Research paper

A MIR4646 associated methylation locus is hypomethylated in adolescent depression

Adrian E. Boström⁎, Diana-Maria Ciuculea, Misty Attwoodb, Regina Krattingerb, Lamia Nikontovic, Olga E. Titovaa, Gerd A. Kullak-Ublick, Jessica Mwinyia, Helgi B. Schiötha

Division of Pharmacology, Department of Neuroscience, Uppsala University, Uppsala, Sweden. Department of Clinical Pharmacology and Toxicology, University Hospital Zurich, University of Zurich, Switzerland.

ARTICLE INFO

Keywords:
MIR4646
Methylome-wide association study
Adolescent depression
Omega-3
MicroRNA

ABSTRACT

Background: Studies of epigenetics and transcriptional activity in adolescents may provide knowledge about possible preventive strategies of depression.

Methods: We present a methylome-wide association study (MWAS) and cohort validation analysis of depression in adolescents, in two separate cohorts: discovery (n=93) and validation data set 1 (n=78). The genome-wide methylation pattern was measured from whole blood using the Illumina 450K array. A second validation cohort, validation data set 2, consists of post-mortem brain biopsies from depressed adults (n=58). We performed a MWAS by robust multiple linear regressions of methylation to a modified risk-score assessment of depression. Methylation levels of candidate CpG sites were correlated with expression levels of the associated gene in an independent cohort of 11 healthy volunteers.

Results: The methylation state of two CpG sites reliably predicted ratings of depression in adolescents (cg13227623 and cg04102384) (p < 10E-06). Cohort validation analysis confirmed cg04102384 – located in the promoter region of microRNA 4646 (MIR4646) – to be hypomethylated in both validation data set 1 and validation data set 2 (p < 0.05). Cg04102384 was inversely correlated to expression levels of MIR4646-3p in healthy controls (p < 0.05).

Limitations: MIR4646 was not differentially expressed in a subset of samples with adolescent depression measured by qRT-PCR measurements.

Conclusion: We identify a specific MIR4646 associated epigenetic risk site to be associated with depression in adolescents. Cg04102384 putatively regulates gene expression of MIR4646-3p. Target gene prediction and gene set overrepresentation analysis revealed involvement of this miRNA in fatty acid elongation, a process related to omega-3 fatty acids, previously associated with depression.

1. Introduction

Depressive disorders are highly prevalent during the adolescence and are associated with a higher risk for suicide, antisocial behavior, substance abuse, significant impairment and an increase of functional disability (Glied and Pine, 2002). Major depression (MDD) is common in young people, affecting up to 5–6% of the adolescents (Costello et al., 2006). Moreover, a reliable detection of early onset depression is of high importance as a solid diagnosis during adolescence would help to efficiently unveil people at risk for depression at later stages of life (Birmaher et al., 2002). However, the exact mechanism that underlies the high risk for depression during the youth is poorly understood. Investigating the relationship between the MDD risk as evaluated by the Development and Well-being Assessment (DAWBA), epigenetic mechanisms and transcriptional activity in adolescents may provide knowledge of preventive strategies of depression among adults.

In recent years, there has been increasing interest to elucidate the role of epigenetic modifications in the pathogenesis of psychiatric disorders (D’Addario et al., 2012; Dell’Osso et al., 2014; Ružicka et al., 2015; Walker et al., 2016). In this context, especially changes in the methylation at CpG dinucleotides within regulatory regions of the DNA were studied, which were shown to be responsive to environmental signals by modifying the transcription of genes (Cordova-Palomera et al., 2015). Notably, the importance of altered gene expression in MDD was recently demonstrated in a study where genes, such as DVL3, CALM1 and NMUR1 were differentially expressed in depressed individuals (Jansen et al., 2015), resulting in a new complementary mechanistic insight. Moreover, DNA methylation has

⁎ Corresponding author.
E-mail address: adrian.bostrom@neuro.uu.se (A.E. Boström).

http://dx.doi.org/10.1016/j.jad.2017.05.017
Received 21 December 2016; Received in revised form 4 April 2017; Accepted 6 May 2017
Available online 17 May 2017
0165-0327/ © 2017 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).
been associated with other genomic functions, such as alternative splicing and promoter usage (Maunakea et al., 2010), which lead as well to a transcriptional modulation of MDD related genes.

Dysregulation of epigenetic control is particularly interesting in the context of MDD development (Menke and Binder, 2014), as the pathogenesis of this disease is characterized by a tight gene-environment interplay that influences neurobiological processes underlying the disease (Nugent et al., 2011; Schmitt et al., 2014). Importantly, depressive symptoms are thought to have an incidence spike during adolescence (Birmaher et al., 2002), suggesting a long-term importance for investigating the epigenetic changes during this time frame. Previous targeted studies suggested that the methylation profile of the BDNF gene could serve as biomarker of late-life depression (Januar et al., 2015a). Furthermore, another study highlighted altered DNA methylation of the serotonin transporter (5-HTT) gene in depressed adolescents (Olsson et al., 2010). A recent genome-wide study identified differentially methylated loci between medication-free depressed individuals and non-psychiatric controls in blood, but the study was not corrected for cell-type heterogeneity (Numata et al., 2015), meriting further research in medication-free subject taking cell heterogeneity into account. The abundance of cell populations is relevant through their distinct signatures of DNA methylation (Reinius et al., 2012) that could induce false differences between depressed individuals and controls. Also, an epigenome-wide study reported differences of PRIMA1 gene methylation between depressed adults and controls. However, the authors were not able to replicate the findings in their additional samples (Sabunciyian et al., 2012).

Another attractive candidate mechanism to induce phenotype-associated changes is related to small non-coding RNAs (miRNA) which play a critical role in post-transcriptional fine-regulation of gene expression in many different tissues, including human brain functioning (Chen and Qin, 2015). There is evidence suggesting an interaction between miRNA expression and psychiatric disorders, possibly mediated by DNA methylation (Villela et al., 2016). Due to the miRNAs role in degrading and/or inhibiting translation, a better understanding of the interaction between shifts in DNA methylation and a changed miRNA expression would provide additional valuable information about mechanistic links between methylation and depression.

Herein, we set out to identify the role of DNA methylation between whole-blood samples of adolescents at high risk for depression and controls in an unbiased methylome-wide approach, using the 450k Illumina array. The identified differentially methylated loci were further replicated in an independent cohort of depressed adolescents. We detect the gene encoding MIR4646 as significantly differentially methylated and investigate its change in expression further in an independent cohort of healthy controls. Lastly, we measure the miRNA levels for a sub-sample of the discovery data set and observe the expression change further in an independent cohort of depressed adults and controls.

2. Methods

2.1. Discovery data set

The study of the Discovery data set was approved by the local ethics committee in Uppsala, Sweden (Regionala etikprövningsnämnden i Uppsala) and included 93 adolescents aged 14–17, recruited in the years 2013–2014. The Discovery data set has been previously published (Ciuculete et al., 2017). Both subjects and their participating parent(s) gave their written informed consent to participating in the study. Subjects were randomly selected from public school in Uppsala County and were included in the study if they exhibited an overall risk for psychiatric diagnoses of 15% or more, as measured by the DAWBA web-based diagnostic interview (described below (Goodman et al., 2011)). Self-reported information pertaining to basic physiological parameters and medication was provided by the participants. Body weight was measured for body mass index (BMI) calculation. The BMI z-scores were calculated and based on these values each subject was stratified into one of four weight category groups (underweight, normal weight, overweight and obese), as defined by the Center for Disease Control and Prevention (CDC). Subjects were grouped into three categories based on their DAWBA risk-score assessment for depression as ‘Controls’ (~0–3%), ‘15% Risk Depression’ (~ 15%) and ‘Depressed’ (~50–75%), respectively.

2.2. Validation data set 1

In the replication stage, 130 samples from the study mentioned above, but characterized and measured earlier in the time frame between November 2012 and January 2013, were studied. The Validation data set 1 has been previously published (Ciuculete et al., 2017). The study was approved by the local ethics committee in Uppsala, Sweden (Regionala etikprövningsnämnden i Uppsala) and all participants and their parent(s) gave their written informed consent. This data set is population-based and the same parameters were recorded as for the Discovery data set. In order to increase the power to detect meaningful differences, we excluded subjects with intermediate risk scores for any psychiatric disorder, and stratified the remaining individuals into one of two categories: ‘Controls’ (~0%) and ‘Depressed’ (~15–50%).

2.3. Psychiatric diagnoses

The DAWBA consists of web-based diagnostic interviews used to evaluate DSM-IV and ICD-10 type diagnoses specifically for individuals in the age range 5–17 years. The questionnaire is completed separately by both a parent and the child, and encompasses the most prevalent behavioral, emotional and hyperactivity type disorders. The covered categories include depression, generalized anxiety disorder, SAD, post-traumatic stress disorder (PTSD), autism and obsessive compulsive disorder (OCD). An algorithm was used to convert questionnaires to probability of diagnosis, ranging from less than 0.1–70% and reflecting the probability that an experienced clinical rater would assign the individual a corresponding DSM-IV or ICD-10 diagnosis (Goodman et al., 2011). No other measure of severity of depressive symptoms was available for the Discovery and Validation data set 1.

2.4. DNA specimens

Body weight was measured and participants answered a questionnaire with questions about living conditions, place of birth and basic physiological parameters (height, age etc.). Venous blood was taken according to standard procedures and stored in six tubes at a total volume of 25 ml comprising two EDTA coated tubes for DNA extraction, two PAX gene tubes for RNA extraction, one lithium-heparin treated tube for plasma and one substrate-free tube for serum. All tubes were kept at minus 80 °C after separation of plasma and serum by centrifugation.

2.5. Methylation profiling

DNA was extracted using the phenol-chloroform method (Sambrook et al., 1989), and bisulfite converted by the EZ DNA Methylation - GoldTM kit (ZymoResearch, USA). Bisulfite converted DNA was hybridized to the Illumina 450k methylation chip (Illumina, San Diego, CA, USA). The Illumina chip measures methylation at 485 777 CpG sites. The Illumina iScan system (Illumina, San Diego, CA, USA) was used for imaging of the array, whereby the methylation level of each CpG site was determined.

2.5.1. Data preprocessing and quality control

Preprocessing of the methylation data was performed by adjustment...
of probe type differences, removal of batch effects and probe exclusion. Subsequently, principal component analysis (PCA) was used to identify sample outliers in the methylation data. Methylation preprocessing steps were performed in using R statistics (www.r-project.org) together with the packages minfi (Aryee et al., 2014), wateRmelon (Schalkwyk et al., 2013), ChAMP (Morris et al., 2014), sva (Leek et al., 2012), and limma (Smyth, 2004) of the Bioconductor project and the FactoMineR (Lê et al., 2008) package of the CRAN project. Concerning adjustment of probe type differences, removal of batch effects, probe exclusion and sample exclusion criteria please see Supplementary material.

2.5.2. Accounting for cellular heterogeneity

DNA methylation measured in whole blood is composed of different cell populations (Reinius et al., 2012). Rask-Andersen et al. demonstrated recently that changes in leukocyte fractions could introduce considerable variability in the DNA methylation pattern which could bias downstream analyses. Thus, it is important to take into account white blood cell type heterogeneity in genome-wide DNA methylation studies (Rask-Andersen et al., 2016). In the Discovery and Validation data set 1, we implemented a mini-f-based statistical procedure of the Houseman algorithm (Aryee et al., 2014), which uses raw intensity DNA methylation files to calculate the relative proportions of B cells, CD4+ and CD8+ T cells, granulocytes, monocytes, and natural killer cells.

2.6. Micro-RNA profiling

2.6.1. Sample inclusion considerations

Candidate microRNAs were investigated by measuring their expression levels in whole blood. For this purpose, we extracted RNA from a group of 27 subjects comprising eleven ‘Controls’, nine ‘15% risk depression’ and seven ‘Depressed’ individuals from the Discovery data set. Samples were selected based on their level of DNA methylation at the associated methylation locus (cg04102384). In order to increase the power to detect meaningful associations, we selected the eleven controls with the lowest methylation levels, and 16 cases (‘15% Risk Depression’ or ‘Depressed’) with the highest methylation levels.

2.6.2. Reverse transcription and quantitative real-time PCR

To measure miRNA expression, 10 ng of extracted RNA was reverse transcribed into cDNA using a TaqMan® miRNA Reverse Transcription Kit (Applied Biosystems, Waltham, MA, USA) and specific stem-loop reverse transcription primers (TaqMan® MicroRNA Assays MIR4646-3p and MIR4646-5p; Life Technologies, Carlsbad, CA, USA). RT-PCR was performed using 0.67 µl cDNA and 9.3 µl RT-PCR Universal Fast Master Mix (Applied Biosystems) including primers. U6snRNA/RNU44 was used as an internal control. All measurements were performed in triplicate.

2.7. Characterisation of the expression data set

Eleven healthy male volunteers aged between 18 and 40 years were recruited from the region of Uppsala, Sweden, between 2013 and 2014. Blood analyses were performed before and after a meal intake. For the purpose of this study, only the non-fasting blood samples were further studied. The data was adjusted for white blood cell type heterogeneity. More details on the cohort and preprocessing of the methylation and RNA specimens have been previously published (Rask-Andersen et al., 2016). MicroRNA expression level was analyzed in RNA samples isolated from eleven healthy individuals in blood. After biotinylated RNA was prepared according to FlashTag Biotin HSR RNA labeling kit, 120 µl of each sample were loaded to the Affymetrix® miRNA 4.1 Array Plates. Finally, the arrays were hybridized, washed and scanned with the GeneTitan® Multi-Channel (MC) Instrument. The raw data was normalized in Expression Console, provided by Affymetrix, using the robust multi-array average method that was first suggested by Li and Wong (Irizarry et al., 2003; Li and Wong, 2001).

2.8. Validation data set 2

The Validation data set 2 includes data from an independent published cohort (Array Express Database, http://www.ebi.ac.uk/arrayexpress/). The data is openly available (E-GEOD-41826) and were originally published by Guintivano et al. (2013). The group measured DNA methylation from post mortem frontal cortex samples of 29 major depression (MDD) subjects and 29 matched controls. After the neuronal proportion extraction, they studied the extent of neuron and glia specific DNA methylation variation independent of disease status (MDD or control). DNA methylation profiles were generated using the Illumina 450 K methylation BeadChip, which have been made available online along with phenotypic information pertaining to age, gender, cell type (glia, neurons, bulk or mixed), diagnosis (depression or control), post mortem interval and race (Asian, African or Caucasian). For more details regarding the cohort and methylation specimens see the original article by Guintivano et al. (2013).

2.9. Statistical analysis

All statistical analyses were performed in using R statistics, version 3.3.0.

2.9.1. Data analysis

Chi-squared tests with Monte Carlo computed p-values were used to detect differences in categorical variables, e.g. gender, BMI group (underweight, normal weight, overweight and obese), use of medications (sleep medications, neuroleptics, contraceptive pills, anxiolytics, ADHD medications and antidepressives) and the DAWBA risk score for any psychiatric disorder. In the Discovery group, ANOVA tests were used to investigate group differences in age. In the Validation data set 1, which was stratified into two DAWBA groups, we used t-tests to investigate group differences in age.

2.9.2. Adjusting for potential confounders

In methylene-wide association studies, hidden confounders such as lifestyle patterns or even prandial states can introduce unknown sources of bias. Drawing on a method for epigenome-wide analyses introduced by Zhagool et al. (Zaghool et al., 2015), PCA-analysis was used to account for potential unmeasured sources of variation in the DNA methylation data using the MethylPCA tool (Chen et al., 2013). The known covariates (including age, gender, weight category, the relative proportions of white blood cell types and the DAWBA overall risk score for any psychiatric disorder) were regressed out prior to PCA. The calculated first ten principal components were considered as additional potential covariates in the methylene-wide analysis.

There were many potential covariates on the association analysis between DNA methylation and depression risk score, e.g. gender, age, self-reported use of medication, the white blood cell coefficients and the first ten principal components. To avoid overfitting by including too many covariates, we investigated each individual covariate against the phenotype of interest in regression models using the ‘lm’ function in R. Covariates were incrementally and independently selected. Using the computed analysis of variance, we tested whether the addition of a particular covariate resulted in a better fit to the model and only included variables with a p-value < 0.05. The best linear model for depression risk score included the CpG sites, risk score for any psychiatric disorder (p < 0.00001), self-reported use of ADHD medication (p = 0.022), the relative proportion of CD4 + T cells (p = 0.017) and the first principal component (p = 0.00081). For statistical analysis, we transformed the beta values to M-values, which have been shown to be statistically more robust (Du et al., 2010).

2.9.3. Methylnote-wide association study

The association between DNA methylation and depression was tested by linear models using the ‘limma’ package for R, applying an
empirical Bayes method based on a moderated t-statistic (Smyth, 2004). We assumed a linear model where the M values of each CpG site were used as a quantitative dependent trait and the phenotype characterizing the risk for depression were used as covariates together with the other optimal covariates.

We used the package 'GenABEL' available for R to calculate the genomic inflation factor lambda for the epigenome-wide analysis (Aulchenko et al., 2007) to evaluate whether a general systemic inflation of significance values was abundant. In accordance to Zaghloul et al. (2015), we first tested whether the calculated optimal covariates resulted in less systemic inflation as compared to limiting the covariates to age, gender, and BMI z-score weight category. Using the optimal covariates in the methylome-wide association analysis between DNA methylation and depression risk score reduced the genomic inflation factor lambda from 1.95 to 1.11. To exclude any potential bias from systemic inflation of significance values (lambda > 1 in this case), all subsequent chi-squared statistics on a set of candidate markers were divided by lambda (Hinrichs et al., 2009). For each CpG site, we thus divided the regression t-value by lambda, and used the adjusted t-values to test for significance based on the t distribution. The inflation factor lambda for the adjusted p-values was estimated to be ~0.90, and systemic inflation of significance values was excluded. We used the bonferroni method correct for multiple testing. Bonferroni and lambda-adjusted p-values < 0.05 were considered significant.

2.9.4. Investigation of candidate CpG-sites in the Validation data set 1

In the Validation data set 1, we performed independent samples t-tests of candidate CpG-sites, contrasting methylation between cases and controls and taking the direction of the methylation change into account.

2.9.5. Correlation analyses between methylation and expression data

CpG sites consistently hyper- or hypomethylated in the Discovery and Validation data set 1 were further investigated with regard to an association with transcriptional expression of the microRNA in focus using the Expression data set. We performed Spearman correlations of methylation M-values to test for significance values (lambda > 1 in this case), all subsequent chi-squared statistics on a set of candidate markers were divided by lambda (Hinrichs et al., 2009). For each CpG site, we thus divided the regression t-value by lambda, and used the adjusted t-values to test for significance based on the t distribution. The inflation factor lambda for the adjusted p-values was estimated to be ~0.90, and systemic inflation of significance values was excluded. We used the bonferroni method correct for multiple testing. Bonferroni and lambda-adjusted p-values < 0.05 were considered significant.

We also measured the miRNA levels of MIR4646-3p and MIR4646-5p by TaqMan analysis in 26 subjects from the Discovery data set and studied methylation-expression correlations by Spearman's rank correlation.

2.9.6. Differential expression analysis of MIR4646-3p/5p in a subset of samples from the Discovery data set

In the 26 subjects for whom miRNA levels were measured, we studied potential differences in normalized expression levels between the defined depression groups by ANOVA models, not taking any covariates into account.

2.9.7. Investigation of candidate CpG-sites in Validation data set 2

In Validation data set 2, we performed independent samples t-tests of candidate CpG-sites, contrasting methylation M-values of different brain cell types between depressed cases and controls and taking the direction of the methylation change into account. Neuronal and glial derived DNA methylation profiles were studied separately. As a second step, we performed binomial logistic regression models of a binary outcome variable (major depression or control) to CpG-site methylation, and taking into account age, gender, post mortem interval and ethnicity as co-variables.

2.10. Functional analysis of the identified CpG site via chromatin states and long-range interactions

In order to illustrate the functional role of the identified CpG site in brain, as well as its potential regulatory effect on other genes, we performed chromatin states and long-range interactions analyses using the ENCODE project. The analysis uses Hidden Markov Models (HMMs), which were applied to seven brain regions, i.e. brain angular gyrus (BrainAG), brain anterior caudate (BrainAC), brain cingulate gyrus (BrainCG), brain hippocampus (BrainHippo), brain inferior temporal lobe (BrainITL), brain substantia nigra (BrainSN) and brain dorsolateral prefrontal cortex (BrainDPC), together with peripheral blood mononuclear primary cells (PBMC). As a result, an 18-state model was obtained, which, for simplicity, was reduced to five regions defining the relevant gene regulatory roles, which were indicated as (1) red, for active/flanking active/bivalent/poised transcription start site (TSS), (2) yellow, for active/bivalent/genic enhancer; orange, for flanking bivalent TSS enhancer, green, for active transcription and grey, for repressed polyComb state (Fig. 3.). The long-range interactions were investigated using chromatin analysis by paired-end tag sequencing (ChIA-PET). Several cell lines were used for the analysis, including erythrocytic leukaemia cells (K562), breast cancer (MCF-7), cervical cancer (HelaS3) and human colon carcinoma (HCT-116) cells, targeting the transcription factors RNA polymerase II and CTCF. Data was downloaded from the WashU Epigenome Browser, 37/hg19 version.

2.11. Target gene prediction and pathway analysis of MIR4646-3p

Candidate CpG-sites associated with microRNAs were further investigated by computationally predicting putative gene targets of the aforementioned miRNAs using the online webtool MiRWalk 2.0 (Dweep et al., 2011), a sophisticated online software tool that documents predictions from several independent prediction algorithms, including Targetscan (Grimson et al., 2007), DIANA-microT-CDS (Paraskevopoulou et al., 2013), mir randel2010 (Betel et al., 2010) and RNAhybrid (Rehmsmeier et al., 2004). miRNA targets were considered as relevant hits when having a seed length ≥7 bases and when located within the 3′-UTR region. Genes identified as putative miRNA targets were further investigated by overrepresentation analysis of KEGG-defined pathways, using the online web tool ‘ConsensusPathDB-human’ (Kamburov et al., 2013).

2.12. In vivo interaction analysis between hsa-miR-4646-3p and its putatively regulated target genes as identified in overrepresented pathway analyses

In the Expression cohort, we performed Pearson correlations of hsa- miR-4646-3p levels with the expression of genes identified in the KEGG-defined pathways. The following genes were studied: Acyl-CoA thioesterase 1 (ACOT1), Acyl-CoA thioesterase 1 (ACOT2), ELOVL Fatty Acid Elongase 2 (ELOVL2), ELOVL Fatty Acid Elongase 5 (ELOVL5), Hydroxysteroid 17-Beta Dehydrogenase 12 (HSD17B12) and Palmitoyl-Protein Thioesterase 1 (PPT1).

2.13. Alignment and synteny conservation across species

A thorough search of miRNA-4646 was performed in the miRBase database (release 21) using the webserver (Kozomara and Griffiths-Jones, 2014). A BLASTn (Boratyn et al., 2013) search was performed through the NCBI Blast webserver (v. 2.5.0) using the NCBI Genomic Reference Sequences dataset with word size 16 and all other parameters default values. The relevant hits were retained considering e-value (> 5e-06), sequence identity and appropriate sequence coverage, i.e., if the seed regions of the mature miRNAs were covered. The sequence region of interest for Homo sapiens (NC_000006.12), Pan troglodytes (NC_006473.4), Pan paniscus (NW_01413975.1), Gorilla gorilla gorilla (NC_018430.1), Pongo abelii (NC_012597.1), Nomascus leucogenys (NC_019816.1), Macaca mulatta (NC_027896.1), Papio anubis (NW_003873063.1), Saimiri boliviensis (NW_003943814.1), Callithrix jacchus (NC_013899.1), Aotus nancymaae (NW_012186114.1), Mandrillus leucophaeus (NW_012106809.1), Rhinopithecus bieti
3. Results

3.1. Behavior of the clinical outcome variables

In the Discovery data set, comprising 93 subjects and in the majority female, we initially aimed to identify CpG-sites, in which modifications of the epigenetic profile are associated with a modified risk score of depression. There were no significant differences between the three DAWBA sub-groups (‘Controls’, ‘15% Risk Depression’ and ‘Depressed’) in age, gender, BMI z-score derived weight categories, in the use of medications or in the relative proportion of white blood cell types (CD4+ and CD8+ T cells, B cells, monocytes, NK cells, and granulocytes). As expected, the ‘Depressed’ subgroup had significantly higher DAWBA general risk score estimates for any psychiatric diagnosis (p < 0.001). The Validation data set 1 of 78 subjects included only female, we initially aimed to identify CpG-sites, in which modifications of the epigenetic profile are associated with a modified risk score of depression. There were no significant differences between the three DAWBA sub-groups (‘Controls’, ‘15% Risk Depression’ and ‘Depressed’) in age, gender, BMI z-score derived weight categories, in the use of medications or in the relative proportion of white blood cell types (CD4+ and CD8+ T cells, B cells, monocytes, NK cells, and granulocytes). As expected, the ‘Depressed’ subgroup had significantly higher DAWBA general risk score estimates for any psychiatric diagnosis (p < 0.001). The Validation data set 1 of 78 subjects included only female subjects. Depressed cases showed a higher BMI (p < 0.01), took more often contraceptive pills (p < 0.01) and scored higher in the DAWBA depression. There were no significant differences between the three DAWBA sub-groups (‘Controls’, ‘15% Risk Depression’ and ‘Depressed’) in age, gender, BMI z-score derived weight categories, in the use of medications or in the relative proportion of white blood cell types (CD4+ and CD8+ T cells, B cells, monocytes, NK cells, and granulocytes). As expected, the ‘Depressed’ subgroup had significantly higher DAWBA general risk score estimates for any psychiatric diagnosis (p < 0.001). The Validation data set 1 of 78 subjects included only female subjects. Depressed cases showed a higher BMI (p < 0.01), took more often contraceptive pills (p < 0.01) and scored higher in the DAWBA general risk score for any psychiatric diagnosis. There were no between-group differences in age, the relative proportion of white blood cell types or in the use of non-contraceptive medications (Table 1). The Validation data set 2 comprises post mortem frontal cortex samples of 29 subjects with major depression and 29 matched controls, measuring both neuronal

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Discovery cohort</th>
<th>Validation data set 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Depressed</td>
<td>15% Risk Depression</td>
</tr>
<tr>
<td>N</td>
<td>22</td>
<td>21</td>
</tr>
<tr>
<td>Men: Women (n(%))</td>
<td>(29.1):20(30.9)</td>
<td>(19.0):17(81.0)</td>
</tr>
<tr>
<td>Age (years, n(%))</td>
<td>14</td>
<td>2(9.1)</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>9(40.9)</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>11(50.0)</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>–</td>
</tr>
<tr>
<td>BMI z-score</td>
<td>Underweight (n(%))</td>
<td>1(4.5)</td>
</tr>
<tr>
<td></td>
<td>Normal weight (n(%))</td>
<td>18(81.8)</td>
</tr>
<tr>
<td></td>
<td>Overweight (n(%))</td>
<td>2(9.1)</td>
</tr>
<tr>
<td></td>
<td>Obese (n(%))</td>
<td>1(4.5)</td>
</tr>
<tr>
<td>Medications</td>
<td>None (n(%))</td>
<td>16(72.7)</td>
</tr>
<tr>
<td></td>
<td>Sleep meds (n(%))</td>
<td>2(9.1)</td>
</tr>
<tr>
<td></td>
<td>Neuroleptics (n(%))</td>
<td>1(4.5)</td>
</tr>
<tr>
<td></td>
<td>Contraceptive pills (n(%))</td>
<td>1(4.5)</td>
</tr>
<tr>
<td></td>
<td>Antixiolytic meds (n(%))</td>
<td>1(4.5)</td>
</tr>
<tr>
<td></td>
<td>ADHD meds (n(%))</td>
<td>0(0.0)</td>
</tr>
<tr>
<td></td>
<td>Antidepressives (n(%))</td>
<td>3(13.6)</td>
</tr>
<tr>
<td>DAWBA risk score for Depression (n(%))</td>
<td>Level 0 (&lt; 0.1%)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Level 1 (~0.5%)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Level 2 (~3%)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Level 3 (~15%)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Level 4 (~50%)</td>
<td>14(63.6)</td>
</tr>
<tr>
<td></td>
<td>Level 5 (&gt;70%)</td>
<td>8(36.4)</td>
</tr>
<tr>
<td>DAWBA risk score for Any Psychiatric Disorder (n(%)</td>
<td>Level 0 (&lt; 0.1%)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Level 1 (~0.5%)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Level 2 (~3%)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Level 3 (~15%)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Level 4 (~50%)</td>
<td>14(63.6)</td>
</tr>
<tr>
<td></td>
<td>Level 5 (&gt;70%)</td>
<td>8(36.4)</td>
</tr>
</tbody>
</table>

Legend: One of six probability bands for individual psychiatric disorders (including depression) were in silico generated from Development and Well-Being Assessment (DAWBA) package measurements. These bands range from less than 0,1–70% and reflect the probability that an experienced clinical rater would assign the individual a corresponding DSM-IC or ICD-10 diagnosis. Modified risk score assessments were generated by us based on these measurements into ‘Depression’ and ‘Controls’. Weight category definitions according to the Centers for Disease Control and Prevention (CDC).

Self-reported use of medications.
and glial DNA methylation profiles. Men and women were equally represented and the mean age was 32 years. Caucasians were over-represented in both study arms. Africans represented around 20% of subjects and there was one Asian subject in the depressed subgroup. There were no between-group differences in age, gender, the post mortem interval or ethnicity (Supplementary Table 1).

### 3.2. Two CpG sites in proximity to ZWIMS and MIR4646 are differentially methylated by depression risk score group

We performed multiple linear regression models of methylation M-values and depression risk group, to study the association between DNA methylation and a three-factorial risk score of depression in the Discovery data set, adjusting for the identified optimal covariates, comprising the DAWBA general risk score assessment of any psychiatric disorder, self-reported use of ADHD medication, the estimated relative proportion CD4+ T-cells and the first principal component (PC1). Coefficients and p-values shown correspond to the depression risk score factor. Adjusted p-values < 0.05 were considered significant. Abbreviations: bp, base pairs; dist, distance; logFC, log fold change; TSS, transcriptional start site.

<table>
<thead>
<tr>
<th>CpG</th>
<th>Position (bp)</th>
<th>Chromosome</th>
<th>Dist. TSS</th>
<th>Transcript</th>
<th>Gene</th>
<th>logFC</th>
<th>P. value</th>
<th>P. value (Bonf.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg13227621</td>
<td>45671360</td>
<td>1</td>
<td>889</td>
<td>NM_020883</td>
<td>ZSWIM5</td>
<td>-0.22</td>
<td>8.93E-09</td>
<td>1.44E-07</td>
</tr>
<tr>
<td>cg04102384</td>
<td>31668621</td>
<td>6</td>
<td>246</td>
<td>NR_039789</td>
<td>MIR4646</td>
<td>-0.59</td>
<td>1.13E-08</td>
<td>1.77E-07</td>
</tr>
<tr>
<td>cg07624948</td>
<td>28962846</td>
<td>16</td>
<td>530</td>
<td>NM_032815</td>
<td>NFATC2IP</td>
<td>0.32</td>
<td>1.32E-05</td>
<td>1.06E-05</td>
</tr>
<tr>
<td>cg14506141</td>
<td>98962104</td>
<td>2</td>
<td>-512</td>
<td>NM_001079878</td>
<td>CNGA3</td>
<td>-0.29</td>
<td>1.44E-06</td>
<td>1.14E-05</td>
</tr>
<tr>
<td>cg10896318</td>
<td>47516037</td>
<td>11</td>
<td>35</td>
<td>NM_198705</td>
<td>CELF1</td>
<td>-0.32</td>
<td>2.19E-06</td>
<td>1.63E-05</td>
</tr>
</tbody>
</table>

Robust multiple linear regressions of methylation to a three factorial depression score, adjusting for risk score assessment of any psychiatric disorder, ADHD medication, white blood cell coefficient CD4T and the first principal component (PC1). Coefficients and p-values shown correspond to the depression risk score factor. Adjusted p-values < 0.05 were considered significant. Abbreviations: bp, base pairs; dist, distance; logFC, log fold change; TSS, transcriptional start site.

*P-values have been adjusted to the genomic inflation factor (–1.11).
A.E. Boström et al.  
Journal of Affective Disorders 220 (2017) 117–128

with depression risk score in the Discovery data set, was con-
set 1
DAWBA score of depression, is also hypomethylated in the Validation data set with higher risk scores for depression as compared to the controls be hypomethylated in whole blood of the group of thirteen subjects fi
3.3. Methylation of cg04102384, which is associated with a modi-
DAWBA score of depression, is also hypomethylated in the Validation data set 1

Hypomethylation of cg04102384, which was found to be associated with depression risk score in the Discovery data set, was confirmed to be hypomethylated in whole blood of the group of thirteen subjects with higher risk scores for depression as compared to the controls (Validation data set 1)(Table 3., p < 0.05). Cg13227621 was not differentially methylated in relation to DAWBA.

3.4. CpG site methylation of cg04102384 is negatively correlated with expression of MIR4646-3p

To evaluate to what extent the methylation of cg04102384 is associated with the expression of the adjacent microRNA, Spearman correlations were performed using the Expression data set of eleven non-fasting healthy controls. The methylation state and the level of transcriptional miRNA expression were compared in a proportional manner to each other, assuming a negative correlation. Cg04102384 significantly inversely correlated with transcriptional levels of MIR4646-3p (p < 0.05), but not with the transcriptional expression of MIR4646-5p or pre-miRNA levels of MIR4646.

To confirm these findings, we performed Taqman analysis, individually measuring the expression levels of MIR4646-3p and MIR4646-5p in 27 subjects from the Discovery data set (eleven ‘Controls’, nine ‘15% Risk Depression’ and seven ‘Depressed’). There were no between-group differences in expression levels of either of the two microRNAs. Nor could we confirm an association between methylation and expression by Spearman’s rank correlation method for either of the two miRNAs and the candidate CpG site.

3.5. In the Validation data set 2, cg04102384 is hypomethylated in post mortem glial cell samples from the frontal cortex of subjects with major depression compared to controls

We further investigated the methylation state of MIR4646-associated CpG-site cg04102384 in the Validation data set 2, to see whether it is hypomethylated in post mortem brain samples from 29 subjects with major depression compared to 29 controls. Neuronal and glial DNA methylation profiles were studied separately. By independent samples t-tests and taking the direction into account, cg04102384 was significantly hypomethylated in the glial cell line (p < 0.01) (Supplementary Table 2.) (Fig. S1.). In the binomial logistic regression models, cg04102384 remained significantly hypomethylated after adjustments were made for age, gender, ethnicity and the post mortem interval (p = 0.0365) (Supplementary Table 3.). No association was found for cg04102384 in the neuronal cell line (data not shown).

3.6. A significant overrepresentation of hsa-miR-4646-3p putative gene targets are involved in biological processes associated with fatty acid elongation and biosynthesis of unsaturated fatty acids

Using the MirWalk2.0 analysis software (Dweep et al., 2011), we retrieved the predicted gene targets for hsa-miR-4646-3p. 552 genes were identified as putative gene targets for hsa-miR-4646-3p and were subsequently investigated by overrepresentation analysis of KEGG-defined pathways. There was a statistically significant overrepresentation of genes associated with fatty acid elongation (6 genes, q-value < 0.01) and biosynthesis of unsaturated fatty acids (5 genes, q-value < 0.05).

3.7. Hsa-miR-4646-3p is significantly correlated with ACOT2 - one of the putatively regulated target genes identified in the ‘fatty acid elongation’ pathway

In the Expression cohort, we correlated hsa-miR-4646-3p levels with the expression of 6 genes identified in the KEGG-defined ‘fatty acid elongation’ pathway. Acyl-Coa Thioesterase 2 (ACOT2) was significantly positively correlated with the expression of hsa-miR-4646-3p (p < 0.01, r = 0.75) (Supplementary Table 4)(Fig. S2). No significant associations were observed for the other 5 genes studied Tables 4 and 5.

Table 3  
Validation data set 1 analysis.

<table>
<thead>
<tr>
<th>CpG</th>
<th>Position (bp)</th>
<th>Chromosome</th>
<th>Transcript</th>
<th>Gene</th>
<th>t</th>
<th>df</th>
<th>p.val</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg13227621</td>
<td>45671360</td>
<td>1</td>
<td>NM_020883</td>
<td>ZSWIM5</td>
<td>-1.27</td>
<td>26.17</td>
<td>ns</td>
</tr>
<tr>
<td>cg04102384</td>
<td>31668621</td>
<td>6</td>
<td>NR_039789</td>
<td>MIR4646</td>
<td>2.10</td>
<td>18.84</td>
<td>2.48E-02</td>
</tr>
</tbody>
</table>

* Independent samples t-tests with the alternative hypothesis that controls have higher methylation M-levels than depressed cases. Abbreviations: bp, base pairs; coef., coefficient; df, degrees of freedom; p.val, p-value; t, t-statistic.
3.8. The CpG site next to the gene MIR4646 may have a regulatory role in brain and even modulate other genes

The investigation of the chromatin regions overlapping our CpG locus, cg04102384, revealed a regulatory role throughout all investigated brain regions, except brain cingulate gyrus. This CpG site is located in an enhancer region in blood, angular gyrus, dorsolateral prefrontal cortex, inferior temporal lobe and anterior caudate, and even in a TSS region in hippocampus and substantia nigra. Furthermore, the long-range interactions for this CpG site, represented by arcs (Fig. 3.), refer to possible regulatory effects on multiple other genes, e.g. CLIC1 and BAG6.

3.9. MIR4646 appears to be evolutionary conserved from the advent of Simiiformes (monkey)

The miRNA-4646 stem-loop region is identified only in Homo sapiens in the mirBase database and a BLASTn search in the NCBI webserver was used to investigate if this region is conserved in other species. Sixteen relevant hits were obtained with particular emphasis on the conservation of the seed regions. The conservation throughout the seed regions and the mature 5p and 3p regions are presented in the multiple sequence alignment (Fig. 4; Part A). All sixteen sequences are predicted to fold into a pre-microRNA folding structure, i.e., hairpin loop. The gene synteny (Fig. 4; Part B) is conserved among 13 of the 16 species investigated, except for variation in the number of LY6G6C/6D/6E/6F genes. Several of the species did not have an available mapped genome through the NCBI map view, or the mapping was incomplete. The reverse complement of the miRNA-4646-3p mature sequence had nine relevant hits with the miR-204 family when searched in the miRBase database. The conservation of the mature miRNA with the

Table 5
Gene set overrepresentation analysis of hsa-miR-4646-3p putative gene targets by KEGG pathways.

<table>
<thead>
<tr>
<th>KEGG Pathway</th>
<th>Set size</th>
<th>Count</th>
<th>%</th>
<th>p-value</th>
<th>q-value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatty acid elongation</td>
<td>25</td>
<td>6</td>
<td>24.0</td>
<td>3.83E−05</td>
<td>4.98E−03</td>
<td>KEGG</td>
</tr>
<tr>
<td>Biosynthesis of unsaturated fatty acids</td>
<td>23</td>
<td>5</td>
<td>21.7</td>
<td>2.88E−04</td>
<td>1.88E−02</td>
<td>KEGG</td>
</tr>
<tr>
<td>Phagosome</td>
<td>153</td>
<td>10</td>
<td>6.5</td>
<td>7.58E−03</td>
<td>3.19E−01</td>
<td>KEGG</td>
</tr>
<tr>
<td>Complement and coagulation cascades</td>
<td>69</td>
<td>6</td>
<td>8.7</td>
<td>9.80E−03</td>
<td>3.19E−01</td>
<td>KEGG</td>
</tr>
</tbody>
</table>

The online web tool MirWalk2.0 was used to computationally predict putative gene targets of hsa-miR-4646-3p. Using the online web tool ConsensusPathDB-human, 552 identified genes were investigated to see if there was a statistically significant abundance of genes involved in specific KEGG-defined pathways. Abbreviations: Count, number of candidate genes in a particular pathway; %, percentage of candidate genes in a particular pathway.

3.8. The CpG site next to the gene MIR4646 may have a regulatory role in brain and even modulate other genes

The investigation of the chromatin regions overlapping our CpG locus, cg04102384, revealed a regulatory role throughout all investigated brain regions, except brain cingulate gyrus. This CpG site is located in an enhancer region in blood, angular gyrus, dorsolateral prefrontal cortex, inferior temporal lobe and anterior caudate, and even in a TSS region in hippocampus and substantia nigra. Furthermore, the long-range interactions for this CpG site, represented by arcs (Fig. 3.), refer to possible regulatory effects on multiple other genes, e.g. CLIC1 and BAG6.

3.9. MIR4646 appears to be evolutionary conserved from the advent of Simiiformes (monkey)

The miRNA-4646 stem-loop region is identified only in Homo sapiens in the mirBase database and a BLASTn search in the NCBI webserver was used to investigate if this region is conserved in other species. Sixteen relevant hits were obtained with particular emphasis on the conservation of the seed regions. The conservation throughout the seed regions and the mature 5p and 3p regions are presented in the multiple sequence alignment (Fig. 4; Part A). All sixteen sequences are predicted to fold into a pre-microRNA folding structure, i.e., hairpin loop. The gene synteny (Fig. 4; Part B) is conserved among 13 of the 16 species investigated, except for variation in the number of LY6G6C/6D/6E/6F genes. Several of the species did not have an available mapped genome through the NCBI map view, or the mapping was incomplete. The reverse complement of the miRNA-4646-3p mature sequence had nine relevant hits with the miR-204 family when searched in the miRBase database. The conservation of the mature miRNA with the

Table 4
Methylation/transcription correlations of cg04102384 and hsa-mir-4646 associated probes.

<table>
<thead>
<tr>
<th>CpG Position (bp)</th>
<th>Chromosome</th>
<th>Transcript</th>
<th>Gene</th>
<th>probe-ID</th>
<th>microRNA</th>
<th>Spearman’s rho statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg04102384</td>
<td>31668621</td>
<td>6</td>
<td>NR_039789</td>
<td>hsa-mir-4646</td>
<td>−0.61</td>
<td>2.59E−02</td>
</tr>
<tr>
<td>cg04102384</td>
<td>31668621</td>
<td>6</td>
<td>NR_039789</td>
<td>hsa-mir-4646</td>
<td>−ns</td>
<td>−ns</td>
</tr>
</tbody>
</table>

* Analysis of spearman’s rho statistic of CpG-site methylation and microRNA expression levels. Abbreviations: bp, base pairs; Coef., spearman’s rho; p.val, p-value.
A. E. Boström et al.  
Journal of Affective Disorders 220 (2017) 117–128

A

H. sapiens
G. gorilla
P. troglodytes
P. paniscus
P. abelii
N. leucogenys
P. abeles
C. ays
M. leucophaeus
M. mulatta
C. drita
C. angolensis
C. capucinus
S. bolivienis
C. pacus
A. nancyymae

<table>
<thead>
<tr>
<th>mature 3p</th>
<th>seed</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

B

mature 5p
seed

miR-4646

CLIC1  DDAH2  C(B)orf47  LY6G6C/6D/6E/6F  ABHD16A  LY6G5/5C  CSNK2B  GPANK1  C(B)orf47  APOM

Fig. 4. Evolutionary conservation of MIR4646. Part A. Conservation of genomic DNA of pre-miRNA-4646 with the mature 5p and 3p regions and the seed positions (2–8) highlighted above the multiple sequence alignment. As mir-4646 is identified only in humans in miRBase, the genomic region was investigated in other species and studied to see if this could be a conserved miRNA. A BLAST search obtained sixteen relevant hits and the MAFFT alignment server (v7) was used to generate the multiple sequence alignment. The multiple sequence alignment includes species within the Simiiformes group. In addition to the sequence alignment, all of the sequences are predicted to fold into a pre-miRNA secondary structure, suggesting that this region may be conserved from Simiiformes (monkeys). Part B. Gene synteny in surrounding genomic region of investigated species. In at least thirteen of the sixteen investigated species, the gene synteny appears to be preserved except for variation in the number of LY6G6C/6D/6E/6F genes, implying that the general structure of this region has been conserved.

We detect hypomethylation of cg04102384 – a methylation locus located within 246 bp of the transcriptional start site for MIR4646 - and the risk of depression in adolescents. By further investigating associations between methylation and expression in healthy individuals, we provide evidence that the level of methylation at this CpG site regulates expression of MIR4646-3p in whole blood, underlying the functional importance of this locus. Our functional analysis revealed that the MIR4646 associated CpG site is located in an enhancer region of the gene throughout all investigated brain tissues, as well as in peripheral blood cells (see Fig. 3). These homogenously findings for both brain tissue and blood are of importance as it underlines the applicability/transferability of epigenetic findings made in blood on changes putatively also abundant in brain in addition to the correlations between blood and brain methylomes (Hannon et al., 2015; Nikolova et al., 2014). The fact that the MIR4646 associated CpG site is also shown by us to be hypomethylated in the brain samples of MDD patients further supports this claim. Moreover, in the MDD context, the whole blood methylation analysis may be relevant (Walker et al., 2016) particularly for the multiple pathophysiological pathways associated with this disease, e.g. inflammatory (Zunszain et al., 2013), immune (Blume et al., 2011) and metabolic processes (Vogelzangs et al., 2014).

Strikingly, MIR4646-3p appears to play a role in the neurobiological context, as detected by KEGG pathway annotation. In our study, we identify that MIR4646-3p is associated with the elongation of fatty acids. Importantly, this process mediates the conversion of omega-3 fatty acids, e.g. the eicosapentaenoic (EPA) and docosahexaenoic (DHA). Essential omega-3 fatty acids cannot be synthesized by the organism and have to be taken up by the diet, from sources as fish, eggs and soy products. They were shown to be involved in brain functioning, particularly in memory and cognitive performance (Ruxton et al., 2004). Interestingly, omega-3 fatty acids appear to be implicated in MDD pathophysiology (Hibbeln, 1998; Logan, 2004; Mamalakis et al., 2002; Peet et al., 1998). A study showed that pretreated mice with fish oil have developed less depressive symptoms through the suppression of neuroinflammation (Numata et al., 2015). Grayson et al. revealed in neuroimaging studies that lack of omega-3 implicate functional dysregulation at prefrontal cortical connectivity in monkeys (Grayson et al., 2014). Notably, while it was shown that adolescents with MDD have a decreased frontal white matter integrity and worse connectivity within frontal lobe cortical networks (Connolly et al., 2013; Ho et al.,

Fig. 5. Alignment of MIR4646-3p and MIR204-3p. Alignment of the Homo sapiens reverse complement (RC) miRNA-4646-3p mature sequence with the miRNA-204-3p family. The reverse complement miRNA-4646-3p sequence aligns to the mir-204-3p family. The reverse complement of miRNAs are speculated to be able to work in subtle regulatory activity and have some functional activity (Shao et al., 2012). The mature miRNA-4646-3p sequence from miRBase was searched against the miRBase database (release 21) and aligned using the MAFFT alignment server (v 7).

Discussion

Using a methylome -wide study approach, we identified two genes, ZSWIM5 and MIR4646, to be differentially methylated in whole blood of 93 adolescents in association with a modeled risk score for MDD. To the best of our knowledge, this is the first study investigating genome-wide DNA methylation shifts in relationship to MDD in a population-based cohort of adolescents (Januar et al., 2015b). We were able to validate our finding in an independent data set of 78 adolescents and could also show that the methylation drift of the MIR4646 associated CpG locus was inversely correlated with the degree of gene expression in healthy individuals. In a second independent data set, the identified methylation locus was further shown to be differentially methylated in post mortem glial cells from the frontal cortex of adult subjects diagnosed with major depression compared to matched controls. Thus, we detect miR-4646 as a novel marker for risk of MDD in adolescence, a critical period for environmental experience, and could add value in planning the preventive and control strategies for adulthood MDD.

seed region emphasized in the multiple sequence alignment (Fig. 5).

We detect hypomethylation of cg04102384 – a methylation locus located within 246 bp of the transcriptional start site for MIR4646 - and the risk of depression in adolescents. By further investigating associations between methylation and expression in healthy individuals, we provide evidence that the level of methylation at this CpG site regulates expression of MIR4646-3p in whole blood, underlying the functional importance of this locus. Our functional analysis revealed that the MIR4646 associated CpG site is located in an enhancer region of the gene throughout all investigated brain tissues, as well as in peripheral blood cells (see Fig. 3). These homogenous findings for both brain tissue and blood are of importance as it underlines the applicability/transferability of epigenetic findings made in blood on changes putatively also abundant in brain in addition to the correlations between blood and brain methylomes (Hannon et al., 2015; Nikolova et al., 2014). The fact that the MIR4646 associated CpG site is also shown by us to be hypomethylated in the brain samples of MDD patients further supports this claim. Moreover, in the MDD context, the whole blood methylation analysis may be relevant (Walker et al., 2016) particularly for the multiple pathophysiological pathways associated with this disease, e.g. inflammatory (Zunszain et al., 2013), immune (Blume et al., 2011) and metabolic processes (Vogelzangs et al., 2014).

Strikingly, MIR4646-3p appears to play a role in the neurobiological context, as detected by KEGG pathway annotation. In our study, we identify that MIR4646-3p is associated with the elongation of fatty acids. Importantly, this process mediates the conversion of omega-3 fatty acids, e.g. the eicosapentaenoic (EPA) and docosahexaenoic (DHA). Essential omega-3 fatty acids cannot be synthesized by the organism and have to be taken up by the diet, from sources as fish, eggs and soy products. They were shown to be involved in brain functioning, particularly in memory and cognitive performance (Ruxton et al., 2004). Interestingly, omega-3 fatty acids appear to be implicated in MDD pathophysiology (Hibbeln, 1998; Logan, 2004; Mamalakis et al., 2002; Peet et al., 1998). A study showed that pretreated mice with fish oil have developed less depressive symptoms through the suppression of neuroinflammation (Numata et al., 2015). Grayson et al. revealed in neuroimaging studies that lack of omega-3 implicate functional dysregulation at prefrontal cortical connectivity in monkeys (Grayson et al., 2014). Notably, while it was shown that adolescents with MDD have a decreased frontal white matter integrity and worse connectivity within frontal lobe cortical networks (Connolly et al., 2013; Ho et al.,
 involve to the mill that the epigenetic contribution to psychopathologies in estimated e, Chhetry et al. found that A.E. Boström et al. face meeting of a clinician and the child, however, it cannot be obtained in a face-to-face meeting of a clinician and the child, however, it cannot be completely excluded that an individual subject received an unrepres- sentative risk score assessment. Our study is also restricted by the limited number of individuals with higher risk-score assessments of MDD in both discovery (n=22) and Validation data set 1 (n=13). Future studies including a high number of individuals would be of great value to confirm the results presented. Furthermore, in the Validation data set 1, there were between-group differences in BMI z-score defined weight categories and in the use of contraceptive pills. Due to power issues, we were not able to adjust for these factors on the confirmatory association analysis between DNA methylation at cg04102384 and depression risk group. As such, this represents a limitation which could be a source of potential bias. Furthermore, there was a predominance of female subjects in both Discovery and Validation data set 1. However, as there were no between-group differences in gender between the three depression subgroups in the discovery data set, it is unlikely that there is a gender biasness in the methylene-wide analysis. This assumption is further supported by the findings in Validation data set 2, showing an equal distribution of men and women and no significant impact of gender on the association between methylation and MDD. An additional point which needs to be addressed is the findings in the Validation data set 2, where cg04102384 was significantly hypomethyl- ated only in the glial cells but not in the neuronal cell lines. This finding is consistent with previous studies suggesting significant brain cellular heterogeneity that bias cell type specific DNA methylation patterns (Guinivan et al., 2013).

Previous genome-wide studies showed evidence regarding altered DNA methylation in depression disorder, but mainly in monozygotic twin pairs, which are not reflecting the normal population (Byrne et al., 2013; Cordova-Palmera et al., 2015; Dempster et al., 2014). In our study, we could take into account both genetics and environmental components as an interconnected biological network, by investigating a population-based cohort. Moreover, the advantages of studying an adolescent population at risk of MDD instead of investigating individ- uals showing already clinical symptoms are numerous. Firstly, as a childhood episode of MDD substantially increases the risk of adverse outcomes in adulthood (Fergusson et al., 2007), it calls attention for early detection of MDD as a marker for severity of the underlying vulnerability. Studying a cohort at risk for MDD may thus provide an excellent platform for identifying epigenetic susceptibility factors that hold true for depressive disorders in general. Secondly, it can be argued that investigating developmental stages of psychopathology allows better isolating of causative factors for disease pathology, whereas analyses performed on adults that have had psychiatric problems for a long time are more likely to be burdened by confounding factors, e.g. antidepressants, alcohol or substance abuse, smoking, cardiovascular complications or even aging, all of which may influence methylation and could produce false outcomes (Breitling et al., 2011; Menke and Binder, 2014; Zampieri et al., 2015).

While DNA methylation is the best characterized mechanism of epigenetic regulation (Novik et al., 2002), different epigenetic mechanisms may act co-dependently and participate in cross-regulating activities (Feil and Fraga, 2012; Iorio et al., 2010; Sun et al., 2013). Klenge et al. suggested that a better understanding of DNA methylation and its link to other epigenetic mechanisms will be important for understanding the epigenetic contribution to psychopathologies (Klenge et al., 2014). In this respect, our findings provide further grist to the mill that the epigenetic contribution to psychopathologies involve co-dependent and cross-regulatory processes of DNA methyla- tion and miRNA activity.

The DAWBA bands have been shown to be similar or identical to clinician-rated diagnoses in estimated effect sizes, significance levels and substantive conclusions regarding risk factor associations (Goodman et al., 2011). As the scores were not obtained in a face-to-face meeting of a clinician and the child, however, it cannot be completely excluded that an individual subject received an unrepres-
Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jad.2017.05.017.

References


