 6 family member D were found to be associated with residential greenness (125). In the partial correlation network draw by Fabbri et al., three proteins were correlated with endocrine disrupting chemicals in children (33). In another UK Biobank study, Pan et al. identified seven significant interactions between air pollutants and plasma proteins in relation to depression prevalence (126). However, comprehensive studies connecting external exposures, internal biomarkers, and depression outcomes together remain limited.

3 Aims of the study

The overarching aim of this thesis was to elucidate the comprehensive effect of the environment (the exposome) on depression in adolescents and young adults and to examine the underlying genetic influences using the special twin design.

Specific thesis aims were:

- 1) To evaluate the effect of the exposome on depression by identifying key environmental exposures in adolescents and young adults (Studies I and II)
- 2) To investigate the association between environmental exposures during mid-adolescence and proteomic markers of mental health (Study III)
- 3) To explore the genetic contribution to the relationship between the exposome and depression (Studies I, II, and III)

4 Material and methods

4.1 Study participants

4.1.1 FinnTwin12

The FinnTwin12 cohort is a population-based prospective cohort among all Finnish twins born between 1983 and 1987. It was designed to explore the relationship between the environment, genetics, and individual health using a multilevel approach. A total of 5184 twins enrolled at wave one when they reached age 11 or 12 as the epidemiological sample. Twins filled in the general age 14 (wave two), age 17 (wave three), young adulthood (wave four), and early midlife (wave five) questionnaires with 92%, 75%, 66%, and 41% retention, respectively.

Moreover, 1035 families with 2070 twins were invited to take part in an intensive study involving psychiatric interviews, biological sampling, and additional questionnaires. Twins in the intensive study were mostly randomly selected from the epidemiology study, while enriched with twins at elevated familial risk for alcoholism (127). They were then invited to participate again when they were young adults (~age 22), and 1347 twins completed psychiatric interviews and 779 provided serum samples.

A recent study has described the updated progress of FinnTwin12 (128). The FinnTwin12 cohort is used in Studies I, II, and III.

4.1.2 WALNUTs

WALNUTs is a multi-school, parallel, controlled, 6-month superiority randomized clinical trial in Barcelona, Spain, with baseline measurement between April 2016 and June 2017 (129). The trial aimed to recruit healthy teenagers: 771 participants were equally randomized into intervention (mean age: 13.9) and control (mean age:

13.8) groups, and 674 were retained through 6 months (130). A part of the participants had plasma samples collected at the baseline. The intervention involved the daily consumption of walnuts in addition to general health eating recommendations, such as eating a piece of fresh fruit each day (129). Exclusion criteria were regular consumption of omega-3 polyunsaturated fats supplements, daily consumption of walnuts, lactose intolerance, or allergy to walnuts, gluten, cereals, dried fruits, peanuts, soy, sesame, or sulfites (129). The WALNUTs cohort is used in Study II

4.2 Ethical approval

The ethics committee of the Department of Public Health of the University of Helsinki (Helsinki, Finland) and the Institutional Review Board of Indiana University (Bloomington, Indiana, USA) approved the FinnTwin12 study protocol from the start of the cohort. The ethics committee of the Helsinki University Central Hospital District (Helsinki, Finland) reviews the study annually, and 2024's statement is number 14/2024, dated May 22nd, 2024. All participants and their parents/legal guardians gave informed written consent to participate in the study. The WALNUTs study was approved by the CEIC Parc Salut Mar Clinical Research Ethics Committee (approval number: 2015/6026). The trial was registered at ClinicalTrials.gov (NCT02590848) on 29/10/2015. Written informed consent to participate in the original WALNUTs study was provided by the participants' legal guardian/next of kin.

All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

4.3 Study measurements

4.3.1 Exposome measurement

First, in FinnTwin12, the age at separation was calculated as a proxy for the entire familial exposome, defined as the age that, for the first time, the twins no longer

resided at the same address, with one or both twins moving to a new address. Age at separation was obtained from the residential history from the register of the Digital and Population Data Service Agency, Finland from birth to 2020. The residential history data included north and east coordinates (EUREF-FIN) and the ages at which participants moved in and out of each address. Age at separation was categorized into four groups: before age 17, age 17 to <19.5, age 19.5 to <22, and after age 22. Separation status (yes/no) before ages 17 and 22 was also created to indicate whether twins had separated before these ages. Age at separation was used in Study I. Participants without valid data or who had not separated by December 2020 were excluded from any analyses involving this variable.

Then, we curated a comprehensive, longitudinal, multi-level external exposome database for FinnTwin12, starting from the prenatal period. The current database includes 236 types of exposures, with many measured repeatedly or across different buffer areas, resulting in a larger number of available exposures. Notably, lifestyle factors are generally excluded from the exposome database, as they are important endpoints in health research as well. The database draws from multiple sources. First, multiple waves of questionnaires in the cohort, conducted with twins, their parents, and teachers, collected detailed information on family, peers, teachers, schools, and related factors. Second, the work reported in this thesis has been intensely involved in an EU-funded exposome project, Equal-Life (www.equal-life.eu), which aims to evaluate the effects of physical and social environmental exposures at different levels on mental health and development in children and adolescents. European-wide collaborators have enriched the database from open-source platforms like the Copernicus program, and a part of air pollution exposures were calculated via land use regression models, as described in previous studies (56, 131). These exposures are available at the geocode level and are linked to participants via their residential history. Third, social indicators were obtained from Statistics Finland. The reference time point for population and household data is the last day or week of the statistical year. Indicators are presented at the postal code or municipality level (2022 version). The map from Esri Finland containing 2022's postal code and municipality information was used to link the twins and social indicators. Table 1 summarizes the exposome data available for FinnTwin12. Due to temporality (e.g., noise) or because some studies were completed before the

exposure enrichment (e.g., some air pollution exposures), studies II and III used a subset of exposures from the database.

Table 1 The Summary of the exposome databased of FinnTwin12

Domain	Number of exposure type	Earliest available time	Source	Measurement level
Air pollution	4	1990	Land use regression (131)	Geocode
Building	3	1990	Copernicus	Geocode
Land use	10	2006	Copernicus	Geocode
Noise	2	2019	Common Noise Assessment Methods in Europe (132)	Geocode
View factors	3	Median: 2019	Google Street View	Geocode
Roads, streets, and traffic	24	2014	OpenStreetMap	Geocode
Green, blue spaces, and parks	41	1984	Copernicus, OpenStreetMap, USGS/NASA Landsat	Geocode
Population density	2	1990	Copernicus, WorldPop	Geocode
Geocoordinates	3	Since birth	Copernicus, Digital and population data services agency	Geocode
Prenatal exposures	7	Prenatal	Cohort collection	Parental questionnaire
Passive smoking	1	< age 1	Cohort collection	Twin or parental questionnaire
Family and parents	51	age 11-12	Cohort collection	Twin or parental questionnaire
Friend and romantic relationships	12	age 11-12	Cohort collection	Twin questionnaire
School and teachers	4	age 11-12	Cohort collection	Twin questionnaire
Stressful life events	16	age 14	Cohort collection	Twin questionnaire
Social indicators	53	1987	Statistics Finland	Postal code, municipality

In Study II, 385 exposures were used and reclassified into 12 domains: air pollution, buildings, blue and green spaces, population density, geocoordinates, prenatal exposures, passive smoking, family and parents, friendships and romantic

relationships, school and teachers, stressful life events, and social indicators. These exposures were collected by or before age 17 or the year 2006 to ensure temporality, and repeated measures were included as well.

In Study III, we selected 60 exposures whose measured periods were as close as possible to mid-adolescence (age 14 in FinnTwin12 and baseline in WALNUTs) to compose the mid-adolescence exposome. They were classified into seven concept areas: built environment, community social environment, family social environment, indoor environmental quality, natural environment, lifestyle, and outdoor environmental quality. The new classification follows the Equal-Life framework (20), which includes lifestyle factors as well. WALNUTs employed a similar approach to collect exposures. As a partner of Equal-Life, collaborators enriched some exposure data in the same way as in FinnTwin12, such as those related to the built environment, using data from the WALNUTs study year or from 2012 to 2015. Exposures from the concept areas of neighborhood social environment for WALNUTs were obtained from the Spanish local authority, either in the study year or in 2011. Exposures related to the family social environment, indoor environmental quality, and lifestyle were collected through baseline questionnaires (56). Harmonization between the two cohorts has been conducted.

For both Studies II and III, categorical exposures were appropriately regrouped. Exposures with over 20% missing were excluded, and the rest missing exposure values were imputed by the median.

4.3.2 Depression measurement

The short-version General Behavior Inventory (GBI) in Finnish was used to assess depressive symptoms at age 17 and in young adulthood (133). This self-reported inventory is designed to identify the presence and severity of mood-related behaviors, including depressive and manic symptoms. The short version consists of 10 items and uses a 4-point Likert scale ranging from 0 (never) to 3 (very often) (see Appendix 1 for questions). The sum scores were obtained, ranging from 0 to 30, with higher scores reflecting more occurrence of depressive symptoms. The changes in GBI scores from age 17 to young adulthood were also calculated. Due to the skewness of the GBI scores at both ages, we added one to the GBI score and log-

transformed it. The GBI was inquired for all the participants in the waves of age 17 and young adulthood in FinnTwin12. The GBI scores were used in Studies I and II.

The Finnish version of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) yields the diagnosis information on MDD in young adulthood, following the criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (134). Among the twins in the intensive study in the wave of young adulthood, the SSAGA interviews were administered by trained nurses and Masters of Health Science or via telephone if a twin could not travel to Helsinki (135). The MDD diagnosis was further used to validate the GBI score; the GBI score in young adulthood strongly predicted MDD diagnosis in a logistic regression analysis ($N=1232$), with an area under the receiver operating characteristic curve of 0.83. Study II used the incidence of MDD.

4.3.3 Assay of proteome (biomarker of mental health)

Blood plasma samples were obtained from the intensive samples in FinnTwin12 (young adulthood) and a part of participants in WALNUTs (baseline, age 11-16) to profile the proteome. The extraction, processing, identification, and quantification of the proteome and further batch correction were conducted by collaborators at the University of Eastern Finland, which has been described in several publications (117, 119, 136). Then, based on three previous proteome studies on both cohorts, 111 proteins have been identified that were associated with adolescent mental health conditions, and these conditions were externalizing and internalizing (including depression) symptoms (117–119). The coding genes for these proteins were further validated using meta-analyzed genome-wide association results from FinnGen, UK Biobank, and Estonia Biobanks (release 11), ensuring biological plausibility with mental health and psychiatric outcomes (137). Validation was confirmed through significant associations ($P\text{-value} < 5 \times 10^{-8}$) between protein-coding genes and psychiatric-related phenotypes, including depression and the use of depression medications. Proteins identified in at least two studies were also considered validated. A total of 35 proteins met this criterion. After excluding proteins that were unavailable or had more than 10% missing data, 26 and 31 proteins were

included in FinnTwin12 and WALNUTs, respectively, as biomarkers of mental health.

4.3.4 Covariates

A group of covariates was identified a priori to control for potential confounders in the population-based epidemiological analysis. In FinnTwin12, age, sex, zygosity, smoking status, study and work status, and parental education levels were included. Age variables were aligned with the timing of mental health or plasma sampling. Sex was assigned at birth. Zygosity (MZ, DZ, unknown) was determined by well-validated physical similarity questions and further validated by DNA testing (138). Smoking status (never, former, occasional, current) was self-reported at age 17 and again in young adulthood. At age 17, when most twins were still in school, study and work status was categorized as: neither study nor work, study only, or work only. In young adulthood, this was split into two variables: work status (full-time, part-time, irregular, or not working) and secondary-level education (vocational, senior high school, or none). Maternal and paternal education levels (missing, low, medium, high) were initially reported by parents in wave one. The two levels were then combined into a new variable: parental education level (139). If both parents had low (less than high school, 12 years) or high (12 or more years) education, the parental education was classified as "limited" or "high," respectively. If parents had different education levels, it was classified as "intermediate." In WALNUTs, only sex and age at plasma sampling were included. The inclusion of covariates, besides sex, zygosity, and age, was based on their correlations with the environment and mental health (140–142). Different sets of covariates are used in different analyses (more information on the usage of covariates in the below section 4.5).

4.4 Hypothesis

For specific aim 1, the working hypothesis was that an earlier age at parental separation is associated with more severe depressive symptoms in adolescence and young adulthood. Additionally, higher levels of exposure are hypothesized to lead to more depressive symptoms, serving as a template hypothesis given a wide range of exposures included. For specific aim 2, the template hypothesis was that higher

levels of exposure were associated with higher levels of proteomic biomarkers of mental health, since the analysis involved a range of exposures and biomarkers. For specific aim 3, generally, we hypothesized a genetic contribution underlying the observed associations between exposures and depression-related outcomes. Although hypothesis-free ExWAS was employed for specific aims 1 and 2, these practical template hypotheses were still made for better understanding and interpretation.

4.5 Data analysis

Due to availability and temporality, the participants and exposures included in each analysis varied. Table 2 presents the characteristics of the analyses in Sections 4.4.1 and 4.4.2.

Table 2 Characteristics of analyses

Study	Cohort	Depression outcome	Age at depression assessment	Number of exposures	Number of individual twins	Analysis
I	FinnTwin12	GBI	Age 17	1 (age at separation)	2725	Linear regression
			Young adulthood	1 (age at separation)	2976	Linear regression
			Age 17, young adulthood	1 (age at separation)	2725 (age 17), 2976 (young adulthood)	MMRM
		Change of GBI	/	1 (age at separation)	2700	Linear regression
II	FinnTwin12	GBI	Age 17	385	4127	ExWAS
			Young adulthood	385	3025	ExWAS
		Incidence of MDD	Young adulthood	286	1236	ExWAS
III	FinnTwin12	26 proteomic biomarkers of mental health	Young adulthood	60	748	ExWAS
	WALNUTs	31 proteomic biomarkers of mental health	Age 11-16 (baseline)	60	187	ExWAS

Abbreviations: general behavior inventory (GBI), major depressive disorder (MDD), mixed model for repeated measures (MMRM), exposome-wide association studies (ExWAS)

4.5.1 Regression between the age at separation and depressive symptoms (Study I)

First, linear regression models were used to assess the association between age at separation and log-transformed GBI scores at age 17, in young adulthood, and changes in GBI scores over time. Covariates were adjusted: sex, zygosity, smoking status, work status, secondary school education, paternal education level, maternal education level, and age (for outcomes of the GBI score in young adulthood and the changes). Participants without information on these variables and residential mobility (a measure in the publication but not considered in the thesis) were dropped. Robust standard errors were applied to account for the twin pair clustering effect.

Second, a mixed model for repeated measures (MMRM) was employed to visualize the trajectory of GBI scores by age at separation. The MMRM, a type of direct-likelihood analysis, uses longitudinal data to estimate the covariance between data from different time points. This method allows us to evaluate the effect of the exposure of interest (age at separation, treated as a fixed effect) on the outcome (GBI in young adulthood) while accounting for baseline GBI at age 17 (random effect). The same set of covariates was adjusted, and robust standard errors were also applied.

The analyses were only conducted on FinnTwin12 using Stata 17.0 (StataCorp, College Station, TX, USA).

4.5.2 Exposome-wide association studies (ExWAS) (Studies II-III)

Generalized linear regression models with Gaussian or binomial distributions were fitted to examine the associations between each exposure and outcomes of log-transformed GBI scores at age 17 and in young adulthood and the incidence of MDD in young adulthood in FinnTwin12. Bonferroni correction by the number of effective tests (calculated by PCA) was used to adjust for multiple testing and account for correlations between exposures (143). In the ExWAS for the log-transformed GBI score at age 17, the models were adjusted for sex, zygosity, parental education level, smoking status reported at age 17, and study and work status. In the ExWASes for the log-transformed GBI score and incidence of MDD in

young adulthood, the covariates included sex, zygosity, parental education level, smoking status reported in young adulthood, work status, secondary education, and age at the time of GBI assessment. Robust standard errors were applied as well.

Then, generalized linear regression models with Gaussian distribution were fitted between each exposure and each protein (ExWAS for each protein). Bonferroni correction by the number of effective tests was used to adjust for multiple testing as well. Because many abovementioned covariates were regarded as exposures in the ExWAS for proteins, age at plasma sampling and sex were the only two covariates included in either cohort. Robust standard errors were applied too. The analysis was performed separately for FinnTwin12 and WALNUTs. Of note, the measurement time of exposures varied, which means that some of the regression models are essentially cross-sectional (i.e., a model between exposures collected from the baseline questionnaire and proteins in WALNUTs). The inclusion criteria of participants in ExWAS were: (1) twins had available residential history, (2) twins and their families had completed at least one questionnaire in any wave, and (3) twins had available outcome and covariate information.

The rexposome package in the R environment (version 4.3.3) was used (144).

4.5.3 Twin analysis (Studies I-III)

In Study I, to evaluate the potential genetic influence between separation and depressive symptoms, intrapair correlation coefficients were calculated at first as an exploration. Differences in the coefficients between zygosity groups indicated the differential contribution of genetic and environmental factors to variances of GBI scores. Stratification by age at separation was conducted to show its effects, such as modification, on the genetic and environmental contribution. Then, a bivariate cross-lagged path model was fitted with separation status and within-pair differences of GBI score (similarity) at both ages with the maximum likelihood estimator (the twin pair as the unit of observation). This gauged the structural relationships of repeatedly measured variables and tested both directions between differences and separation, which could suggest a potential role of genetics (145). If the within-pair difference leads to separation, genetic effects are suggested, while if separation leads to the within-pair difference, we should consider more

environmental roles on the phenotype. Three covariates were adjusted for sex and zygosity combination (male monozygotic, female monozygotic, male dizygotic, female dizygotic, and opposite-sex dizygotic), paternal education, and maternal education. Twin pairs whose zygosity was unknown or who had missing information on separation status and within-pair differences in GBI score at both ages were excluded. Stata 17.0 (StataCorp, College Station, TX, USA) and Mplus 8.6 (Muthen & Muthen, Los Angeles, CA, USA) were used.

In Study II, before twin modeling, confirmatory factor analysis (CFA) was conducted based on the significant exposures identified (associated with depression) to estimate the exposome score for each outcome. Following the concept of the exposome, a one-factor structure was specified as theory-driven. The CFA assumed the correlations between exposures due to the exposome score, which were verified using structural equation modeling. The score was estimated using maximum likelihood, and model fit was evaluated with the standardized root mean square residual. The cluster effect was controlled. For categorical exposures, the entire exposure was included instead of subgroups. To maximize power, all available twins were included at each time point in the CFA. Stata 17.0 (StataCorp, College Station, TX, USA) was used.

This twin model decomposed the variance of traits into additive genetic (A), dominant genetic (D), common environmental (C), or unique environmental (E) factors (146). The intrapair correlation coefficients for GBI scores in DZ (0.22 in young adulthood and 0.16 at age 17) and MZ pairs (0.52 in young adulthood and 0.51 at age 17) supported the use of an ADE model. After restricting the analysis to full twin pairs, the saturated model via constrained means and variances suggested that the assumptions of equal means and variances for twin order and zygosity were generally met, and further sex limitation in the saturated model highlighted a significant sex difference. Among univariate twin models (ADE, AE, and E models) for the log-transformed GBI score in four scenarios (two time-points and two sexes), AE models showed the best fit in general by Akaike information criterion. Consequently, the bivariate Cholesky AE model was applied for the exposome score and log-transformed GBI score (146). Latent factors, Aexposome and Eexposome, influence both the exposome score through pathways a_{11} and e_{11} and the log-

transformed GBI score through pathways a_{21} and e_{21} . Latent factors, AGBI and EGBI, influence only the log-transformed GBI score by pathways a_{22} and e_{22} . The percentage of covariance between the exposome score and log-transformed GBI scores by unique environmental factors could be calculated as: $(e_{21} * e_{11}) / (e_{21} * e_{11} + a_{21} * a_{11})$. Age, reported in the young adulthood survey, was adjusted in univariate and bivariate models for GBI in young adulthood. Sex stratification was performed. The analysis was performed using the OpenMx package in R (version 4.2.3) (147).

In Study III, based on the significant exposure—protein sets, within-pair fixed-effect linear regression models were employed, representing associations conditional on exposures shared within pairs and controlling for unobserved individual heterogeneity from genetics and shared environments (148). The model removes variation between cotwins as the fixed effects. Age at plasma sampling and sex were controlled for. Standard errors were robusted for the design effect of the twin cohort. After only keeping full twin pairs, cross-twin cross-trait (CTCT) correlations (149), stratified by zygosity, were calculated to show the magnitude that exposure in cotwin 1 was correlated with the corresponding protein level in cotwin 2. It shows the extent to which covariations between exposures and proteins are attributed to genetic and environmental factors (149). If the CTCT correlations were higher for MZ pairs than for DZ twin pairs, especially two-fold higher, it indicates a strong influence of genetic effects on the association between the exposure and protein. Stata 17.0 (StataCorp, College Station, TX, USA) and OpenMX package in the R environment (version 4.3.3) were used (147).

5 Results

5.1 Characteristic of participants included in each study

Study I included 3071 twins in the final analysis, consisting of 1284 full twin pairs and 503 individual twins. Most participants were female (56.4%), dizygotic (60.6%), and reported never smoking in young adulthood (53.5%). Most twins reported that they had received a senior high school education (60.2%) and were engaging in full-time work (51.5%) in the young adulthood follow-up.

In Study II, 3025 and 4127 individual twins were included in ExWASes for outcomes of GBI in young adulthood and at age 17, respectively (the samples were not completely overlapping). Most participants were female, were from dizygotic pairs, and had parents with limited education (less than high school). At age 17, 25.4% of twins reported being current smokers, and 82.6% were full-time students not working. In young adulthood, 25.3% reported being current smokers, and 51.4% were employed full-time. Besides, 1236 of 3025 individual twins were in ExWAS for the outcome of the incidence of MDD in young adulthood.

In Study III, 748 individuals were included from the FinnTwin12 cohort (40.6% female) and 187 individuals from the WALNUTs cohort (55.6% female), with mean ages of 22.3 years (standard deviation (SD): 0.6) and 13.3 years (SD: 1.0), respectively, at the time of plasma sampling for proteome assessment. In FinnTwin12, most fathers (61.2%) and mothers (68.3%) had received a medium level of education. However, in WALNUTs, more fathers (56.7%) and mothers (64.7%) had received a high level of education.

5.2 Association between age at separation and depressive symptoms (Study I results)

In linear regression adjusted for covariates (Table 3), twins who separated between ages 17 and 19.5, between ages 19.5 and 22, and after age 22 had significantly lower log-transformed GBI scores at age 17, compared to those who separated before age 17. Additionally, twins who separated between ages 19.5 and 22 and after age 22 also had lower log-transformed GBI scores in young adulthood. There was no significant association between age at separation and change of GBI.

Table 3 Association of age at separation with depressive symptoms (GBI) using linear regression

Age at separation	log-transformed GBI score at age 17		log-transformed GBI score in young adulthood		Change of GBI	
	N. (%)	Adjusted β (95% CI) ^a	N. (%)	Adjusted β (95% CI) ^b	N. (%)	Adjusted β (95% CI) ^b
Before age 17	110 (4.04)	Ref.	131 (4.40)	Ref.	107 (3.96)	Ref.
Age 17 to <19.5	967 (35.49)	-0.16 (-0.31, -0.02)*	1078 (36.22)	-0.17 (-0.35, 0.00)	959 (35.52)	-0.22 (-1.19, 0.75)
Age 19.5 to <22	1146 (42.06)	-0.22 (-0.37, -0.08)*	1228 (41.26)	-0.19 (-0.37, -0.02)*	1133 (41.96)	0.04 (-0.93, 1.00)
After age 22	502 (18.42)	-0.44 (-0.60, -0.28)*	539 (18.11)	-0.25 (-0.44, -0.06)*	501 (18.56)	0.67 (-0.34, 1.67)

^a Adjusted for sex, zygosity, smoking, work status, secondary level school, paternal education, and maternal education; for number of moves and total distance of moving before 17

^b Adjusted for sex, zygosity, smoking, work status, secondary level school, paternal education, maternal education, and age when twins provided the GBI assessment in young adulthood

*P < 0.05

In MMRM (Table 4), after controlling for log-transformed GBI score at age 17 and covariates, twins who separated between ages 17 and 19.5, between ages 19.5 and 22, and after age 22 scored significantly lower on the GBI (log-transformed) in young adulthood than those who separated before age 17.

Table 4 Association of age at separation with depressive symptoms (GBI) using MMRM

Age at separation	Mean (SD)		Adjusted β (95% CI) ^a
	GBI at age 17	GBI in young adulthood	Log-transformed GBI score in young adulthood
Before age 17	6.77 (5.02)	6.20 (5.81)	Ref.
Age 17 to <19.5	5.91 (5.28)	4.81 (4.89)	-0.19 (-0.34, -0.03)*
Age 19.5 to <22	4.91 (4.63)	4.17 (4.37)	-0.21 (-0.36, -0.05)*
After age 22	3.65 (3.98)	3.79 (4.50)	-0.26 (-0.42, -0.09)*

a Adjusted for sex, zygosity, smoking, work status, secondary level school, paternal education, maternal education, and age when twins provided the GBI assessment in young adulthood

*P < 0.05

5.3 Association between the exposome and depressive symptoms (Study II results)

After correcting for multiple testing, the significance thresholds for p-value were 3.09×10^{-4} , 3.39×10^{-4} , and 3.63×10^{-4} for the ExWAS with outcomes of GBI in young adulthood, the incidence of MDD in young adulthood, and GBI at age 17, respectively.

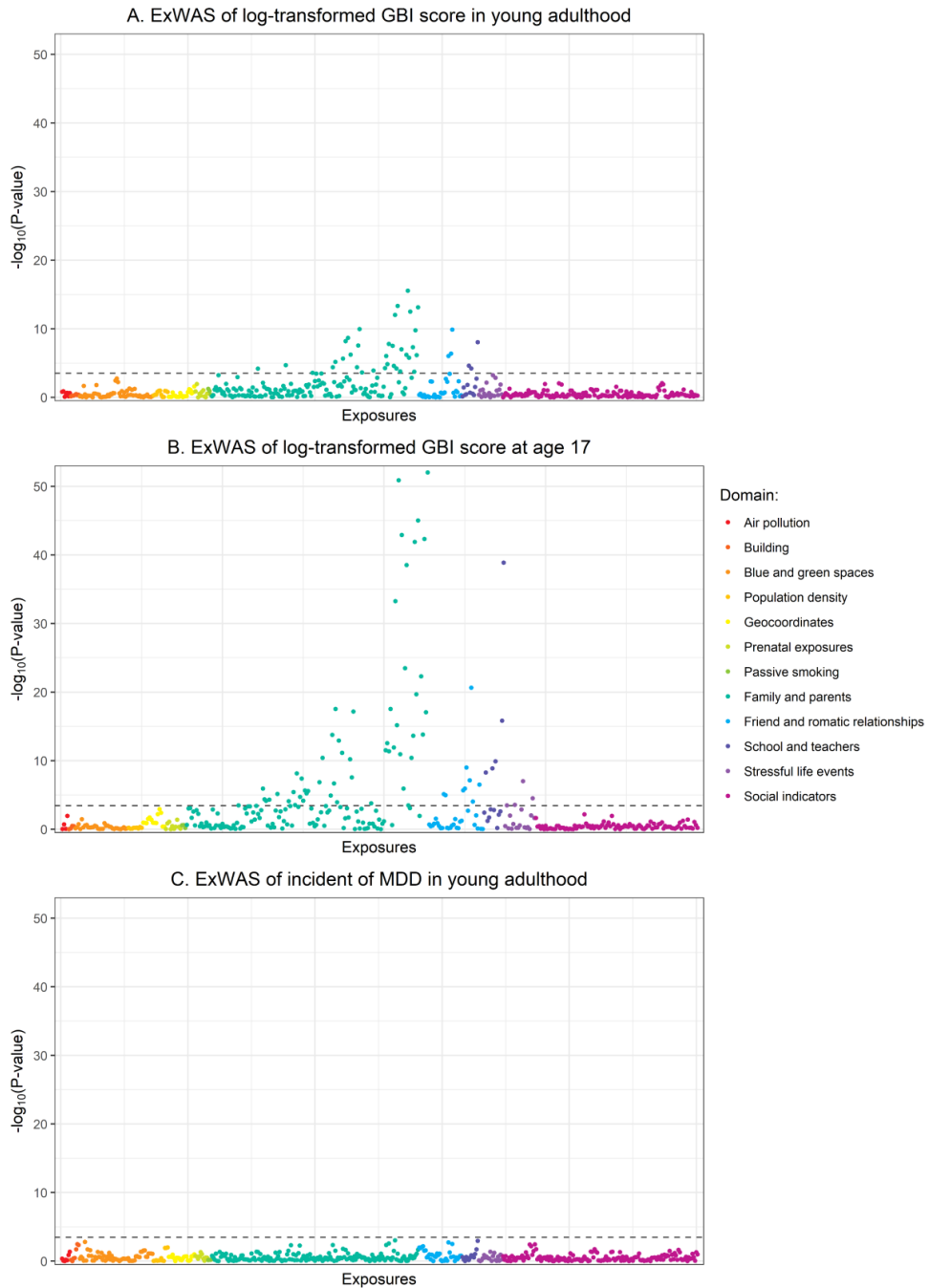
Out of 385 exposures examined, 29 were significantly associated with log-transformed GBI scores in young adulthood (Figure 1A). These exposures belonged to three domains: family and parents (24), friendships and romantic relationships (2), and school and teachers (3). The most protective exposure was the perception of fairness in the home environment at age 17. Twins who felt their home environment was "not at all unfair" had a 0.40 lower log-transformed GBI score (95% CI: -0.50, -0.31) compared to those who perceived it as "completely, quite, or somewhat unfair." The most harmful exposure was dissatisfaction with friendships at age 14. Twins who were "somewhat, mainly not, or not at all satisfied" with their relationships with friends had a 0.42 higher log-transformed GBI score (95% CI: 0.29, 0.55) compared to those who were "completely satisfied."

There were 46 of 286 exposures significantly associated with log-transformed GBI scores at age 17 (Figure 1B). These exposures were from the domains of family and parents (2), friendships and romantic relationships (6), school and teachers (4), and

stressful life events (4). The most protective exposure at age 17 remained the same as in young adulthood; twins who perceived their home environment as "not at all unfair" had a -0.50 lower log-transformed GBI score (95% CI: -0.57, -0.43) compared to those who felt it was "completely, quite, or somewhat unfair." The most harmful exposure at age 17 was dissatisfaction with success in work or studies. Twins who were "mainly not or not at all satisfied" with their success at work or studies had a 0.65 higher log-transformed GBI score (95% CI: 0.55, 0.74) compared to those who were "completely satisfied."

No exposure showed a significant association with the incidence of MDD in young adulthood (Figure 1C).

Picture 1 Associations of the exposome with log-transformed GBI score in young adulthood (individual twin n = 3025) (A), at age 17 (individual twin n = 4127) (B), and incidence of MDD in young adulthood (individual twin n = 1236) (C) in Study II (modified from Fig. 2, Extended Data Fig. 2 and 3 of the publication of the Study II)



Note: Manhattan association plots for exposures in relation to three depression outcomes in Study II. The y-axis shows statistical significance as $-\log_{10}(\text{P value})$ and the dashed gray horizon line presents the significance threshold adjusted for multiple testing. The significance thresholds for p-

values were 3.09×10^{-4} , 3.39×10^{-4} , and 3.63×10^{-4} for the ExWAS with outcomes of GBI in young adulthood, the incidence of MDD in young adulthood, and GBI at age 17, respectively. The adjusted covariates for outcomes in young adulthood were sex, zygosity, parental education, smoking in young adulthood, work status in young adulthood, secondary-level school in young adulthood, and age when twins provided the GBI assessment in young adulthood. The adjusted covariates for the outcome at age 17 were sex, zygosity, parental education, smoking at age 17, and study and working status at age 17. Abbreviations: general behavior inventory (GBI), major depressive disorder (MDD), exposome-wide association studies (ExWAS).

5.4 Association between the mid-adolescence exposome and proteomic markers of mental health (Study III results)

After correcting for multiple testing, the significance thresholds for p-value were 1.84×10^{-3} in the FinnTwin12 cohort and 2.42×10^{-3} in the WALNUTs cohort.

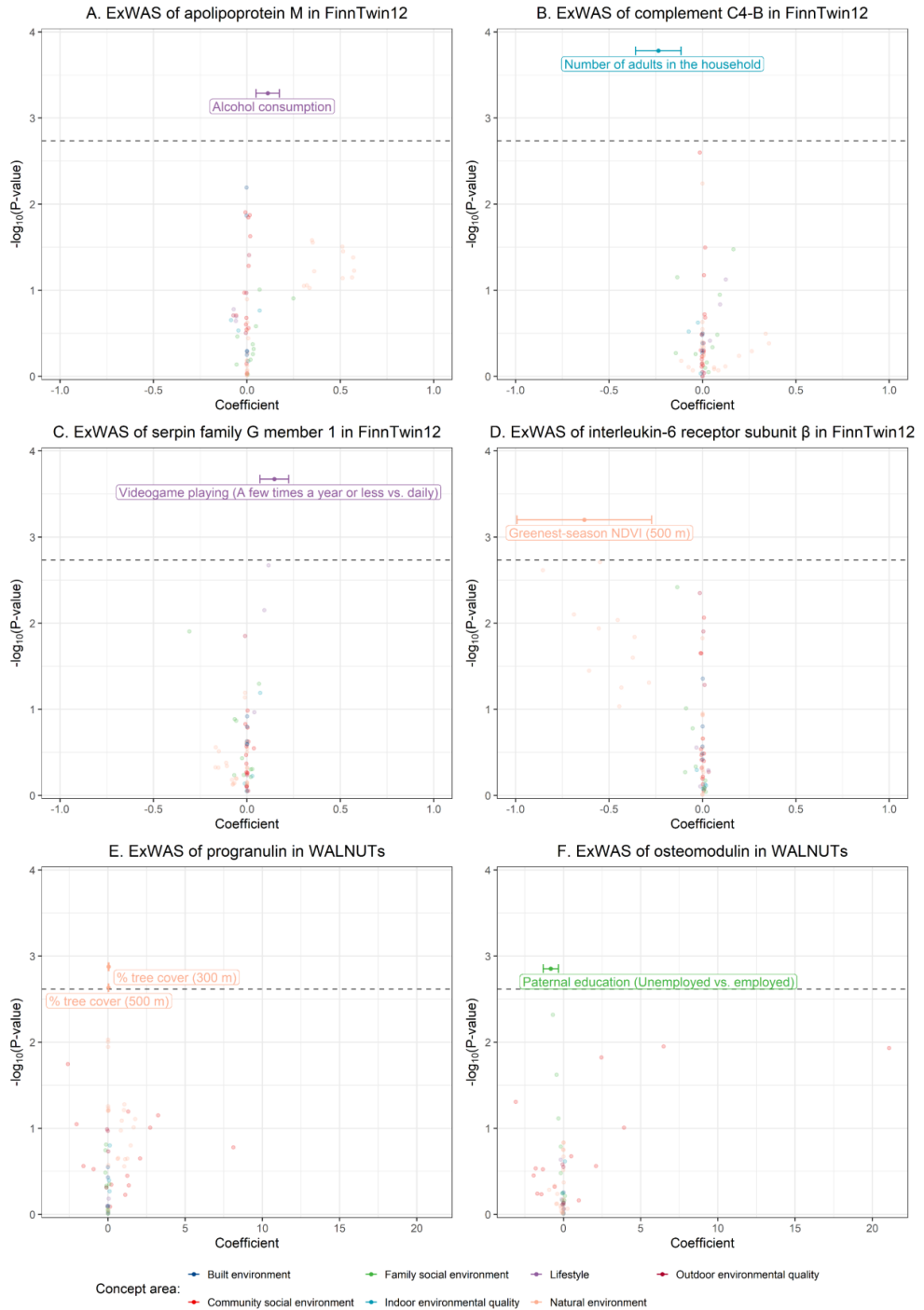
In FinnTwin12, four exposures from three concept areas (two in lifestyle, one in indoor environmental quality, and one in natural environment) were significantly associated with four proteins. Participants who reported alcohol consumption at age 14 had higher levels of apolipoprotein M compared to non-consumers (coefficient: 0.11, 95% CI: 0.05, 0.17) (Figure 2A). A higher number of adults in the household (larger household size) at age 11–12 was associated with a low level of complement C4-B (coefficient: -0.34, 95% CI: -0.36, -0.11) (Figure 2B). Twins who played video games daily at age 14 had a higher level of serpin family G member 1 compared to those who played a few times a year or less (coefficient: 0.15, 95% CI: 0.07, 0.22) (Figure 2C). A higher level of greenest-season Normalized Difference Vegetation Index (NDVI) within a 500 m buffer was associated with a lower level of interleukin-6 receptor subunit β (coefficient: -0.63, 95% CI: -0.99, -0.27) (Figure 2D).

In WALNUTs, higher percentages of the area covered by trees within 300 m (Beta: 0.04, 95% CI: 0.01, 0.06) and 500 m (coefficient: 0.03, 95% CI: 0.01, 0.05) buffers, from the concept area of the natural environment, were associated with higher levels of progranulin (Figure 2E). Participants whose fathers were unemployed had a lower level of osteomodulin (coefficient: -0.82, 95% CI: -1.32, -0.32) compared to

those whose fathers were employed (concept area of the family social environment) (Figure 2F).

Further interpretation of the exposure's effect on mental health needs to integrate with previous proteome studies (117–119). The complete interpretation can be found in the Discussion section.

Picture 2 Associations of the exposome with apolipoprotein M (A), complement C4-B (B), serpin family G member 1 (C), interleukin-6 receptor subunit β (D) in FinnTwin12, and progranulin (E) and osteomodulin (F) in WALNUTS in Study III (Modified from Fig. 2 of the preprint version of Study III)



Note: Only ExWASes with significant associations were plotted. The y-axis shows statistical significance as $-\log_{10}(p\text{-value})$ for the adjustment for multiple testing and x-axis represent the

estimated coefficient. Adjustments were made for sex and age at plasma sampling for proteome assessment. The significance thresholds for p-values were 1.84×10^{-3} for FinnTwin12 and 2.42×10^{-3} for WALNUTs (the dark gray dash line). The error bars indicate 95% CI for significant exposures only. Abbreviations: exposome-wide association studies (ExWAS), Normalized Difference Vegetation Index (NDVI).

5.5 Outputs from twin analysis

5.5.1 Relationship between separation and within-pair similarity on depressive symptoms (Study I)

The overall intrapair correlation coefficients for GBI scores among MZ and DZ pairs were 0.56 and 0.14 at age 17 and 0.52 and 0.22 in young adulthood, respectively. Regardless of the age at separation, the intrapair correlation coefficients for GBI scores at age 17 in MZ pairs were consistently twice as high as those in DZ pairs. In twin pairs who separated after age 17, the intrapair correlation coefficients for GBI scores in young adulthood in MZ pairs were also always higher in DZ pairs. However, the coefficients did not differ too much between MZs (0.55) and DZs (0.53) in twin pairs who separated before age 17.

In the cross-lagged path model with only full twin pairs, within-pair differences in GBI at age 17 were significantly associated with separation status before age 22 (coefficient: 0.01) (Figure 3A).

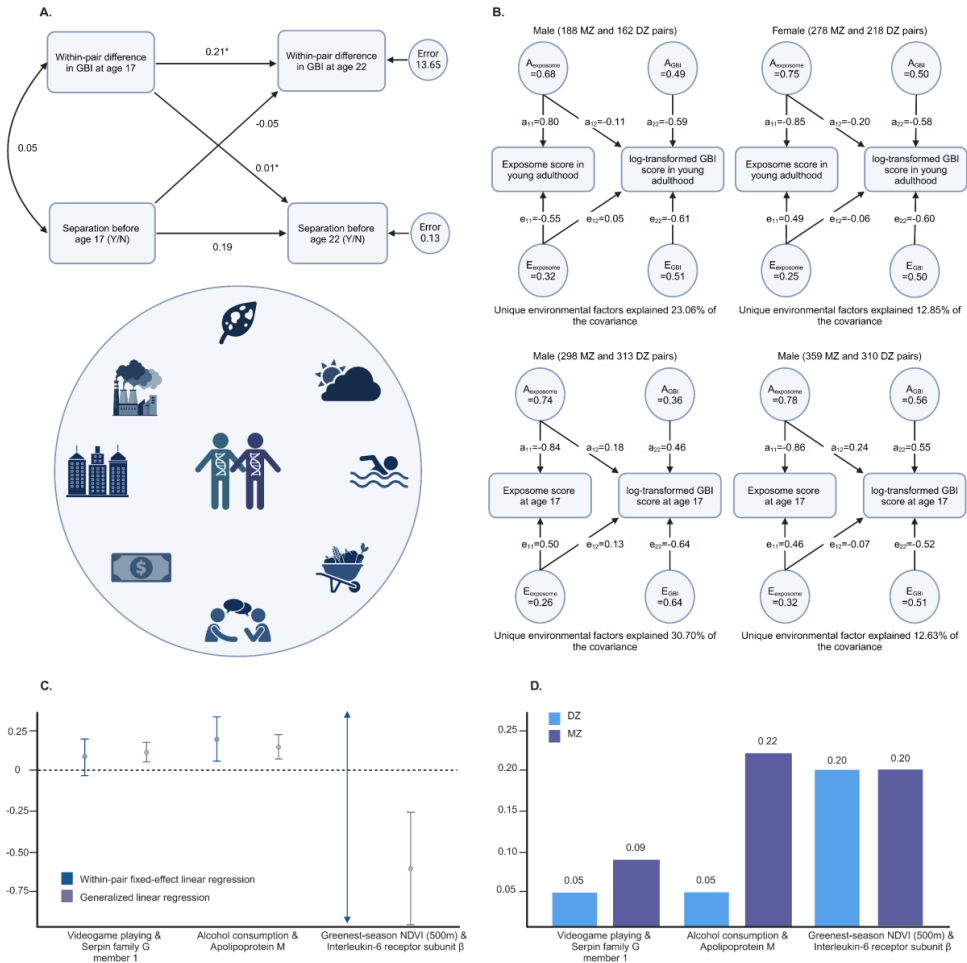
5.5.2 Bivariate twin modeling between the exposome score and depressive symptoms (Study II)

The standardized root mean square residuals of CFA models of exposome score in young adulthood and at age 17 were 0.100 and 0.078, respectively, indicating acceptable fit. Among 350 male and 496 female twin pairs, unique environmental factors explained 23% and 13 % of the covariance between the exposome score and log-transformed GBI score in young adulthood for males and females, respectively (Figure 3B). Among 611 male and 669 female pairs, unique environmental factors accounted for 31% and 13% of the covariance between the exposome score and log-transformed GBI score at age 17 in males and females, respectively (Figure 3B).

5.5.3 Twin analysis between significant exposure—protein sets (Study III)

In the within-pair fixed-effect regression model, videogame playing at age 14 was still significantly associated with serpin family G member 1 (coefficient: 0.19, 95% CI: 0.05, 0.34) (Figure 3C). After keeping only full twin pairs, the CTCT correlation coefficient between videogame playing and serpin family G member 1 in MZ pairs (0.22) was nearly 5 times higher than in DZ pairs (0.05) (Figure 3D). The CTCT correlation coefficients between alcohol consumption and apolipoprotein M were 0.09 in MZ pairs and 0.05 in DZ pairs (Figure 3D). The CTCT correlation coefficients between greenest-season NDVI within a 500 m buffer and interleukin-6 receptor subunit β were similar in MZ (0.20) and DZ pairs (0.20) (Figure 3D). Household size was not included in the analysis, since it was reported by parents (no variation between cotwins).

Picture 3 Twin analyses (integrated from the figures involving twin analyses of Studies I-III)



Note: **A:** the cross-lagged path model for within-pair differences in depressive symptoms (GBI) and separation status (1080 twin pairs) in Study I. Sex and zygosity combination were adjusted. Numbers over the lines indicate the regression coefficients. *: P-value<0.05. **B:** bivariate Cholesky AE model for the exposome score and log-transformed GBI score in young adulthood and at age 17 in Study II. A, standardized variance of additive genetic effect; E, standardized variance of unique environmental effect. The a and e stand for pathway coefficients from A and E, respectively, to both the exposome score and log-transformed GBI score. Age, reported in the young adulthood survey, was adjusted in models involving GBI in young adulthood. **C:** the forest plot based on within-pair fixed-effect and generalized linear regression models in Study III (749 individual twins). The comparison was between participants playing video games every day and a few times a year or less (ref.) for the association between videogame playing and serpin family G member 1. **D:** the histogram plot for cross-twin cross-trait correlation in Study III (325 twin pairs). Categorical

exposure of videogame playing was regarded as a complete variable in the correlation calculation.
Abbreviations: general behavior inventory (GBI), monozygotic (MZ), dizygotic (DZ).

6 Discussion

6.1 Main findings of the thesis

Under the exposome framework, this thesis provides a systematic and comprehensive investigation of environmental influences on depression in late adolescence and young adulthood, based on the FinnTwin12 cohort. A pronounced effect from family and parent (based on the proxy of the familial exposome, age at separation, and a range of related exposures) on the depressive symptoms was revealed. Exposures related to friendships, romantic relationships, school environments, and stressful life events were significantly associated with depressive symptoms, as well. Supported by the additional WALNUTs cohort, mid-adolescence exposures related to lifestyle, indoor environmental quality, the natural environment, and the family social environment were linked to proteomic biomarkers of mental health, while the effect from greenspaces was consistently observed in both countries. Leveraging the twin study design, the thesis also uncovered the role of genetics in the separation of cotwins and in shaping the covariation between the environment and depression. The genetic contribution varied by time, sex, and type of exposure. Genetic and environmental effects on depression coexisted and interrelated, which should be considered in future large-scale research.

6.2 Interpretation of the results and comparison with previous literature

The thesis underscores the significant impact of familial and parental environments on depression, first reflected through age at separation. In twin studies, age at separation is typically used as a proxy for the common environment, which is expanded hereby to the length of time staying with siblings and parents to reflect the broader familial exposome. In Ohio, US, young adults (ages 18–24) who

continued living with their parents reported fewer depressive symptoms than those living independently or who had returned to the parental home (150). In a small study among women in New Zealand, early home leaving predicted conduct disorder but not depression at age 21 (151), while there can be comorbidity between conduct disorder and depression later in life (152). In Germany, Seiffe-Krenke has observed that those who left home "in time" at age 21 (female) and age 23 (male) had worse psychological status than those who stayed at home (153). Separation itself is also a hallmark of youth, as the transition from a familiar environment to a new phase. In this study, most twins separated between ages 17 and 22, often associated with significant life changes, such as entering university or becoming employed. These transitions unavoidably influence mental health negatively or positively depending on the familial context. For instance, after separation, individuals with warm and supportive sibling relationships may experience minimal emotional loss, whereas siblings with conflict may feel relief (154).

Some familial factors like closeness were linked to the timing of leaving home (153, 155). Under the growing environment proxied by age at separation, ExWAS investigated the effects from numerous familial factors thoroughly. Consistent with our findings, a 2014 systematic review has identified factors such as authoritarian parenting, autonomy granting, and aversive behaviors associated with depression among adolescents aged 12–18, drawn from 164 studies (156). A large survey among university students in China has found that familial factors such as cohesion, conflict, and control correlated with depressive symptoms (157). Unfavorable family dynamics, such as less family triangulation (e.g., parent–child alliances) and dissatisfaction, were associated with more depressive symptoms from childhood to late adolescence, demonstrating a developmental trajectory (158, 159). Perceived family fairness, a dimension of parentification referring to equity, reciprocity, and acknowledgment, was identified as a critical protective factor for adolescent mental health (160–162), which the thesis found as the most protective factor for depressive symptoms at both ages. One major difference between the present thesis and previous family studies is the use of single items to represent different aspects rooted in the scale instead of a composite score, which helps to detect new correlations and rank their importances. Additionally, the frequency of visiting family and friends was associated with a lower odd of depression from ExWAS in

the general UK population, with causal inference (9). In the biomarker analysis in WALNUTs, a lower paternal education level was associated with a lower level of osteomodulin, which could further be associated with more emotional and behavioral problems among adolescents (after connecting with previous literature) (118). It aligns with experimental findings in marmosets that hand-reared infants exhibited altered osteomodulin gene expression compared to those reared by parents, underscoring the importance of parenting in early life (163).

Exposures from the domain of friend and romantic relationships, school and teachers, and stressful life events are also important. A large systematic review of over 200 studies has found small but significant concurrent and longitudinal associations between negative youths' friendship experiences (e.g., fewer friends, poor friendship quality) and more depressive symptoms (164). Some studies suggested that friendships may protect against adverse exposures like bullying, but the evidence is not consistent (165). Peer support has been linked to reduced depression among university students and non-student young adults (166). In the thesis, popularity among the opposite sex reduced depressive symptoms. In Canadian adolescents, the longitudinal depression profiles were predicted by adolescents' relations with their opposite-sex peers (167). Based on a representative sample of in-school students aged 15-24 in Switzerland, self-reported easiness of making opposite-sex friends was associated with risk behaviors, which risk behaviors are highly correlated with depression (168). Adolescents and young adults spend a lot of time in school. A review by Kidger et al. has identified teacher support, school connectedness, happiness at school, social facilitation (e.g., supportive discussions among peers and with teachers), and other related factors affecting adolescent depression (169). Stressors like changes in habits and academic performance were significantly associated with depressive symptoms in American college students (170). A number of studies have shown that stressful life events were associated with depressive symptoms in adolescents (171–174). Temporality may be one of the reasons why the events that happened before or during late adolescence were not further associated with depressive symptoms in young adulthood. In a study among Caucasian female same-sex twins in the US, the association between stressful life events and onsets of depressive symptoms was usually strongest in the month of occurrence and might be extended for up to 6

months depending on the severity of events (175). The identified stressful live events often involve family, parents, and friends, reinforcing the previous discussion.

Alcohol consumption in mid-adolescence was shown to link with higher apolipoprotein M level, a lipocalin protein associated with high-density lipoprotein (HDL), which can be harmful for mental health in adolescence (119). In addition, a Danish case-control study has found apolipoprotein M to dysregulate the fibrosis stages of adult patients with alcohol-related liver disease compared to healthy controls (176). A systematic review has reported a positive correlation between alcohol consumption and all HDL subfractions (177). In the US, alcohol use interacted with the relationship between platelet–HDL ratio and depressive symptoms in adults (178). The connection between alcohol consumption and depression, as well as broader mental health conditions, is well-documented (179, 180).

Other associations in our biomarker analysis suggest that exposures may influence depression and mental health through the immune system. Complement C4-B, an immune effector, was found to be negatively associated with household size, indicating smaller household sizes may benefit mental health (118). A WHO report suggested crowded living conditions to worsen mental health outcomes (181). The difference between previous literature and the thesis may be because of unmeasured factors such as room availability (181); household size is not equivalent to crowding, and smaller households may shallow a close family. Moreover, levels of complement C4 proteins have been higher in MDD patients compared to healthy controls, although they did not correlate with depressive symptoms (182, 183). A higher frequency of videogame playing in mid-adolescence was associated with a higher level of serpin family G member 1 and subsequently more mental health problems (119). This protein is involved in the regulation of the complement cascade (a part of the immune system), where elevated blood pressure activates the complement system and produces complement factors (184), which was also identified as a key biomarker of MDD in a transcriptomic study (185). A cross-sectional study of 282 overweight or obese Canadian adolescents linked videogame playing with higher blood pressure, indicating blood pressure can be a potential link

between videogame use, serpin family G member 1, and mental health (186, 187). Some serpin family proteins engage in blood pressure regulation (188, 189).

Associations were consistently observed between greenspaces and protein levels in both Finland and Spain. A US study has found that distance-weighted vegetated land cover correlated with serum interleukin-8, which is in the same family as the interleukin-6 receptor subunit β (finding in FinnTwin12) (190). Interleukin-6 receptor subunit β and progranulin (finding in WALNUTs) take part in immune reactions in the neuro system and have a functional role in the neurodevelopment of depression, which could participate in both pro- and anti-inflammatory processes (191–193). Therefore, a possible mechanism is through the indoor microbial environment from residential greenspaces (194), which enhances immune modulation, such as increasing levels of interleukin family proteins (195), and further reduces inflammation and promotes mental health (196). However, after integrating results with previous protein—mental health research, contrary observations appeared in the two countries. In Finland, greater greenspace exposure (higher NDVI) was associated with lower levels of interleukin-6 receptor subunit β , which, in turn, has been linked to increased emotional and behavioral problems at age 14 (118). In Spain, higher tree cover was associated with increased progranulin levels, and higher levels of progranulin have been linked to fewer emotional and behavioral problems at age 14 (118). These results suggest anti-inflammatory roles for both interleukin-6 receptor subunit β and progranulin. Interleukin-6 classic signaling can exert anti-inflammatory effects by shifting energy metabolism to oxidative phosphorylation, a mechanism that may support better mental health (197). Progranulin deficiency, on the other hand, has been associated with frontotemporal dementia-like behaviors and astrocyte activation in the nucleus accumbens, promoting the release of inflammatory cytokines and contributing to neuroinflammation and depression (191). The opposition between two cohorts may come from the characteristics of greenspace such as type and quality and other contextual factors like sporting culture. A recent systematic review has found that higher-quality greenspaces are generally associated with better mental health outcomes (198). Better methods to assess greenspaces and more detailed investigations are needed.

Only a few twin studies have employed the exposome concept targeted depression or broader phenotype. Another FinnTwin12 study I was involved in has shown the similarity of the external exposome and omics profile between cotwins altered the differences in lifestyle and substance use (199). Except for the abovementioned study in Section 2.3 (112), another study based on the Netherlands Twin Registry has found that the general and specific (parent and self-reports of psychosocial factors) external exposome could predict adult well-being using machine learning models, while the addition of genome did not improve the prediction performance (200). Neither Dutch twin study has integrated the twin design. In the Detroit metro area, US, researchers identified one key component of each domain of built (6 exposures), social (4 exposures), and toxicant (3 exposures) neighborhood conditions by PCA in the Twin Study of Behavioral and Emotional Development (201). Built environments (e.g., more nonprofits, churches, alcohol outlets, and fewer vacant properties) were linked to conduct problems, while toxicant exposures (e.g., industrial areas, toxin releases, and pre-1978 structures) were associated with oppositional defiance (201). In the community-based Washington State Twin Registry, 5 of 86 built environmental exposures from four domains (regional context, neighborhood composition, destinations, and transportation) were significantly associated with physical activity, but these exposures did not improve the ability to predict physical activity compared to a base model of sociodemographic factors (202). Irregular physical activity is a risk factor for depression (203). Nevertheless, researchers only used robust standard errors to account for the design effect in these two US twin studies. Krapohal et al. drew unrelated individuals from a twin cohort and examined the covariations between the trait (education, body mass index, and schizophrenia)-associated polygenetic variation and a range of early environmental exposures such as substance use during pregnancy, lifestyle, and familial income (204). The great potential of twin design in the exposome study should gain more attention, and corresponding analytical methods could be developed more.

6.3 Complex reality and implication

The first type of complexity was within the environment. Many important exposures have been identified, but their effect sizes and directions were different,

implying the possibility of multi-factor synergistic effects. As Chung et al. suggested, ExWAS should serve as an initial screening tool in exposome research to identify key candidates (68). Moving forward, co-exposure analytic platforms using machine learning approaches can help explore these complex effects more thoroughly. In addition, some exposures only exerted effects at a specific time point, and some exposures only affected depressive symptoms measured at a specific time point. Combining findings within different settings such as family, friends, and school, spatiotemporality of environmental impacts is highlighted by longitudinal observation. Interactions between different contextual settings were also raised as “meso-system” (e.g., family-school interaction) (205), and some studies have observed their effects on depressive symptoms and mental health (206, 207). Research focuses like interaction, contextualization, longitudinality, or more are encouraged to be organized together to enable more systematic environmental health impact assessments. Besides, previous ExWAS studies targeting broader mental health outcomes had some homogeneous or heterogeneous results to ours (9, 67, 208, 209). Albeit the divergent findings, the accumulation of exposome research across countries and populations has enriched the growing knowledge of the exposome effect on depression (as well as broader mental health). The present thesis is one series-study in Europe, and by collective efforts, critical implications for the prevention and treatment of depression can be drawn in the future.

The second type of complexity involves the interplay between environmental and genetic factors. In Swedish twins, age at separation, as the indicator of the common environment, was correlated with within-pair differences in HDL levels among MZ twins (210) but was not associated with some depression-related phenotypes, such as tobacco use (211) or cognitive abilities (212). Through comparing MZ twins raised together or apart, there were similar intrapair correlations for several psychophysiological traits, indicating a limited influence of common environments (213). These studies, although somewhat inconsistent, suggested that both environmental and genetic factors contribute to depression. In our findings, a minor genetic influence on separation behavior was observed, hinting that genetics may shape the environments twins choose to live in later as gene-environment correlation (214). In addition to depression, genetic variances in proteome (215, 216), peer difficulty (217), alcohol consumption (218, 219), and videogame playing

(220) have been detected by univariate twin modeling. Bivariate twin models revealed that genetic factors accounted for more of the covariance between exposome scores and depressive symptoms than unique environmental factors. Studies in UK and US twins have shown gene-environment correlations involving stressful life events on adolescent depression (221, 222). In another UK twin study using polygenic risk scores, parents were likely to shape their children's academic environments based partly on their own genetic predispositions rather than on their children's traits (223). There were some studies showing gene-environment interaction on the depressive symptoms (224).

Just as few diseases are purely genetic (Mendelian), few exposures should be assumed to be purely environmental. Even in the case of extreme events like natural disasters such as earthquakes or ones caused by humans such as economic recessions, family and genetic factors correlate with coping strategies and mental health outcomes. The relationship between genes and environment is not an either-or opposition but an intertwined one. The role of genetics is highly suggested to be considered in any environmental research. There are a few possible types of study design. One direct way is to use molecular genetic data by sequencing, such as creating an interaction with the polygenetic risk score (224). Natural experiment designs, such as twin and family studies (like this thesis), can be used to control for genetic heterogeneity or estimate the genetic contribution to the living environment people selected (224). Additionally, multiethnic population cohorts can be employed to examine trait differences among diverse ethnic groups sharing the same environment (225).

6.4 Limitations

Several limitations should be acknowledged. First, the sample size was relatively small, which could limit the statistical power, increase the risk of type I error, and lead to null findings. Although Chung et al. suggest that a sample size between 1,795 and 3,625 is power-sufficient when using Bonferroni correction (226), this study did not include stratification. Limited power also obstructed the detection of critical small-effect exposures, as the ubiquity of the exposome (that low-grade exposures are highly prevalent). Second, only two basic covariates were included in the

ExWAS of proteins, because lifestyle factors were regarded as exposures. Therefore, I stated that this analysis is an exploratory one, and subsequent network analysis could be helpful in addressing potential confounding effects. Third, no causality assessment or pathway analysis was conducted to link external exposures, biomarkers, and depression. High-dimensional mediation analysis techniques may provide an opportunity for such evaluations in future studies (227). The interpretation and translation of the findings should be cautious for further policymaking and intervention, particularly regarding ethical considerations, due to lack of causality (228). Fourth, there is the concern of reverse direction, where the mental health status of children and adolescents might shape their family environment. A sensitivity analysis observed some correlations between mental health outcomes at age 14 and age at separation; nevertheless, when we included the depression endpoint at age 14 as a covariate in main regression models, many of the significant findings remained unchanged. The reciprocal relationships between individuals' mental health and their environmental profiles should be noted to clarify these dynamics in the future. Fifth, most of the ExWAS, CFA, and twin modeling were based on FinnTwin12, and there were not many overlapping results between FinnTwin12 and WALNUTs. Replication with additional twin cohorts would reduce the risk of model overfitting and enhance reliability. Sixth, expert opinions were used to group similar exposures into domains (concept areas), and domains were interpreted more than each specific exposure in the discussion. While this approach made it manageable to cover hundreds of exposures within a single study, it may have overlooked subtle yet important differences between exposures, such as buffering effects (229). Seventh, the exposome set included in biomarker analyses included only exposures available in both cohorts, leading to potential bias in the selection of exposures. Since adolescents and young adults spend lots of time in school or perhaps at work, physical exposures also should be measured at the school or workplace. The biomarkers represent a small fraction of potential biomarkers, most of which were not assessed by the omics platforms that were used. A broader range of exposures and omics in future studies could help to systematically investigate environmental influences on biological mechanisms.

7 Conclusion

The findings of this thesis advance our understanding of the comprehensive impact of the exposome, the totality of the environment, on depression in adolescents and young adults. Studies I and II highlighted the significant role of family, reflected by age at separation and a wide range of family-related variables, in shaping depressive symptoms during late adolescence and young adulthood. In addition to family and parental influences, friendship and romantic relationships, and stressful life events were also critical aspects. Study III identified seven mid-adolescence exposures related to lifestyle, indoor quality, and the natural and social family environment as being associated with six proteomic biomarkers of mental health, which provided insights into how the environment influences mental health, even before clinical symptom onset. Notably, greenspaces exerted opposite effects on mental health in Finland and Spain. By twin analysis, a minor genetic influence was revealed on cotwin separation. A substantial genetic contribution was noted in the covariation between the composite exposome score and depressive symptoms, as well as between exposures and proteomic biomarkers of mental health. These findings advocate the systematic evaluation of environmental influences on depressive symptoms and the indispensable role of genetics. Despite the limitations, interventions targeting the family environment and potentially the broader interpersonal social environment are encouraged to develop for reducing the depression burden, while greenspaces are recommended to consider in European-wide policy planning for mental health promotion. Future studies with broader exposure assessments and diverse populations are essential to develop a robust evidence base for improving youth mental health.

Author contribution

Study I: I am the first author. I participated in conceiving and conceptualizing the research idea and developed the analysis plan. I conducted data analysis and drafted the original draft of the manuscript. After review and approval by coauthors, I submitted the manuscript to the journal.

Study II: I am the first author. I participated in conceiving and conceptualizing the research idea. I developed the analysis plan. A part of exposure data was acquired by me. I conducted data analysis with the support of another coauthor and drafted the original draft of the manuscript. After review and approval by coauthors, I submitted the manuscript to the journal.

Study III: I am the first author. I participated in conceiving and conceptualizing the research idea. I organized regular meetings to discuss the study and develop the preliminary analysis plan. I conducted data analysis on the FinnTwin12 part and wrote the protocol and code for the partner in Spain to conduct analysis in their dataset. I drafted the original draft of the manuscript. After review and approval by coauthors, I submitted the manuscript to the journal.

Acknowledgements

The work for this thesis was performed at the University of Helsinki in the Institute for Molecular Medicine Finland, HiLIFE. I would like to thank the directors and staff for providing an excellent work environment and facilities. My sincere thanks for the funding that allowed this thesis to be performed: funding from the Equal-Life project and the funding position from the Doctoral Programme in Population Health. I would also like to acknowledge the twins and their families for their continued participation in the cohort, and people who contributed to developing and maintaining the Finnish Twin cohorts, as well as all the funding bodies. I stood on the shoulders of giants.

I would like to appreciate Prof. Sylvain Sebert for acting as the opponent and Dr. Jaana Halonen and Dr. Léa Maitre for being the pre-examiners; your comments and suggestions have substantially improved my thesis. I would like to thank Prof. Antti Latvala for being the faculty representative and Prof. Samuli Ripatti for being the academic coordinator. To my thesis committee members, Prof. Juha Pekkanen and Prof. Petri Pellikka, I greatly appreciate your suggestions and support for these years' journey. I would like to appreciate Prof. Andrea Ganna for being the custos and secondary supervisor; your support is precious.

The thesis has been intensively involved in the EU exposome project "Equal-Life". I would like to pay tribute to the project coordinator, Dr. Irene van Kamp, whose leadership and excellent attitude in the project have enabled us to sail in the ocean of exposomes. I would also like to thank all my colleagues, whose support has given me room to explore the exposomes and whose guidance has helped me to correct mistakes, improve my ability, and continue to grow in the face of challenges.

My group teammates: Gabin, Hannes, Aino, Mikaela, Tianyu, Sari, and there are many more people I may not mention one by one are the precious wealth in my life. Your passion for research and your sincerity towards life have always deeply

touched me and made me a better person outside of work. I will never forget the afternoons and evenings we chatted together.

I would like to express my most sincere gratitude to my supervisor, Prof. Jaakko Kaprio. These years have been profoundly meaningful to me. You provided not only a supportive environment but also a clear and inspiring direction with potential. Your trust and guidance helped me refine both my research ideas and scientific thinking. More than that, you have been a true role model. From you, I learned how to do rigorous research, how to do elegant science, and, most importantly, how to become a good scientist. I've come to understand the meaning of "Beauty is Truth." When the 50th anniversary of the Finnish Twin Cohorts was celebrated last year, I imagined creating a cohort of my own and one day giving a speech at its anniversary. Perseverance, I now realize, is the greatest romance.

Love from my parents has given me the strength to pursue my dreams abroad on my own, knowing that I always have a safe and dependable harbor to return to. My soul could be truly free. Words cannot fully capture the depth of your love. I want you to know that I am truly doing my best to be a good person, and I am sure that you are proud of me.

Finally, I want to thank my country. The goal of the great nation rejuvenation has always been the motivation to me, and traditional Chinese virtues such as diligence and courage are always kept in mind.

This is not the end but a new starting point. In the midst of great changes that have not been seen in a century, let us unite and work together to build the community of common destiny.

Zhiyang Wang 汪之洋

Helsinki, May, 2025

Reference

1. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR)* (American Psychiatric Association, 2000).
2. R. M. Lerner, C. D. Brindis, M. Batanova, R. W. Blum, “Adolescent Health Development: A Relational Developmental Systems Perspective BT - Handbook of Life Course Health Development” in N. Halfon, C. B. Forrest, R. M. Lerner, E. M. Faustman, Eds. (Springer International Publishing, 2018), pp. 109–121.
3. A. Thapar, O. Eyre, V. Patel, D. Brent, Depression in young people. *Lancet* **400**, 617–631 (2022).
4. P. F. Sullivan, M. C. Neale, K. S. Kendler, Genetic Epidemiology of Major Depression: Review and Meta-Analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
5. C. P. Wild, The exposome: from concept to utility. *Int. J. Epidemiol.* **41**, 24–32 (2012).
6. A. Gururajan, G. Clarke, T. G. Dinan, J. F. Cryan, Molecular biomarkers of depression. *Neurosci. Biobehav. Rev.* **64**, 101–133 (2016).
7. A. J. MacGregor, H. Snieder, N. J. Schork, T. D. Spector, Twins: novel uses to study complex traits and genetic diseases. *Trends Genet.* **16**, 131–134 (2000).
8. J. Xu, *et al.*, Effects of urban living environments on mental health in adults. *Nat. Med.* **29**, 1456–1467 (2023).

9. K. W. Choi, *et al.*, An Exposure-Wide and Mendelian Randomization Approach to Identifying Modifiable Factors for the Prevention of Depression. *Am. J. Psychiatry* **177**, 944–954 (2020).
10. C. P. Wild, Complementing the Genome with an “Exposome”: The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **14**, 1847–1850 (2005).
11. K. Tolkunova, *et al.*, Transgenerational and intergenerational effects of early childhood famine exposure in the cohort of offspring of Leningrad Siege survivors. *Sci. Rep.* **13**, 11188 (2023).
12. P. González-Rodríguez, J. Füllgrabe, B. Joseph, The hunger strikes back: an epigenetic memory for autophagy. *Cell Death Differ.* **30**, 1404–1415 (2023).
13. J. R. Blaustein, M. J. Quisel, N. M. Hamburg, S. Wittkopp, Environmental Impacts on Cardiovascular Health and Biology: An Overview. *Circ. Res.* **134**, 1048–1060 (2024).
14. A. Ventriglio, *et al.*, Environmental pollution and mental health: a narrative review of literature. *CNS Spectr.* **26**, 51–61 (2021).
15. J. W. J. Beulens, *et al.*, Environmental risk factors of type 2 diabetes—an exposome approach. *Diabetologia* **65**, 263–274 (2022).
16. J. F. Dawson, A. W. Richter, Probing three-way interactions in moderated multiple regression: Development and application of a slope difference test. *J. Appl. Psychol.* **91**, 917–926 (2006).
17. C. Gennings, *et al.*, The impact of exposure to a mixture of eighteen polyhalogenated aromatic hydrocarbons on thyroid function: Estimation of an interaction threshold. *J. Agric. Biol. Environ. Stat.* **12**, 96–111 (2007).
18. M. Vrijheid, The exposome: a new paradigm to study the impact of environment on health. *Thorax* **69**, 876 LP – 878 (2014).

19. S. Scholz, *et al.*, The Eco-Exposome Concept: Supporting an Integrated Assessment of Mixtures of Environmental Chemicals. *Environ. Toxicol. Chem.* **41**, 30–45 (2022).
20. H. Gudi-Mindermann, *et al.*, Integrating the social environment with an equity perspective into the exposome paradigm: A new conceptual framework of the Social Exposome. *Environ. Res.* **233**, 116485 (2023).
21. J. Kaprio, M. Koskenvuo, Genetic and Environmental Factors in Complex Diseases: The Older Finnish Twin Cohort. *Twin Res.* **5**, 358–365 (2002).
22. C. Jiang, *et al.*, Dynamic Human Environmental Exposome Revealed by Longitudinal Personal Monitoring. *Cell* **175**, 277-291.e31 (2018).
23. P. Gao, *et al.*, Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. *Genome Res.* **32**, 1199–1214 (2022).
24. P. de Prado-Bert, *et al.*, The early-life exposome and epigenetic age acceleration in children. *Environ. Int.* **155**, 106683 (2021).
25. C. Kenner, “Fetal Development: Environmental Influences and Critical Periods” in (Springer Publishing Company), pp. 1–2019.
26. N. Prabhu, *et al.*, First trimester maternal tobacco smoking habits and fetal growth. *Thorax* **65**, 235 LP – 240 (2010).
27. P. Clarke, *et al.*, Cumulative Exposure to Neighborhood Context: Consequences for Health Transitions Over the Adult Life Course. *Res. Aging* **36**, 115–142 (2013).
28. H. Ohanyan, *et al.*, Associations between the urban exposome and type 2 diabetes: Results from penalised regression by least absolute shrinkage and selection operator and random forest models. *Environ. Int.* **170**, 107592 (2022).
29. H. Ohanyan, *et al.*, Machine learning approaches to characterize the obesogenic urban exposome. *Environ. Int.* **158**, 107015 (2022).

30. E. Traini, *et al.*, A prospective exploration of the urban exposome in relation to headache in the Dutch population-based Occupational and environmental health cohort study (AMIGO). *Environ. Int.* **188**, 108776 (2024).
31. L. Houweling, *et al.*, The effect of the urban exposome on COVID-19 health outcomes: A systematic review and meta-analysis. *Environ. Res.* **240**, 117351 (2024).
32. A. Pronk, *et al.*, Applying the exposome concept to working life health: The EU EPHOR project. *Environ. Epidemiol.* **6** (2022).
33. L. Fabbri, *et al.*, Childhood exposure to non-persistent endocrine disrupting chemicals and multi-omic profiles: A panel study. *Environ. Int.* **173**, 107856 (2023).
34. B. Loef, *et al.*, Using random forest to identify longitudinal predictors of health in a 30-year cohort study. *Sci. Rep.* **12**, 10372 (2022).
35. T. M. Lam, *et al.*, Development of a neighborhood obesogenic built environment characteristics index for the Netherlands. *Obesity* **31**, 214–224 (2023).
36. D. R. Healy, *et al.*, Longitudinal associations of an exposome score with serum metabolites from childhood to adolescence. *Commun. Biol.* **7**, 890 (2024).
37. L.-K. Pries, *et al.*, Estimating Exposome Score for Schizophrenia Using Predictive Modeling Approach in Two Independent Samples: The Results From the EUGEI Study. *Schizophr. Bull.* **45**, 960–965 (2019).
38. E. Buelow, *et al.*, Hospital discharges in urban sanitation systems: Long-term monitoring of wastewater resistome and microbiota in relationship to their eco-exposome. *Water Res. X* **7**, 100045 (2020).
39. L.-K. Pries, *et al.*, Estimating the Association Between Exposome and Psychosis as Well as General Psychopathology: Results From the ABCD Study. *Biol. Psychiatry Glob. Open Sci.* **2**, 283–291 (2022).

40. B. Vandeninden, *et al.*, Cluster pattern analysis of environmental stressors and quantifying their impact on all-cause mortality in Belgium. *BMC Public Health* **24**, 536 (2024).
41. A. P. Keil, *et al.*, A quantile-based g-computation approach to addressing the effects of exposure mixtures. *Environ. Health Perspect.* **128**, 47004 (2020).
42. G. M. Christensen, *et al.*, The complex relationship of air pollution and neighborhood socioeconomic status and their association with cognitive decline. *Environ. Int.* **167**, 107416 (2022).
43. S. H. Liu, *et al.*, Bayesian varying coefficient kernel machine regression to assess neurodevelopmental trajectories associated with exposure to complex mixtures. *Stat. Med.* **37**, 4680–4694 (2018).
44. R. Juwel, *et al.*, Joint effects of time-varying air pollutant mixtures on health outcome: An extension of the Quantile G-computation method. *ISEE Conf. Abstr.* **2024** (2024).
45. P. Zang, *et al.*, Nonlinear Effects of the Built Environment on Light Physical Activity among Older Adults: The Case of Lanzhou, China. *Int. J. Environ. Res. Public Health* **19** (2022).
46. Y. Chen, *et al.*, Examining the importance of built and natural environment factors in predicting self-rated health in older adults: An extreme gradient boosting (XGBoost) approach. *J. Clean. Prod.* **413**, 137432 (2023).
47. T. J. Haight, Y. Wang, M. J. van der Laan, I. B. Tager, A cross-validation deletion–substitution–addition model selection algorithm: Application to marginal structural models. *Comput. Stat. Data Anal.* **54**, 3080–3094 (2010).
48. H. Hu, *et al.*, A spatial and contextual exposome-wide association study and polyexposomic score of COVID-19 hospitalization. *Exposome* **3**, osad005 (2023).
49. E. Y. Lee, *et al.*, Questionnaire-based exposome-wide association studies (ExWAS) reveal expected and novel risk factors associated with

- cardiovascular outcomes in the Personalized Environment and Genes Study. *Environ. Res.* **212**, 113463 (2022).
50. A. Lydiane, *et al.*, A Systematic Comparison of Linear Regression–Based Statistical Methods to Assess Exposome-Health Associations. *Environ. Health Perspect.* **124**, 1848–1856 (2016).
 51. L. Maitre, *et al.*, State-of-the-art methods for exposure-health studies: Results from the exposome data challenge event. *Environ. Int.* **168**, 107422 (2022).
 52. C. Warembourg, *et al.*, Statistical Approaches to Study Exposome-Health Associations in the Context of Repeated Exposure Data: A Simulation Study. *Environ. Sci. Technol.* **57**, 16232–16243 (2023).
 53. A. Anguita-Ruiz, *et al.*, Beyond the single-outcome approach: A comparison of outcome-wide analysis methods for exposome research. *Environ. Int.* **182**, 108344 (2023).
 54. J. Barrera-Gómez, *et al.*, A systematic comparison of statistical methods to detect interactions in exposome-health associations. *Environ. Heal.* **16**, 74 (2017).
 55. P. Vineis, *et al.*, The contribution to policies of an exposome-based approach to childhood obesity. *Exposome* **3**, osado06 (2023).
 56. I. van Kamp, *et al.*, Early environmental quality and life-course mental health effects: The Equal-Life project. *Environ. Epidemiol.* **6** (2022).
 57. T. Sonnenschein, S. Scheider, G. A. de Wit, C. C. Tonne, R. Vermeulen, Agent-based modeling of urban exposome interventions: prospects, model architectures, and methodological challenges. *Exposome* **2**, osaco09 (2022).
 58. M. Pristner, B. Warth, Drug–Exposome Interactions: The Next Frontier in Precision Medicine. *Trends Pharmacol. Sci.* **41**, 994–1005 (2020).

59. N. J. Lennon, *et al.*, Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nat. Med.* **30**, 480–487 (2024).
60. I. S. Chan, G. S. Ginsburg, Personalized medicine: Progress and promise. *Annu. Rev. Genomics Hum. Genet.* **12**, 217–244 (2011).
61. H. Wang, N. Liu, J. Chen, S. Guo, The Relationship Between Urban Renewal and the Built Environment: A Systematic Review and Bibliometric Analysis. *J. Plan. Lit.* **37**, 293–308 (2021).
62. Y. Chen, G. Liu, T. Zhuang, Evaluating the Comprehensive Benefit of Urban Renewal Projects on the Area Scale: An Integrated Method. *Int. J. Environ. Res. Public Health* **20** (2023).
63. C. Vert, *et al.*, Health Benefits of Physical Activity Related to an Urban Riverside Regeneration. *Int. J. Environ. Res. Public Health* **16** (2019).
64. X. Xue, R. Zhang, X. Zhang, R. J. Yang, H. Li, Environmental and social challenges for urban subway construction: An empirical study in China. *Int. J. Proj. Manag.* **33**, 576–588 (2015).
65. G. Xing-lei, Z. Qian, Y. Zhen-lei, L. Zhong-hui, S. Chang-zheng, Research on the impact of China's urban rail transit on economic growth: Based on PSM-DID model. *Front. Environ. Sci.* **11** (2023).
66. S. M. Rappaport, Implications of the exposome for exposure science. *J. Expo. Sci. Environ. Epidemiol.* **21**, 5–9 (2011).
67. B. D. Lin, *et al.*, Nongenetic Factors Associated With Psychotic Experiences Among UK Biobank Participants: Exposome-Wide Analysis and Mendelian Randomization Analysis. *JAMA Psychiatry* **79**, 857–868 (2022).
68. M. K. Chung, *et al.*, Decoding the exposome: data science methodologies and implications in exposome-wide association studies (ExWASs). *Exposome* **4**, osae001 (2024).

69. C. Cappell, R. B. Heiner, The intergenerational transmission of family aggression. *J. Fam. Violence* **5**, 135–152 (1990).
70. N. Powdthavee, A. Vignoles, Mental Health of Parents and Life Satisfaction of Children: A Within-Family Analysis of Intergenerational Transmission of Well-Being. *Soc. Indic. Res.* **88**, 397–422 (2008).
71. B. M. Schlee, A. K. Mullis, M. Shriner, Parents social and resource capital: Predictors of academic achievement during early childhood. *Child. Youth Serv. Rev.* **31**, 227–234 (2009).
72. M. McGue, A. Skytthe, K. Christensen, The nature of behavioural correlates of healthy ageing: a twin study of lifestyle in mid to late life. *Int. J. Epidemiol.* **43**, 775–782 (2014).
73. K. S. KENDLER, J. H. BAKER, Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* **37**, 615–626 (2007).
74. A. Smits, R. I. Van Gaalen, C. H. Mulder, Parent–Child Coresidence: Who Moves in With Whom and for Whose Needs? *J. Marriage Fam.* **72**, 1022–1033 (2010).
75. American Psychiatric Association, *Diagnostic and statistical manual of mental disorders: DSM-5TM*, 5th ed. (American Psychiatric Association, 2013).
76. J. L. CAMERON, Interrelationships between Hormones, Behavior, and Affect during Adolescence: Understanding Hormonal, Physical, and Brain Changes Occurring in Association with Pubertal Activation of the Reproductive Axis. Introduction to Part III. *Ann. N. Y. Acad. Sci.* **1021**, 110–123 (2004).
77. A. Colver, G. Dovey-Pearce, “The Anatomical, Hormonal and Neurochemical Changes that Occur During Brain Development in Adolescents and Young Adults BT - Health Care Transition: Building a Program for Adolescents and Young Adults with Chronic Illness and Disability” in A. C. Hergenroeder, C. M. Wiemann, Eds. (Springer International Publishing, 2018), pp. 15–19.

78. D. R. Singla, *et al.*, Implementation and effectiveness of adolescent life skills programs in low- and middle-income countries: A critical review and meta-analysis. *Behav. Res. Ther.* **130**, 103402 (2020).
79. M. G. Landazabal, Psychopathological Symptoms, Social Skills, and Personality Traits: A Study with Adolescents. *Span. J. Psychol.* **9**, 182–192 (2006).
80. M. Eskin, K. Ertekin, H. Demir, Efficacy of a Problem-Solving Therapy for Depression and Suicide Potential in Adolescents and Young Adults. *Cognit. Ther. Res.* **32**, 227–245 (2008).
81. A. Thapar, S. Collishaw, D. S. Pine, A. K. Thapar, Depression in adolescence. *Lancet* **379**, 1056–1067 (2012).
82. E. J. Costello, B. Maughan, Annual Research Review: Optimal outcomes of child and adolescent mental illness. *J. Child Psychol. Psychiatry* **56**, 324–341 (2015).
83. T. H. MELTON, P. E. CROARKIN, J. R. STRAWN, S. M. MCCLINTOCK, Comorbid Anxiety and Depressive Symptoms in Children and Adolescents: A Systematic Review and Analysis. *J. Psychiatr. Pract.* **22** (2016).
84. M. R. Khan, *et al.*, Depression, Sexually Transmitted Infection, and Sexual Risk Behavior Among Young Adults in the United States. *Arch. Pediatr. Adolesc. Med.* **163**, 644–652 (2009).
85. D. J. Smith, D. H. R. Blackwood, Depression in young adults. *Adv. Psychiatr. Treat.* **10**, 4–12 (2004).
86. GBD 2019 Mental Disorders Collaborators, 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**, 137–150 (2022).
87. S. Shorey, E. D. Ng, C. H. J. Wong, Global prevalence of depression and elevated depressive symptoms among adolescents: A systematic review and meta-analysis. *Br. J. Clin. Psychol.* **61**, 287–305 (2022).

88. M. A. Villarroel, E. P. Terlizzi, “Symptoms of depression among adults: United States, 2019. NCHS Data Brief, no 379” (2020).
89. M. Daly, Prevalence of Depression Among Adolescents in the U.S. From 2009 to 2019: Analysis of Trends by Sex, Race/Ethnicity, and Income. *J. Adolesc. Heal.* **70**, 496–499 (2022).
90. A. Steffen, J. Thom, F. Jacobi, J. Holstiege, J. Bätzing, Trends in prevalence of depression in Germany between 2009 and 2017 based on nationwide ambulatory claims data. *J. Affect. Disord.* **271**, 239–247 (2020).
91. K. E. Cairns, M. B. H. Yap, P. D. Pilkington, A. F. Jorm, Risk and protective factors for depression that adolescents can modify: A systematic review and meta-analysis of longitudinal studies. *J. Affect. Disord.* **169**, 61–75 (2014).
92. J. B. Sander, C. A. McCarty, Youth Depression in the Family Context: Familial Risk Factors and Models of Treatment. *Clin. Child Fam. Psychol. Rev.* **8**, 203–219 (2005).
93. R. Bevan Jones, *et al.*, Psychoeducational interventions in adolescent depression: A systematic review. *Patient Educ. Couns.* **101**, 804–816 (2018).
94. K. L. Kumpfer, C. Magalhães, Strengthening Families Program: An Evidence-Based Family Intervention for Parents of High-Risk Children and Adolescents. *J. Child Adolesc. Subst. Abuse* **27**, 174–179 (2018).
95. K. Stirling, J. W. Toumbourou, B. Rowland, Community factors influencing child and adolescent depression: A systematic review and meta-analysis. *Aust. New Zeal. J. Psychiatry* **49**, 869–886 (2015).
96. N. Rautio, S. Filatova, H. Lehtiniemi, J. Miettunen, Living environment and its relationship to depressive mood: A systematic review. *Int. J. Soc. Psychiatry* **64**, 92–103 (2017).
97. H. Xie, *et al.*, Affective disorder and brain alterations in children and adolescents exposed to outdoor air pollution. *J. Affect. Disord.* **331**, 413–424 (2023).

98. I. Tran, O. Sabol, J. Mote, The Relationship Between Greenspace Exposure and Psychopathology Symptoms: A Systematic Review. *Biol. Psychiatry Glob. Open Sci.* **2**, 206–222 (2022).
99. S. A. Kuwabara, B. W. Van Voorhees, J. K. Gollan, G. C. Alexander, A qualitative exploration of depression in emerging adulthood: disorder, development, and social context. *Gen. Hosp. Psychiatry* **29**, 317–324 (2007).
100. R. W. Roeser, J. S. Eccles, “Schooling and Mental Health BT - Handbook of Developmental Psychopathology” in A. J. Sameroff, M. Lewis, S. M. Miller, Eds. (Springer US, 2000), pp. 135–156.
101. Z. Zajkowska, *et al.*, A systematic review of the association between biological markers and environmental stress risk factors for adolescent depression. *J. Psychiatr. Res.* **138**, 163–175 (2021).
102. F. Rice, Genetics of childhood and adolescent depression: insights into etiological heterogeneity and challenges for future genomic research. *Genome Med.* **2**, 68 (2010).
103. A. L. Glowinski, P. A. F. Madden, K. K. Bucholz, M. T. Lynskey, A. C. Heath, Genetic epidemiology of self-reported lifetime DSM-IV major depressive disorder in a population-based twin sample of female adolescents. *J. Child Psychol. Psychiatry* **44**, 988–996 (2003).
104. Y.-M. Hur, Sex Differences in Genetic and Environmental Contributions to Depression Symptoms in South Korean Adolescent and Young Adult Twins. *Twin Res. Hum. Genet.* **11**, 306–313 (2008).
105. M. G. Nivard, *et al.*, Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. *Psychol. Med.* **45**, 1039–1049 (2015).
106. A. M. McIntosh, C. M. Lewis, M. J. Adams, Psychiatric Genomics Consortium Major Depressive Disorder Working, Genome-wide study of half a million individuals with major depression identifies 697 independent associations,

- infers causal neuronal subtypes and biological targets for novel pharmacotherapies. *medRxiv*, 2024.04.29.24306535 (2024).
107. L. Xia, S. Yao, The Involvement of Genes in Adolescent Depression: A Systematic Review. *Front. Behav. Neurosci.* **9** (2015).
 108. D. M. Ciuculete, *et al.*, meQTL and ncRNA functional analyses of 102 GWAS-SNPs associated with depression implicate HACE1 and SHANK2 genes. *Clin. Epigenetics* **12**, 99 (2020).
 109. J. Y. F. Lau, T. C. Eley, Disentangling gene-environment correlations and interactions on adolescent depressive symptoms. *J. Child Psychol. Psychiatry* **49**, 142–150 (2008).
 110. J. J. Li, M. S. Berk, S. S. Lee, Differential susceptibility in longitudinal models of gene–environment interaction for adolescent depression. *Dev. Psychopathol.* **25**, 991–1003 (2013).
 111. P. Meng, *et al.*, Associations between genetic loci, environment factors and mental disorders: a genome-wide survival analysis using the UK Biobank data. *Transl. Psychiatry* **12**, 17 (2022).
 112. M. P. van de Weijer, *et al.*, Capturing the well-being exposome in poly-environmental scores. *J. Environ. Psychol.* **93**, 102208 (2024).
 113. X. Farré, *et al.*, Decoding depression by exploring the exposome-genome edge amidst COVID-19 lockdown. *Sci. Rep.* **14**, 13562 (2024).
 114. M. Rejek, B. Misiak, The Associations of Exposome Score with Various Domains of Psychopathology: A Network Analysis in a Non-Clinical Sample. *Brain Sci.* **14** (2024).
 115. I. Pelgrims, *et al.*, Association between urban environment and mental health in Brussels, Belgium. *BMC Public Health* **21**, 635 (2021).
 116. L. C. Silva-Costa, P. T. Carlson, P. C. Guest, V. de Almeida, D. Martins-de-Souza, “Proteomic Markers for Depression BT - Reviews on Biomarker

Studies in Psychiatric and Neurodegenerative Disorders” in P. C. Guest, Ed. (Springer International Publishing, 2019), pp. 191–206.

117. A. M. Afonin, *et al.*, Proteomic insights into mental health status: plasma markers in young adults. *Transl. Psychiatry* **14**, 55 (2024).
118. A. M. Whipp, G. Drouard, R. J. Rose, L. Pulkkinen, J. Kaprio, Protein associations and protein–metabolite interactions with depressive symptoms and the p-factor. *Transl. Psychiatry* **15**, 128 (2025).
119. I. de Sousa Maciel, *et al.*, Plasma proteomics discovery of mental health risk biomarkers in adolescents. *Nat. Ment. Heal.* **1**, 596–605 (2023).
120. J. Kang, *et al.*, Plasma proteomics identifies proteins and pathways associated with incident depression in 46,165 adults. *Sci. Bull.* **70**, 573–586 (2025).
121. U. Bhattacharyya, *et al.*, Circulating Blood-Based Proteins in Psychopathology and Cognition: A Mendelian Randomization Study. *JAMA Psychiatry* (2025) <https://doi.org/10.1001/jamapsychiatry.2025.0033>.
122. X. Xue, *et al.*, Sex differences in plasma proteomic markers in late-life depression. *Psychiatry Res.* **334**, 115773 (2024).
123. D. Braconi, G. Bernardini, A. Santucci, Linking protein oxidation to environmental pollutants: Redox proteomic approaches. *J. Proteomics* **74**, 2324–2337 (2011).
124. S. Guillotin, N. Delcourt, Studying the Impact of Persistent Organic Pollutants Exposure on Human Health by Proteomic Analysis: A Systematic Review. *Int. J. Mol. Sci.* **23** (2022).
125. J. Guo, *et al.*, Association between greenness and dyslipidemia in patients with coronary heart disease: A proteomic approach. *Ecotoxicol. Environ. Saf.* **231**, 113199 (2022).

126. C. Pan, *et al.*, Large-scale plasma proteomics uncovers novel targets linking ambient air pollution and depression. *Mol. Psychiatry* (2025) <https://doi.org/10.1038/s41380-025-02953-x>.
127. A. M. Whipp, *et al.*, Early adolescent aggression predicts antisocial personality disorder in young adults: a population-based study. *Eur. Child Adolesc. Psychiatry* **28**, 341–350 (2019).
128. M. E. Cooke, *et al.*, Alcohol use in Early Midlife: Findings from the Age 37 Follow-Up Assessment of the FinnTwin12 Cohort. *Behav. Genet.* (2025) <https://doi.org/10.1007/s10519-024-10212-y>.
129. J. Julvez, *et al.*, Walnuts, Long-Chain Polyunsaturated Fatty Acids, and Adolescent Brain Development: Protocol for the Walnuts Smart Snack Dietary Intervention Trial. *Front. Pediatr.* **9** (2021).
130. A. Pinar-Martí, *et al.*, Effect of walnut consumption on neuropsychological development in healthy adolescents: a multi-school randomised controlled trial. *eClinicalMedicine* **59**, 101954 (2023).
131. K. de Hoogh, *et al.*, Spatial PM_{2.5}, NO₂, O₃ and BC models for Western Europe – Evaluation of spatiotemporal stability. *Environ. Int.* **120**, 81–92 (2018).
132. S. Kephelopoulou, *et al.*, Advances in the development of common noise assessment methods in Europe: The CNOSSOS-EU framework for strategic environmental noise mapping. *Sci. Total Environ.* **482–483**, 400–410 (2014).
133. K. Kokko, L. Pulkkinen, Unemployment and Psychological Distress: Mediator Effects. *J. Adult Dev.* **5** (1998).
134. K. K. Bucholz, *et al.*, A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J. Stud. Alcohol* **55**, 149–158 (1994).
135. R. J. Rose, *et al.*, FinnTwin12 Cohort: An Updated Review. *Twin Res. Hum. Genet.* **22**, 302–311 (2019).

136. A. Afonin, A.-K. Piironen, J. Julvez, I. van Kamp, K. M. Kanninen, Plasma proteome demonstrates sex-specific associations with mental health risks in adolescents. *medRxiv*, 2024.12.16.24319062 (2024).
137. M. I. Kurki, *et al.*, FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
138. A. Jelenkovic, *et al.*, Genetic and environmental influences on growth from late childhood to adulthood: A longitudinal study of two Finnish twin cohorts. *Am. J. Hum. Biol.* **23**, 764–773 (2011).
139. C. Huppertz, *et al.*, The effects of parental education on exercise behavior in childhood and youth: a study in Dutch and Finnish twins. *Scand. J. Med. Sci. Sports* **27**, 1143–1156 (2017).
140. K. Salmela-Aro, *et al.*, Depressive Symptoms and Career-Related Goal Appraisals: Genetic and Environmental Correlations and Interactions. *Twin Res. Hum. Genet.* **17**, 236–243 (2014).
141. C. Edelbrock, R. Rende, R. Plomin, L. A. Thompson, A Twin Study of Competence and Problem Behavior in Childhood and Early Adolescence. *J. Child Psychol. Psychiatry* **36**, 775–785 (1995).
142. J. K. Richard J. Rose, Genes, Environments, and Adolescent Substance Use: Retrospect and Prospect from the FinnTwin Studies. *Acta Psychol. Sin.* **40**, 1062–1072 (2008).
143. M.-X. Li, J. M. Y. Yeung, S. S. Cherny, P. C. Sham, Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131**, 747–756 (2012).
144. C. Hernandez-Ferrer, *et al.*, Comprehensive study of the exposome and omic data using rexposome Bioconductor Packages. *Bioinformatics* **35** (2019).
145. S. F. Posner, L. Baker, A. Heath, N. G. Martin, Social contact, social attitudes, and twin similarity. *Behav. Genet.* **26**, 123–133 (1996).

146. E. Røysamb, K. Tambs, The beauty, logic and limitations of twin studies. *Nor. Epidemiol.* **26** (2016).
147. M. C. Neale, *et al.*, OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika* **81**, 535–549 (2016).
148. L. C. Gurrin, J. B. Carlin, J. A. C. Sterne, G. S. Dite, J. L. Hopper, Using Bivariate Models to Understand Between- and Within-Cluster Regression Coefficients, with Application to Twin Data. *Biometrics* **62**, 745–751 (2006).
149. P. J. Visser, A. Braber, M. Campo, C. E. Teunissen, B. M. Tijms, Proteomics correlates of amyloid beta and tau in CSF: A monozygotic twin study. *Alzheimer's Dement.* **19**, e073824 (2023).
150. J. E. Copp, P. C. Giordano, M. A. Longmore, W. D. Manning, Living With Parents and Emerging Adults' Depressive Symptoms. *J. Fam. Issues* **38**, 2254–2276 (2015).
151. A. M. Bardone, T. E. Moffitt, A. Caspi, N. Dickson, P. A. Silva, Adult mental health and social outcomes of adolescent girls with depression and conduct disorder. *Dev. Psychopathol.* **8**, 811–829 (1996).
152. M. Cerdá, A. Sagdeo, J. Johnson, S. Galea, Genetic and environmental influences on psychiatric comorbidity: A systematic review. *J. Affect. Disord.* **126**, 14–38 (2010).
153. I. Seiffe-Krenke, Leaving home or still in the nest? Parent-child relationships and psychological health as predictors of different leaving home patterns. *Dev. Psychol.* **42**, 864–876 (2006).
154. K. J. Conger, W. M. Little, Sibling relationships during the transition to adulthood. *Child Dev. Perspect.* **4**, 87–94 (2010).
155. L. van den Berg, M. Kalmijn, T. Leopold, Stepfamily Effects on Early Home-Leaving: The Role of Conflict and Closeness. *J. Marriage Fam.* **83**, 305–321 (2021).

156. M. B. H. Yap, P. D. Pilkington, S. M. Ryan, A. F. Jorm, Parental factors associated with depression and anxiety in young people: A systematic review and meta-analysis. *J. Affect. Disord.* **156**, 8–23 (2014).
157. Y. Yu, *et al.*, The Role of Family Environment in Depressive Symptoms among University Students: A Large Sample Survey in China. *PLoS One* **10**, e0143612 (2015).
158. L. Wang, D. R. Crane, The Relationship Between Marital Satisfaction, Marital Stability, Nuclear Family Triangulation, and Childhood Depression. *Am. J. Fam. Ther.* **29**, 337–347 (2001).
159. V. Stavropoulos, H. Lazaratou, E. Marini, D. Dikeos, Low Family Satisfaction and Depression in Adolescence: The Role of Self-Esteem. *J. Educ. Dev. Psychol.* **5**, 109–118 (2015).
160. L. M. Hooper, S. A. Wallace, Evaluating the Parentification Questionnaire: Psychometric Properties and Psychopathology Correlates. *Contemp. Fam. Ther.* **32**, 52–68 (2010).
161. G. P. Kuperminc, N. J. Wilkins, G. J. Jurkovic, J. L. Perilla, Filial responsibility, perceived fairness, and psychological functioning of Latino youth from immigrant families. *J. Fam. Psychol.* **27**, 173–182 (2013).
162. A. Kowal, L. Kramer, J. L. Krull, N. R. Crick, Children’s perceptions of the fairness of parental preferential treatment and their socioemotional well-being. *J. Fam. Psychol.* **16**, 297–306 (2002).
163. H. Shinohara, *et al.*, Early parental deprivation during primate infancy has a lifelong impact on gene expression in the male marmoset brain. *Sci. Rep.* **14**, 330 (2024).
164. R. A. Schwartz-Mette, J. Shankman, A. R. Dueweke, S. Borowski, A. J. Rose, Relations of friendship experiences with depressive symptoms and loneliness in childhood and adolescence: A meta-analytic review. *Psychol. Bull.* **146**, 664–700 (2020).

165. R. L. Miller-Slough, J. C. Dunsmore, Parent and Friend Emotion Socialization in Adolescence: Associations with Psychological Adjustment. *Adolesc. Res. Rev.* **1**, 287–305 (2016).
166. J. Richard, *et al.*, Scoping review to evaluate the effects of peer support on the mental health of young adults. *BMJ Open* **12**, e061336 (2022).
167. M. Brendgen, B. Wanner, A. J. S. Morin, F. Vitaro, Relations with Parents and with Peers, Temperament, and Trajectories of Depressed Mood During Early Adolescence. *J. Abnorm. Child Psychol.* **33**, 579–594 (2005).
168. C. Akre, A. Berchtold, Y. Barrense-Dias, J.-C. Suris, Characteristics defining perceived popularity among same-sex and opposite-sex peers. *Int. J. Adolesc. Med. Health* **33**, 227–234 (2021).
169. J. Kidger, R. Araya, J. Donovan, D. Gunnell, The Effect of the School Environment on the Emotional Health of Adolescents: A Systematic Review. *Pediatrics* **129**, 925–949 (2012).
170. L. Acharya, L. Jin, W. Collins, College life is stressful today – Emerging stressors and depressive symptoms in college students. *J. Am. Coll. Heal.* **66**, 655–664 (2018).
171. X. GE, M. N. NATSUAKI, R. D. CONGER, Trajectories of depressive symptoms and stressful life events among male and female adolescents in divorced and nondivorced families. *Dev. Psychopathol.* **18**, 253–273 (2006).
172. J. S. Carter, J. Garber, Predictors of the first onset of a major depressive episode and changes in depressive symptoms across adolescence: Stress and negative cognitions. *J. Abnorm. Psychol.* **120**, 779–796 (2011).
173. K. G. Askeland, *et al.*, Life events and adolescent depressive symptoms: Protective factors associated with resilience. *PLoS One* **15**, e0234109 (2020).
174. R. H. Friis, H.-U. Wittchen, H. Pfister, R. Lieb, Life events and changes in the course of depression in young adults. *Eur. Psychiatry* **17**, 241–253 (2002).

175. K. S. KENDLER, L. M. KARKOWSKI, C. A. PRESCOTT, Stressful Life Events and Major Depression: Risk Period, Long-Term Contextual Threat, and Diagnostic Specificity. *J. Nerv. Ment. Dis.* **186** (1998).
176. L. Niu, *et al.*, Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat. Med.* **28**, 1277–1287 (2022).
177. T. L. Wilkens, K. Tranæs, J. N. Eriksen, L. O. Dragsted, Moderate alcohol consumption and lipoprotein subfractions: a systematic review of intervention and observational studies. *Nutr. Rev.* **80**, 1311–1339 (2022).
178. J. Ni, P. Wu, X. Lu, C. Xu, Examining the cross-sectional relationship of platelet/high-density lipoprotein cholesterol ratio with depressive symptoms in adults in the United States. *BMC Psychiatry* **24**, 427 (2024).
179. J. Jeanblanc, Comorbidity Between Psychiatric Diseases and Alcohol Use Disorders: Impact of Adolescent Alcohol Consumption. *Curr. Addict. Reports* **2**, 293–301 (2015).
180. K. Ning, D. Gondek, P. Patalay, G. B. Ploubidis, The association between early life mental health and alcohol use behaviours in adulthood: A systematic review. *PLoS One* **15**, e0228667 (2020).
181. World Health Organization, “WHO housing and health guidelines” (2018).
182. H. Yu, *et al.*, Association of the plasma complement system with brain volume deficits in bipolar and major depressive disorders. *Psychol. Med.* **53**, 6102–6112 (2023).
183. H. Shatri, *et al.*, The correlation of anxiety and depression with C3 and C4 Levels and systemic lupus erythematosus activity. *BMC Psychiatry* **23**, 785 (2023).
184. U. O. Wenzel, C. Kemper, M. Bode, The role of complement in arterial hypertension and hypertensive end organ damage. *Br. J. Pharmacol.* **178**, 2849–2862 (2021).

185. A. Bouzid, *et al.*, Integrative bioinformatics and artificial intelligence analyses of transcriptomics data identified genes associated with major depressive disorders including NRG1. *Neurobiol. Stress* **26**, 100555 (2023).
186. G. S. Goldfield, *et al.*, Video Game Playing Is Independently Associated with Blood Pressure and Lipids in Overweight and Obese Adolescents. *PLoS One* **6**, e26643 (2011).
187. H. L. Schaare, *et al.*, Associations between mental health, blood pressure and the development of hypertension. *Nat. Commun.* **14**, 1953 (2023).
188. G. A. Kelly-Robinson, *et al.*, The Serpin Superfamily and Their Role in the Regulation and Dysfunction of Serine Protease Activity in COPD and Other Chronic Lung Diseases. *Int. J. Mol. Sci.* **22** (2021).
189. S. Janciauskiene, U. Lechowicz, M. Pelc, B. Olejnicka, J. Chorostowska-Wynimko, Diagnostic and therapeutic value of human serpin family proteins. *Biomed. Pharmacother.* **175**, 116618 (2024).
190. A. I. Egorov, *et al.*, Vegetated land cover near residence is associated with reduced allostatic load and improved biomarkers of neuroendocrine, metabolic and immune functions. *Environ. Res.* **158**, 508–521 (2017).
191. J. Wang, *et al.*, Progranulin from different gliocytes in the nucleus accumbens exerts distinct roles in FTD- and neuroinflammation-induced depression-like behaviors. *J. Neuroinflammation* **19**, 318 (2022).
192. G. E. Hodes, C. Ménard, S. J. Russo, Integrating Interleukin-6 into depression diagnosis and treatment. *Neurobiol. Stress* **4**, 15–22 (2016).
193. A. H. Miller, C. L. Raison, The role of inflammation in depression: from evolutionary imperative to modern treatment target. *Nat. Rev. Immunol.* **16**, 22–34 (2016).
194. H. Lai, E. J. Flies, P. Weinstein, A. Woodward, The impact of green space and biodiversity on health. *Front. Ecol. Environ.* **17**, 383–390 (2019).

195. M. Saarenpää, *et al.*, Urban indoor gardening enhances immune regulation and diversifies skin microbiota — A placebo-controlled double-blinded intervention study. *Environ. Int.* **187**, 108705 (2024).
196. R. Rieder, P. J. Wisniewski, B. L. Alderman, S. C. Campbell, Microbes and mental health: A review. *Brain. Behav. Immun.* **66**, 9–17 (2017).
197. S. Xu, *et al.*, Interleukin-6 classic and trans-signaling utilize glucose metabolism reprogramming to achieve anti- or pro-inflammatory effects. *Metabolism* **155**, 155832 (2024).
198. H. Park, C. D. Brown, A. L. Pearson, A systematic review of audit tools for evaluating the quality of green spaces in mental health research. *Health Place* **86**, 103185 (2024).
199. G. Drouard, *et al.*, Lifestyle differences between co-twins are associated with decreased similarity in their internal and external exposome profiles. *Sci. Rep.* **14**, 21261 (2024).
200. D. H. M. Pelt, *et al.*, Building machine learning prediction models for well-being using predictors from the exposome and genome in a population cohort. *Nat. Ment. Heal.* **2**, 1217–1230 (2024).
201. A. L. Pearson, E. A. Shewark, S. A. Burt, Associations between neighborhood built, social, or toxicant conditions and child externalizing behaviors in the Detroit metro area: a cross-sectional study of the neighborhood ‘exposome.’ *BMC Public Health* **22**, 1064 (2022).
202. A. V. Moudon, *et al.*, Probabilistic walking models using built environment and sociodemographic predictors. *Popul. Health Metr.* **17**, 7 (2019).
203. M. Pearce, *et al.*, Association Between Physical Activity and Risk of Depression: A Systematic Review and Meta-analysis. *JAMA Psychiatry* **79**, 550–559 (2022).
204. E. Krapohl, *et al.*, Widespread covariation of early environmental exposures and trait-associated polygenic variation. *Proc. Natl. Acad. Sci.* **114**, 11727–11732 (2017).

205. U. Bronfenbrenner, "Ecological systems theory." in *Six Theories of Child Development: Revised Formulations and Current Issues.*, (Jessica Kingsley Publishers, 1992), pp. 187–249.
206. G. F. Moore, *et al.*, School, Peer and Family Relationships and Adolescent Substance Use, Subjective Wellbeing and Mental Health Symptoms in Wales: a Cross Sectional Study. *Child Indic. Res.* **11**, 1951–1965 (2018).
207. L. Bond, *et al.*, Social and School Connectedness in Early Secondary School as Predictors of Late Teenage Substance Use, Mental Health, and Academic Outcomes. *J. Adolesc. Heal.* **40**, 357.e9-357.e18 (2007).
208. M. Y. Ni, *et al.*, Determinants of physical, mental and social well-being: a longitudinal environment-wide association study. *Int. J. Epidemiol.* **49**, 380–389 (2020).
209. M. P. van de Weijer, *et al.*, Expanding the environmental scope: an environment-wide association study for mental well-being. *J. Expo. Sci. Environ. Epidemiol.* **32**, 195–204 (2022).
210. X. Chen, *et al.*, Dominant Genetic Variation and Missing Heritability for Human Complex Traits: Insights from Twin versus Genome-wide Common SNP Models. *Am. J. Hum. Genet.* **97**, 708–714 (2015).
211. K. S. Kendler, L. M. Thornton, N. L. Pedersen, Tobacco Consumption in Swedish Twins Reared Apart and Reared Together. *Arch. Gen. Psychiatry* **57**, 886–892 (2000).
212. N. L. Pedersen, *et al.*, The Swedish Adoption Twin Study of Aging: An Update. *Acta Genet. medicae Gemellol. twin Res.* **40**, 7–20 (1991).
213. T. J. Bouchard, D. T. Lykken, M. McGue, N. L. Segal, A. Tellegen, Sources of human psychological differences: The minnesota study of twins reared apart. *Science (80-.)*. **250**, 223–228 (1990).
214. P. D. Quinn, B. M. D'Onofrio, "Nature Versus Nurture" in J. B. B. T.-E. of I. and E. C. D. (Second E. Benson, Ed. (Elsevier, 2020), pp. 373–384.

215. F. A. Hagenbeek, *et al.*, Maximizing the value of twin studies in health and behaviour. *Nat. Hum. Behav.* **7**, 849–860 (2023).
216. G. Drouard, *et al.*, Twin study provides heritability estimates for 2,321 plasma proteins and assesses missing SNP heritability. *medRxiv*, 2024.04.24.24306270 (2024).
217. M. Boivin, *et al.*, Strong Genetic Contribution to Peer Relationship Difficulties at School Entry: Findings From a Longitudinal Twin Study. *Child Dev.* **84**, 1098–1114 (2013).
218. V. Penninkilampi-Kerola, J. Kaprio, I. Moilanen, R. J. Rose, Co-Twin Dependence Modifies Heritability of Abstinence and Alcohol Use: A Population-Based Study of Finnish Twins. *Twin Res. Hum. Genet.* **8**, 232–244 (2005).
219. S. Virtanen, J. Kaprio, R. Viken, R. J. Rose, A. Latvala, Birth cohort effects on the quantity and heritability of alcohol consumption in adulthood: a Finnish longitudinal twin study. *Addiction* **114**, 836–846 (2019).
220. T. Hassan, Genetic and Environmental Influences on Playing Video Games. *Media Psychol.* **26**, 530–550 (2023).
221. F. Rice, G. T. Harold, A. Thapar, Negative life events as an account of age-related differences in the genetic aetiology of depression in childhood and adolescence. *J. Child Psychol. Psychiatry* **44**, 977–987 (2003).
222. J. D. Boardman, K. B. Alexander, M. C. Stallings, Stressful Life Events and Depression among Adolescent Twin Pairs. *Biodemography Soc. Biol.* **57**, 53–66 (2011).
223. Q. Zhou, *et al.*, Gene-environment correlation: the role of family environment in academic development. *Mol. Psychiatry* (2024) <https://doi.org/10.1038/s41380-024-02716-0>.
224. D. M. Dick, Gene-environment interaction in psychological traits and disorders. *Annu. Rev. Clin. Psychol.* **7**, 383–409 (2011).

225. A. R. Baldassari, *et al.*, Multi-Ethnic Genome-Wide Association Study of Decomposed Cardioelectric Phenotypes Illustrates Strategies to Identify and Characterize Evidence of Shared Genetic Effects for Complex Traits. *Circ. Genomic Precis. Med.* **13**, e002680 (2020).
226. M. K. Chung, G. M. Buck Louis, K. Kannan, C. J. Patel, Exposome-wide association study of semen quality: Systematic discovery of endocrine disrupting chemical biomarkers in fertility require large sample sizes. *Environ. Int.* **125**, 505–514 (2019).
227. J. A. Goodrich, *et al.*, Integrating Multi-Omics with environmental data for precision health: A novel analytic framework and case study on prenatal mercury induced childhood fatty liver disease. *Environ. Int.* **190**, 108930 (2024).
228. S. N. G. Jansen, *et al.*, Ethics of early detection of disease risk factors: A scoping review. *BMC Med. Ethics* **25**, 25 (2024).
229. J. G. Su, P. Dadvand, M. J. Nieuwenhuijsen, X. Bartoll, M. Jerrett, Associations of green space metrics with health and behavior outcomes at different buffer sizes and remote sensing sensor resolutions. *Environ. Int.* **126**, 162–170 (2019).

Appendices

Appendix 1. General Behavior Inventory (GBI)

Questions are formulated to the twin at the third (age 17) and fourth (young adulthood) waves. The Finnish version and English translation of the questionnaire and answer choices are presented here

Questions:

Vastaa seuraaviin kysymyksiin ympyröimällä itseäsi parhaiten kuvaava vaihtoehto.

Answer the following questions by circling the option that best describes you.

1. Oletko ollut surullinen, masentunut tai ärtyisä päiväkausien ajan tai pitempään käsittämättä miksi?

Have you felt sad, depressed, or irritable for days or longer without understanding why?

2. Onko ollut aikoja, jolloin et juurikaan ole tuntenut mielenkiintoa sellaisia asioita kohtaan, joista tavallisesti nautit (esim. ystävät, harrastukset, työ)?

Have there been times when you had little to no interest in things you usually enjoy (e.g., friends, hobbies, work)?

3. Onko Sinulla ollut monta päivää kestäneitä tai pidempiä kausia, jolloin suuri osa ajastasi on mennyt Sinulle sattuneiden ikävien asioiden hautomiseen?

Have you had periods lasting days or longer when you spent much of your time dwelling on unpleasant things that happened to you?

4. Onko Sinulla ollut kausia, joiden aikana olet ollut niin allapäin, että Sinusta on tuntunut vaikealta ryhtyä puhumaan tai että puhumiseen olisi kulunut liikaa voimia?

Have you had times when you felt so down that it was difficult to start a conversation, or it felt like talking would take too much effort?

5. Oletko kokenut päiväkausia tai pidempään kestäneitä jaksoja, jolloin olet ollut allapäin ja masennuksissa, ja Sinua on lisäksi vaivannut fyysinen levottomuus etkä ole kyennyt istumaan aloillasi, vaan Sinun on pitänyt olla jatkuvasti liikkeessä tai siirtyä yhdestä puuhasta toiseen?

Have you experienced days-long or longer periods of feeling down and depressed, combined with physical restlessness, making it hard to sit still and forcing you to stay active or shift from one task to another?

6. Onko Sinulla ollut useita päiviä kestäneitä tai pidempään jatkuneita surumielisyyden ja masennuksen kausia, joiden kestäessä olet iltaisin joutunut odottamaan unen tuloa yli tunnin, vaikka oletkin ollut jo hyvin väsynyt?

Have you had periods lasting several days or more where you felt down or depressed and struggled to fall asleep for over an hour at night, even when very tired?

7. Onko Sinulla ollut useita päiviä tai kauemmin kestäneitä jaksoja, jolloin olet ollut tyytymätön itseesi ja omissa silmissäsi mitätön ihminen?

Have you had episodes lasting several days or more when you felt dissatisfied with yourself and viewed yourself as worthless?

8. Onko Sinulla ollut useita päiviä tai kauemmin kestäneitä jaksoja, jolloin olosi on ollut raskas etkä ole kyennyt liikkumaan yhtä nopeasti kuin tavallisesti?

Have you had several days or longer periods where you felt heavy and unable to move as quickly as usual?

9. Onko Sinulla ollut pitkiä masennuskausia, joiden kestäessä mielialasi on kuitenkin ajoittain kohentunut normaaliksi?

Have you experienced long depressive periods during which your mood occasionally improved to a normal level?

10. Onko Sinulla ollut kausia, jolloin mielesi on vallannut synkkyys, joka ei ole tahtonut hellittää?

Have you had periods of pervasive gloom that did not seem to lift?

Answer choices:

1. ei koskaan (never)
2. silloin tällöin (occasionally)
3. usein (often)
4. hyvin usein (very often)

RESEARCH

Open Access



Age at separation, residential mobility, and depressive symptoms among twins in late adolescence and young adulthood: a FinnTwin12 cohort study

Zhiyang Wang¹, Alyce M. Whipp^{1,2}, Marja Heinonen-Guzejev¹ and Jaakko Kaprio^{1,2*}

Abstract

Background Separating with close siblings and leaving the parental home at an early age represents a major life event for an adolescent (reflected by age at separation in a twin pair) and may predispose them to poor mental health. This study aims to examine the association of age at separation and residential mobility on depressive symptoms in late adolescence and young adulthood and to explore possible underlying genetic effects.

Methods Residential mobility consisted of the number and total distance of moves before age 17. Based on 3071 twins from the FinnTwin12 cohort, we used linear regression to assess the association of age at separation and residential mobility with General Behavior Inventory (GBI) scores at age 17 and in young adulthood. A higher GBI score indicated more depressive symptoms occurred. Then, the mixed model for repeated measures (MMRM) was used to visualize the scores' trajectory and test the associations, controlling for "baseline" state. Twin analyses with a bivariate cross-lagged path model were performed between the difference in GBI scores, between cotwins, and separation status for the potential genetic influence.

Results Compared to twins separated before age 17, twins who separated later had significantly lower GBI scores at age 17 and in young adulthood. In MMRM, separation at a later age and a higher number of moves were associated with a higher GBI score in young adulthood. A small genetic effect was detected wherein GBI within-pair differences at age 17 were associated with separation status before age 22 (coefficient: 0.01).

Conclusion The study provides valid evidence about the influence of siblings and family on depressive symptoms in later adolescence and young adulthood while finding some evidence for a reverse direction effect. This suggests more caution in the interpretation of results. A strong association between residential mobility and depressive symptoms was affirmed, although further detailed research is needed.

Keywords Twins, Residential mobility, Depressive symptom, Adolescent

Background

The global burden of major depressive disorder (MDD) is getting progressively worse since the last century [1]. High-income areas had a higher prevalence of MDD [2]. In 2019, according to the Global Burden of Diseases, the age-standardized prevalence of depressive disorder was 3851 per 100,000 people in Western Europe and was

*Correspondence:

Jaakko Kaprio

jaakko.kaprio@helsinki.fi

¹ Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, PL 20 (Tukholmankatu 8), Helsinki FI-00014, Finland

² Department of Public Health, University of Helsinki, Helsinki, Finland



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

higher than the global level [3]. Data for the general US population shows the prevalence of depression in adolescents and young adults has been increasing since 2005 [4, 5]. Multiple risk factors for depression in childhood and adolescence have been identified such as low-quality education, stressful life events, and family influences [6].

During childhood, moving to a new location exposes children to greater social network loss than adults and their parents are less able to give them the attention they need during the process. Moving to a new location has been regarded as a type of negative life experience that can be associated with later depression [7, 8]. Residential instability or mobility describes those children who have an unstable life and are more frequently exposed to this deleterious experience. Numerous studies have suggested that high residential instability is linked to a larger risk of many psychological or psychiatric outcomes, including depression [7, 9–11]. However, the measure of residential mobility is often limited by recall bias from children's or their parents' questionnaires [9]. Longitudinal residential records from the national system could quantify the cumulative effect of mobility to enhance the assessment of the strength of this relationship [11].

In late adolescence and young adulthood, people will gradually part with their siblings and move to a new place due to various reasons. Because of the similar age and same generation, siblings spend more time together and have more mutual influence than parents. Long-term daily companionship from siblings promotes to a strong emotional connection, which has a significant impact on children's development [12]. Supportive and beneficial sibling relationships, such as spending more time with each other, appears to be interconnected with fewer depressive symptoms and higher well-being levels [13]. Positive sibling bonding is also capable of relieving detrimental parental influences [14]. For a relatively extreme example in China, AIDS orphans who were separated from their siblings suffered a higher level of anxiety, depression, anger, and dissociation [15]. As a special case of siblings, in general, twins have a closer relationship and reciprocity with each other than non-twin siblings [16, 17]. The separation between cotwins, alternatively termed "relational shift", breaks the intense intimacy and dependency and can be accompanied by emotional turmoil and confusion due to individuation [18]. Separation is an important challenge for twins during childhood and adolescence.

Moreover, separating with siblings/twins in adolescence or young adulthood may imply leaving from the parental home. Anxiety driven by early separation from parents has been connected to a series of negative mental health conditions such as complicated grief, more severe symptoms of depression, or later panic disorder

[19–21]. Parental support acts like a "safety net" to enhance the resilience in coping with challenges and to sustain the successful transition of the child to adulthood [22, 23]. Family closeness, support, attentiveness, and other positive factors are associated with a later timing of leaving, while teenagers living with step-parents or incomplete families tend to leave early due to a lack of resources [24–26].

The age at separation between cotwins reflects the length of time that twins stay together and also probably with their parents. It is likely associated with the amount of family support, degree of mental maturity, and preparation time for twins to address the separation challenge, connecting to later mental health. On the other hand, we also assume the time of separation is when at least one of the cotwins leaves the parental home. Unlike questions about the family environment in surveys, age at separation by residential history in twins could be used as an objective measure to describe the separation time, as well as familial influence, validly with less information bias. Therefore, in this study, we hypothesize that older age at separation is a protective factor and residential mobility is a risk factor and aim to determine: 1) the association between age at separation and depressive symptoms and the sex-specific effect, 2) the association between residential mobility before age 17 and depressive symptoms in late adolescence and young adulthood, and 3) the genetic influence behind the association between age at separation and depressive symptoms, based on the FinnTwin12 twin cohort.

Methods

This is a longitudinal twin cohort study.

Participants

The study participants were from the FinnTwin12 cohort, which is a population-based prospective cohort among all Finnish twins born between 1983 and 1987. Participants were divided into two parts: the overall cohort and the intensive study. A total of 5184 initial twins enrolled at wave one (the overall cohort) when they reached age 11, and twins filled in the general age 14 (wave two), age 17 (wave three), and young adult (wave four) questionnaires with 92%, 75%, and 66% retention, respectively. In the intensive study, 1035 families were additionally invited when the twin was at age 14 (wave two) for psychiatric interviews, some biological samples, and additional questionnaires, and 1854 twins participated. They were invited back for new interviews and other measurements as young adults (wave four). Interviews ($n=1347$) in young adulthood were completed for 73% of twins in the intensive study. The twins in the intensive study were mostly randomly selected from the overall epidemiology

study, while enriched with twins at elevated familial risk for alcoholism [27]. There was no evidence for selection for family type, parental age, residence, twin sex, and zygosity after carefully assessing the non-responders at each stage [28]. The flowchart of the general FinnTwin12 cohort is presented in supplemental Fig. 1. The updated review of this cohort was published in 2019 [29].

Measures

Age at separation

Age at separation was defined as the age that, for the first time, the twins no longer resided at the same address, with one or both twins moving to a new address as described elsewhere [30]. Reunification might occur sometimes after separation but was not considered in this study. Age at separation was generated from the residential records linking to the register of national Digital and Population Data Service Agency (DVV) in Finland since birth. The DVV, an official government agency, promotes digitalization, secures the availability of information, and provides various services including address services and moving notices. It updates the residential records rapidly ensuring accuracy. Separation indicated that the DVV residential records differed between cotwins. The residential records included the information on the north and east coordinates (EUREF-FIN), and the age moving in and out of each address. Then, we categorized it into four groups: separated before age 17, 17 to <19.5, 19.5 to <22, and age 22 or older. Subsequently, we created

two binary variables for separation status (yes, no) before ages 17 and 22 based on the age at separation to show whether twins separated at that age. Twins without valid information on age at separation (44 individuals) or twin pairs who never separated by the end of the follow-up in December 2020 (13 pairs) were excluded from the analysis with age at separation.

Residential mobility

Residential mobility consisted of two variables: the number of moves (never, once, twice, three or more times) and the total distance of moves (in kilometers) before age 17 accumulated across all moves. Both were generated from the residential records. If twins did not have address information due to missing addresses or living abroad for any period, their total distance of moves was invalid and excluded from the corresponding analysis (301 individual twins). Overall, the mean total distance of moves before age 17 was 48.4 (SD: 153.2) kilometers, and the total distance of moves was distributed in a right skewed fashion.

Depressive symptoms

We used the short-version General Behavior Inventory (GBI) to evaluate the depressive symptoms among twins [31]. It is a self-reported inventory designed to identify mood-related behaviors such as depressive and manic symptoms [32]. The short-scale version has 10 items and uses a 4-point Likert scale from 0 (never) to 3 (very often) to query the occurrence of depressive symptoms. The

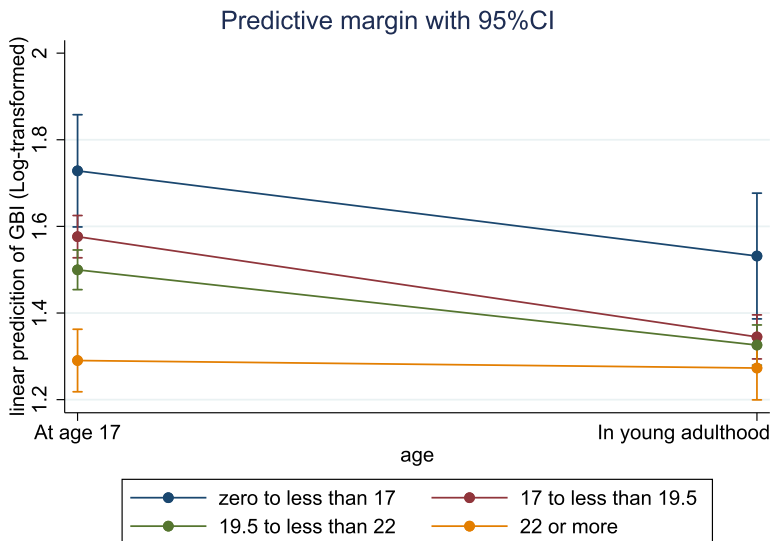


Fig. 1 Trajectory of log-transformed GBI score by the age at separation

total score ranges from 0 to 30. This reliable measure has been used in many Finnish studies including FinnTwin12 [31, 33, 34]. To validate the GBI, we compared it to a DSM-IV diagnosis of MDD assessed by the Semi-Structured Assessment for the Genetics of Alcoholism from the intensive study [35]. In a logistic regression analysis based on 1232 individual twins (mean age: 22.4), the GBI score in young adulthood strongly predicted MDD, with an area under the receiver operating characteristic curve (AUC) of 0.83. The GBI was assessed at age 17 and in young adulthood. We also calculated the change of GBI score between the two stages.

Another two variables for the within-pair differences in GBI score were created at both stages. The variables were constructed by subtracting the GBI score of one cotwin from the score of the other cotwin, and then we took the absolute values. Higher values indicated a higher level of within-pair difference.

Covariates

There were seven covariates defined a priori: sex (male, female), zygosity (monozygotic (MZ), dizygotic (DZ), unknown), smoking (never, quit, occasional, current) in young adulthood, work status (full time, part-time, irregular, not working) in young adulthood, secondary level school (senior high school, vocational, none) in young adulthood, paternal education, maternal education and age when twins provided the GBI assessment in young adulthood. Sex was assigned at birth because of biological dimensions. Paternal and maternal education (missing, low, medium, high) was reported by parents in wave one. These covariates were selected due to their relationship to either age at separation, mental health, or both, according to previous literature [30, 36, 37]. Additionally, the mean age when twins provided the GBI assessment in young adulthood was 24.18 years (SD: 1.68), which was only adjusted for in the analysis involving GBI in young adulthood.

Statistical analysis

A total of 3071 individual twins was included in analyses at final, of whom 2769 provided data at the age 17 survey and 3045 in young adulthood. Because of the differential missing information pattern, sample sizes in different analyses were slightly different.

Due to the skewness of the GBI scores at both ages, we added one to the GBI score and log-transformed it. Then, linear regression was used to examine the association of age at separation and residential mobility with log-transformed GBI scores at ages 17 and in young adulthood and change of GBI score. The pre-specified covariates were adjusted in the models. We used the univariate analysis of variance to determine the statistical differences between

the mean GBI score for the groups for categorical covariates. Additionally, the separation before age 17 (no/no information/reared apart, yes) was further adjusted for in the analysis with residential mobility.

Further, since we had two measures of GBI score, the mixed model for repeated measures (MMRM) was chosen to visualize the trajectory of GBI score by age at separation for twins at age 17 and in young adulthood. The MMRM method is a type of direct-likelihood analysis and uses longitudinal data to estimate the covariance between data from different time points [38]. This technique is able to consider the influence on the outcome (GBI in young adulthood) by the exposure of interest (fixed effect) and “baseline” GBI at age 17 (random effect) [39]. It also allowed us to handle the unevenness of missing information at different ages to increase the power. Covariates were adjusted for. Stratification analyses by sex were also performed, given that depression is more common in females than in males after puberty.

To evaluate the potential genetic influence between separation and depressive symptoms, intraclass correlation coefficients (r) were calculated at first, as an exploration. This was done to decompose overall variance into the genetic variances: additive genetic (A) and dominant genetic (D) components and environmental variances: common environmental (C), and unique environmental (E) components for depressive symptoms at both ages preliminarily. Based on the difference of r between the MZ and DZ twin pairs, the most appropriate model (ACE or ADE) would be selected, and then we calculated variance components based on intraclass correlations [40]. Differences in r between zygosity indicated the differential contribution of genetic and environmental variances of the overall variance in GBI scores, and stratification by age at separation showed its effects, such as modification, on the genetic and environmental variances. Furthermore, a bivariate cross-lagged path model was fitted with separation status and within-pair differences, similarity, in GBI score at both ages with the maximum likelihood estimator (twin pairs were analyzed as the unit of observation). This gauged the structural relationships of repeatedly measured variables and tested both directions between differences and separation, which would could suggest a potential role of genetics [41]. If within-pair difference (instead of at the individual level) leads to separation, genetic effects are suggested, while if separation leads to within-pair difference, we should considered the more environmental roles on the phenotype. Three covariates were adjusted for: sex and zygosity combination (male monozygotic (MMZ), female monozygotic (FMZ), male dizygotic (MDZ), female dizygotic (FDZ), and opposite-sex dizygotic (OSDZ)), paternal education, and maternal education. Twin pairs whose zygosity was

unknown or who had missing information on separation status and within-pair differences in GBI score at both ages were excluded from the cross-lagged path model.

Due to the possibility of reverse causality, wherein mental health conditions might influence twins to separate or leave home, we conducted a sensitivity linear regression to examine the association of age at separation with several mental health conditions at age 14, adjusting for sex, zygosity, paternal education, and maternal education. There were five subscales — depression, social anxiety, hyperactivity, aggression, and inattention — from the Multidimensional Peer Nomination Inventory (MPNI) was used to assess twins’ mental health at age 14. Each subscale contained a varying number of items (4-point Likert scale), and we used the mean of item responses as the outcome. The MPNI scale has been validated and showed predictive power in the previous study [27]. Due to different instruments to measure depressive symptoms between different ages, depressive symptoms from the MPNI at age 14 were not added to the MMRM. Then, we conducted another sensitivity analysis based on the linear regression for the potential confounding by mental health at age 14. The depression from MPNI was additionally adjusted as a covariate.

The first two hypotheses in the study were tested in the same types of models, but the age at separation and residential mobility were tested independently in models as the exposure of interest. Additionally, we controlled for the cluster effect by twin pair in the regression analyses through “robust” standard errors. Regression coefficients and 95% confidence intervals (CIs) were reported. A two-tailed $P < 0.05$ was considered statistically significant. All statistical analyses were performed using Stata 17.0 (StataCorp, College Station, TX, USA) and Mplus 8.6 (Muthen & Muthen, Los Angeles, CA, USA). The code used in this research can be found at Github (<https://github.com/doge73/age-at-separation-GBI>).

Results

Demographic characteristics

A total of 3071 twins were included in the final analysis including 1284 twin pairs with both twins included and another 503 individual twins (i.e. their cotwin had not participated or otherwise excluded as defined above). As shown in Table 1, the mean score for GBI assessment at age 17 was 5.1, and in young adulthood was 4.4. The majority of twins were female (56.4%), dizygotic (60.6%), and reported never smoking in young adulthood (53.5%). For socioeconomic status, work status, secondary level school, maternal education had significant between-group differences in either or both GBI scores at age 17 and young adulthood. Supplemental

Table 1 Demographic, socioeconomic characteristics, and depressive symptoms for individual twins

Characteristics	Individual n. (%)	Mean (SD) or P-value ^a	
		GBI at age 17 ^b	GBI in young adulthood ^c
Overall	3071	5.1 (4.9)	4.4 (4.68)
Sex		< 0.0001	< 0.0001
Male	1339 (43.6)	3.7 (3.9)	3.6 (4.3)
Female	1732 (56.4)	6.2 (5.2)	5.0 (4.9)
Zygosity		0.00	0.06
Monozygotic	1066 (34.7)	4.6 (4.6)	4.1 (4.5)
Dizygotic	1861 (60.6)	5.3 (5.0)	4.5 (4.7)
Unknown	144 (4.7)	5.7 (5.2)	4.8 (5.6)
Smoking		< 0.0001	< 0.0001
Never	1642 (53.5)	4.4 (4.4)	3.7 (4.1)
Quit	340 (11.1)	6.2 (5.4)	4.7 (5.0)
Occasional	309 (10.1)	5.7 (5.0)	4.5 (4.7)
Current	780 (25.4)	5.9 (5.3)	5.7 (5.4)
Work		0.11	< 0.0001
Full-time work	1582 (51.5)	4.9 (4.8)	4.0 (4.3)
Part-time work	394 (12.8)	5.1 (4.4)	4.5 (4.3)
Irregular work	372 (12.1)	5.1 (4.6)	4.4 (4.4)
Not working	723 (23.5)	5.5 (5.2)	5.3 (5.7)
Secondary level school		< 0.0001	< 0.0001
Senior high school	1850 (60.2)	4.9 (4.6)	4.1 (4.4)
Vocational	1042 (33.9)	5.0 (4.8)	4.4 (4.6)
None	179 (5.8)	7.9 (7.2)	7.7 (6.7)
Paternal education		0.06	0.39
Missing	496 (16.2)	5.5 (5.4)	4.6 (4.9)
Low	678 (22.1)	5.4 (5.0)	4.5 (4.8)
Medium	1496 (48.7)	4.9 (4.7)	4.3 (4.5)
High	401 (13.1)	5.0 (4.7)	4.5 (4.7)
Maternal education		0.00	0.03
Missing	225 (7.3)	6.1 (5.3)	5.4 (6.2)
Low	517 (16.8)	5.7 (5.3)	4.8 (4.9)
Medium	1863 (60.7)	5.0 (4.7)	4.3 (4.5)
High	466 (15.2)	4.7 (4.7)	4.4 (4.8)

^a P value was calculated from the analysis of variance
^b 302 individual twins were missing because they did not provide GBI assessment at age 17
^c 26 individual twins were missing because they did not provide GBI assessment in young adulthood

Table 1 shows the comparison of some characteristics between individual twins included in the analyses and was available at baseline ($n = 5104$, consented to participate in FinnTwin12 currently), and a similar distribution in both groups suggested a low risk of selection bias.

Association of age at separation with GBI

The sample sizes of each linear regression analysis are presented in Table 2. After adjustment for covariates, compared to twins separated before age 17, there were significant lower log-transformed GBI scores at age 17 in all three groups of twins, those who separated between ages 17 and 19.5 (coefficient: -0.16, 95% CI: -0.31, -0.02), between ages 19.5 and 22 (coefficient: -0.22, 95% CI: -0.37, -0.08), and after age 22 (coefficient: -0.44, 95% CI: -0.60, -0.28). Moreover, twins who separated between ages 19.5 and 22 and after age 22 also had lower log-transformed GBI scores in young adulthood (coefficient: -0.19, 95% CI: -0.37, -0.02 and coefficient: -0.25, 95% CI: -0.44, -0.06, respectively), compared to twins separated before age 17.

Figure 1 depicts the trajectory of the predictive marginal mean of log-transformed GBI score from age 17 to young adulthood. The three age-at-separation groups before age 22 decreased in their predictive marginal mean of log-transformed GBI scores, but the latest

age at separation group seems not to change notably. The results of the MMRM are presented in Table 3 and included 2712 twins at age 17 and 2962 twins in young adulthood. After controlling for log-transformed GBI score at age 17 and pre-specified covariates, twins who separated between ages 17 and 19.5 (coefficient: -0.19, 95% CI: -0.34, -0.04), between ages 19.5 and 22 (coefficient: -0.21, 95% CI: -0.36, -0.05), and after age 22 (coefficient: -0.26, 95% CI: -0.42, -0.09) scored significantly lower on the GBI (log-transformed) in young adulthood than those who separated before age 17.

Then we performed sex-stratified MMRM of age at separation (Supplemental Table 2). Regardless of the age at separation, males' mean GBI scores were notably lower than those of females. In females, after controlling for log-transformed GBI score at age 17 and pre-specified covariates, twins who separated after age 22 had lower log-transformed GBI scores in young adulthood (coefficient: -0.29, 95% CI: -0.52, -0.06), compared to twins separated before age 17.

Table 2 Association of age at separation and residential mobility with depressive symptoms (GBI) using linear regression

Characteristics	log-transformed GBI score at age 17			log-transformed GBI score in young adulthood			Change of GBI		
	Individual n. (%)	Unadjusted coefficient (95% CI)	Adjusted coefficient (95% CI) ^a	Individual n. (%)	Unadjusted coefficient (95% CI)	Adjusted coefficient (95% CI) ^b	Individual n. (%)	Unadjusted coefficient (95% CI)	Adjusted coefficient (95% CI) ^b
Age at separation	2725			2976			2700		
zero to less than 17	110 (4.04)	Ref	Ref	131 (4.40)	Ref	Ref	107 (3.96)	Ref	Ref
17 to less than 19.5	967 (35.49)	-0.18 (-0.34, -0.01)*	-0.16 (-0.31, -0.02)*	1078 (36.22)	-0.21 (-0.40, -0.02)*	-0.17 (-0.35, 0.00)	959 (35.52)	-0.41 (-1.41, 0.60)	-0.22 (-1.19, 0.75)
19.5 to less than 22	1146 (42.06)	-0.34 (-0.50, -0.18)*	-0.22 (-0.37, -0.08)*	1228 (41.26)	-0.31 (-0.50, -0.13)*	-0.19 (-0.37, -0.02)*	1133 (41.96)	-0.05 (-1.03, 0.94)	0.04 (-0.93, 1.00)
22 or more	502 (18.42)	-0.62 (-0.79, -0.44)*	-0.44 (-0.60, -0.28)*	539 (18.11)	-0.43 (-0.63, -0.23)*	-0.25 (-0.44, -0.06)*	501 (18.56)	0.65 (-0.36, 1.66)	0.67 (-0.34, 1.67)
Number of moves before age 17	2769			3045			2743		
None	736 (26.58)	Ref	Ref	776 (25.48)	Ref	Ref	728 (26.54)	Ref	Ref
Once	733 (26.47)	0.05 (-0.05, 0.15)	0.05 (-0.04, 0.13)	809 (26.57)	0.06 (-0.03, 0.16)	0.06 (-0.03, 0.15)	726 (26.47)	-0.10 (-0.58, 0.38)	-0.07 (-0.55, 0.40)
Twice	549 (19.83)	0.01 (-0.09, 0.12)	-0.00 (-0.10, 0.10)	602 (19.77)	0.12 (0.01, 0.23)*	0.09 (-0.01, 0.19)	544 (19.83)	0.56 (0.04, 1.08)*	0.52 (-0.00, 1.04)
Three times or more	751 (27.12)	0.07 (-0.03, 0.16)	0.00 (-0.09, 0.10)	858 (28.18)	0.15 (0.06, 0.25)*	0.08 (-0.01, 0.18)	745 (27.16)	0.37 (-0.16, 0.89)	0.40 (-0.15, 0.94)
Total distance of moving (per 100 km) before age 17	2505	0.01 (-0.02, 0.03)	0.01 (-0.02, 0.03)	2747	0.01 (-0.02, 0.03)	0.01 (-0.02, 0.03)	2482	0.00 (-0.12, 0.12)	0.00 (-0.12, 0.12)

^a Adjusted for sex, zygosity, smoking, work status, secondary level school, paternal education, and maternal education; for number of moves and total distance of moving before 17, further adjusted for separation before age 17

^b Adjusted for sex, zygosity, smoking, work status, secondary level school, paternal education, maternal education, and age when twins provided the GBI assessment in young adulthood; for number of moves and total distance of moving before 17, further adjusted for separation before age 17

* $P < 0.05$

Table 3 Association of age at separation with depressive symptoms (GBI) using MMRM

Age at separation	Mean (SD)		Adjusted coefficient (95% CI) ^a
	GBI at age 17 (individual n. = 2725)	GBI in young adulthood (individual n. = 2976)	Log-transformed GBI score in young adulthood
zero to less than 17	6.77 (5.02)	6.20 (5.81)	Ref
17 to less than 19.5	5.91 (5.28)	4.81 (4.89)	-0.19 (-0.34, -0.04)*
19.5 to less than 22	4.91 (4.63)	4.17 (4.37)	-0.21 (-0.36, -0.05)*
22 or more	3.65 (3.98)	3.79 (4.50)	-0.26 (-0.42, -0.09)*

^a Adjusted for sex, zygosity, smoking, work status, secondary level school, paternal education, maternal education, and age when twins provided the GBI assessment in young adulthood

* $P < 0.05$

The sensitivity analysis revealed significant correlations of age at separation with depression, hyperactivity, and inattention at age 14 (Supplemental Table 3), suggesting the presence of reverse causality. After further adjusting depression at age 14, compared to those who separated before age 17, twins who separated after age 22 had lower log-transformed GBI scores at age 17 and in young adulthood (Supplemental Table 4). Besides, twins who separated between age 17 and 19.5 also had lower log-transformed GBI scores in young adulthood. After further adjusting depression at age 14, MMRM showed that twins who separated between ages 17 and 19.5 (coefficient: -0.22, 95% CI: -0.42, -0.02), and after age 22 (coefficient: -0.27, 95% CI: -0.50, -0.05) had lower log-transformed GBI scores in young adulthood than those who separated before age 17 (Supplemental Table 5).

Association of residential mobility with GBI

In the linear regression, after adjustment for covariates, we did not observe any significant association of the number and total distance of moves before age 17 with any of the GBI variables (Table 2). Supplemental Fig. 2 depicts the trajectory of the predictive marginal mean of log-transformed GBI score from age 17 to young adulthood by the number of moves before age 17. The predictive marginal means decreased in all numbers of moves groups and reductions were more substantial in twins who never moved or moved once than in other groups. In MMRM, after controlling for baseline values and covariates, twins who moved twice or moved three or more times had higher GBI (log-transformed) scores in young adulthood than those who never moved (coefficient: 0.10, 95% CI: 0.01, 0.19 and coefficient: 0.09, 95% CI: 0.00, 0.17, respectively) (Supplemental Table 6). After further adjusting depression at age 14, associations attenuated to null (Supplemental Table 4).

Relationship between separation and within-pair difference

The overall intraclass correlation coefficients of GBI scores among MZ and DZ pairs were 0.56 and 0.14 at age 17 and 0.52 and 0.22 in young adulthood, respectively (Supplemental Table 7). The ADE model was selected (as $r_{MZ} > 2r_{DZ}$). At age 17, the genetic (A + D) and environmental (E) variances were 0.56 and 0.44, respectively. In young adulthood, the genetic and environmental variances were 0.52 and 0.48, respectively. Notably, for the GBI score in young adulthood, coefficients did not differ between MZs (0.55) compared to DZs (0.53) in twin pairs who separated before age 17.

Supplemental Table 8 presents the distribution of 1280 twin pairs, including in the cross-lagged path model, over the separation status and demographic characteristics. The within-pair difference in GBI scores decreased from age 17 to age 22 overall and in almost every subgroup, except MMZ and FDZ. Within-pair differences in GBI at age 17 were significantly associated with separation status before age 22 (coefficient: 0.01), with a small effect size. Moreover, there were other significant associations between the within-pair differences in GBI score at ages 17 and in young adulthood (coefficient: 0.22) and between separation status before age 17 and age 22 (coefficient: 0.19) (Supplemental Fig. 3).

Discussion

With a total of 3071 twins from the FinnTwin12 cohort included in this study, we found that later separation in twin pairs was associated with fewer depressive symptoms in later adolescence and young adulthood. In addition, higher residential mobility was linked to more depressive symptoms in young adulthood. Although females had more depressive symptoms than males, there were no obvious sex differences in the association between age at separation and depressive symptoms.

However, this assessment was shadowed by reverse causality to some extent, and interpretation and translation should be more cautious. Furthermore, we found a small genetic influence on the separation and fairly minor and inconsistent effects of separation on the difference between cotwins, indicating that age at separation is not a major modifier of sibling difference in depressive symptoms.

Parting from the family of origin is considered as a hallmark of becoming an adult, in which the youth is separated from their parents and siblings, loses their original familial network, and enters a completely new living environment. The majority of the age at separation of the twins in this study was between ages 17 and 22, which also implied life transitions such as entering university to some extent. The transition unavoidably leads to potential negative or positive mental health effects, which are complicated by sibling and familial contexts. Individuals with warm and encouraging siblings relationships may feel small loss, while siblings with relationships marked by conflict may feel mutual relief after separation [42]. Seiffe-Krenke found that the rate of family conflict was higher in adolescents who left home at the age of 21 and 23 years for females and males, respectively, than those still in the nest, and the in-time leaver pattern correlated with a worse level of psychological health [43]. Higher individual and family income were associated with higher odds of home leaving [44, 45], and lower financial stress was mutually correlated with fewer depressive symptoms at the stage of emerging adulthood [46]. With our results, we observed that longer or more family support, indicated by an older age at separation, as well as more mutual support by cotwins, could help adolescents cope with the mental health burden of separation from parents and siblings (same-aged twin sibling in our case) and leaving home. Separation is regarded as a part of the social/family exposome [47]. Extended family support including more relatives has been associated with lower rates of MDD and fewer depressive symptoms [48, 49]. Longer and more support could be extrapolated to better support. Better family involvement and sibling relationships including warmth, attention, and praise from parents predicted less MDD in adulthood and are aspects of corresponding interventions [50–52]. Familial cohesion (emotional bonding) and satisfactory social support were also shown to be helpful in reducing depressive symptoms in lower-middle-class communities in the US [53]. A favorable family background and relationships could be protective factors to support early adulthood transition, and stable transitions were associated with fewer depressive symptoms [54].

Given the evidence for a reverse effect of mental health on age at separation in this study, we further propose a

more dynamic effect between family, separation, and mental health. For example, a previous study showed, a later age of leaving was associated with closer residence to aging parents and more frequent contact, compared with siblings who moved out on time, depicting a positive effect from extended coresidence [55]. A UK longitudinal study suggested that parents' relationship quality and early childhood behavioral difficulties were reciprocally linked [56]. Additionally, there was a bidirectional influence of parenting behavior and conduct problems from childhood to adolescence among boys shown in a US study [57]. Furthermore, Stice and Barrera found a full reciprocal relationship between adolescent substance use, but not externalizing, and family support and control [58], while null evidence also exists [59]. On the other hand, age at separation reflected the shared environment between cotwins instead of the separation event itself, thus, to some degree, it also included the familial environment induced by twins' mental health. This means more complex interactions between individuals, siblings, and families. Future studies require more sophisticated models and a larger number of familial variables, and a lifecourse perspective is recommended.

Residential mobility involves the constant change of the built environment and could be regarded as a part of the physical exposome. Instable residence reduces family interaction, fragments education, cuts social connections, and leads to emotional stress [60]. A solid body of evidence has been established that a strong association between higher childhood residential mobility and poorer mental health at later ages across countries [7, 9, 11]. One study in North Ireland used census-based record linkage [61], while our findings could strengthen this association through longitudinal analysis with rapid updates. Solís et al. developed the biological plausibility that broad early negative experience increased physiological wear-and-tear, measured by allostatic load, which largely explained the health behaviors and education level in later ages [62]. Nevertheless, we should be aware that the effect could be complicated by children's emotionality, familial context, and reason of moves. Our null results regarding the distance of moves may be due to heterogeneous motivation, wherein a shorter distance could help to maintain the social network and longer distances could be related to work repositioning [7]. Physical moving itself does not necessarily correlate with the separation of family, and company with family is persisting. In a sample of 70 children from rural areas of Georgia, US., children with intense emotions had a surprising reduction in depressive symptoms as a function of increased residential mobility, while children with non-intense emotions experienced the opposite [10]. Additionally, familial support was able to buffer the harmful effect of

mobility on children's education and career achievement [63]. In our study, we had to exclude twins who moved abroad at least once, limiting our analysis by selection bias to some extent. Multiple studies have shown that non-Western immigrant adolescents have a higher risk of mental health problems compared to native European adolescents [64, 65].

Extensive research has investigated the direction between similarity and contact, and it seems that both directions exist, but the magnitude varies by age, sex, and traits [41, 66–68]. Separation is a turning point of contact, since the degree of contact inevitably decreases after separation. Based on the small effect size in the cross-lagged longitudinal analyses, the genetic influence on the separation was minimally observed. Age at separation might modify the similarity of depressive symptoms to a minor degree. The intraclass correlation coefficients were almost always higher in MZs than in DZs at all ages at separation in our study, but the difference decreased in young adulthood and with later separation, showing a slightly decreasing genetic effect. Between ages 17 and 22, the environmental variance increased, which was consistent with increasing E with age reported in a previous study [69]. This was also corroborated by the nearly identical coefficients in the earliest separation group between MZ and DZ in young adulthood. Moreover, as abovementioned, age at separation was reflected as a proxy to the entire familial exposome. Given the heritability of 37% for depressive symptoms in the previous analysis [70], there was vast room for environmental effects including the familial exposome, and our results suggested that genes play a role in the interaction of environmental effects with depressive symptom. A previous twin study in the UK highlighted the genetic influence in shaping a depressogenic environment in middle childhood as a gene–environment correlation [71]. Through DNA methylation, Carreras-Gallo et al. illustrated a gene–environment interaction mechanism through DNA methylation, elucidating its connection with the early-life exposome [72]. Thus, instead of implying any causality, our findings on the role of genetics suggested that age at separation should be considered in refining the assessment of both genetic and environmental influences, revealing the intricate relationships in future studies.

Our study was strengthened by the objective measure of the length that twins stayed in their parental home by age at separation through the national registry, which helps to avoid recall bias and provide more accurate information such as the date of moves. Moreover, in our previous study, the sensitivity and specificity of self-reported separation status were not ideal compared to the age at separation provided through the national

system [30]. In addition, the two variables on residential mobility helped us quantify its cumulative and longitudinal effect on the development of twins. Furthermore, the analysis, including repeated measures, could characterize the critical phase of depressive symptom development from late adolescence to young adulthood.

There were also several limitations in this study. First, the sensitivity analysis revealed the probability of reverse causality, suggesting a complex relationship between separation and mental health, especially considering the role of genetics and complex interpersonal network. Although, after adjusting for mid-adolescence depression, our main finding about age at separation was still valid, we should be cautious in interpreting and translating our findings. Second, the 10-item GBI was a non-diagnostic self-report measure of depressive symptoms. However, previous studies have shown high reliability and validity of the full-scale GBI from which this shorter version is derived [73]. As we had interview-based diagnostic information on MDD on the same participants, we could show a high validity of this short GBI. Third, the twin study may lead to the concern of low generalizability or representativeness. Most traits in adult twins including adaptive behavior, depression, and anxiety did not differ from singletons, which were suggested to be generalizable [74–76]. Fourth, when Finnish males are doing their mandatory military service, their residential records do not change to their service location, which may lead to information bias. Additionally, military service could possibly affect mental health [77], which suggests potential roles as a confounder or modifier. Compared to many other countries, the duration of Finnish military service is relatively short, but it is obligatory for all men. Cotwins usually start simultaneously and can spend weekends and holidays together at home. We believe that the influence of military service was neither random nor very substantial, and performed the sex-stratification analysis to investigate this further, but there was no obvious sex-specific effect. Fifth, some potential confounders were not included due to availability such as family wealth or personality. Physical environmental components, such as greenery, as a reflection of social segregation, during adolescence should be also considered in a lifecourse perspective [78]. Although parent's education could reflect the general family condition to some extent, future access to Finnish social and educational registers may help to address this concern. Finally, the overlap between personality, perhaps some other phenotypes, and depressive symptoms may interfere with the estimation [79], and we would like to further disentangle their complicated relationships with separation in the future.

Conclusions

Overall, the results provided a novel and valid estimation of the effect of age at separation on depressive symptoms in later adolescence and early adulthood. It could reflect the influence of the length and amount of support from the cotwin and family for twins on addressing this separation challenge and depressive symptoms, however, we should consider a more dynamic relationship between family, separation, and mental health. Moreover, we confirmed the association between residential mobility and mental health with longitudinal analysis. Further intervention targeting the family environment, could reduce the incidence of depressive symptoms.

Abbreviations

A	Additive genetic component
AUC	Area under the receiver operating characteristic curve
C	Common environmental component
CI	Confidence intervals
D	Dominant genetic component
DZ	Dizygotic
E	Unique environmental component
FDZ	Female dizygotic
FMZ	Female monozygotic
GBI	General Behavior Inventory
MDZ	Male dizygotic
MMZ	Male monozygotic
MZ	Monozygotic
OSDZ	Opposite-sex dizygotic
r	Intraclass correlation coefficient

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12889-024-19734-w>.

Supplementary Material 1

Acknowledgements

Not applicable.

Authors' contributions

All authors conceived and conceptualized the research idea and developed the analysis plan. Z.W. conducted data analysis. Z.W. drafted the original draft of the manuscript, and A.M.W., M.H.-G., and J.K. reviewed the manuscript and provided critical comments. J.K. gave critical supervision and guidance. All authors approved the manuscript.

Funding

Open Access funding provided by University of Helsinki (including Helsinki University Central Hospital). This research was partly funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 8747724 (Equal-Life). Equal-Life is part of the European Human Exposome Network. Data collection in FinnTwin12 has been supported by the National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to Richard J. Rose, and AA015416 to Jessica Salvatore) and the Academy of Finland (grants 100499, 205585, 118555, 141054, 264146, 308248, 312073, and 336823 to Jaakko Kaprio). Jaakko Kaprio acknowledges support by the Academy of Finland (grants 265240, 2632278).

Availability of data and materials

The datasets generated and analysed during the current study are not publicly available due to the restrictions of informed consent but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Informed consent has been obtained from the participants, their parents and legally authorized representatives in this study. The ethical committee of the Department of Public Health of the University of Helsinki, the ethics committee of the Helsinki University Central Hospital District, and Indiana University's Institutional Review Board approved the ethics of the study. All twins and their parents provided informed consent or written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 6 April 2023 Accepted: 8 August 2024

Published online: 17 August 2024

References

- COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet*. 2021;398(10312):1700–12. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7).
- Bromet E, Andrade LH, Hwang I, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med*. 2011;9(1):90. <https://doi.org/10.1186/1741-7015-9-90>.
- GBD 2019 Mental Disorders Collaborators. 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. *Lancet Psychiatry*. 2022;9(2):137–50. [https://doi.org/10.1016/S2215-0366\(21\)00395-3/ATTACHMENT/13E52F88-5D65-478F-843A-2CF574C4DF4/MMC1.PDF](https://doi.org/10.1016/S2215-0366(21)00395-3/ATTACHMENT/13E52F88-5D65-478F-843A-2CF574C4DF4/MMC1.PDF).
- Mojtabai R, Olfson M, Han B. National Trends in the Prevalence and Treatment of Depression in Adolescents and Young Adults. *Pediatrics*. 2016;138(6):e20161878. <https://doi.org/10.1542/peds.2016-1878>.
- Weinberger AH, Gbedemah M, Martinez AM, Nash D, Galea S, Goodwin RD. Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. *Psychol Med*. 2018;48(8):1308–15. <https://doi.org/10.1017/S0033291717002781>.
- Parker G, Roy K. Adolescent depression: a review. *Aust New Zeal J Psychiatry*. 2001;35(5):572–80. <https://doi.org/10.1080/0004867010060504>.
- Morris T, Manley D, Sabel CE. Residential mobility: towards progress in mobility health research. *Prog Hum Geogr*. 2016;42(1):112–33. <https://doi.org/10.1177/0309132516649454>.
- Cuijpers P, Smit F, Unger F, Stikkelbroek Y, ten Have M, de Graaf R. The disease burden of childhood adversities in adults: a population-based study. *Child Abuse Negl*. 2011;35(11):937–45. <https://doi.org/10.1016/j.chiabu.2011.06.005>.
- Jelleyman T, Spencer N. Residential mobility in childhood and health outcomes: a systematic review. *J Epidemiol Community Health*. 2008;62(7):584 LP – 592. <https://doi.org/10.1136/jech.2007.060103>.
- Stoneman Z, Brody GH, Churchill SL, Winn LL. Effects of Residential instability on head start children and their relationships with older siblings: influences of child emotionality and conflict between family caregivers. *Child Dev*. 1999;70(5):1246–62. <https://doi.org/10.1111/1467-8624.00090>.
- Simsek M, Costa R, de Valk HAG. Childhood residential mobility and health outcomes: A meta-analysis. *Health Place*. 2021;71: 102650. <https://doi.org/10.1016/j.healthplace.2021.102650>.
- Feinberg ME, Solmeyer AR, McHale SM. The Third Rail of Family Systems: Sibling Relationships, Mental and Behavioral Health, and Preventive Intervention in Childhood and Adolescence. *Clin Child Fam Psychol Rev*. 2012;15(1):43–57. <https://doi.org/10.1007/s10567-011-0104-5>.
- Yuan ASV. Sibling relationships and adolescents' mental health: the inter-relationship of structure and quality. *J Fam Issues*. 2009;30(9):1221–44. <https://doi.org/10.1177/0192513X09334906>.

14. Shumaker DM, Miller C, Ortiz C, Deutsch R. The forgotten bonds: the assessment and contemplation of sibling attachment in divorce and parental separation. *Fam Court Rev.* 2011;49(1):46–58. <https://doi.org/10.1111/j.1744-1617.2010.01352.x>.
15. Gong J, Li X, Fang X, et al. Sibling separation and psychological problems of double AIDS orphans in rural China – a comparison analysis. *Child Care Health Dev.* 2009;35(4):534–41. <https://doi.org/10.1111/j.1365-2214.2009.00969.x>.
16. Fortuna K, Goldner I, Knafo A. Twin relationships: a comparison across monozygotic twins, dizygotic twins, and nontwin siblings in early childhood. *Fam Sci.* 2010;1(3–4):205–11. <https://doi.org/10.1080/19424620.2010.569367>.
17. Mark KM, Pike A, Latham RM, Oliver BR. Using Twins to Better Understand Sibling Relationships. *Behav Genet.* 2017;47(2):202–14. <https://doi.org/10.1007/s10519-016-9825-z>.
18. Adelman MB, Siemon M. Communicating the relational shift: separation among adult twins. *Am J Psychother.* 1986;40(1):96–109. <https://doi.org/10.1176/appi.psychotherapy.1986.40.1.96>.
19. Silove D, Manicavasagar V, Curtis J, Blaszczyński A. Is early separation anxiety a risk factor for adult panic disorder? A critical review. *Compr Psychiatry.* 1996;37(3):167–79. [https://doi.org/10.1016/S0010-440X\(96\)90033-4](https://doi.org/10.1016/S0010-440X(96)90033-4).
20. Vanderwerker LC, Jacobs SC, Parkes CM, Prigerson HG. An Exploration of Associations Between Separation Anxiety in Childhood and Complicated Grief in Later Life. *J Nerv Ment Dis.* 2006;194(2). https://journals.lww.com/jonmd/Fulltext/2006/02000/An_Exploration_of_Associations_Between_Separation.8.aspx
21. Aschenbrand SG, Kendall PC, Webb A, Safford SM, Flannery-Schroeder E. Is childhood separation anxiety disorder a predictor of adult panic disorder and agoraphobia? A Seven-year longitudinal study. *J Am Acad Child Adolesc Psychiatry.* 2003;42(12):1478–85. <https://doi.org/10.1097/00004583-200312000-00015>.
22. Swartz TT, Kim M, Uno M, Mortimer J, O'Brien KB. Safety nets and scaffolds: parental support in the transition to adulthood. *J Marriage Fam.* 2011;73(2):414–29. <https://doi.org/10.1111/j.1741-3737.2010.00815.x>.
23. Newman T, Blackburn S. Transitions in the Lives of Children and Young People: Resilience Factors. Edinburgh: Scottish Executive Education Department; 2002.
24. Gillespie BJ. Adolescent intergenerational relationship dynamics and leaving and returning to the parental home. *J Marriage Fam.* 2020;82(3):997–1014. <https://doi.org/10.1111/jomf.12630>.
25. van den Berg L, Kalmin J, Leopold T. Family structure and early home leaving: a mediation analysis. *Eur J Popul.* 2018;34(5):873–900. <https://doi.org/10.1007/s10680-017-9461-1>.
26. Egondi T, Kabiru C, Beguy D, Kanyiva M, Jessor R. Adolescent home-leaving and the transition to adulthood: a psychosocial and behavioural study in the slums of Nairobi. *Int J Behav Dev.* 2013;37(4):298–308. <https://doi.org/10.1111/0165025413479299>.
27. Whipp AM, Korhonen T, Raevuori A, et al. Early adolescent aggression predicts antisocial personality disorder in young adults: a population-based study. *Eur Child Adolesc Psychiatry.* 2019;28(3):341–50. <https://doi.org/10.1007/s00787-018-1198-9>.
28. Kaprio J, Pulkkinen L, Rose RJ. Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. *Twin Res Hum Genet.* 2002;5(5):366–71. <https://doi.org/10.1375/twin.5.5.366>.
29. Rose RJ, Salvatore JE, Aaltonen S, et al. FinnTwin12 Cohort: An Updated Review. *Twin Res Hum Genet.* 2019;22(5):302–11. <https://doi.org/10.1017/thg.2019.83>.
30. Wang Z, Whipp A, Heinonen-Guzejev M, Kaprio J. Age at Separation of Twin Pairs in the FinnTwin12 Study. *Twin Res Hum Genet.* 2022;25(2):67–73. <https://doi.org/10.1017/thg.2022.17>.
31. Kokko K, Pulkkinen L. Unemployment and psychological distress: mediator effects. *J Adult Dev.* 1998;5(4):205–17. <https://doi.org/10.1023/A:1021450208639>.
32. Depue RA, Slater JF, Wolfstetter-Kausch H, Klein D, Goplerud E, Farr D. A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: a conceptual framework and five validation studies. *J Abnorm Psychol.* 1981;90(5):381–437. <https://doi.org/10.1037/0021-843X.90.5.381>.
33. Ranjit A, Korhonen T, Buchwald J, et al. Testing the reciprocal association between smoking and depressive symptoms from adolescence to adulthood: a longitudinal twin study. *Drug Alcohol Depend.* 2019;200:64–70. <https://doi.org/10.1016/j.drugalcdep.2019.03.012>.
34. Salmela-Aro K, Read S, Vuoksimaa E, et al. Depressive Symptoms and Career-Related Goal Appraisals: Genetic and Environmental Correlations and Interactions. *Twin Res Hum Genet.* 2014;17(4):236–43. <https://doi.org/10.1017/thg.2014.33>.
35. Bucholz KK, Cadoret R, Cloninger CR, et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J Stud Alcohol.* 1994;55(2):149–58. <https://doi.org/10.15288/jsa.1994.55.149>.
36. Baggio S, Iglesias K, Deline S, et al. Not in education, employment, or training status among young swiss men. Longitudinal associations with mental health and substance use. *J Adolesc Heal.* 2015;56(2):238–43. <https://doi.org/10.1016/j.jadohealth.2014.09.006>.
37. Fakhrunnisak D, Patria B. The positive effects of parents' education level on children's mental health in Indonesia: a result of longitudinal survey. *BMC Public Health.* 2022;22(1):949. <https://doi.org/10.1186/s12889-022-13380-w>.
38. Mallinckrodt CH, Lane PW, Schnell D, Peng Y, Mancuso JP. Recommendations for the primary analysis of continuous endpoints in longitudinal clinical trials. *Drug Inf J Drug Inf Assoc.* 2008;42(4):303–19. <https://doi.org/10.1177/009286150804200402>.
39. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. *JAMA.* 2016;315(4):407–8. <https://doi.org/10.1001/jama.2015.19394>.
40. van Dongen J, Slagboom PE, Draisma HHM, Martin NG, Boomsma DI. The continuing value of twin studies in the omics era. *Nat Rev Genet.* 2012;13(9):640–53. <https://doi.org/10.1038/nrg3243>.
41. Posner SF, Baker L, Heath A, Martin NG. Social contact, social attitudes, and twin similarity. *Behav Genet.* 1996;26(2):123–33. <https://doi.org/10.1007/BF02359890>.
42. Conger KJ, Little WM. Sibling relationships during the transition to adulthood. *Child Dev Perspect.* 2010;4:87–94. <https://doi.org/10.1111/j.1750-8606.2010.00123.x>.
43. Seiffe-Krenke I. Leaving home or still in the nest? Parent-child relationships and psychological health as predictors of different leaving home patterns. *Dev Psychol.* 2006;42(5):864–76. <https://doi.org/10.1037/0012-1649.42.5.864>.
44. Avery R, Goldscheider F, Speare A Jr. Feathered nest/gilded cage: Parental income and leaving home in the transition to adulthood. *Demography.* 1992;29(3):375–88. <https://doi.org/10.2307/2061824>.
45. Mulder CH, Clark WAV. Leaving home and leaving the State: evidence from the United States. *Int J Popul Geogr.* 2000;6(6):423–37. [https://doi.org/10.1002/1099-1220\(200011/12\)6:6<423::AID-IJPG199>3.0.CO;2-R](https://doi.org/10.1002/1099-1220(200011/12)6:6<423::AID-IJPG199>3.0.CO;2-R).
46. Cao H, Zhou N, Li X, Serido J, Shim S. Temporal dynamics of the association between financial stress and depressive symptoms throughout the emerging adulthood. *J Affect Disord.* 2021;282:211–8. <https://doi.org/10.1016/j.jad.2020.12.166>.
47. van Kamp I, Persson Wayne K, Kanninen K, et al. Early environmental quality and life-course mental health effects: The Equal-Life project. *Environ Epidemiol.* 2022;6(1). https://journals.lww.com/enviroepidem/Fulltext/2022/02000/Early_environmental_quality_and_life_course_mental.2.aspx.
48. Taylor RJ, Chae DH, Lincoln KD, Chatters LM. Extended family and friendship support networks are both protective and risk factors for major depressive disorder and depressive symptoms among African-Americans and black Caribbeans. *J Nerv Ment Dis.* 2015;203(2):132–40. <https://doi.org/10.1097/NMD.0000000000000249>.
49. Johnson JE, Esposito-Smythers C, Miranda R, Rizzo CJ, Justus AN, Clum G. Gender, Social Support, and Depression in Criminal Justice–Involved Adolescents. *Int J Offender Ther Comp Criminol.* 2010;55(7):1096–109. <https://doi.org/10.1177/0306624X10382637>.
50. Ehrenreich JT, Santucci LC, Weiner CL. Separation anxiety disorder in youth: phenomenology, assessment, and treatment. *Psicol Conductual.* 2008;16(3):389–412. <https://doi.org/10.1901/jaba.2008.16-389>.
51. Waldinger RJ, Vaillant GE, Orav EJ. Childhood sibling relationships as a predictor of major depression in adulthood: a 30-year prospective study. *Am J Psychiatry.* 2007;164(6):949–54. <https://doi.org/10.1176/ajp.2007.164.6.949>.
52. Chen P, Harris KM. Association of Positive Family Relationships With Mental Health Trajectories From Adolescence to Midlife. *JAMA Pediatr.* 2019;173(12):e193336–e193336. <https://doi.org/10.1001/jamapediatrics.2019.3336>.

53. Reinherz HZ, Stewart-Berghauer G, Pakiz B, Frost AK, Moeykens BA, Holmes WM. The relationship of early risk and current mediators to depressive symptomatology in adolescence. *J Am Acad Child Adolesc Psychiatry*. 1989;28(6):942–7. <https://doi.org/10.1097/00004583-19891000-00021>.
54. Booth A, Rustenbach E, McHale S. Early family transitions and depressive symptom changes from adolescence to early adulthood. *J Marriage Fam*. 2008;70(1):3–14. <https://doi.org/10.1111/j.1741-3737.2007.00457.x>.
55. Leopold T. The legacy of leaving home: long-term effects of coresidence on parent – child relationships. *J Marriage Fam*. 2012;74(3):399–412. <https://doi.org/10.1111/j.1741-3737.2012.00964.x>.
56. Panico L, Bécaries L, Webb EA. Exploring household dynamics: the reciprocal effects of parent and child characteristics. *Longit Life Course Stud*. 2014;5(1):42–55. <https://doi.org/10.14301/LLCS.V5I1.259>.
57. Pardini DA, Fite PJ, Burke JD. Bidirectional associations between parenting practices and conduct problems in boys from childhood to adolescence: the moderating effect of age and African-American ethnicity. *J Abnorm Child Psychol*. 2008;36(5):647–62. <https://doi.org/10.1007/s10802-007-9162-z>.
58. Stice E, Barrera M. A longitudinal examination of the reciprocal relations between perceived parenting and adolescents' substance use and externalizing behaviors. *Dev Psychol*. 1995;31(2):322–34. <https://doi.org/10.1037/0012-1649.31.2.322>.
59. Kim Y, Richards JS, Oldehinkel AJ. Self-control, mental health problems, and family functioning in adolescence and young adulthood: between-person differences and within-person effects. *J Youth Adolesc*. 2022;51(6):1181–95. <https://doi.org/10.1007/s10964-021-01564-3>.
60. Baker E, Pham NTA, Daniel L, Bentley R. How does household residential instability influence child health outcomes? A quantile analysis. *Int J Environ Res Public Health*. 2019;16(21):4189. <https://doi.org/10.3390/ijerph16214189>.
61. Tseliou F, Maguire A, Donnelly M, O'Reilly D. The impact of childhood residential mobility on mental health outcomes in adolescence and early adulthood: a record linkage study. *J Epidemiol Community Health*. 2016;70(3):278 LP – 285. <https://doi.org/10.1136/jech-2015-206123>.
62. Barboza Solís C, Kelly-Irving M, Fantin R, et al. Adverse childhood experiences and physiological wear-and-tear in midlife: findings from the 1958 British birth cohort. *Proc Natl Acad Sci*. 2015;112(7):E738–46. <https://doi.org/10.1073/pnas.1417325112>.
63. Hagan J, MacMillan R, Wheaton B. New kid in town: social capital and the life course effects of family migration on children. *Am Sociol Rev*. 1996;61(3):368–85. <https://doi.org/10.2307/2096354>.
64. Duinhof EL, Smid SC, Vollebergh WAM, Stevens GWJM. Immigration background and adolescent mental health problems: the role of family affluence, adolescent educational level and gender. *Soc Psychiatry Psychiatr Epidemiol*. 2020;55(4):435–45. <https://doi.org/10.1007/s00127-019-01821-8>.
65. McMahon EM, Corcoran P, Keeley H, et al. Mental health difficulties and suicidal behaviours among young migrants: multicentre study of European adolescents. *BJPsych Open*. 2017;3(6):291–9. <https://doi.org/10.1192/bjpo.bp.117.005322>.
66. Kaprio J, Koskenvuo M, Rose RJ. Change in cohabitation and intrapair similarity of monozygotic (MZ) cotwins for alcohol use, extraversion, and neuroticism. *Behav Genet*. 1990;20(2):265–76. <https://doi.org/10.1007/BF01067794>.
67. Lykken DT, McGue M, Bouchard TJ, Tellegen A. Does contact lead to similarity or similarity to contact? *Behav Genet*. 1990;20(5):547–61. <https://doi.org/10.1007/BF01065871>.
68. Rose RJ, Kaprio J, Williams CJ, Viken R, Obrenski K. Social contact and sibling similarity: Facts, issues, and red herrings. *Behav Genet*. 1990;20(6):763–78. <https://doi.org/10.1007/BF01065919>.
69. Nivard MG, Dolan CV, Kendler KS, et al. Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. *Psychol Med*. 2015;45(5):1039–49. <https://doi.org/10.1017/S003329171400213X>.
70. Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. *Am J Psychiatry*. 2000;157(10):1552–62. <https://doi.org/10.1176/appi.ajp.157.10.1552>.
71. Wilkinson PO, Trzaskowski M, Haworth CMA, Eley TC. The role of gene–environment correlations and interactions in middle childhood depressive symptoms. *Dev Psychopathol*. 2013;25(1):93–104. <https://doi.org/10.1017/S0954579412000922>.
72. Carreras-Gallo N, Cáceres A, Balagué-Dobón L, et al. The early-life exposome modulates the effect of polymorphic inversions on DNA methylation. *Commun Biol*. 2022;5(1):455. <https://doi.org/10.1038/s42003-022-03380-2>.
73. Pendegast LL, Youngstrom EA, Brown C, Jensen D, Abramson LY, Alloy LB. Structural invariance of General Behavior Inventory (GBI) scores in black and white young adults. *Psychol Assess*. 2015;27(1):21–30. <https://doi.org/10.1037/pas0000020>.
74. Munn-Chernoff MA, von Ranson KM, Culbert KM, Larson CL, Burt SA, Klump KL. An examination of the representativeness assumption for twin studies of eating pathology and internalizing symptoms. *Behav Genet*. 2013;43(5):427–35. <https://doi.org/10.1007/s10519-013-9603-0>.
75. Pulkkinen L, Vaalamo I, Hietala R, Kaprio J, Rose RJ. Peer reports of adaptive behavior in twins and singletons: is twinship a risk or an advantage? *Twin Res*. 2003;6(2):106–18. <https://doi.org/10.1375/twin.6.2.106>.
76. Willemsen G, Odintsova V, de Geus E, Boomsma DI. Twin-Singleton Comparisons Across Multiple Domains of Life BT - Twin and Higher-order Pregnancies. In: Khalil A, Lewi L, Lopriore E, eds. Springer International Publishing; 2021:51–71. https://doi.org/10.1007/978-3-030-47652-6_4.
77. Appelqvist-Schmidlechner K, Upanne M, Henriksson M, Parkkola K, Stengård E. Young men exempted from compulsory military or civil service in Finland — a group of men in need of psychosocial support? *Scand J Public Health*. 2010;38(2):168–76. <https://doi.org/10.1177/1403494809357103>.
78. Di Stefano G, Cyr F. Child adjustment following parental separation: the role of maternal well-being, parenting quality, and household income. *J Child Custody*. 2014;11(1):5–24. <https://doi.org/10.1080/15379418.2014.892802>.
79. Smith MM, Sherry SB, Rnic K, Saklofske DH, Enns M, Gralnick T. Are perfectionism dimensions vulnerability factors for depressive symptoms after controlling for neuroticism? A meta-analysis of 10 longitudinal studies. *Eur J Pers*. 2016;30(2):201–12. <https://doi.org/10.1002/per.2053>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

The effect of environment on depressive symptoms in late adolescence and early adulthood: an exposome-wide association study and twin modeling

Received: 6 April 2023

Accepted: 9 August 2023

Published online: 25 September 2023

 Check for updatesZhiyang Wang¹, Stephanie Zellers¹, Alyce M. Whipp^{1,2}, Marja Heinonen-Guzejev², Maria Foraster^{3,4,5,6}, Jordi Júlvez^{4,7}, Irene van Kamp⁸ & Jaakko Kaprio^{1,2}✉

The exposome represents the totality of environmental effects, but systematic evaluation between it and depressive symptoms is scant. Here we sought to comprehensively identify the association of the exposome with depressive symptoms in late adolescence and early adulthood and determine genetic and environmental covariances between them. Based on the FinnTwin12 cohort (3,025 participants in young adulthood and 4,127 at age 17), the exposome-wide association study (ExWAS) design was used to identify significant exposures from 12 domains. Bivariate Cholesky twin models were fitted to an exposome score and depressive symptoms. In ExWASes, 29 and 46 exposures were significantly associated with depressive symptoms in young adulthood and at age 17, respectively, and familial exposures were the most influential. Twin models indicated considerable genetic and environmental covariances between the exposome score and depressive symptoms with sex differences. The findings underscore the systematic approach of the exposome and the consideration of relevant genetic effects.

Depressive symptoms are a type of chronic mental health condition with complex etiology, and major depressive disorder (MDD) is the clinical disorder diagnosed when depressive symptoms reach a threshold of severity and duration. Depressive symptoms and MDD lead to a serious public health burden. The updated Global Burden of Diseases study showed that the age-standardized prevalence of MDD was 4% (3,951 per 100,000 people) in Western Europe, higher than the global level, and underlined the heavy burden on people aged between 15 and 24 (ref. 1). Among adolescents, a 2021 systematic review indicated that

the pooled prevalence of self-reported depressive symptoms was 34% and of MDD was 5% from the studies between 2001 to 2020, and the prevalence is increasing². The COVID-19 pandemic exacerbated the already growing trend of hardship. Given a growing body of evidence on the environmental effect on depressive symptoms and MDD^{3,4}, more systematic investigation is urgently needed, especially among youth.

The concept of the ‘exposome’, which depicts the dynamic totality of the environment that an individual experiences, was raised in 2005⁵. The exposome is divided into three parts—specific external, general

¹Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, Helsinki, Finland. ²Clinicum, Department of Public Health, University of Helsinki, Helsinki, Finland. ³PHAGEX Research Group, Blanquerna School of Health Science, Universitat Ramon Llull (URL), Barcelona, Spain. ⁴ISGlobal, Barcelona, Spain. ⁵Universitat Pompeu Fabra (UPF), Barcelona, Spain. ⁶CIBER Epidemiología y Salud Pública (CIBEResp), Madrid, Spain. ⁷Clinical and Epidemiological Neuroscience (NeuroEpi), Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Spain. ⁸Centre for Sustainability, Environment and Health, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands.

✉ e-mail: jaakko.kaprio@helsinki.fi

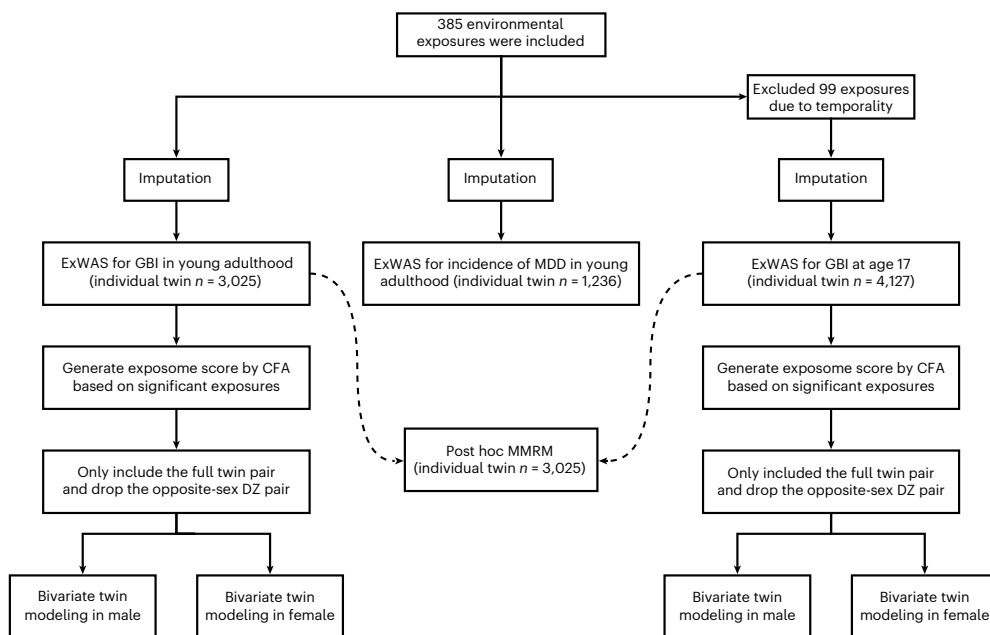


Fig. 1 | Flowchart of the analysis pipeline. Flowchart of the analysis pipeline demonstrating the path from the choice of exposures to ExWAS analysis and ending with bivariate twin modeling. The full path was used for depressive symptoms (GBI) at two ages. Only ExWAS was completed for MDD.

external, and internal exposomes—and the external exposome could be further subdivided into the familial, social, built exposome, and so on. Instead of studying a single or small group of exposures, an exposome study aims to investigate the overall effect of the environment while, unavoidably, complexities such as interaction or ubiquity increase the difficulty⁶. An exposome- (environmental, exposure) wide association study (ExWAS), like other ‘WAS’ studies, denotes an agnostic and systematic method for hypothesis generating, which is comparatively appropriate to the exposome’s spatiotemporal variabilities and multi-level structure⁷. Several ExWAS studies have targeted mental health^{8–10}, and Choi et al.¹¹ used clinical incident depression as the outcome and identified multiple modifiable factors. As it is the early warning sign of MDD, focusing on depressive symptoms in adolescence or young adulthood could be easier to guide translational intervention as early as possible, which would be more cost effective.

Despite the benefits of the exposome approach, there are some other hindrances. First, under the current technique, we cannot measure every possible exposure (far from reaching ‘1-genome’), and the exposome keeps updating, expanding, and enriching. Moreover, some studies have emphasized exposures’ non-genetic properties, which ignores how the environment interacts with genetics through multiple mechanisms among many traits, including depression^{12,13}. Medda and colleagues, on the basis of the Italian Twin Registry, demonstrated the substantial genetic role in exogenous metallomics, where the estimations of standardized genetic variance, as a proportion of total variance of the measured exposures, ranged from 0.15 (arsenic) to 0.79 (zinc)¹⁴. As a natural experiment, twin and family studies provide a method to evaluate genetic and environmental relationships between traits and exposures. This design decomposes the variance of traits into additive genetic (A), dominant genetic (D), common environmental (C), and unique environmental (E) components, which contain the distinct features of the exposome as the overall environmental effect. Such indirect evidence of genetic effects based on genetic relationships of

family members is an efficient way to demonstrate the presence (or lack of) genetic effects. Thus, the combination of exposome and twin studies could advance our knowledge of the complexities between genes and environments, improve our understanding of existing deficiencies in exposome measures, and produce further research questions. A natural extension is then to include measured genotypes, either targeting specific genes such as those involved in the metabolism of external compounds or more broad-based genome-wide approaches to derive polygenic scores of genetic susceptibility.

In this study, based on the FinnTwin12 cohort, we aim to (1) comprehensively and systematically determine exposures that are significantly associated with depressive symptoms and MDD in late adolescence and early adulthood through three ExWASes and (2) estimate to what extent the exposome and depressive symptoms share the same genetic and environmental risk factors.

Results

Characteristics of the study, participants, and exposures

Figure 1 shows the flowchart of the analysis pipeline, which consisted of three ExWASes and the following bivariate twin modeling. Per the FinnTwin12 cohort, there were 3,025, 1,236, and 4,127 individual twins included in three separate ExWASes with the outcomes of general behavior inventory (GBI) score in young adulthood (primary), the incidence of MDD in young adulthood, and GBI score at age 17, respectively. The characteristics of each ExWAS are shown in Table 1.

For individual twins included in ExWASes of all outcomes (Table 2), the majority were female and from dizygotic (DZ) pairs, and their parental education levels were limited (less than high school). At age 17, 25.4% of individual twins reported being current smokers, and 82.6% were full-time students and not working. In young adulthood, 25.3% of individual twins reported that they were currently smoking, and 51.4% had a full-time job. The mean GBI scores at age 17 and in young adulthood were 5.0 (s.d.: 4.9) and 4.4 (s.d.: 4.7), respectively, and the

Table 1 | Characteristics of ExWASes

Outcome	Number of individual twins	Number of exposures	Number of P values	Significant threshold ($-\log_{10}(P\text{value})$)
GBI in young adulthood	3,025	385	501	3.51
Incidence of MDD in young adulthood	1,236	385	501	3.47
GBI at age 17	4,127	286	394	3.44

two measures correlated with 0.49. The incidence of lifetime MDD in young adulthood was 12.3%.

Exposures' code names, description, and statistics based on twins included in the ExWAS of GBI in young adulthood (before imputation) are presented in Supplementary Table 1. There are 12 domains of exposures, colored in the following plots: air pollution, building, blue and green spaces, population density, geocoordinates, prenatal exposures, passive smoking, family and parents, friend and romantic relationships, school and teachers, stressful life events, and social indicators. In principal component analysis (PCA), the first principal component (PC1) attributed only 10.93% and 10.66% to the total variability of all included exposures in young adulthood and at age 17, respectively (Extended Data Fig. 1). From the scatter plots of PC1 and PC2, we identified some potential clusters of exposures from domains of building, blue and green spaces, and social indicators via visual assessment.

ExWAS of depressive symptoms and MDD in young adulthood
The adjusted coefficient and $-\log_{10}(P\text{value})$ of all exposures included for both adult outcomes are presented in Supplementary Table 2. There were 40 significant P values in 29 exposures, which were associated with log-transformed GBI score in young adulthood, identified from 385 exposures (Fig. 2a). There were 24, 2, and 3 exposures belonging to the domains of family and parents, friend and romantic relationships, and school and teachers, respectively. For the most protective exposure, compared with twins who felt their home environment was completely unfair, quite unfair, or somewhat unfair at age 17 (unfair_A17), twins who felt it was not at all unfair at age 17 were associated with a 0.40 lower log-transformed GBI score (95% confidence interval (CI): $-0.50, -0.31$) (Fig. 2b). For the most harmful exposure, compared with twins who were completely satisfied with their relationships with friends at age 14 (sat_friend_A14), twins who felt somewhat satisfied, mainly not satisfied, or not at all satisfied at age 14 were associated with a 0.42 higher log-transformed GBI score (95% CI: 0.29, 0.55) (Fig. 2b). By contrast, none of the exposures showed a significant association with MDD (Extended Data Fig. 2).

ExWAS of depressive symptoms at age 17
The adjusted coefficient and $-\log_{10}(P\text{value})$ for the age 17 outcome are presented in Supplementary Table 2. There were 71 significant P values in 46 exposures, which were significantly associated with log-transformed GBI score, identified from 286 exposures (Extended Data Fig. 3a). There were 32, 6, 4, and 4 exposures belonging to the domains of family and parent, friend and romantic relationship, school and teachers, and stressful life events, respectively. For the most harmful exposures, compared with twins who were completely satisfied with their success at work or studies at age 17 (sat_studywork_A17), twins who felt mainly not satisfied or not at all satisfied at age 17 were associated with a 0.65 higher log-transformed GBI score (95% CI: 0.55, 0.74) (Extended Data Fig. 3b). For the most protective exposure, the same as the result in young adulthood, compared with twins who felt their home environment was completely unfair, quite unfair, or somewhat unfair at age 17 (unfair_A17), twins who felt it was not at all unfair at age 17 were associated with a 0.50 lower log-transformed GBI score

(95% CI: $-0.57, -0.43$) (Extended Data Fig. 3b). There are 27 exposures that are significantly associated with log-transformed GBI scores both in young adulthood and at age 17, and 22 exposures belong to the domain of family and parents.

Twin modeling of depressive symptoms with exposome scores
Before the bivariate modeling, the best-fit univariate AE model (had the lowest Akaike information criterion compared with ADE and E models) indicated E explained 61% of the variance of depressive symptoms in males and 45% in females at age 17, and the numbers slightly reduced to 59% and 42%, respectively, in young adulthood (Supplementary Table 3). The exposome score was created by confirmatory factor analysis (CFA) based on the significant exposures from ExWASes. The standardized root mean square residual of models in young adulthood and at age 17 were 0.100 and 0.078, respectively, indicating acceptable model fit. MDD was not included in the CFA or following twin modeling due to the smaller sample size and no significant exposure being identified. Then we used the exposome score to conduct bivariate twin modeling between the exposome score and depressive symptoms. Given the sex differences in the prevalence of depressive symptoms, the differences in heritability, and the fact that sex-limited bivariate models also indicated significant sex differences (Supplementary Table 4) at both age points, we ran the bivariate models separately for males and females.

Figure 3 and Supplementary Table 5 show the path coefficients for the model for exposome score and log-transformed GBI score in young adulthood (mean age: 23.9). Unique environmental factors accounted for 23% and 13% of the covariances in males and females, respectively, while additive genetic factors accounted for 77% in males and 87% in females. In males, standardized variances of E_{exposome} and E_{GBI} were 0.32 (95% CI: 0.26, 0.39) and 0.51 (95% CI: 0.42, 0.62); the numbers reduced to 0.25 (95% CI: 0.21, 0.30) and 0.50 (95% CI: 0.42, 0.58) in females. The remaining share of variance was accounted for by additive genetic effects.

Extended Data Fig. 4 and Supplementary Table 5 show the path coefficients for the model for exposome score and log-transformed GBI score at age 17. Unique environmental factors accounted for 31% and 13% of the covariances in males and females, respectively. Additive genetic factors accounted for 69% in males and 87% in females. The standardized variances of E_{exposome} at age 17 are similar to E_{exposome} in young adulthood regardless of sex. The standardized variance of E_{exposome} is 0.26 (95% CI: 0.22, 0.30) and 0.22 (95% CI: 0.19, 0.25) and of E_{GBI} is 0.64 (95% CI: 0.55, 0.73) and 0.44 (95% CI: 0.38, 0.50) in males and females, respectively. The remaining share of variance was accounted for by additive genetic effects.

Post hoc mixed model repeated measures. On the basis of the longitudinal design and 27 significant exposures selected by both ExWASes of log-transformed GBI score, after adjusting for covariates and baseline effect, all the exposures were still significantly associated with log-transformed GBI score in young adulthood. The results are presented in Supplementary Table 6.

Discussion
Using data on depressive symptoms and diagnosed MDD from the FinnTwin12 study and a wide range of exposures from multiple sources, we applied a two-stage analysis to first screen the exposome and then estimate the environmental sources of correlation between the exposome and depressive symptoms via twin modeling. First, multiple exposures by self-report have been identified across domains of family and parents, friend and romantic relationships, school and teachers, and stressful life events, which were significantly associated with depressive symptoms in young adulthood and at age 17. By contrast, none of the exposures correlated with the incidence of MDD in young adulthood. Second, after generating an exposome score based on significantly associated exposures, the best-fitting bivariate AE models indicated

Table 2 | Characteristics of included twins according to the ExWAS

Characteristics	N (%) / mean (s.d.)		
	Participants included in the ExWAS of		
	GBI (individual twin n=3,025)	Incidence of MDD (individual twin n=1,236)	GBI (individual twin n=4,127)
	In young adulthood		At age 17
GBI score	4.4 (4.7)	—	5.0 (4.9)
MDD incidence			
Yes	—	152 (12.3)	—
No	—	1,084 (87.7)	—
Sex			
Male	1,318 (43.6)	564 (45.6)	1,988 (48.2)
Female	1,707 (56.4)	672 (54.4)	2,139 (51.8)
Zygoty			
MZ	1,050 (34.7)	513 (41.5)	1,362 (33.0)
DZ	1,833 (60.6)	721 (58.3)	2,577 (62.4)
Unknown	142 (4.7)	2 (0.2)	188 (4.6)
Parental education			
Limited	1,743 (57.6)	672 (54.4)	2,392 (58.0)
Intermediate	666 (22.0)	305 (24.7)	950 (23.0)
High	616 (20.4)	259 (21.0)	785 (19.0)
Smoking			
Never	1,617 (53.5)	614 (49.7)	1,218 (29.5)
Former	339 (11.2)	115 (9.3)	1,418 (34.4)
Occasional	304 (10.1)	132 (10.7)	445 (10.8)
Current	765 (25.3)	375 (30.3)	1,046 (25.4)
Work (young adulthood)			
Full-time work	1,556 (51.4)	497 (40.2)	—
Part-time work	388 (12.8)	236 (19.1)	—
Irregular work	368 (12.2)	338 (27.4)	—
Not working	713 (23.6)	165 (13.4)	—
Secondary-level school (young adulthood)			
Vocational	1,025 (33.9)	377 (30.5)	—
Senior high school	1,826 (60.4)	778 (62.9)	—
None	174 (5.8)	81 (6.6)	—
Age (young adulthood)	24.2 (1.7)	22.4 (0.7)	—
Study and work status (age 17)			
Neither study nor work	—	—	150 (3.6)
Only study	—	—	3,406 (82.5)
Both study and work	—	—	571 (13.8)

that unique environmental effects accounted for a marked fraction of the covariance between the exposome score and depressive symptoms. This environmental fraction was higher in males than in females, suggesting a notable sex difference. Our result implies that environmental effects are more impactful compared with genetic effects in males than in females.

Influence from the familial component of the social exposome, especially from the familial atmosphere, was demonstrated by our evidence as having the most substantial impact on depressive symptoms

in late adolescence and early adulthood and their trajectory. A large Chinese survey also found that familial factors such as cohesion, conflict, and control correlated with the occurrence of depressive symptoms among university students¹⁵. Other studies have revealed the connection of family triangulation (parent–child coalition and alliance) and satisfaction with depressive symptoms from childhood to late adolescence across countries^{16,17}. Fairness (largest protective effect size of GBI at both age points), as a dimension of parentification, was demonstrated as a unique predictor of mental health symptoms¹⁸. These existing conventional investigations were consistent with ours, while our ExWAS more systematically evaluated a wide range of exposures and reduced the chance of type I error without any pre-identified hypothesis. Moreover, instead of traditional scales for assessing familial and interpersonal relationships, we treated each scale component as an ‘independent’ exposure in models, which helped us to identify new correlations, detect the relative importance, and prepare for further analysis of more intricate relationships between different components and depressive symptoms.

Results from bivariate twin modeling reveal a complex relationship among genes, environments, and depressive symptoms. Although the unique environmental factor explains a notable amount of covariance between exposome score and depressive symptoms, the additive genetic factor explained relatively more. Many significant exposures were chosen under the guidance of the exposome paradigm, but it does not necessarily imply a pure environmental effect. Many familial influences are considered ‘inheritable factors’ between generations to a certain extent, according to the intergenerational transmission theory. Such effects can be transmitted from parents to children through shared genes but also by shared environments. Early studies have found that life satisfaction or family violence from parents and origin families led to an important impact on the development of subsequent similar familial environments among offspring^{19,20}. Moreover, we should consider the existence of the gene–environment interaction (G×E), which suggests the different effects of a genotype on disease risk in persons with different environmental exposures²¹. Choi et al¹¹ stratified the ExWAS by polygenic risk scores of major depression and found that some significant factors in the full sample became null in the genetically at-risk sample. Another study suggested the multiple modulation pathways by exposure to DNA methylation, through numerous testing, regarded as the G×E-WAS²². In addition, previous twin studies found geographic confounding in the assessment of A, C, and E variances, possibly attributable to differences in genetic ancestry. Results from the Netherlands Twin Register found 1.8% of the variance in children’s height was captured by regional clustering²³. In the Netherlands, there were strong genetic differentiations between the north and south, between the east and west, and between the middle band and the rest of the country by PCA on genome-wide data²⁴. In the Finnish population, also a substantial population structure difference is observed between the east and west parts of country²⁵. In brief, the hidden heritable and genetic factors critically influence the association between the exposome and depressive phenotype through various mechanisms, which potentially lead to a propensity to weak associations in our findings.

Furthermore, exposures from the more external domains, particularly in the physical exposome, also showed, at most, weak connections with depressive symptoms. While it may be the case that the relative importance of the physical exposome is much less than that of the social and familial exposome with respect to depressive symptoms, there are possibly other explanations. First, a more complex structure of the exposome, such as the interaction or correlation between individual exposures and external exposome, may exist. Some previous exposome analyses have indicated this^{26,27}, but the ExWAS design cannot characterize it. For example, the social exposome is an explaining part of the physical exposome, which could not be completely separated. We aim to investigate the complicated effect of the depressive phenotype in the pluralistic platform like machine learning on the

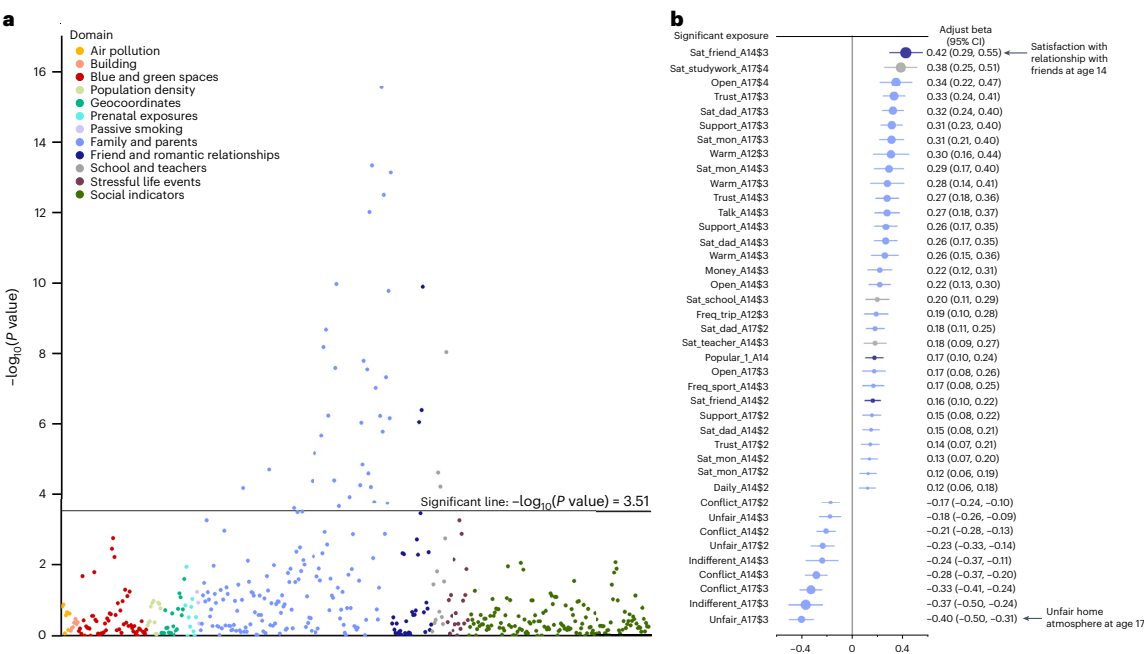


Fig. 2 | Association results between exposure and log-transformed GBI score in young adulthood, adjusted for covariates (individual twin $n = 3,025$), using generalized linear regression^a. **a, Manhattan association plot for exposures in relation to log-transformed GBI score in young adulthood. The y axis is showing statistical significance as $-\log_{10}(P \text{ value})$ for the adjustment for multiple testing. **b**, Forest plot for the adjusted beta for significant exposures in descending order from top to bottom (from harmful to protective). The center**

dot and bar present the effect size (coefficient of linear regression) and 95% CI, and the sizes of the dots present the effect size relatively. The color legend applies to both **a** (Manhattan association plot) and **b** (forest plot). The adjusted covariates were sex, zygosity, parental education, smoking in young adulthood, work status in young adulthood, secondary-level school in young adulthood, and age when twins provided the GBI assessment in young adulthood.

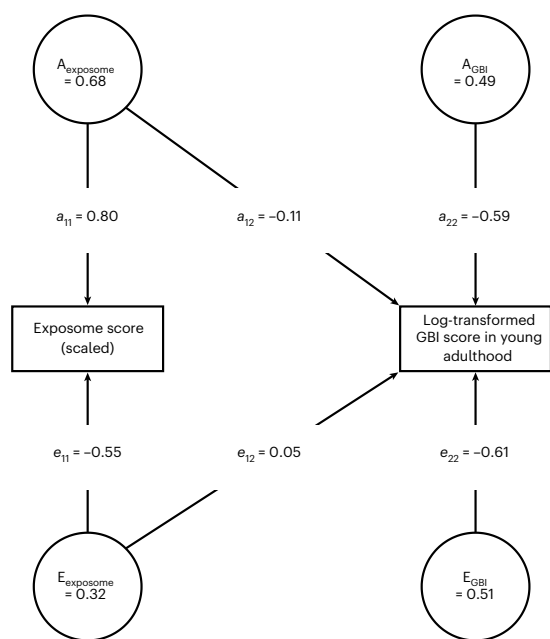
basis of our findings in the future. Second, Finland has been ranked very high in the beneficial environmental effect on the child by UNICEF (United Nations Children's Fund), providing environments with low air pollution, high greenness, safe water, and other constructive aspects relatively equally to most residents in childhood and adolescence²⁸. It could explain null results with external living environments due to a lack of individual variation in exposures. Another matter contributing to large familial effects is the overlap between interpersonal relationships and depressive symptoms. In a Swedish twin study among females, interpersonal relationships contributed between 18% and 31% of the variance for depressive symptoms²⁹. Some personality disorders are tightly connected with interpersonal relationships, for example, borderline, avoidant, and paranoid personality disorders' liability factors overlapped substantially with MDD's, in particular, clusters among Norwegian young adults³⁰. This overlapping may have led to an overestimation of the importance of interpersonal relationships.

For social indicators, besides the critical period, various risk models such as accumulation or trajectory may exist, which may also explain the null results. Morrissey and Kinderman confirmed the hypothesis that accumulation of adverse financial hardship negatively affects mental health, but not the hypothesis of critical periods³¹, while our risk model is the 'critical period'. Another study demonstrated the complicated effect among changes in racial composition, neighborhood socioeconomic status, and depressive symptoms³². The social indicators derived from Statistics Finland's (stat.fi/tilastotieto) registers are at the postal code or municipality level, which leads to some concern about the inaccurate measurement of an individual's exposure (information bias).

Several previous ExWAS studies linking the exposome to mental health had some similar or heterogeneous results to ours. van de Weijer et al.¹⁰, identified several social indicators such as safety and income being linked to mental well-being, but the links were weak in our analysis. This may be due to using different outcomes, the older age in their samples, and different statistical methods between the two countries' authorities¹⁰. Although the ExWAS of Choi et al. was on the general population in the United Kingdom, they also found that a higher frequency of visits with family/friends reduced the odds of depression incidence, and Mendelian randomization reinforced the causality of this association¹¹. However, we do not have many common variables with Choi et al.¹¹, in which they included many lifestyle factors (specific external exposome), while we have more general external exposome variables. Another ExWAS on psychotic experiences identified many stressful life-event factors, a result that was similar to our study⁸. Despite the divergent findings, the accumulation of ExWAS findings from different countries, populations, and age groups enhances our understanding of growing concepts of the exposome on depression, as well as broad mental health. The inclusion of a large number of exposures about interpersonal and person-societal relationships is also an important addition to the existing evidence. Notably, some of the information was provided by the parents, not only the twins. Furthermore, some scientists have raised the concept of an 'eco-exposome' to thoroughly assess the internal exposome, including molecules affected by exogenous exposures³³, which could be assimilated into further research.

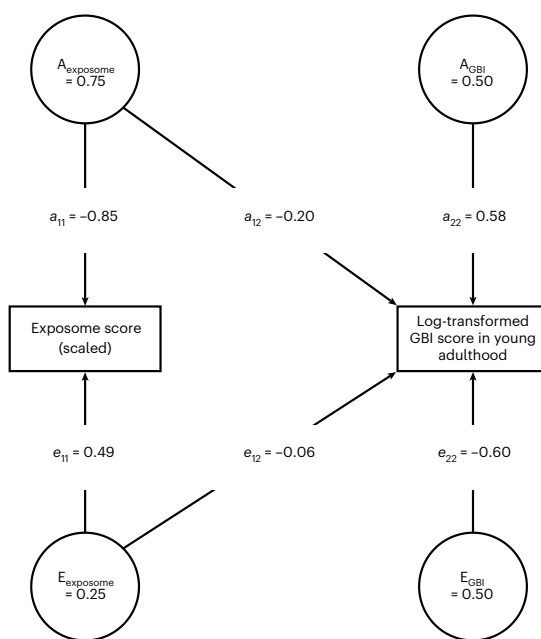
The sex difference is notable. Our previous study found that male twins tend to stay together longer, implying more exposure to any familial impact³⁴. In a Swedish study, family structure, conflict, and

Male (188 MZ and 162 DZ pairs)



Unique environmental factor explained 23.06% of the covariance

Female (278 MZ and 218 DZ pairs)



Unique environmental factors explained 12.85% of the covariance

Fig. 3 | Bivariate Cholesky AE model for the exposome score and log-transformed GBI score in young adulthood (twin pair $n = 846$).

A, standardized variance of additive genetic effect; E, standardized variance of unique environmental effect. The a and e stand for pathway coefficients from

A and E, respectively, to both the exposome score and log-transformed GBI score. The 95% CIs of standardized variances and pathway coefficients are presented in Supplementary Table 4.

child disclosure of information to parents were associated with offending behavior in boys, while only one factor was salient in girls³⁵. Another British study found that boys in detrimental familial environments were increasingly disadvantaged in school achievement compared with girls³⁶. The evidence hints that males are more easily affected by the family environment, which could explain the higher contribution of E on the covariance between the exposome and depressive symptoms in males. This inference is not certain, and there is contrary evidence³⁷. Moreover, sex differences exist in many biological mechanisms regarding how the body neurophysiologically reflects the external environment. Several sex-differentially expressed neurotransmitters or hormones, such as progesterone in females, are involved in systemic dysregulation, inducing depression³⁸. Furthermore, environmental endocrine-disrupting chemicals are able to alter neurodevelopment with sex-specific effects at very early developmental stages³⁹. In the future, integrating with the internal exposome such as metabolites and other omics will help us advance the study of sex-difference mechanisms on the relationship between the exposome and depressive phenotype.

As a part of the European Human Exposome Network, our overarching goal is to evaluate the impact of the exposome on human health across various age groups and with respect to multiple outcomes. The present analysis represents one individual analysis, and by pooling our collective efforts, important implications for clinical practice can be drawn in the future. Our findings suggest that studies on the familial component of social exposome should be noticed and investigated in the improvement of current therapy. It does not mean that we should ignore the physical exposure group, due to ubiquity, even though their

relevance is not salient⁴⁰. In addition, it is imperative to incorporate the consideration of familial effects and genetic liability at the same time for a more thorough understanding in future studies.

There are some other limitations in our study. First, compared with other ExWASes, our sample size is relatively small. Although Chung et al. indicated that a sample size between 1,795 and 3,625 participants is adequate when using the Bonferroni correction⁴¹, we did not stratify the ExWAS by sex due to the sample size being reduced by half. Second, we did not further assess the causality. Causal inferences are critical for further policymaking and intervention. Mendelian randomization in larger samples is a future direction. Third, the ExWAS, CFA, and twin modeling were all performed on the basis of the FinnTwin12 cohort, which raises concerns about model overfitting and leakage. Different models with different purposes, hypotheses, and methodologies in two stages reduce the risk of overfitting and leakage. ExWAS was used to identify salient exposure, while CFA and twin modeling were used to explore. The observational unit was each twin pair in twin modeling, while in ExWAS and CFA, it is each individual twin. Replication on other twin cohorts and in family datasets is warranted.

Conclusion

This study applied a two-stage analysis. First, in ExWAS, we identified that exposures from family and parents, friend and romantic relationships, school and teachers, and stressful life events were significantly associated with depressive symptoms in late adolescence and young adulthood. The family and parent exposures were the most influential. Second, twin modeling between the exposome and depressive

symptoms uncovered a complex relationship among genes, environments, and depressive symptoms with sex differences. The findings underline the importance of systematic evaluation of the environmental effects on depressive symptoms and recommend the consideration of genetic effects in future studies.

Methods

Study participants

The participants came from the FinnTwin12 cohort, which is a nationwide prospective cohort among all Finnish twins born between 1983 and 1987. First, the overall epidemiological study consisted of all 5,184 twins who responded (age 11–12) at wave 1, and there are three general following waves at ages 14, 17, and in young adulthood (mean age: 21.9). Moreover, 1,035 families with 2,070 twins were invited to take part in an intensive study with psychiatric interviews, some biological samples, and additional questionnaires⁴². At age 14 (wave 2), 1,854 twins participated. They were then invited to participate again as young adults (wave 4) of the study. Psychiatric interviews in young adulthood were completed for 1,347 twins in the intensive study, including assessment of MDD using the Semi-Structured Assessment for Genetics of Alcohol based on Diagnostic and Statistical Manual of Mental Disorders IV criteria^{43,44}. The twins also completed questionnaires on health, health behaviors, work, and multiple psychological scales. The flowchart of general FinnTwin12 cohort is presented in Extended Data Fig. 5. An updated review has been published⁴⁵.

The ethics committee of the Department of Public Health of the University of Helsinki and the Institutional Review Board of Indiana University approved the FinnTwin12 study protocol from the start of the cohort. The ethical approval of the ethics committee of the Helsinki University Central Hospital District (HUS) is the most recent and covers the most recent data collection (wave 4) (HUS/2226/2021). The HUS reviews the study annually, and 2023's statement is number 4/2023, dated 1 February 2023. All participants and their parents/legal guardians gave informed written consent to participate in the study. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

Measures

The primary outcome is the short-version GBI scores in young adulthood. It is a self-reported inventory to evaluate the occurrence of depressive symptoms, which is composed of ten Likert-scale questions⁴⁶. The total score ranges from 0 to 30, and a higher score implies more depressive symptoms occurred. There are two secondary outcomes: GBI scores at age 17 and incidence of MDD in young adulthood.

In total, we curated 385 environmental exposures under the concept of the Equal-Life project⁴⁷ from multiple sources and grouped them into 12 domains. Air pollution exposures came from the annual average air quality of each observation station from the Finnish Meteorological Institute. Domains of building, blue and green spaces, population density, and a part of geocoordinates were from Equal-Life enrichment. Their description can be found in a previous study⁴⁸ and is presented in Supplementary Note 1. Exposures from prenatal exposures, passive smoking, family and parents, friend and romantic relationships, school and teachers, and stressful life events domains were from FinnTwin12 questionnaires by self-report or parent report and are described in a published review⁴⁵. Social indicators were from Statistics Finland and are described in Supplementary Note 1. Except for FinnTwin12 questionnaires, exposures from other sources were linked to individual twins via EUREF-FIN geocoordinates. The full residential history of the twins from birth onward until 2020 was obtained as geocoordinates and dates of moving in and out of specific addresses from the Digital and Population Data Services Agency in Finland³⁴. The types of exposures are continuous, binary, and categorical. Considering the

temporality, we included repeated exposures for the critical-period risk model, and Extended Data Fig. 6 presents the timeline of the study. There are three exposure inclusion criteria: (1) twins have available residential history, (2) twins and their family completed at least one questionnaire at any wave, and (3) the percentage of missing values is less than 20% in ExWAS. The code names of each exposure were developed from the description as closely as possible, and their domains, resources, and dates are presented in Supplementary Table 1. The missing patterns of each exposure in each ExWASes are presented in Supplementary Table 7

For analysis of outcomes in young adulthood, we *a priori* identified seven covariates: sex (male, female), zygosity (monozygotic (MZ), DZ, unknown), parental education (limited, intermediate, high)⁴⁹, smoking (never, former, occasional, current), work status (full-time, part-time, irregular, not working), secondary-level school (vocational, senior high school, none), and age. The latter four variables were reported by twins as young adults (wave 4). For analysis of outcome at age 17, sex, zygosity, parental education, smoking (reported at age 17) remained. Study and working status (neither study nor work, only study, only work) were included when most participants were in school at age 17 (wave 3). The inclusion of covariates, besides sex, zygosity, and age, was based on the previous literature, which shows correlations with the environment and depressive symptoms^{50–52}. Parental education was adjusted for to represent the family resources and resilience⁴⁹.

Data pre-processing and descriptive statistics

Participants missing information on outcome or covariates were excluded from the corresponding age's analyses. Due to the skewness of the GBI score, we added one to the GBI score and log-transformed it. Appropriate regrouping was conducted for categorical exposures, and then we used simple imputation by the median to replace the missing values of exposures. As a dimension reduction technique, PCA was utilized to measure the proportion of total variability of all included exposures attributed to each PC and visually assess the potential clusters of exposures (correlated) on the basis of the two-dimensional coordinate with the first and second components. It was conducted only for outcomes of GBI at age 17 and in young adulthood, not for the incidence of MDD.

Exposome-wide association study

To conduct the ExWAS, a generalized linear regression model with Gaussian distribution (essentially linear regression) for the outcomes of log-transformed GBI score was repeatedly performed for each exposure. We used Bonferroni correction by the number of effective tests (calculated by PCA) to adjust for multiple testing and account for correlation between exposures⁵³. Covariates were adjusted and the cluster effect of sampling based on families of twin pairs was controlled for by the robust standard error. For the outcome of the incidence of MDD, the distribution was switched to be binomial. The number of included exposures of secondary outcomes was smaller due to the third exposure inclusion criteria, and the sample size varied; thus, the *P*-value thresholds varied. Due to categorical exposures, the number of *P* values was higher than the number of exposures. We used the *rexp* package in the R environment (version 4.2.3)⁵⁴.

We further calculated power using the R package WebPower (R environment, version 4.2.3) for ExWASes of the log-transformed GBI score at both age points. These calculations were based on the smallest absolute effect size among significant results (0.12 in young adulthood and 0.10 at age 17), sample size (3,025 in young adulthood and 4,127 at age 17), number of predictor variables in a single model (8 in young adulthood and 6 at age 17), and significant thresholds (3.09×10^{-4} in young adulthood and 3.63×10^{-4} at age 17). The powers were 1 for ExWASes both in young adulthood and at age 17, indicating adequate sample sizes in this study.

Generating exposome score

Based on the significant exposures selected from the ExWAS, CFA was used to estimate an exposome score, preparing for the following twin modeling. According to the concept of the environment's totality, we indicated a one-factor structure for the exposome. The CFA assumes the correlation between exposures due to the exposome score and verifies it based on structural equation modeling as theory driven. We used maximum likelihood to estimate the score and standardized root mean square residual to evaluate the model fit⁵⁵. The cluster effect was controlled like before. Due to multiple subgroups in categorical exposures, we included the whole exposure variable when there was at least one subgroup that was significant compared with the reference in ExWAS. The coefficients of significant exposures were presented in Supplementary Tables 8 and 9 for outcomes of GBI in young adulthood and at age 17, respectively. In addition, we conducted exploratory factor analysis (EFA) estimated by maximum likelihood with 100 optimizations, whereas a large number of retained factors indicated potential overfitting of EFA. The CFA and EFA were performed using Stata 18.0 (StataCorp), and package sem was used.

Twin modeling

In twin modeling, the genetic effect is usually divided into additive and dominant genetic effects⁵⁶. Since MZ twins are roughly genetically identical and DZ twins share roughly half of their segregating genes, the correlation of A is set to 1.0 and 0.5 and of D is set to 1.0 and 0.25 within MZ and DZ twin pairs, respectively. The epistatic effect is a part of A. The environmental effect is also divided into two components: common environment, whose correlation is assumed to be 1.0 regardless of zygosity, and unique environment (no correlation), which includes unmeasured errors. The use of the twin model assumes the absence of assortative mating for the trait under study among the parents and equal effects of the environment by zygosity.

The intrapair correlations of GBI in DZ ($\rho = 0.22$ in young adulthood and 0.16 at age 17) and MZ ($\rho = 0.52$ in young adulthood and 0.51 at age 17) indicated to use an ADE model initially, instead of the ACE model ($\rho_{MZ} > 2\rho_{DZ}$). Due to using only the twin pair design, instead of the extended family design, we could not use an ACDE model. The saturated twin model was performed to test the assumptions of equal means and variances for twin order and for zygosity, via constraint means and variances, and to detect the sex difference via sex limitation. In the saturated model (Supplementary Table 10), the Akaike information criterion and likelihood ratio test between models suggested that the assumptions were basically met. Results of the sex-limitation saturated model (Supplementary Table 10) indicated a notable sex difference.

Finally, to assess how the current exposome score explains the variance of depressive symptoms, we employed the bivariate Cholesky AE model to fit the exposome score and log-transformed GBI score (Extended Data Fig. 7) at both age points, which efficiently decomposes the phenotypic correlation and offers the attribution (%) to genetic and environmental factors⁵⁷. Two latent factors (A_{exposome} and E_{exposome}) influence both the exposome score (a_{11} and e_{11}) and log-transformed GBI score (a_{21} and e_{21}), and another two latent factors (A_{GBI} and E_{GBI}) influence only the log-transformed GBI score (a_{22} and e_{22}). The overall correlation between the exposome score and GBI could be calculated as $a_{11} \times a_{12} + e_{11} \times e_{12}$. Variances of A_{exposome} , E_{exposome} , A_{GBI} , and E_{GBI} were calculated as $a_{11}^2 + a_{12}^2$, $e_{11}^2 + e_{12}^2$, a_{22}^2 , and e_{22}^2 , respectively. We also re-assess the sex difference via an additional sex-limited saturated bivariate twin model.

Only full MZ and DZ twin pairs were included in the twin modeling. We dropped the opposite-sex DZ pairs and stratified the univariate and bivariate twin models by sex. The characteristics of included and excluded individual twins in the twin modeling are presented in Supplementary Table 11, and we did not observe a large difference, suggesting low selection bias risk due to sex, zygosity, and twin pair. Age, reported in the young adulthood survey, was adjusted in univariate and bivariate

models for the outcome in young adulthood. We used the OpenMx package in the R environment (version 4.2.3)⁵⁸.

Post hoc mixed models for repeated measures. On the basis of the exposures significantly associated with GBI at both time points, we performed the mixed models for repeated measures (MMRM) as a post hoc analysis to further explore the effects on the trajectory of depressive symptoms. This method analyzes the influence on the log-transformed GBI in young adulthood by both exposures of interest (fixed effect) and 'baseline' log-transformed GBI at age 17 (random effect)⁵⁹. The sample size and covariates of the MMRM were the same as in the ExWAS of log-transformed GBI score in young adulthood. The cluster effect was controlled by the robust standard error. The multiple testing was controlled by the false discovery rate (Q value < 0.05 was considered statistically significant). These post hoc analyses were performed using Stata 18.0 (StataCorp).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The FinnTwin12 data are not publicly available due to the restrictions of informed consent. However, the FinnTwin12 data are available through the Institute for Molecular Medicine Finland (FIMM) Data Access Committee (DAC) (fimm-dac@helsinki.fi) for authorized researchers who have IRB/ethics approval and an institutionally approved study plan. To ensure the protection of privacy and compliance with national data protection legislation, a data use/transfer agreement is needed, the content and specific clauses of which will depend on the nature of the requested data. Requests will be addressed in a reasonable time frame (generally two to three weeks), and the primary mode of data access is by either personal visit or remote access to a secure server.

Code availability

No new software, package, or algorithm was developed. All code for data cleaning and analysis associated with the current submission is available upon reasonable request to the corresponding author or Z.W. (zhiyang.wang@helsinki.fi).

References

- GBD 2019 Mental Disorders Collaborators. 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. *Lancet Psychiatry* **9**, 137–150 (2022).
- Shorey, S., Ng, E. D. & Wong, C. H. J. Global prevalence of depression and elevated depressive symptoms among adolescents: a systematic review and meta-analysis. *Br. J. Clin. Psychol.* **61**, 287–305 (2022).
- Köhler, C. A. et al. Mapping risk factors for depression across the lifespan: an umbrella review of evidence from meta-analyses and Mendelian randomization studies. *J. Psychiatr. Res.* **103**, 189–207 (2018).
- Mair, C., Roux, A. V. D. & Galea, S. Are neighbourhood characteristics associated with depressive symptoms? A review of evidence. *J. Epidemiol. Community Health* **62**, 940 LP–940946 (2008).
- Wild, C. P. Complementing the genome with an 'exposome': the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol. Biomarkers Prev.* **14**, 1847–1850 (2005).
- Guloksuz, S., van Os, J. & Rutten, B. P. F. The exposome paradigm and the complexities of environmental research in psychiatry. *JAMA Psychiatry* **75**, 985–986 (2018).

7. Zheng, Y. et al. Design and methodology challenges of environment-wide association studies: a systematic review. *Environ. Res.* **183**, 109275 (2020).
8. Lin, B. D. et al. Nongenetic factors associated with psychotic experiences among UK biobank participants: exposome-wide analysis and Mendelian randomization analysis. *JAMA Psychiatry* **79**, 857–868 (2022).
9. Ni, M. Y. et al. Determinants of physical, mental and social well-being: a longitudinal environment-wide association study. *Int. J. Epidemiol.* **49**, 380–389 (2020).
10. van de Weijer, M. P. et al. Expanding the environmental scope: an environment-wide association study for mental well-being. *J. Expo. Sci. Environ. Epidemiol.* **32**, 195–204 (2022).
11. Choi, K. W. et al. An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression. *Am. J. Psychiatry* **177**, 944–954 (2020).
12. Leffers, H. C. B., Lange, T., Collins, C., Ulf-Møller, C. J. & Jacobsen, S. The study of interactions between genome and exposome in the development of systemic lupus erythematosus. *Autoimmun. Rev.* **18**, 382–392 (2019).
13. Wichers, M. et al. Mechanisms of gene–environment interactions in depression: evidence that genes potentiate multiple sources of adversity. *Psychol. Med.* **39**, 1077–1086 (2009).
14. Medda, E. et al. The response to oxidative stress and metallomics analysis in a twin study: the role of the environment. *Free Radic. Biol. Med.* **97**, 236–243 (2016).
15. Yu, Y. et al. The role of family environment in depressive symptoms among university students: a large sample survey in China. *PLoS ONE* **10**, e0143612 (2015).
16. Stavropoulos, V., Lazaratou, H., Marini, E. & Dikeos, D. Low family satisfaction and depression in adolescence: the role of self-esteem. *J. Educ. Dev. Psychol.* **5**, 109–118 (2015).
17. Wang, L. & Crane, D. R. The relationship between marital satisfaction, marital stability, nuclear family triangulation, and childhood depression. *Am. J. Fam. Ther.* **29**, 337–347 (2001).
18. Hooper, L. M. & Wallace, S. A. Evaluating the parentification questionnaire: psychometric properties and psychopathology correlates. *Contemp. Fam. Ther.* **32**, 52–68 (2010).
19. Cappell, C. & Heiner, R. B. The intergenerational transmission of family aggression. *J. Fam. Violence* **5**, 135–152 (1990).
20. Powdthavee, N. & Vignoles, A. Mental health of parents and life satisfaction of children: a within-family analysis of intergenerational transmission of well-being. *Soc. Indic. Res.* **88**, 397–422 (2008).
21. Ottman, R. Gene–environment interaction: definitions and study design. *Prev. Med.* **25**, 764–770 (1996).
22. Carreras-Gallo, N. et al. The early-life exposome modulates the effect of polymorphic inversions on DNA methylation. *Commun. Biol.* **5**, 455 (2022).
23. Tamimy, Z. et al. Multilevel twin models: geographical region as a third level variable. *Behav. Genet.* **51**, 319–330 (2021).
24. Abdellaoui, A. et al. Population structure, migration, and diversifying selection in the Netherlands. *Eur. J. Hum. Genet.* **21**, 1277–1285 (2013).
25. Kerminen, S. et al. Fine-scale genetic structure in Finland. *G3* **7**, 3459–3468 (2017).
26. Ohanyan, H. et al. Machine learning approaches to characterize the obesogenic urban exposome. *Environ. Int.* **158**, 107015 (2022).
27. Ohanyan, H. et al. Associations between the urban exposome and type 2 diabetes: results from penalised regression by least absolute shrinkage and selection operator and random forest models. *Environ. Int.* **170**, 107592 (2022).
28. *Places and Spaces: Environments and Children's Well-Being* (UNICEF IRC, 2022).
29. Spotts, E. L. et al. Accounting for depressive symptoms in women: a twin study of associations with interpersonal relationships. *J. Affect. Disord.* **82**, 101–111 (2004).
30. Reichborn-Kjennerud, T. et al. Major depression and dimensional representations of DSM-IV personality disorders: a population-based twin study. *Psychol. Med.* **40**, 1475–1484 (2010).
31. Morrissey, K. & Kinderman, P. The impact of childhood socioeconomic status on depression and anxiety in adult life: testing the accumulation, critical period and social mobility hypotheses. *SSM Popul. Health* **11**, 100576 (2020).
32. Lee, H. & Estrada-Martínez, L. M. Trajectories of depressive symptoms and neighborhood changes from adolescence to adulthood: latent class growth analysis and multilevel growth curve models. *Int. J. Environ. Res. Public Health* **17**, 1829 (2020).
33. Scholz, S. et al. The eco-exposome concept: supporting an integrated assessment of mixtures of environmental chemicals. *Environ. Toxicol. Chem.* **41**, 30–45 (2022).
34. Wang, Z., Whipp, A., Heinonen-Guzejev, M. & Kaprio, J. Age at separation of twin pairs in the FinnTwin12 study. *Twin Res. Hum. Genet.* <https://doi.org/10.1017/thg.2022.17> (2022).
35. Nilsson, E.-L. Analyzing gender differences in the relationship between family influences and adolescent offending among boys and girls. *Child Indic. Res.* **10**, 1079–1094 (2017).
36. Mensah, F. K. & Kiernan, K. E. Gender differences in educational attainment: influences of the family environment. *Br. Educ. Res. J.* **36**, 239–260 (2010).
37. Blaauboer, M. & Mulder, C. H. Gender differences in the impact of family background on leaving the parental home. *J. Hous. Built Environ.* **25**, 53–71 (2010).
38. Sikes-Keilp, C. & Rubinow, D. R. In search of sex-related mediators of affective illness. *Biol. Sex Differ.* **12**, 55 (2021).
39. Rebuli, M. E. & Patisaul, H. B. Assessment of sex specific endocrine disrupting effects in the prenatal and pre-pubertal rodent brain. *J. Steroid Biochem. Mol. Biol.* **160**, 148–159 (2016).
40. Maria, F. et al. Association of long-term exposure to traffic-related air pollution with blood pressure and hypertension in an adult population-based cohort in Spain (the REGICOR study). *Environ. Health Perspect.* **122**, 404–411 (2014).
41. Chung, M. K., Buck Louis, G. M., Kannan, K. & Patel, C. J. Exposome-wide association study of semen quality: systematic discovery of endocrine disrupting chemical biomarkers in fertility require large sample sizes. *Environ. Int.* **125**, 505–514 (2019).
42. Whipp, A. M. et al. Early adolescent aggression predicts antisocial personality disorder in young adults: a population-based study. *Eur. Child Adolesc. Psychiatry* **28**, 341–350 (2019).
43. Bucholz, K. K. et al. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. *J. Stud. Alcohol* **55**, 149–158 (1994).
44. Bell, C. C. DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. *JAMA* **272**, 828–829 (1994).
45. Rose, R. J. et al. FinnTwin12 cohort: an updated review. *Twin Res. Hum. Genet.* **22**, 302–311 (2019).
46. Kokko, K. & Pulkkinen, L. Unemployment and psychological distress: mediator effects. *J. Adult Dev.* **5**, 205–217 (1998).
47. van Kamp, I. et al. Early environmental quality and life-course mental health effects: the Equal-Life project. *Environ. Epidemiol.* **6**, e183 (2022).
48. Julvez, J. et al. Early life multiple exposures and child cognitive function: a multi-centric birth cohort study in six European countries. *Environ. Pollut.* **284**, 117404 (2021).
49. Huppertz, C. et al. The effects of parental education on exercise behavior in childhood and youth: a study in Dutch and Finnish twins. *Scand. J. Med. Sci. Sports* **27**, 1143–1156 (2017).

50. Salmela-Aro, K. et al. Depressive symptoms and career-related goal appraisals: genetic and environmental correlations and interactions. *Twin Res. Hum. Genet.* **17**, 236–243 (2014).
51. Edelbrock, C., Rende, R., Plomin, R. & Thompson, L. A. A twin study of competence and problem behavior in childhood and early adolescence. *J. Child Psychol. Psychiatry* **36**, 775–785 (1995).
52. Rose, R. J. & Kaprio, J. Genes, environments, and adolescent substance use: retrospect and prospect from the FinnTwin studies. *Acta Psychol. Sinica* **40**, 1062–1072 (2008).
53. Li, M.-X., Yeung, J. M. Y., Cherny, S. S. & Sham, P. C. Evaluating the effective numbers of independent tests and significant *P*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131**, 747–756 (2012).
54. Hernandez-Ferrer, C. et al. Comprehensive study of the exposome and omic data using rexpome bioconductor packages. *Bioinformatics* **35**, 5344–5345 (2019).
55. Kline, R. B. *Principles and Practice of Structural Equation Modeling* 4th edn (Guilford Publications, 2015).
56. Polderman, T. J. C. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
57. Røysamb, E. & Tambs, K. The beauty, logic and limitations of twin studies. *Nor. Epidemiol.* <https://doi.org/10.5324/nje.v26i1.2.2014> (2016).
58. Neale, M. C. et al. OpenMx 2.0: extended structural equation and statistical modeling. *Psychometrika* **81**, 535–549 (2016).
59. Detry, M. A. & Ma, Y. Analyzing repeated measurements using mixed models. *JAMA* **315**, 407–408 (2016).

Acknowledgments

We acknowledge A. Pia and H. Heidi from the Finnish Meteorological Institute for providing the annual summary of air quality data in Finland. We appreciate A. Ambros and B. Raimbault from ISGlobal for the calculation of the exposome for this study. We thank X. Escribà-Montagut from ISGlobal for technical support on the R package rexpome. FinnTwin12 thanks all participating twins, their parents, and teachers. This research was partly funded by the European Union's Horizon 2020 research and innovation program under grant agreement no. 874724 (Equal-Life). Equal-Life is part of the European Human Exposome Network. Data collection in FinnTwin12 has been supported by the National Institute on Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to R. J. Rose) and the Academy of Finland (grants 100499, 205585, 118555, 141054, 264146, 308248, 312073, 336823, and 352792 to J.K.). J.K. acknowledges support by the Academy of Finland (grants 265240, 263278).

Author contributions

All authors conceived and conceptualized the research idea and developed the analysis plan. Z.W., M.F., J.J., and J.K. acquired the

data. Z.W. and S.Z. conducted data analysis. Z.W. drafted the original draft of the paper. S.Z., A.M.W., M.H.-G., M.F., J.J., I.v.K., and J.K. reviewed the paper and provided critical comments. I.v.K. and J.K. gave critical supervision and guidance. All authors approved the paper.

Funding

Open Access funding provided by University of Helsinki including Helsinki University Central Hospital.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s44220-023-00124-x>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44220-023-00124-x>.

Correspondence and requests for materials should be addressed to Jaakko Kaprio.

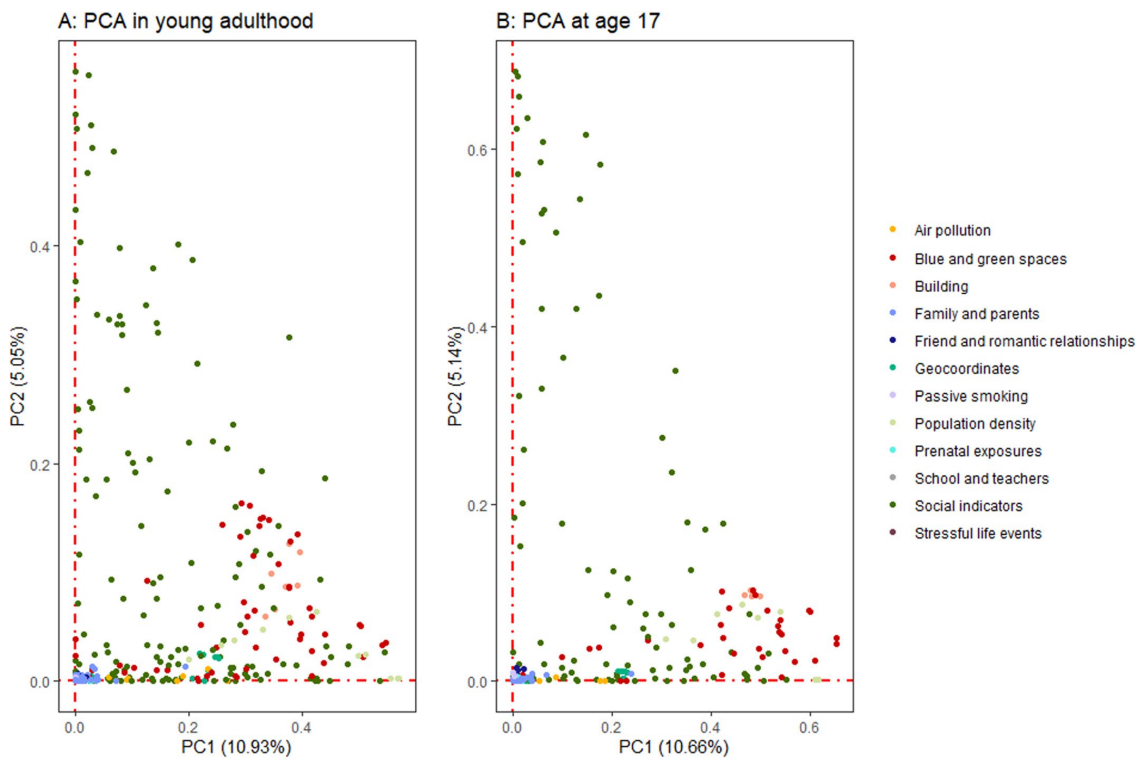
Peer review information *Nature Mental Health* thanks Ran Barzilay, Sinan Guloksuz, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

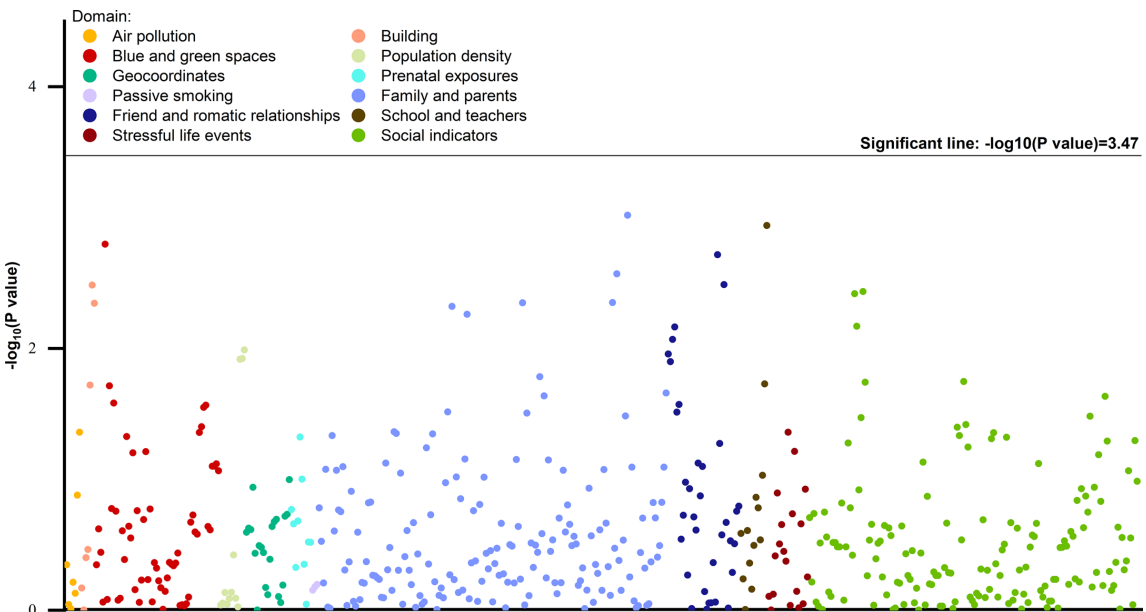
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2024

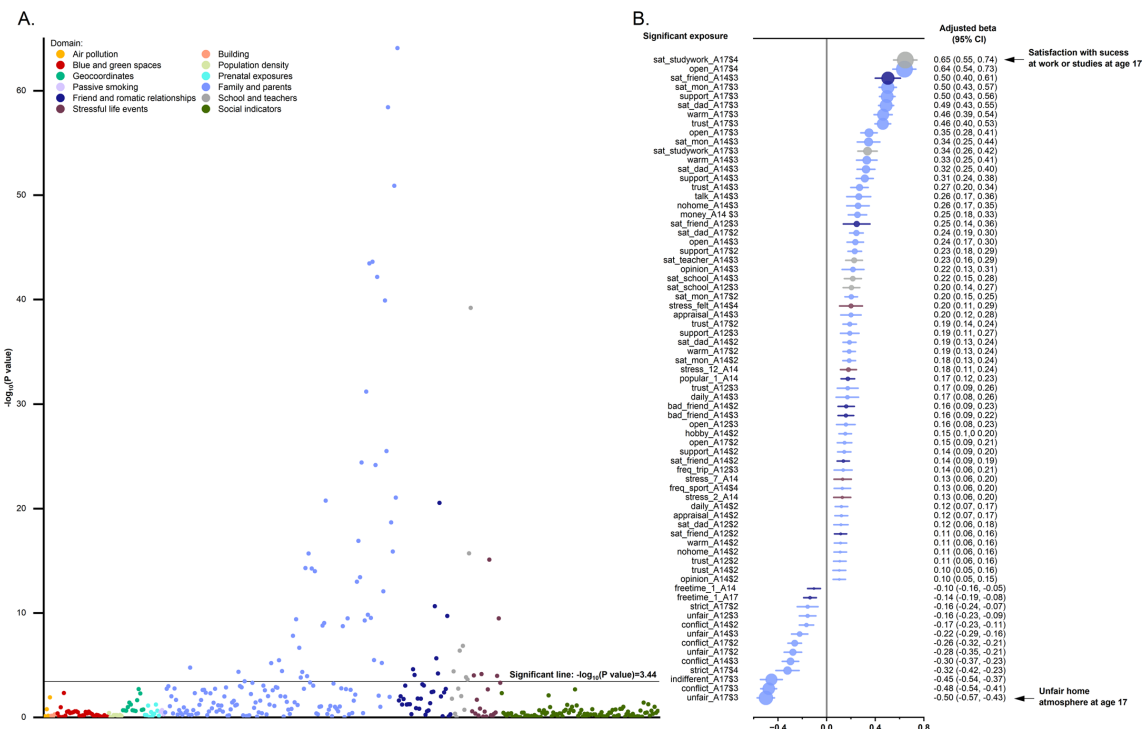


Extended Data Fig. 1 | Principal component analysis for exposures. A, In young adulthood (individual twin n = 3025). B, At age 17 (individual twin n = 1236).



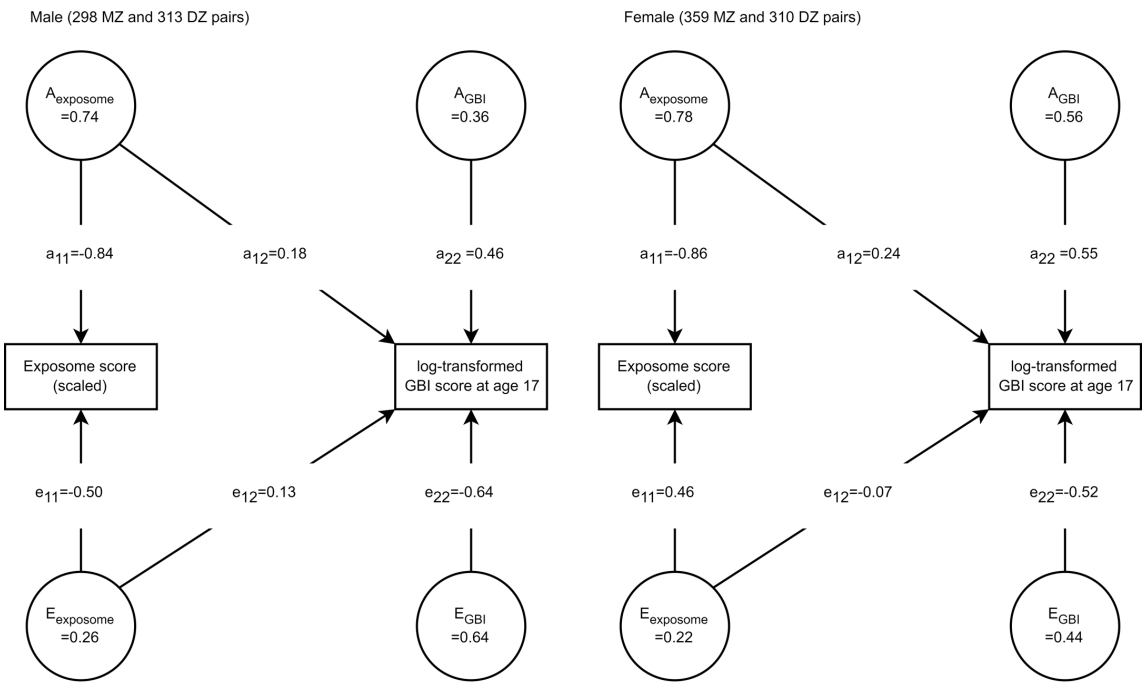
Extended Data Fig. 2 | Association results between exposure and incidence of MDD, adjusted for covariates (individual twin n = 1236), using generalized binomial regression. The adjusted covariates were: sex, zygosity, parental

education, smoking in young adulthood, work status in young adulthood, secondary level school in young adulthood, and age when twins provided the GBI assessment in young adulthood.



Extended Data Fig. 3 | Association results between exposure and log-transformed GBI score at age 17, adjusted for covariates (individual twin $n = 4127$), using generalized linear regression*. Panel A is a Manhattan association plot for exposures in relation to log-transformed GBI score at age 17. The y-axis is showing statistical significance as $-\log_{10}(P \text{ value})$ for the adjustment for multiple testing. Panel B presents the adjusted beta for significant

exposures in descending order from top to bottom (from harmful to protective). In panel B, the center dot and bar present the effect size (coefficient of linear regression) and 95% confidence interval, and the size of the dots presents the effect size relatively. The color legend applies to both Panel A (Manhattan association plot) and B (forest plot). The adjusted covariates were: sex, zygosity, parental education, smoking at age 17, and study and working status at age 17.

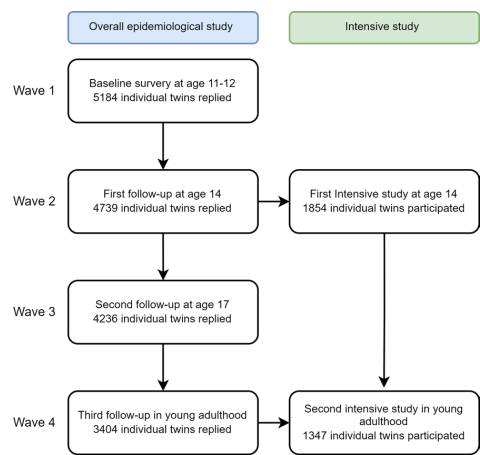


Unique environmental factor explained 30.70% of the covariance

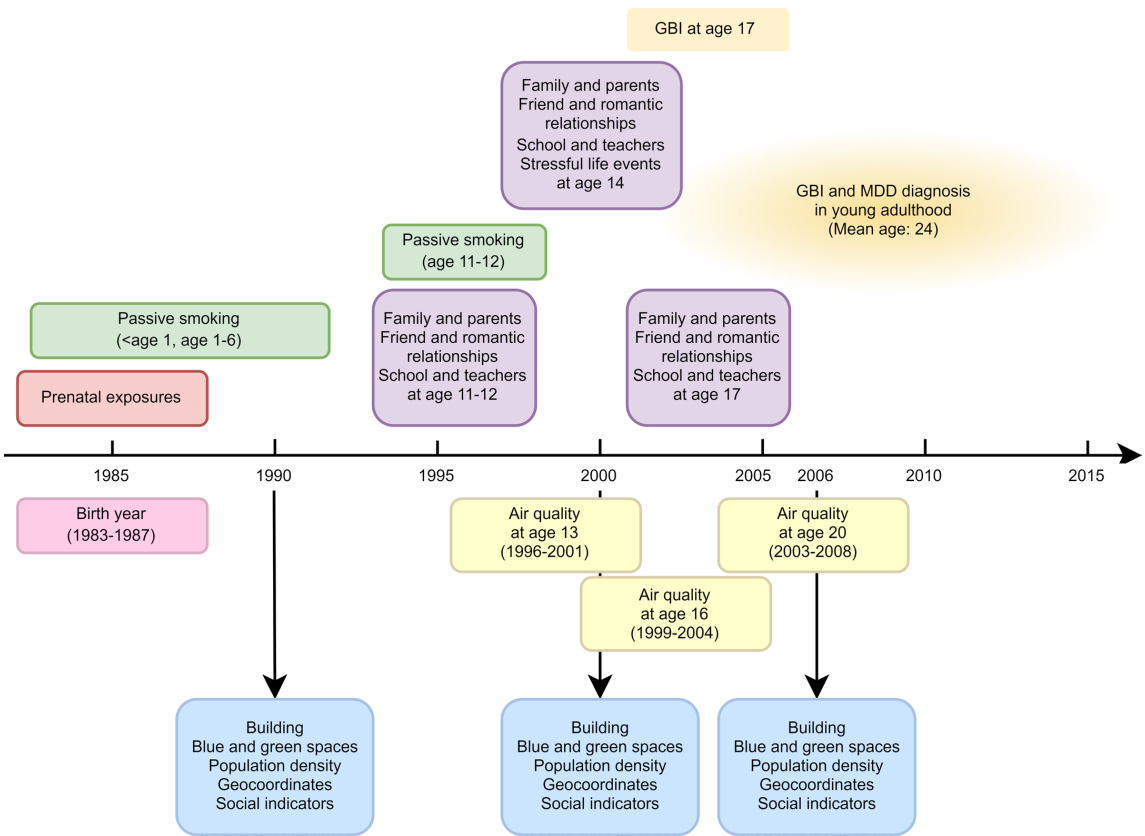
Extended Data Fig. 4 | Bivariate Cholesky AE model for the exposome score and log-transformed GBI score at age 17 (twin pair n = 1000). A stands for standardized variance of additive genetic effect, E stands for standardized variance of unique environmental effect. MZ and DZ stand for monozygotic and

Unique environmental factors explained 12.63% of the covariance

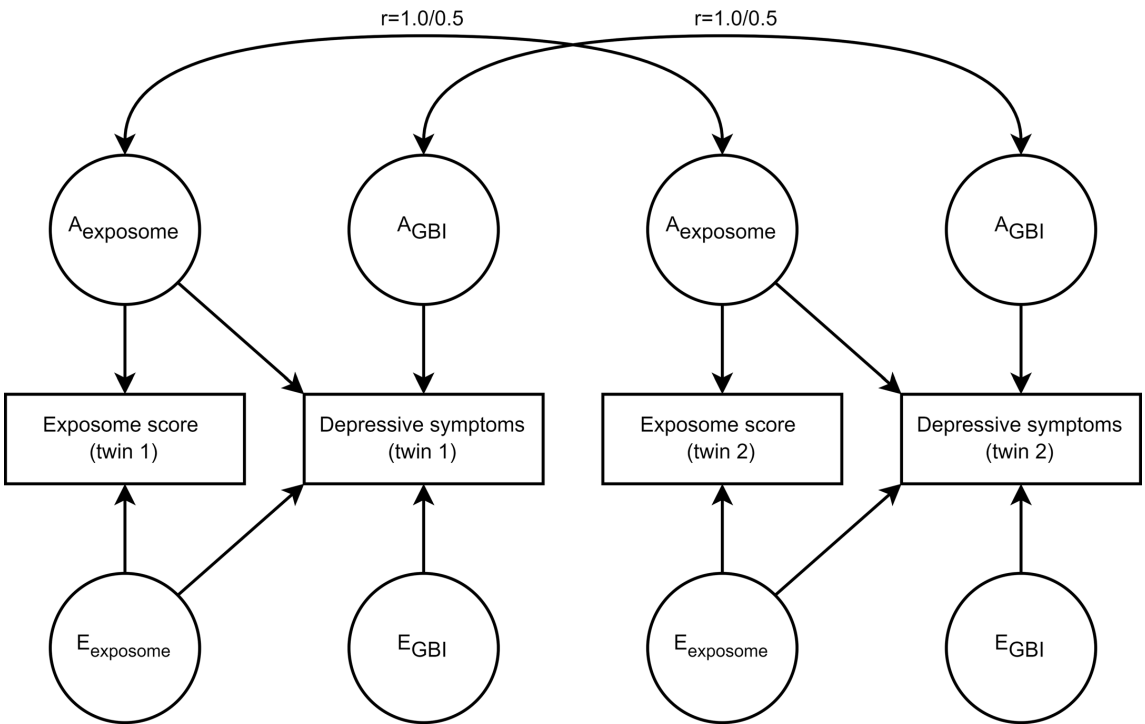
dizygotic twin pairs, respectively. The a and e stand for pathway coefficients from A and E, respectively, to both the exposome score and log-transformed GBI score. The 95% confidence intervals of standardized variances and pathway coefficients are presented in Extended Supplementary Table 4.



Extended Data Fig. 5 | Flowchart of general FinnTwin12 cohort. The cohort is composed of the overall epidemiological study with one baseline and three general follow-ups and the intensive study with two detailed interviews.



Extended Data Fig. 6 | Calendar timeline of included exposures and outcomes. The information started to be recorded in 1983 during the pregnancy of twins' mothers and until 2015.



Extended Data Fig. 7 | Diagram of bivariate Cholesky AE decomposition model (r : correlation). Note: circles denote latent factors of additive genetic (A) and unique environmental (E) components, while rectangles denote measured/calculated variables.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | | |
|-------------------------------------|--|
| n/a | Confirmed |
| ☐ | ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ A description of all covariates tested |
| ☐ | ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☐ | ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

- | | |
|-----------------|--|
| Data collection | No new software, package, and algorithm were developed. All code for data cleaning and analysis associated with the current submission is available upon reasonable request to the corresponding author. Contact: Zhiyang Wang, zhiyang.wang@helsinki.fi |
| Data analysis | No new software, package, and algorithm were developed. All code for data cleaning and analysis associated with the current submission is available upon reasonable request to the corresponding author. Contact: Zhiyang Wang, zhiyang.wang@helsinki.fi |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The FinnTwin12 data is not publicly available due to the restrictions of informed consent. However, the FinnTwin12 data is available through the Institute for Molecular Medicine Finland (FIMM) Data Access Committee (DAC) (fimm-dac@helsinki.fi) for authorized researchers who have IRB/ethics approval and an

institutionally approved study plan. To ensure the protection of privacy and compliance with national data protection legislation, a data use/transfer agreement is needed, the content and specific clauses of which will depend on the nature of the requested data. Requests will be addressed in a reasonable timeframe (generally 2-3 weeks), and the primary mode of data access is either by personal visit or remote access to a secure server.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	In this study, we collected sex information (biological attribute).
Population characteristics	The demographic characteristics was presented in Table 2. Twins born between 1983 and 1987. There was two time points for outcome: age 17 and young adulthood (age 24). This cohort is sex balanced.
Recruitment	The target population was all twins in Finland born between 1983-1987. They and their parents were asked to participate in the study when the twins were aged 11-12, and they have been asked to take part in consequent waves of data collection at ages 14, 17 and as young adults. For details see: Rose RJ, Salvatore JE, Aaltonen S, Barr PB, Bogl LH, Byers HA, Heikkilä K, Korhonen T, Latvala A, Palviainen T, Ranjit A, Whipp AM, Pulkkinen L, Dick DM, Kaprio J. FinnTwin12 Cohort: An Updated Review. Twin Res Hum Genet. 2019;22(5):302-311. doi: 10.1017/thg.2019.83. . PMID: 31640839
Ethics oversight	The ethics committee of the Department of Public Health of the University of Helsinki (Helsinki, Finland), the ethics committee of the Helsinki University Central Hospital District (Helsinki, Finland)(HUS/2226/2021), and the Institutional Review Board of Indiana University (Bloomington, Indiana, USA) approved the FinnTwin12 study protocol. HUS reviews the study annually, and 2023's statement is number 4/2023, dated February 1, 2023.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The study was based on the FinnTwin12 cohort, which is a nationwide prospective cohort among all Finnish twins born between 1983 and 1987
Research sample	FinnTwin12 cohort was a national represented twin cohort aimed to target to overall Finnish population. There is no choice on sex, and other characteristics.
Sampling strategy	We sampled all twin pairs born in Finland between 1983-1987, with both twins resident at the same address at age 11. They have since been asked to participate in follow-up studies at ages 14, 17 and as young adults (PMID 31640839)
Data collection	We curated 385 environmental exposures from multiple sources: cohort questionnaires, Finnish Meteorological Institute, Equal-life enrichment, and Statistics Finland
Timing and spatial scale	The time scale of the environment exposures started from the pregnancy period and twins born between 1983 and 1987. In FinnTwin12, there were 1 baseline at age 11/13 and 3 follow-ups at age 14, 17, and in young adulthood. The choice of temporality of exposure was according to the date of outcomes: age 17 and in young adulthood. So the stop date of data collection in this study was around between 2004 and 2008.
Data exclusions	There are three exposure exclusion criteria: 1) twins without available residential history; 2) twins and their family do not completed any questionnaire at any wave; and 3) the percentage of missing values is over than 20% in ExWAS.
Reproducibility	We acknowledged that there is no replication in this study, which is a limitation. Replication on other twin cohorts and in family data sets is warranted.
Randomization	NA
Blinding	NA
Did the study involve field work?	☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems		Methods	
n/a	Involved in the study	n/a	Involved in the study
☒	☐ Antibodies	☒	☐ ChIP-seq
☒	☐ Eukaryotic cell lines	☒	☐ Flow cytometry
☒	☐ Palaeontology and archaeology	☒	☐ MRI-based neuroimaging
☒	☐ Animals and other organisms		
☒	☐ Clinical data		
☒	☐ Dual use research of concern		

Associations between the mid-adolescent external exposome and proteomic biomarkers of mental health

Zhiyang Wang¹, Gabin Drouard¹, Aleksei Afonin², Núria Botella^{3,4}, Carmen Peuters^{3,5,6}, Aino-Kaisa Piironen², Alyce. M. Whipp¹, Boris Cheval^{3,7,8}, Libor Šulc⁹, Marja Heinonen-Guzejev¹, Maria Foraster^{3,10}, John Gulliver¹¹, Jenny Selander⁹, Payam Dadvand^{3,5,6}, Jordi Júlvez^{3,12}, Irene van Kamp¹³, Katja M. Kanninen², Jaakko Kaprio¹, Equal-Life Scientific Team

¹ Institute for Molecular Medicine Finland, Helsinki Institute of Life Science, University of Helsinki, 00260, Helsinki, Finland

² A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Finland

³ ISGlobal, Barcelona Biomedical Research Park (PRBB), C/ Doctor Aiguader 88, Barcelona, Spain

⁴ Department of Basic Clinical Practice, Faculty of Medicine, University of Barcelona, 08007, Barcelona, Spain.

⁵ Universitat Pompeu Fabra (UPF), C/ Doctor Aiguader 80, Barcelona, Spain

⁶ CIBER Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, c/ Monforte de Lemos 3-5, Madrid, Spain

⁷ Department of Sport Sciences and Physical Education, École normale supérieure de Rennes, Bruz, France

⁸ VIPS2 Laboratory, University of Rennes, France

⁹ Institute of Environmental Medicine, Karolinska Institutet, Nobels väg 13, 171 77 Stockholm, Sweden

¹⁰ PHAGEX Research Group, Blanquerna School of Health Science, Universitat Ramon Llull (URL), Barcelona, Spain

¹¹ Population Health Research Institute, City St George's, University of London, London, United Kingdom.

¹² Clinical and Epidemiological Neuroscience (NeuroÈpia), Institut d'Investigació Sanitària Pere Virgili (IISPV), Reus, Spain

¹³ National Institute for Public Health and the Environment, Centre for sustainability, environment and health, the Netherlands

Equal-Life Scientific Team:

Bruno Raimbault, Ioar Rivas (Barcelona Institute for Global Health, ISGlobal, Barcelona, Spain);

Maarten Hornikx, Maud Dohmen, Michail Evangelos Terzakis (Built Environment, Technical University Eindhoven, Eindhoven, The Netherlands);

Christin Belke, Dirk Schreckenberger, Julia Kuhlmann, Sarah Leona Benz (Centre for Applied Psychology, Environmental and Social Research (Zeus GmbH), Hagen, Germany);

Anna Hansell , Calvin Jephcote, Kathryn Adams, YingXin Chen (Centre for Environmental Health and Sustainability & School of Geography, Geology and the Environment, University of Leicester, Leicester, United Kingdom);

Albert Wong, Anneke Blokstra, Gerda Doornbos, Hendrick Boshuizen , Jolanda Boer, Markus Viljanen, Jurriaan Hoekstra, Elise van Kempen, Harm van Wijnen, Kim White, Rik Bogers, Sammie Jansen, Sandra Boekhold, Sanne Meijering, Wim Swart (National Institute for Public Health and the Environment, Bilthoven, The Netherlands);

Dick Botteldooren, Luc Dekoninck, Nele De Poortere, Sarah Verhulst, Timothy van Renterghem (Department of Information technology, Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium);

Tanja Vrijkotte (Department of Public and Occupational Health, Amsterdam Public Health research institute, Amsterdam, The Netherlands);

Achilleas Psyllidis, Alessandro Bozzon, Gerd Kortuem, Roos Teeuwen, Vasileios Miliadis (Department of Sustainable Design Engineering, Faculty of Industrial Design Engineering, Delft University of Technology, Delft, The Netherlands);

Karin Loh, Janina Fels, Julia Seitz (Institute for Hearing Technology and Acoustics (IHTA), RWTH Aachen University, Aachen, Germany);

Angel Dzhambov, Peter Lercher (Institute for Highway Engineering and Transport Planning, Graz University of Technology, Graz, Austria);

Gabriele Bolte, Jenny Ahrens (Institute of Public Health and Nursing Research, University of Bremen, Bremen, Germany);

Gordana Ristovska, Igor Spiroski, Mihail Kochubovski (Institute of Public Health of the Republic of North Macedonia, Skopje, North Macedonia);

Peter van den Hazel (International Network for Children's Health, Environment and Safety (INCHES));

Miriam Weber (Municipality of Utrecht, Netherlands);

Sonja Jeram (National Institute of Public Health, Ljubljana, Slovenia);

Anna Sandionigi, Marco Balduini, Marco Brambilla (Quantia Consulting srl, Milan, Italy);

Christoph Giehl, Jan Spilski, Larissa Leist, Maria Klatte, Thomas Lachmann (RPTU, Rheinland-Pfälzische Technische Universität, Germany);

Kerstin Persson Waye, Michael Smith, Yun Chen, Natalia Vincens, Jesper Löve (School of Public Health and Community Medicine, Gothenburg University, Gothenburg, Sweden);

Charlotte Clark (City St George's, University of London, London, United Kingdom.);

Andrei Pyko, Nestor Sanchez Martinez, Sven Bölte (Unit of Occupational Medicine at the Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden);

Ella Braat-Eggen (University of Applied Science, Eindhoven, Netherlands).

Correspondence to: Jaakko Kaprio: jaakko.kaprio@helsinki.fi; +358-503715419; address: Institute for Molecular Medicine, University of Helsinki, PL 20 (Tukholmankatu 8), FI-00014, Helsinki, Finland

Abstract

The exposome, encompassing all external and internal environmental factors, provides a comprehensive understanding of the complex etiology of mental health. This study investigated the relationship between the external exposome during mid-adolescence and plasma proteomic biomarkers of mental health in 935 participants from the FinnTwin12 (Finland) and WALNUTs (Spain) cohorts. The mid-adolescent external exposome included 60 exposures. Biomarkers included 26 plasma proteins (mean age 22) for FinnTwin12 and 31 (mean age 13) for WALNUTs. The exposome-wide proteome-wide analysis revealed seven exposures (related to lifestyle, indoor environmental quality, natural environment, and family environment) linked to six proteins. Greenspace-related exposures were associated with proteins in both cohorts. However, these exposures showed no direct association with concurrent measures: the psychopathology factor or depressive symptoms. Twin analyses indicated genetic influences in some covariations between exposures and proteins. These findings provide evidence for mid-adolescents about the association between external environments and proteomic biomarkers of mental health.

Main

Mental health conditions have emerged as a significant contributor to the global burden of disease¹. Approximately one in eight people worldwide suffers from a mental health condition¹, and the prevalence in adolescents is slightly higher (approx. one in seven)². During adolescence, major physiological changes in the nervous and endocrine systems facilitate the development of coping strategies, interpersonal skills, and individual risk perception³. However, suboptimal timing or deviating magnitude of these changes can increase the risk of mental health conditions, making adolescence a vulnerable period^{4,5}. In view of the complexity and overlap of mental health conditions, a general risk factor, known as the psychopathology factor (p-factor), was introduced for nonspecific mental health conditions by Caspi et al.^{6,7}. A broad range of items measuring psychiatric symptoms was hereby aggregated into various distinct diagnoses, covering two overarching domains: Externalizing and Internalizing, and finally into the general latent dimension⁶. The p-factor unites all mental health conditions and has a neurobiological root⁶.

Besides being self-reported or reported by close relatives or professionals, an individual's mental health status can also be reflected in molecules in blood, other body fluids, or tissues. Omics refers to the collective characterization and quantification of these molecules. Studies have identified molecules as biomarkers from different omics layers (e.g., proteomics) that are associated with mental health conditions⁸⁻¹⁰. Biomarkers are further categorized based on their function. For example, diagnostic biomarkers help identifying patients with specific conditions and improve disease classification, which is crucial in mental health care due to its high heterogeneity, low inter-rater reliability, and common comorbidities¹¹. Safety biomarkers, measured before and after exposure to environmental agents, can monitor adverse events, toxicity, and progression¹¹. Overall, biomarker research is able to advance personalized psychiatry that develops preventive strategies, improves diagnosis and treatment, and understands the underlying mechanisms of mental health conditions.

Previous literature has underscored the role of environmental factors in adolescent mental health¹². Departing from the traditional “one exposure, one disease” model, the exposome framework has been adopted, referring to the totality of environmental factors from a variety of external (not only the physical environment but also the socioeconomic context) and internal exposures individuals ever experienced¹³. Internal exposures are biomarkers of effect proven to be associated with external exposures. For example, cortisol is recognized as a biomarker for mental and psychiatric disorders, and some chemical exposures, such as pesticides, have been shown to relate with cortisol levels^{14,15}. Thus, cortisol can be considered as an internal exposure. Under the exposome framework, various external exposures have been identified to associate with mental health including in the period of adolescence¹⁶⁻²⁰. Although some studies have explored the relationship between the external exposome and multiomic profile²¹⁻²⁵, research linking the external exposome, internal exposome/biomarkers, and mental health conditions remains relatively limited. Such research is crucial for understanding the complex biological networks through which environmental exposures connect to physiological changes.

When discussing the relationship between the environment and mental health, genetics should be considered, as many mental health conditions show heritability to different degrees²⁶. The evaluation of interplay between genetic and environmental factors could help to elucidate individual differences in vulnerability and resilience to environmental stressors in relation with mental health. Research using gene-encoded proteins has further explored this interplay and revealed the potential of mental health biomarkers in such investigation²⁷. Our previous research also highlighted a notable genetic contribution to the covariance between the exposome and depressive symptoms in adolescents and young adults¹⁶.

The main aim of this study was to evaluate the associations between the mid-adolescent external exposome and plasma proteomic biomarkers of mental health. Two types of exposome-wide proteome-wide analyses were used: multivariate Bayesian model with shrinkage priors (MBSP) for the one-to-multiple relationship and exposome-wide association study (ExWAS) for each protein for the one-to-one relationship. Second, we aimed to examine to what extent proteins explain the association between exposures and mental health and to investigate the genetic influences underlying significant exposure—protein associations using twin design. The analysis process is outlined in Figure 1.

Results

Description of participants and materials

Participants included 748 individuals from FinnTwin12 (40.6% female) and 187 from WALNUTs (55.6% female) cohorts, with mean ages of 22.3 years (SD: 0.6) and 13.3 years (SD: 1.0), respectively, at plasma sampling for proteome assessment. Given that the included participants were a subsample of each cohort, Supplemental Table 1 presents the comparison of the distribution of sex and parental education level between the included and baseline-available samples.

The mid-adolescent external exposome set consisted of 60 exposures categorized into seven concept areas (Supplemental Table 2): built environment (5), social environment (community level) (14), social environment (family level) (6), indoor environmental quality (3), lifestyle (3), natural environment (25), and outdoor environmental quality (4).

Biomarkers for mental health (Supplemental Table 3) included 26 proteins in FinnTwin12 and 31 in WALNUTs, of which 25 were common for both cohorts. These protein biomarkers were selected with strong biological plausibility to mental health, and the detailed selection process is presented in the Method section.

One-to-multiple relationships between exposures and proteins

Based on the MBSP²⁸ adjusting for age at plasma sampling for proteome assessment and sex, we did not identify any exposure being associated with proteins in either cohort (by 95% posterior credible intervals). After further adjusting for body mass index (BMI), still, no significant associations were found in either cohort.

One-to-one relationships between exposures and proteins

After Bonferroni correction by the number of effective test for multiple testing, the significant threshold for the P value was 1.87×10^{-3} in FinnTwin12 and 2.42×10^{-3} in WALNUTs. The overall results of ExWASes¹⁶ to each protein biomarkers are presented in Supplemental Tables 4 (FinnTwin12) and 5 (WALNUTs). Volcano plots in Figure 2 present those that have significant associations. Age at plasma sampling and sex were adjusted for.

In FinnTwin12, four exposures in three concept areas (two in lifestyle, one in indoor environmental quality, and one in natural environment) were significantly associated with four different proteins. Participants who reported consuming alcohol at age 14 had higher levels of apolipoprotein M compared to non-consumers (Beta: 0.11, 95% confidence interval (CI) 0.05, 0.17) (Figure 2A). A higher number of adults in the household (household size) when participants were 11–12 years old was associated with lower levels of complement C4-B (Beta: -0.34, 95% CI: -0.36, -0.11) (Figure 2B). Participants who played videogames daily at age 14 had higher levels of serpin family G member 1 compared to those who played a few times a year or less (Beta: 0.15, 95% CI: 0.07, 0.22) (Figure 2C). Moreover, the greenest-season Normalized Difference Vegetation Index (NDVI) within a 500 m buffer was negatively associated with interleukin-6 receptor subunit β (Beta: -0.63, 95% CI: -0.99, -0.27) (Figure 2D).

In WALNUTs, three exposures in two concept areas (two in natural environment, one in social environment at the family level) were significantly associated with two proteins. The percentages of the area covered by trees within 300 (Beta: 0.04, 95% CI: 0.01, 0.06) and 500 (Beta: 0.03, 95% CI: 0.01, 0.05) m buffers were positively associated with progranulin (Figure 2E). Participants whose fathers were unemployed had lower levels of osteomodulin (Beta: -0.82, 95% CI: -1.32, -0.32) compared to those whose fathers were employed (Figure 2F).

The statistical power of significant associations were all 1.00 in FinnTwin12 and ranged from 0.15 to 1.00 in WALNUTs.

Secondary analysis

First, we assessed to what extent the protein explains the association between exposure and mental health (measured at the same time as plasma sampling) based on the significant ExWAS results. A total of 733 out of 748 participants in FinnTwin12 and all included participants in WALNUTs provided valid mental health measures of depressive symptoms based on the General Behavior Inventory (GBI) and the p-factor based on the Strengths and Difficulties Questionnaire (SDQ), respectively. In both FinnTwin12 and WALNUTs, none of these exposures was significantly associated with either depressive symptoms or the p-factor, and the protein explained only a very small portion of weak associations between exposures and mental health measures (Extended Figure 1).

Second, we assessed the genetic contribution by twin analyses. In the within-pair fixed-effect regression model, only the association between videogame playing and serpin family G member 1 remained significant (Beta: 0.19, 95% CI: 0.05, 0.34) (Extended Figure 2A, Supplemental Table 6). After keeping only full pairs, the cross-twin cross-trait (CTCT) correlation coefficient between videogame playing and serpin family G member 1 in monozygotic (MZ) twins was almost 5 times higher than in dizygotic (DZ) twins. The CTCT correlation coefficient between alcohol consumption and apolipoprotein M was nearly two times higher in MZ pairs than in DZ pairs (Extended Figure 2B). Supplemental Note 1 illustrated more on results and methodology.

Sensitivity analysis

The sensitivity ExWASes that further adjusted for BMI were presented in Supplemental Note 2 and Extended Figure 3. Four exposures in the concept area of the social environment (community level) were significantly associated with desmoglein-2 in FinnTwin12, while one exposure in the same concept area was significantly associated with fibulin-1 in WALNUTs.

The sensitivity ExWASes based on the young adulthood external exposome (50 exposures) were shown in Supplemental Note 2 and Extended Figure 4. Alcohol consumption frequency reported in young adulthood was significantly associated with six proteins including apolipoprotein M. Videogame playing was associated with transferrin. Five exposures in the concept area of the natural environment were associated with lysosome-associated membrane glycoprotein 1. The annual mean concentration of black carbon was associated with IgGFc-binding protein.

Due to the circannual dramatic difference in daylight in Finland, we assessed the interaction effect of the season when participants responded to the questionnaire on the association between videogame playing and serpin family G member 1. However, no significant interaction was found (Supplemental Note 2 and Supplemental Table 7).

Discussion

This study, involving nearly a thousand participants from Finland and Spain, systematically explored the relationship between various external environmental exposures during mid-adolescence and plasma proteomic biomarkers of mental health. Seven exposures in the concept areas of lifestyle, indoor environmental quality, natural environment, and social environment (family level) were significantly associated with six different proteins. Notably, greenspace-related exposures consistently showed associations with proteins in both countries. None of these seven exposures were further significantly associated with mental health conditions. Additionally, twin analysis revealed some genetic influence on the relationship between exposures and proteins, particularly between videogame playing and serpin family G member 1.

This study confirms previous findings regarding associations between external exposures and plasma proteins. First, apolipoprotein M, a lipocalin protein linked to high-density lipoprotein (HDL), was found to be associated with alcohol consumption in mid-adolescence, as well as with consumption frequency in young adulthood. Consumption frequency in young adulthood was moreover correlated with more proteomic biomarkers. A Danish case-control study suggested it significantly dysregulated the fibrosis stage among adult patients with alcohol-related liver disease compared to healthy controls²⁹. A systematic review highlighted a positive correlation between alcohol consumption and all HDL subfractions³⁰. A US cohort study further demonstrated the interaction effect of alcohol use on the relationship between platelet-HDL ratio and depressive symptoms in adults³¹. Early onset of alcohol use and misuse is also a predictor of mental health conditions in adulthood^{32,33}. Osteomodulin was correlated with parental education level. In FinnGen, the gene coding for osteomodulin, OMD, shows a strongest association with height, which height is an important determinant of higher educational level, though the mechanism is not fully known^{34,35}.

Next, we found significant associations between exposures and two proteins involved in the complement system, potentially contributing to mental health conditions through immune dysfunction. Complement C4-B, an immune effector, was negatively associated with household size. Infants in larger households were observed to have higher levels of regulatory cells and C-reactive protein during their first year, contributing to nonspecific immune responses³⁶. Our findings suggest that smaller household sizes are associated with higher C4-B levels, which may alleviate mental health conditions, while contradicting the WHO reporting that crowded living conditions exacerbated mental health conditions^{37,38}. The difference could be explained by unmeasured factors such as room availability. Household size is not equivalent to crowding, and the beneficial effect may be rooted in the close family atmosphere. Besides, Complement C4 genes are strongly associated with schizophrenia³⁹ and predict suicide among schizophrenia patients⁴⁰, suggesting them as a promising biomarker for psychiatric diseases^{41,42}. Serpin family G member 1 was found to be associated with videogame playing in mid-adolescence. This protein engages in the regulation of the complement cascade, where elevated blood pressure activates the complement system and produces complement factors⁴³. A cross-sectional study of 282 overweight or obese Canadian adolescents found that videogame playing was associated with higher blood pressure⁴⁴. This rise in blood pressure may serve as a critical link between videogame playing, serpin family G member 1, and mental health conditions⁴⁵. In addition, the serpin proteins are regarded as effective regulators of immune response and involved in neuronal signaling.

In the present study, exposures related to greenspaces were associated with protein levels in both countries, implying a universal connection. A US cross-sectional study found that distance-weighted vegetated land cover correlated with serum interleukin-8, which belongs to the same family as interleukin-6 receptor subunit β (finding in FinnTwin12)⁴⁶. Both interleukin-6 receptor subunit β and progranulin, identified in our study, play roles in the immune system, indicating a potential biological mechanism behind this association. Although a proteome study about NDVI and coronary heart disease did not include overlapping proteins with our results, some proteins they identified were related to immune response and inflammation as well⁴⁷. The

microbial input from greenspaces helps the immune system to distinguish between harmful and beneficial antigens, which may affect mental health through reducing systemic inflammation⁴⁸. However, combined with previous protein—mental health research, in brief, increased greenspaces were linked to worse mental health in Finland, while the opposite was observed in Spain^{38,49}. This discrepancy may be due to differences in greenspaces' coverage, type, or quality, as well as other contextual differences^{50,51}. A recent systematic review indicated that higher quality greenspaces are generally associated with better mental health outcomes⁵². Better measurements to characterize greenspaces are desired. Our findings endorse the importance of greenspace in urban planning and encourage allowing children and adolescents to spend time in green environments.

None of the significant exposures were further associated with mental health conditions. Biomarkers, as early indicators, may reflect subclinical changes that have not yet progressed to noticeable symptoms or a diagnosable condition. A global meta-analysis found that the peak onset for mental health conditions such as primary psychotic states and mood disorders occurs around age 20⁵³. This suggests that symptoms or conditions may develop later than our measurement time, especially during mid-adolescence in WALNUTs. The very small PERE value further suggested that these exposures were pleiotropic, which might be linked to mental health via multiple proteins or other molecules. For example, research using phenome-wide association studies and Mendelian randomization has shown that polymorphisms in genes encoding alcohol-metabolizing enzymes is correlated with various health domains, including mental health, implying the involvement of multiple proteins⁵⁴. Our current proteomic analysis is specific to mental health, and the pathway was through single selected proteins, therefore future studies should expand the proteomic scope and integrate other omics layers to construct a broader picture.

Genetic influences are notable in the associations between exposure and proteins. After adjusting for genetic effects, videogame playing remained associated with serpin family G member 1, while their CTCT correlation coefficient was much higher in MZ than in DZ. Although not all genetic and shared environmental effects were controlled, this finding indicates that environmental factors coexist with genetic influences. Univariate twin modeling has indicated the genetic variance in the proteome^{55,56}, alcohol consumption^{57,58}, and videogame playing⁵⁹. This finding highlights that genes and environment are not an either-or opposition but intertwined. In twin studies, the genetic effect is referred to as the heritable effect passed down through generations within families. Considering family as part of the exposome, it emphasizes the inseparability of the gene and environment and their intricate connection with phenotype¹⁶. Twin studies applying proteomics further elucidate the genetic nurture hypothesis, which explains how parental environmental influences shape offspring outcomes, providing a biological rationale⁶⁰.

Some limitations were noted. First, the relatively small sample size could have limited the statistical power with higher risks of type I error and likelihood of null findings. Although most of the significant associations from ExWASes have adequate statistical power, the weak power caused by small effect size in WALNUTs'

findings highlights the concern of the ubiquity of the exposome, wherein numerous exposures may play small yet significant roles that we may have overlooked. Collaboration with more cohorts is encouraged. Second, calculating the PERE by protein represents an exploratory approach aimed at linking exposures, proteins, and mental health outcomes but lacks causal inference capability. The interpretation should be cautious. High-dimensional mediation analysis techniques are encouraged for complexities of omics data. Third, our analyses did not account for co-exposure scenarios, which may raise concerns on confounding, interaction, or other issues. Future research integrating various exposures into a single analytical framework could assess the mixture, joint, or combined effects of these exposures to provide a more comprehensive understanding of the effect of the environmental totality. Fourth, the exposome set included in our study comprised only those exposures available in both cohorts, which may not encompass all potential exposures, and a larger inclusion of exposure can enhance the systemic investigation of the environmental influence in the future.

Conclusion

Seven external exposures during mid-adolescence related to lifestyle, indoor environmental quality, natural environment, and social environment at the family level were associated with six proteomic biomarkers of mental health. Effects from greenspaces were observed in both Finland and Spain. Although these biomarkers have been proven to be related to mental health with strong biological plausibility, these exposures were not directly associated with depressive symptoms or the p-factor measured simultaneously with plasma sampling. Twin analysis highlighted the interplay between genetic and environmental factors in shaping these associations. These findings offer more comprehensive insights into how environmental factors are linked to proteomic biomarkers of mental health, even before the onset of mental health conditions.

Methods

Participants

This study was based on harmonized data from two cohorts, FinnTwin12 and WALNUTs. FinnTwin12 is a population-based prospective cohort of all Finnish twins born between 1983 and 1987, consisting of one baseline and four follow-ups currently. When twins reached age 14, 1035 families with 2070 twins were invited to an intensive sub-study with psychiatric interviews, biological sampling, and additional questionnaires, and 1854 twins participated. They were then invited to participate again as young adults (mean age 21.9), with 73% completed. The subsample was mostly randomly selected from the full sample, while enriched with twins at elevated familial risk for alcoholism. The updated overview of FinnTwin12 has been published⁶¹. The WALNUTs cohort consists of 771 adolescents who participated in a multi-school, parallel, controlled, 6-month superiority randomized clinical trial in Barcelona, Spain, with baseline measurement between April 2016 and June 2017⁶². Inclusion criteria were attending regular secondary

education (age 11-16) in Barcelona, Spain. Exclusion criteria were regular consumption of omega-3 PUFAs supplements, daily consumption of walnuts, lactose intolerance, or allergy to walnuts, gluten, cereals, dried fruits, peanuts, soy, sesame, or sulfites.

The inclusion criteria of this study were: 1) available plasma sample and untargeted analysis of plasma proteome, 2) data on residential history, and 3) data on mental health and exposure from at least one cohort questionnaire. A total of 748 participants from FinnTwin12 and 187 participants from WALNUTs were included. Comparing the included and full baseline sample, the distribution of sex were similar in both cohort (Supplemental Table 1). A higher proportion of females in FinnTwin12 and males in WALNUTs were included. In FinnTwin12, this may be because we non-randomly included twins at elevated familial risk for alcoholism in the intensive sub-study⁶³. In WALNUTs, we only included participants with available proteome assay⁶².

The ethics committee of the Department of Public Health of the University of Helsinki (Helsinki, Finland) and the Institutional Review Board of Indiana University (Bloomington, Indiana, USA) approved the FinnTwin12 study protocol from the start of the cohort. The ethical approval of the ethics committee of the Helsinki University Central Hospital District (Helsinki, Finland) (HUS) is the most recent and covers the most recent data collection (wave 4) (HUS/2226/2021). The HUS reviews the study annually, and 2024's statement is number 14/2024, dated May 22nd, 2024. All participants and their parents/legal guardians gave informed written consent to participate in the study. The WALNUTs study was approved by the CEIC Parc Salut Mar Clinical Research Ethics Committee (approval number: 2015/6026). The trial was registered at ClinicalTrials.gov (NCT02590848) on 29/10/2015. Written informed consent to participate in the original WALNUTs study was provided by the participants' legal guardian/next of kin. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

External exposome

A total of 60 external exposures were curated from both cohorts and classified into seven categories. The description and data source can be found in Supplemental Table 2. We only included exposures with less than 20% missing data in a practical way to balance between losing too much information and maintaining the integrity of the dataset. We selected the exposure period closest to the mid-adolescence (age 14 in FinnTwin12 and age 11-16 in WALNUTs). Exposures in the concept areas of social environment (family level), indoor environmental quality, and lifestyle were collected via questionnaires in each cohort. Social environment (community level) exposures were derived from national statistics agencies at the postal code level in Finland and at the census section level in Barcelona. Exposures related to the built environment, natural environment, and outdoor environmental quality were obtained from open-source platforms such as the Copernicus program or were calculated based on land use regression model, described elsewhere^{13,16,64}.

For example, NDVI quantified vegetation greenness and characterize the vegetation density, derived from U.S. Geological Survey Landsat data via Google Earth Engine⁶⁵. These exposures were at the geocode level. Residential history from the national registry up to 2020 in FinnTwin12 and the current and one-year previous residential address inquired from parents during baseline in WALNUTs was used to integrate exposures into the cohorts, if these exposures were not collected via questionnaires. We imputed missing values with the median in each cohort.

Proteomic biomarkers of mental health

We used plasma samples obtained in young adulthood (mean age: 22) in FinnTwin12 and at the baseline (age 11-16) in WALNUTs to profile the proteome. The detailed procedure for sample handling is presented in previous studies^{49,66}. For this study the raw MS/MS files were re-analysed with Spectronaut software (Biognosys; version 18.0.2). The direct DIA approach was used to identify proteins. The main data analysis parameters in Spectronaut were: (i) Enzyme: Trypsin/P; (ii) up to 2 missed cleavages; (iii) Fixed modifications: Carbamidomethyl (cysteine); (iv) Variable modifications: Acetyl (protein N-terminus) and oxidation (methionine); (v) Precursor FDR Cutoff: 0.01; (vi) Protein FDR Cutoff: 0.01; (vii) Quantification MS level: MS2; (viii) Quantification type: Area under the curve within integration boundaries for each targeted ion; (ix) Protein database: Homo sapiens Swiss-Prot reference proteome (Uniprot release 2022_01_07_HUMAN)⁶⁷ and Universal Protein Contaminant database⁶⁸. As both datasets were run in multiple batches, batch correction was performed using the proBatch package. The batches of FinnTwin12 and of WALNUTs have been described elsewhere^{66,69}.

Then, based on three previous proteome studies on both cohorts, 111 proteins were identified that were associated with adolescent mental health^{38,49,66}. The coding genes of these proteins were validated using meta-analyzed genetic association results from FinnGen, UK Biobank, and Estonia Biobanks (release 11) via FinnGen's browser, ensuring biological plausibility with mental health and psychiatric outcome³⁴. Validation was determined by significant ($P\text{-value} < 5 \times 10^{-8}$) associations of the protein's coding gene with phenotypes falling into five categories: 1) comorbidities of neurological endpoints, 2) neurological endpoints, 3) diseases of the nervous system, 4) mental and behavioral disorders, and 5) psychiatric endpoints. Proteins identified by at least two studies were also considered validated. A total of 35 proteins met this criterion. After excluding proteins that were unavailable or had over 10% missing data, 26 and 31 proteins were included as biomarkers of mental health in FinnTwin12 and WALNUTs, respectively (Supplemental Table 3). For proteins with less than 10% missing data, they were imputed by the minimum observed value and log2-transformed¹³⁸.

Mental health measures

The p-factor was the main mental health outcome in this study, which summarizes occurrences of internalizing and externalizing symptoms and the propensity to develop any and all forms of common

psychopathologies⁶. In WALNUTs, the p-factor was derived from the 10 internalizing and 10 externalizing problem items from the Spanish version of the Strengths and Difficulties Questionnaire (SDQ), rated on a 3-point Likert scale to assess emotional and behavioral difficulties⁷⁰. We first calculated the mean scores for internalizing and externalizing problem items, allowing for up to four missing values per category (eight in total). Next, we added one to the mean score, applied a logarithmic transformation, and standardized it to z-scores, resulting in the p-factor.

However, there was no assessment of the p-factor in the FinnTwin12 study when participants were young adults. Instead, the short-version General Behavior Inventory (GBI) score, which assesses the occurrence of depressive symptoms, was used as a substitute for the p-factor⁷¹. The self-reported GBI consists of ten Likert-scale items, with total scores ranging from 0 to 30. Higher scores indicate a greater occurrence of depressive symptoms. The sum score was standardized too. Since depression is one key dimension of the p-factor, GBI serves as a suitable replacement in this context.

Multivariate Bayesian model with shrinkage priors (MBSP)

The MBSP was employed to identify exposures affecting the majority of proteins (one-to-multiple relationship). It is a multivariate Bayesian technique with global–local shrinkage priors to generate a sparse model of predictors affecting multiple outcomes⁷². The model has a good performance and selects variables by examining the 95% posterior credible interval²⁸. There were nine sets of different priors and tau inputted into models, and the Deviance Information Criterion was used to identify the model with the best goodness-of-fit⁷². Two covariates were added: age at plasma sampling and sex. MBSP were performed separately for the two cohorts. The MBSP package in the R environment (version 4.3.3) was used²⁸.

Exposome-wide association studies (ExWASes)

Generalized linear regression models with Gaussian distribution were fitted between each exposure and protein, as ExWASes, to explore one-to-one relationships. Bonferroni correction by the number of effective tests was used to adjust for multiple testing, accounting for correlation in exposures⁷³. The models were adjusted for the same covariates as the MBSP. To adjust for the clustered nature of the data (twins in FinnTwin12, schools in WALNUTs), the robust standard error was calculated. ExWASes were performed separately for the two cohorts. We used the rexposome package in the R environment (version 4.3.3)⁷⁴.

Percentage excess risk explained (PERE)

To assess the extent to which proteins explained the association between exposure and mental health measures, PERE was calculated based on generalized linear regression. We performed two models with the following independent variables of 1) exposure only and 2) exposure and protein⁷⁵. The PERE formula is:

$$PERE = \left| \frac{\text{Coefficient of exposure (model 1)} - \text{Coefficient of exposure (model 2)}}{\text{Coefficient of exposure (model 1)} - 1} \right|$$

The PERE value could be interpreted as the involvement of proteins on the association between exposure and mental health measures, regardless of underlying mechanisms such as mediation and moderation⁷⁵. PEREs were calculated based on the exposure—protein associations that were identified in ExWASes. The models were adjusted for age at plasma assessment and sex. Participants missing the mental health measures were excluded (7 in FinnTwin12). The standard error was robusted. The models were performed using Stata 18.0 (StataCorp).

Twin analysis

The twin analysis was based on the significant exposure—protein associations identified in FinnTwin12. There were two parts: within-pair fixed-effect linear regression models (based on all included participants) and the cross-twin cross-trait (CTCT) correlations (based on the full twin pairs). The detailed methodology is presented in Supplemental Note 2.

Sensitivity analysis

First, due to the significant influence of BMI on both the exposome⁷⁶ and proteome⁷⁷, we further added BMI at plasma sampling as another covariate in the MBSP and ExWAS. Participants without BMI information were excluded (N=19 in FinnTwin12), since BMI is the focus of sensitivity.

Second, given the gap between the timing of the exposome assessment (mid-adolescence) and plasma sampling (early adulthood) in FinnTwin12, we assessed 50 exposures to characterize the young adulthood exposome close to age 22, which were repeated measures of mid-adolescence exposures. Videogame playing was not inquired about in young adulthood and we replaced it with the videogame playing reported at age 17. We applied the young adulthood exposome to conduct ExWAS based on data from 768 participants.

Acknowledgement

We would like to thank Dr. Stephanie Zeller, University of Helsinki, and Dr. Augusto Anguita-Ruiz, ISGlobal, for his support in the outcome-wide analysis.

Funding

This research was partly funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 874724 (Equal-Life). Equal-Life is part of the European Human Exposome Network. Data collection in FinnTwin12 has been supported by the National Institute on Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to Richard J. Rose) and the Academy of Finland (grants 100499, 205585, 118555, 141054, 264146, 308248, 312073, 336823, and 352792 to Jaakko Kaprio). Jaakko Kaprio acknowledges support by the Academy of Finland (grants 265240, 263278). The WALNUTs project was supported by the Instituto de Salud Carlos III through the projects CP14/00108 and PI16/00261 (co-funded by European Union Regional Development Fund 'A way to make Europe').

Author contribution

All authors conceived and conceptualized the research idea and developed the analysis plan. Z.W., M.F., J.G., J.J., and J.K. acquired the cohort or exposure data. A.A., A.P., and K.K. performed the proteome assessment. Z.W. led the statistical analysis and performed the analysis in FinnTwin12. N.B. and C.P. contributed to the statistical design and performed the analyses for WALNUTs. Z.W. drafted the original draft of the paper, and all authors reviewed the paper and provided critical comments. I.v.K., J.S., K.K., J.J., and J.K. gave critical supervision and guidance. All authors approved the paper.

Competing interests

The authors declare no competing interest.

Data availability

The FinnTwin12 data are not publicly available due to the restrictions of informed consent. However, the FinnTwin12 data are available through the Institute for Molecular Medicine Finland (FIMM) Data Access Committee (DAC) (fimm-dac@helsinki.fi) for authorized researchers who have IRB/ethics approval and an institutionally approved study plan. To ensure the protection of privacy and compliance with national data protection legislation, a data use/transfer agreement is needed, the content and specific clauses of which will depend on the nature of the requested data. Requests will be addressed in a reasonable time frame (generally two to three weeks), and the primary mode of data access is by either personal visit or remote access to a secure server.

The WALNUTs data are not publicly available due to the restrictions of informed consent. The data contain personal information on children and, according to the ethical approval, they should be kept confidential. Data can be made available upon reasonable request to Jordi Júlvez (jordi.julvez@iispv.cat) for researchers

who meet the criteria for access to confidential data. A data-use/transfer agreement is needed to ensure the protection of privacy and compliance with national data protection legislation, the content and specific clauses of which will depend on the nature of the requested data.

Code availability

The main code for major analysis was uploaded to Github (https://github.com/doge73/exposome_proteome_mental/).

Reference

1. GBD 2019 Mental Disorders Collaborators. 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**, 137–150 (2022).
2. World Health Organization. *World mental health report: Transforming mental health for all.* (2022).
3. Rudolph, K. D., Troop-Gordon, W. & Ye, Z. Adolescence, Physiological Adaptation, and the Development of Stress Responses. in *The Cambridge Handbook of the Development of Coping* (eds. Skinner, E. A. & Zimmer-Gembeck, M. J.) 276–302 (Cambridge University Press, 2023). doi:DOI: 10.1017/9781108917230.015.
4. Paus, T., Keshavan, M. & Giedd, J. N. Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* **9**, 947–957 (2008).
5. Pfeifer, J. H. & Allen, N. B. Puberty Initiates Cascading Relationships Between Neurodevelopmental, Social, and Internalizing Processes Across Adolescence. *Biol. Psychiatry* **89**, 99–108 (2021).
6. Caspi, A. *et al.* The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2013).
7. Adam, D. “P factor” could open an important window on core attributes of mental health maladies. *Proc. Natl. Acad. Sci.* **120**, e2316297120 (2023).
8. Drzymalla, E. *et al.* Epigenome-wide association studies of prenatal maternal mental health and infant epigenetic profiles: a systematic review. *Transl. Psychiatry* **13**, 377 (2023).
9. Fernandes, B. S., Dai, Y., Jia, P. & Zhao, Z. Charting the proteome landscape in major psychiatric disorders: From biomarkers to biological pathways towards drug discovery. *Eur. Neuropsychopharmacol.* **61**, 43–59 (2022).
10. Zhou, Y. *et al.* Suicide and suicide behaviors: A review of transcriptomics and multiomics studies in psychiatric disorders. *J. Neurosci. Res.* **98**, 601–615 (2020).
11. García-Gutiérrez, M. S. *et al.* Biomarkers in Psychiatry: Concept, Definition, Types and Relevance to the Clinical Reality. *Front. Psychiatry* **11**, (2020).
12. Ferschmann, L., Bos, M. G. N., Herting, M. M., Mills, K. L. & Tamnes, C. K. Contextualizing adolescent structural brain development: Environmental determinants and mental health outcomes. *Curr. Opin. Psychol.* **44**, 170–176 (2022).
13. van Kamp, I. *et al.* Early environmental quality and life-course mental health effects: The Equal-Life project. *Environ. Epidemiol.* **6**, (2022).

14. Yehuda, R. & Seckl, J. Minireview: Stress-Related Psychiatric Disorders with Low Cortisol Levels: A Metabolic Hypothesis. *Endocrinology* **152**, 4496–4503 (2011).
15. Kongtip, P. *et al.* Disruption of the Diurnal Cortisol Hormone Pattern by Pesticide Use in a Longitudinal Study of Farmers in Thailand. *Ann. Work Expo. Heal.* **65**, 406–417 (2021).
16. Wang, Z. *et al.* The effect of environment on depressive symptoms in late adolescence and early adulthood: an exposome-wide association study and twin modeling. *Nat. Ment. Heal.* **1**, 751–760 (2023).
17. Lin, B. D. *et al.* Nongenetic Factors Associated With Psychotic Experiences Among UK Biobank Participants: Exposome-Wide Analysis and Mendelian Randomization Analysis. *JAMA Psychiatry* **79**, 857–868 (2022).
18. Ni, M. Y. *et al.* Determinants of physical, mental and social well-being: a longitudinal environment-wide association study. *Int. J. Epidemiol.* **49**, 380–389 (2020).
19. van de Weijer, M. P. *et al.* Expanding the environmental scope: an environment-wide association study for mental well-being. *J. Expo. Sci. Environ. Epidemiol.* **32**, 195–204 (2022).
20. Choi, K. W. *et al.* An Exposure-Wide and Mendelian Randomization Approach to Identifying Modifiable Factors for the Prevention of Depression. *Am. J. Psychiatry* **177**, 944–954 (2020).
21. Fabbri, L. *et al.* Childhood exposure to non-persistent endocrine disrupting chemicals and multi-omic profiles: A panel study. *Environ. Int.* **173**, 107856 (2023).
22. Gao, P. *et al.* Precision environmental health monitoring by longitudinal exposome and multi-omics profiling. *Genome Res.* **32**, 1199–1214 (2022).
23. Robinson, O. *et al.* Associations of four biological age markers with child development: A multi-omic analysis in the European HELIX cohort. *Elife* **12**, e85104 (2023).
24. Maitre, L. *et al.* Multi-omics signatures of the human early life exposome. *Nat. Commun.* **13**, 7024 (2022).
25. Drouard, G. *et al.* Lifestyle differences between co-twins are associated with decreased similarity in their internal and external exposome profiles. *Sci. Rep.* **14**, 21261 (2024).
26. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
27. Wermter, A.-K. *et al.* From nature versus nurture, via nature and nurture, to gene $\times$ environment interaction in mental disorders. *Eur. Child Adolesc. Psychiatry* **19**, 199–210 (2010).
28. Bai, R. & Ghosh, M. High-dimensional multivariate posterior consistency under global–local

- shrinkage priors. *J. Multivar. Anal.* **167**, 157–170 (2018).
29. Niu, L. *et al.* Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat. Med.* **28**, 1277–1287 (2022).
30. Wilkens, T. L., Tranæs, K., Eriksen, J. N. & Dragsted, L. O. Moderate alcohol consumption and lipoprotein subfractions: a systematic review of intervention and observational studies. *Nutr. Rev.* **80**, 1311–1339 (2022).
31. Ni, J., Wu, P., Lu, X. & Xu, C. Examining the cross-sectional relationship of platelet/high-density lipoprotein cholesterol ratio with depressive symptoms in adults in the United States. *BMC Psychiatry* **24**, 427 (2024).
32. Pascale, A. *et al.* Exploring the relationships between adolescent alcohol misuse and later life health outcomes. *Alcohol. Clin. Exp. Res.* **46**, 1753–1765 (2022).
33. Stephenson, M. *et al.* Patterns and predictors of alcohol misuse trajectories from adolescence through early midlife. *Dev. Psychopathol.* 1–17 (2024) doi:DOI: 10.1017/S0954579424000543.
34. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
35. Jelenkovic, A. *et al.* Genetic and environmental influences on human height from infancy through adulthood at different levels of parental education. *Sci. Rep.* **10**, 7974 (2020).
36. Hadley, C. & Decaro, J. A. Testing hypothesized predictors of immune activation in tanzanian infants and children: Community, household, caretaker, and child effects. *Am. J. Hum. Biol.* **26**, 523–529 (2014).
37. World Health Organization. *WHO housing and health guidelines.* (2018).
38. Whipp, A. M., Drouard, G., Rose, R. J., Pulkkinen, L. & Kaprio, J. Protein associations and protein–metabolite interactions with depressive symptoms and the pfactor. *medRxiv* 2024.10.30.24316418 (2024) doi:10.1101/2024.10.30.24316418.
39. Sekar, A. *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
40. Ebrahimi, M. *et al.* Association study of the complement component C4 gene and suicide risk in schizophrenia. *Schizophrenia* **10**, 14 (2024).
41. Severance, E. G. *et al.* Prospects and Pitfalls of Plasma Complement C4 in Schizophrenia: Building a Better Biomarker. *Dev. Neurosci.* **45**, 349–360 (2023).
42. Nimmo, J. *et al.* The complement system in neurodegenerative diseases. *Clin. Sci.* **138**, 387–412

(2024).

43. Wenzel, U. O., Kemper, C. & Bode, M. The role of complement in arterial hypertension and hypertensive end organ damage. *Br. J. Pharmacol.* **178**, 2849–2862 (2021).
44. Goldfield, G. S. *et al.* Video Game Playing Is Independently Associated with Blood Pressure and Lipids in Overweight and Obese Adolescents. *PLoS One* **6**, e26643 (2011).
45. Schaare, H. L. *et al.* Associations between mental health, blood pressure and the development of hypertension. *Nat. Commun.* **14**, 1953 (2023).
46. Egorov, A. I. *et al.* Vegetated land cover near residence is associated with reduced allostatic load and improved biomarkers of neuroendocrine, metabolic and immune functions. *Environ. Res.* **158**, 508–521 (2017).
47. Guo, J. *et al.* Association between greenness and dyslipidemia in patients with coronary heart disease: A proteomic approach. *Ecotoxicol. Environ. Saf.* **231**, 113199 (2022).
48. Lai, H., Flies, E. J., Weinstein, P. & Woodward, A. The impact of green space and biodiversity on health. *Front. Ecol. Environ.* **17**, 383–390 (2019).
49. de Sousa Maciel, I. *et al.* Plasma proteomics discovery of mental health risk biomarkers in adolescents. *Nat. Ment. Heal.* **1**, 596–605 (2023).
50. Fuller, R. A. & Gaston, K. J. The scaling of green space coverage in European cities. *Biol. Lett.* **5**, 352–355 (2009).
51. Wang, Z. *et al.* The urban physical exposome and leisure-time physical activity in early midlife: a FinnTwin12 study. *medRxiv* 2024.06.09.24308658 (2024) doi:10.1101/2024.06.09.24308658.
52. Park, H., Brown, C. D. & Pearson, A. L. A systematic review of audit tools for evaluating the quality of green spaces in mental health research. *Health Place* **86**, 103185 (2024).
53. Solmi, M. *et al.* Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry* vol. 27 (2022).
54. Jennings, M. V. *et al.* A phenome-wide association and Mendelian randomisation study of alcohol use variants in a diverse cohort comprising over 3 million individuals. *eBioMedicine* **103**, (2024).
55. Hagenbeek, F. A. *et al.* Maximizing the value of twin studies in health and behaviour. *Nat. Hum. Behav.* **7**, 849–860 (2023).
56. Drouard, G. *et al.* Twin study provides heritability estimates for 2,321 plasma proteins and assesses missing SNP heritability. *medRxiv* 2024.04.24.24306270 (2024) doi:10.1101/2024.04.24.24306270.

57. Penninkilampi-Kerola, V., Kaprio, J., Moilanen, I. & Rose, R. J. Co-Twin Dependence Modifies Heritability of Abstinence and Alcohol Use: A Population-Based Study of Finnish Twins. *Twin Res. Hum. Genet.* **8**, 232–244 (2005).
58. Virtanen, S., Kaprio, J., Viken, R., Rose, R. J. & Latvala, A. Birth cohort effects on the quantity and heritability of alcohol consumption in adulthood: a Finnish longitudinal twin study. *Addiction* **114**, 836–846 (2019).
59. Hassan, T. Genetic and Environmental Influences on Playing Video Games. *Media Psychol.* **26**, 530–550 (2023).
60. Kong, A. *et al.* The nature of nurture: Effects of parental genotypes. *Science (80-.).* **359**, 424–428 (2018).
61. Rose, R. J. *et al.* FinnTwin12 Cohort: An Updated Review. *Twin Res. Hum. Genet.* **22**, 302–311 (2019).
62. Julvez, J. *et al.* Walnuts, Long-Chain Polyunsaturated Fatty Acids, and Adolescent Brain Development: Protocol for the Walnuts Smart Snack Dietary Intervention Trial. *Front. Pediatr.* **9**, (2021).
63. Sihvola, E. *et al.* Early-onset depressive disorders predict the use of addictive substances in adolescence: a prospective study of adolescent Finnish twins. *Addiction* **103**, 2045–2053 (2008).
64. de Hoogh, K. *et al.* Spatial PM2.5, NO2, O3 and BC models for Western Europe – Evaluation of spatiotemporal stability. *Environ. Int.* **120**, 81–92 (2018).
65. Landsat Missions. Landsat Normalized Difference Vegetation Index | U.S. Geological Survey. <https://www.usgs.gov/landsat-missions/landsat-normalized-difference-vegetation-index>.
66. Afonin, A. M. *et al.* Proteomic insights into mental health status: plasma markers in young adults. *Transl. Psychiatry* **14**, 55 (2024).
67. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
68. Frankenfield, A. M., Ni, J., Ahmed, M. & Hao, L. Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics. *J. Proteome Res.* **21**, 2104–2113 (2022).
69. Afonin, A., Piironen, A.-K., Julvez, J., van Kamp, I. & Kanninen, K. M. Plasma proteome demonstrates sex-specific associations with mental health risks in adolescents. *medRxiv* 2024.12.16.24319062 (2024) doi:10.1101/2024.12.16.24319062.
70. Ortuño-Sierra, J., Fonseca-Pedrero, E., Paino, M., Sastre i Riba, S. & Muñiz, J. Screening mental

health problems during adolescence: Psychometric properties of the Spanish version of the Strengths and Difficulties Questionnaire. *J. Adolesc.* **38**, 49–56 (2015).

71. Kokko, K. & Pulkkinen, L. Unemployment and Psychological Distress: Mediator Effects. *J. Adult Dev.* **5**, (1998).
72. Anguita-Ruiz, A. *et al.* Beyond the single-outcome approach: A comparison of outcome-wide analysis methods for exposome research. *Environ. Int.* **182**, 108344 (2023).
73. Li, M.-X., Yeung, J. M. Y., Cherny, S. S. & Sham, P. C. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131**, 747–756 (2012).
74. Hernandez-Ferrer, C. *et al.* Comprehensive study of the exposome and omic data using rexposome Bioconductor Packages. *Bioinformatics* **35**, (2019).
75. Richiardi, L., Bellocco, R. & Zugna, D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int. J. Epidemiol.* **42**, 1511–1519 (2013).
76. Haddad, N., Andrianou, X., Parrish, C., Oikonomou, S. & Makris, K. C. An exposome-wide association study on body mass index in adolescents using the National Health and Nutrition Examination Survey (NHANES) 2003–2004 and 2013–2014 data. *Sci. Rep.* **12**, 8856 (2022).
77. Rodriguez-Muñoz, A. *et al.* A Systematic Review of Proteomics in Obesity: Unpacking the Molecular Puzzle. *Curr. Obes. Rep.* (2024) doi:10.1007/s13679-024-00561-4.

Fig. 1: Flowchart of the study pipeline.

Flowchart of the study demonstrating the path from the selection of external exposome and proteome to exposome-wide proteome-wide analysis and ending with twin analysis and pathway exploration toward mid-adolescent external mental health.

Fig. 2: Volcano plots of exposome-wide association studies (ExWASes) for apolipoprotein M (A), complement C4-B (B), serpin family G member 1 (C), interleukin-6 receptor subunit β (D) in FinnTwin12 and progranulin (E) and osteomodulin (F) in WALNUTS.

Only ExWASes with significant associations were plotted. The y-axis shows statistical significance as $-\log_{10}(\text{p-value})$ for the adjustment for multiple testing. Adjustments were made for sex and age at plasma sampling for proteome assessment. The significance thresholds for p-values were 1.84×10^{-3} for FinnTwin12 and 2.42×10^{-3} for WALNUTs (the dark gray dash line). The light gray dash line indicated the significance threshold by a restricted Bonferroni correction ($\text{p-value} < 8.33 \times 10^{-4}$).

Extended Fig. 1: Bar plots for protein-explained excess risk for exposure–mental health association based on the identified exposure–protein associations

a. Percentage excess risk explained (PERE) by apolipoprotein M in the association between alcohol consumption and GBI score. **b.** PERE by complement C4-B in the association between the number of adults in the household and GBI score. **c.** PERE by serpin family G member 1 in the association between videogame playing and GBI score. **d.** PERE by interleukin-6 receptor subunit β in the association between greenest-season NDVI (500m) and GBI score. **e.** PERE by progranulin in the association between % tree cover (300 m) and p-factor. **f.** PERE by progranulin in the association between % tree cover (500 m) and p-factor. **g.** PERE by osteomodulin in the association between paternal education and p-factor. The comparison was between participants playing video games every day and a few times a year or less (ref.) for the association between video game playing and GBI score and was between participants whose fathers received low and high (ref.) levels of education for the association between paternal education and p-factor.

Extended Fig. 2: Twin analyses for the identified exposure–protein associations.

a. Forest plot based on within-pair fixed-effect linear regression models. The comparison was between participants playing video games every day and a few times a year or less (ref.) for the association between videogame playing and serpin family G member 1. **b.** Histogram plot for cross-twin cross-trait correlation. Categorical exposure of videogame playing was regarded as a complete variable in the correlation calculation. Since the number of adults in the household remained constant between cotwins, no twin analysis was conducted to examine its association with complement C4-B.

Extended Fig. 3: Volcano plots of sensitivity exposome-wide association studies (ExWASes) for apolipoprotein M (A), complement C4-B (B), serpin family G member 1 (C), interleukin-6 receptor subunit β (D), and desmoglein-2 (E) in FinnTwin12 and progranulin (F), osteomodulin (G), and fibulin-1(H) in WALNUTs

Only ExWASes with significant associations were plotted. Adjustments were made for sex and age and BMI at plasma sampling for proteome assessment. The y-axis shows statistical significance as $-\log_{10}(\text{p-value})$ for the adjustment for multiple testing.

Extended Fig. 4: Volcano plots of sensitivity exposome-wide association studies (ExWASes) for the young adulthood exposome for apolipoprotein M (A), transferrin (B), serpin family A member 6 (C), clusterin (D), lysosome-associated membrane glycoprotein 1 (E), fibulin-1 (F), serpin family A member 10 (G), and IgGfC-binding protein (H).

Only ExWASes with significant associations were plotted. The young adulthood exposome were used instead. Adjustments were made for sex and age and at plasma sampling for proteome assessment. The y-axis shows statistical significance as $-\log_{10}(\text{p-value})$ for the adjustment for multiple testing.

