

# Epigenome-Wide Association Study of Wellbeing

Bart M. L. Baselmans,<sup>1,2</sup> Jenny van Dongen,<sup>1,2</sup> Michel G. Nivard,<sup>1</sup> Bochao D. Lin,<sup>1</sup> BIOS Consortium,<sup>3</sup> Nuno R. Zilhão,<sup>1</sup> Dorret I. Boomsma,<sup>1,2,4</sup> and Meike Bartels<sup>1,2,4</sup>

<sup>1</sup>Department of Biological Psychology, VU University, Amsterdam, the Netherlands

<sup>2</sup>EMGO<sup>+</sup> Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands

<sup>3</sup>The Biobank-Based Integrative Omics Study (BIOS) Consortium

<sup>4</sup>Neuroscience Campus Amsterdam, Amsterdam, the Netherlands

Wellbeing (WB) is a major topic of research across several scientific disciplines, partly driven by its strong association with psychological and mental health. Twin-family studies have found that both genotype and environment play an important role in explaining the variance in WB. Epigenetic mechanisms, such as DNA methylation, regulate gene expression, and may mediate genetic and environmental effects on WB. Here, for the first time, we apply an epigenome-wide association study (EWAS) approach to identify differentially methylated sites associated with individual differences in WB. Subjects were part of the longitudinal survey studies of the Netherlands Twin Register (NTR) and participated in the NTR biobank project between 2002 and 2011. WB was assessed by a short inventory that measures satisfaction with life (SAT). DNA methylation was measured in whole blood by the Illumina Infinium HumanMethylation450 BeadChip (HM450k array) and the association between WB and DNA methylation level was tested at 411,169 autosomal sites. Two sites (cg10845147,  $p = 1.51 \times 10^{-8}$  and cg01940273,  $p = 2.34 \times 10^{-8}$ ) reached genome-wide significance following Bonferroni correction. Four more sites (cg03329539,  $p = 2.76 \times 10^{-7}$ ; cg09716613,  $p = 3.23 \times 10^{-7}$ ; cg04387347,  $p = 3.95 \times 10^{-7}$ ; and cg02290168,  $p = 5.23 \times 10^{-7}$ ) were considered to be genome-wide significant when applying the widely used criterion of a FDR  $q$  value  $< 0.05$ . Gene ontology (GO) analysis highlighted enrichment of several central nervous system categories among higher-ranking methylation sites. Overall, these results provide a first insight into the epigenetic mechanisms associated with WB and lay the foundations for future work aiming to unravel the biological mechanisms underlying a complex trait like WB.

■ **Keywords:** wellbeing, satisfaction with life, DNA methylation, epigenetics, EWAS, 450k, twins

Because of its strong associations, in individuals and in society, with physical and mental health as well as economic development, WB has become a topic of interest across different scientific disciplines (Seaford, 2011; Steptoe et al., 2014; Stiglitz et al., 2009). In general, WB is conceptualized to include a continuous spectrum of positive feelings and subjective life assessment that can be assessed with a series of measures, such as SAT, happiness (HAP), and quality of life (QoL).

Twin-family studies report that in the general population, a substantial part of the variation in the different measures of WB is explained by genetic differences between individuals (Stubbe et al., 2005). A large meta-analysis (Bartels, 2015) showed that the weighted average heritability for WB was 36% (95% CI: 34–38, and for SAT 32% (95% CI: 29–35). A multivariate twin-sibling study exploring the etiology of the covariance among multiple WB indices revealed that the genetic variance in the different measures was explained by one underlying set of genes (Bartels &

Boomsma, 2009). These results also emphasize the importance of environmental factors in the variation of WB, and a dynamic interplay between genes and environment is to be expected for a complex trait like WB.

Epigenetic regulation of gene expression by mechanisms such as DNA methylation may mediate the interplay between the genetic make-up of individuals and their exposure to the environment (Bernstein et al., 2007; Rakyan et al., 2011; Reik, 2007). Methylation changes can be caused by external conditions, such as long-term stress exposure (Klengel et al., 2014; Romens et al., 2015), (prenatal, maternal) smoking exposure (Allione et al., 2015; Lee et al., 2015)

RECEIVED 11 September 2015; ACCEPTED 21 October 2015. First published online 1 December 2015.

ADDRESS FOR CORRESPONDENCE: Bart Baselmans, Department of Biological Psychology, VU University Amsterdam, Van der Boerchorststraat 1, 1081BT, Amsterdam, the Netherlands. E-mail: [b.m.l.baselmans@vu.nl](mailto:b.m.l.baselmans@vu.nl)

and dietary modifications at conception (Dominguez-salas et al., 2014). There are no epigenetic studies of the association between DNA methylation and WB, but some epigenetic studies have been performed involving complex traits related to WB. From twin studies, it is known that there is a negative association between psychopathology and WB, with the strongest association between WB and anxiety/depression (Bartels et al., 2013). Epigenetic differences in candidate genes related to major depressive disorder (MDD) have been reported in multiple studies (Dalton et al., 2014; Oh et al., 2015). Additionally, a DNA methylation EWAS of monozygotic twins discordant for adolescent depression (Dempster et al., 2014) found two reproducible differentially-methylated probes (DMPs) that were located within *STK32C*, which encodes a serine/threonine kinase of unknown function.

Here, we describe the first EWAS for WB performed in a population-based sample from the NTR. Our aim was to identify genomic locations where differences in DNA methylation in blood level are associated with differences in WB.

## Methods

### Subjects and Samples

The subjects in this EWAS participated in longitudinal survey studies conducted by the NTR (Van Beijsterveldt et al., 2013; Willemsen et al., 2013) and in the NTR biobank project (Willemsen et al., 2010). Peripheral blood samples were drawn from the NTR participants in the morning after an overnight fast, and for biomarkers studies and for DNA and RNA isolation (see Willemsen et al. (2013)). In 3,264 peripheral blood samples from 3,221 participants, genome-wide methylation probes were assessed. After quality control (QC) of the methylation data, 3,089 samples from 3,057 participants were retained (for a detailed description of the QC procedures, see van Dongen, 2015, this issue). For the present study, we included participants if the following information was available: SAT score, and good quality methylation data and data on white blood cell counts, leaving 2,519 samples from 2,456 subjects for the final analyses. The dataset included 606 complete MZ and 291 complete DZ pairs, 102 fathers of twins, 112 mothers of twins, 15 siblings, and 2 spouses of twins.

All subjects provided written informed consent and study protocols were approved by the Central Ethics Committee on Research, involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB-2991 under Federal wide Assurance-3703: IRB/institute codes, NTR 03-180).

### Wellbeing

WB was assessed by a short inventory that measures SAT (Diener et al., 1985). Data on SAT were collected in multi-

**TABLE 1**

**Satisfaction with Life Scale (Diener et al., 1985)**

Item
SAT item 1 In most ways my life is close to ideal
SAT item 2 The conditions of my life are excellent
SAT item 3 I am satisfied with my life
SAT item 4 So far I have gotten the most important things I want in life
SAT item 5 If I could live my life over, I would change almost nothing

ple NTR surveys. For the current study, data from surveys 6 (2002), 8 (2009), and 10 (2013) were analyzed. The SAT scale consists of five items that have to be answered on a 7-point scale ranging from 1 = *strongly disagree* to 7 = *strongly agree*. A typical question included in this questionnaire is 'If I could live my life over, I would change almost nothing' (for all items, see Table 1). Internal consistency of the scale was good with a Chronbach's alpha of 0.86 and test-retest scores in the range of 0.24 (over 16 years), to 0.54 (over 4 years) to 0.84 for a period of 2 weeks to 1 month (Fujita & Diener, 2005). Within this NTR sample, the test-retest scores were 0.53 between surveys 6 and 8 (7-year interval), 0.48 between surveys 6 and 10 (11-year interval), and 0.63 between surveys 8 and 10 (4-year interval), and the phenotypic correlation of SAT with an overall WB factor score was 0.97. For individuals who completed surveys 6, 8, and 10, the WB score closest to the moment of blood draw was selected.

### Infinium HumanMethylation450 BeadChip Data

DNA methylation was assessed with the Infinium HumanMethylation450 BeadChip Kit (Illumina, Inc.) (Bibikova et al., 2011). 500ng of genomic DNA from whole blood was treated by bisulfite using the Zymo Research EZ DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA) following the standard protocol for Illumina 450K microarrays, by the department of Molecular Epidemiology from the Leiden University Medical Center (LUMC), the Netherlands. Subsequent steps (i.e., sample hybridization, staining, scanning) were performed by the Erasmus Medical Center micro-array facility, Rotterdam, the Netherlands. QC and processing of the blood methylation dataset has been described in detail previously (Van Dongen et al., 2015). A number of sample-level and probe-level quality checks were performed. Sample-level QC was performed using MethyIAid (van Iterson et al., 2014). Probes were set to missing in a sample if they had an intensity value of exactly zero, or a detection  $p$  value  $> .01$ , or a bead count  $< 3$ . After these steps, probes that failed based on the above criteria in  $> 5\%$  of the samples were excluded from all samples (only probes with a success rate  $\geq 0.95$  were retained). Probes were also excluded from all samples if they mapped to multiple locations in the genome (Chen et al., 2013), or if they had a SNP within the CpG site (at the C or G position) irrespective of minor allele frequency in the Dutch population

(Genome of the Netherlands Consortium, 2014). Only autosomal methylation sites were analyzed in the EWAS. The methylation data were normalized with functional normalization (Fortin, 2014), and normalized intensity values were converted into beta ( $\beta$ ) values. The  $\beta$  value represents the methylation level at a site, ranging from 0 to 1, and is calculated as:

$$\beta = \frac{M}{M + U + 100}$$

where  $M$  = Methylated signal,  $U$  = Unmethylated signal, and 100 represents a correction term to control the  $\beta$  value of probes with very low overall signal intensity. After QC and normalization, principal component analysis was conducted on genome-wide methylation sites.

### Covariates

White blood cell percentages were included as covariates in the EWAS to account for variation in cellular composition between whole blood samples. The following subtypes of white blood cells were counted in blood cells: neutrophils, lymphocytes, monocytes, eosinophils, and basophils (Willemsen et al., 2013). Because of its strong correlation with neutrophil counts ( $r = -0.93$ ), lymphocyte counts were not included in the model, while basophil counts were not included because they showed little variation between subjects (many subjects having 0% of basophils in their blood). Inspection of the PCs that were computed on the genome-wide methylation data indicated that PC1 reflected, as expected, sex ( $r = 0.99$ ), PC2 showed a strong correlation with lymphocyte percentage ( $r = -0.8$ ) and neutrophil percentage ( $r = 0.79$ ). Additionally, PC3 showed a modest correlation with age ( $r = -0.41$ ) and a weak correlation with white blood cell percentages (absolute  $r \sim 0.1$ ). However, it is possible that this PC is reflective of unmeasured white blood cell subtypes and was therefore included in the model. Because of their correlation with several lab procedures, such as sample plate and order of processing, PC4 and PC5 were included to account for technical variability (for a graphical representation of the included PCs, see Supplemental Figure S1).

### Epigenome-Wide Association Analysis

Using generalized estimation equation (GEE) models, we tested whether DNA methylation was associated with WB for each methylation site, with DNA methylation  $\beta$  value as outcome variable and the following predictors: WB score, sex, age at blood sampling, age squared, monocyte percentage, eosinophil percentage, neutrophil percentage, HM450k array row, and principal components (PCs) 3, 4, and 5 from the methylation data. Age squared was included as a covariate as several studies suggest a U-shaped relationship between WB and age, with the lowest point approximately in midlife (Blanchflower & Oswald, 2008; Layard, 2010). GEE models were fitted with the R package `gee`, with the

following specifications: Gaussian link function (for continuous data), 100 iterations, and the ‘exchangeable’ option to account for the correlation structure within families and within persons.

FDR  $q$  value was computed with the R package `qvalue` with default settings. The genomic inflation factor ( $\lambda$ ) was calculated with the default regression method from the R package `GenABEL`. In all analyses, an FDR  $q$  value  $< 0.05$  was considered statistically significant (Benjamini & Hochberg, 1995). Additionally, a more stringent Bonferroni correction was applied by dividing 0.05 by the number of observations ( $N = 411,169$ ). Follow-up analyses, including a test for enrichment of genomic locations and gene ontologies, were performed based on the output from the EWAS.

### Genomic Annotation

As described in detail by Sliker et al. (2013), methylation sites were mapped to genomic features and DNase I hypersensitive sites (DHS). These genomic features consists of five gene-centric regions: (1) intergenic regions ( $> 10$  kb to  $-1.5$  kb from the nearest transcription start site [TSS]), (2) proximal promotor ( $-1.5$  kb to  $+500$  kb from the nearest TSS), (3) distal promotor ( $-10$  kb to  $-1.5$  kb from the nearest TSS), (4) gene body ( $+500$ bp to  $3'$  end of the gene) and (5) downstream region ( $3'$  end to  $+5$  kb from  $3'$  end). Additionally, CpG were mapped to CG island (CGIs; CG content  $> 50\%$ , length  $> 200$ bp, and observed/expected ratio of CpGs  $> 0.6$ ), CGI shore (2 kb region flanking CGI), CGI shelf (2 kb region flanking CGI shore), or non-CGI regions. Locates were obtained from the UCSC genome browser (Kent et al., 2002). DHS locations, mapped by the ENCODE project (The ENCODE Project Consortium, 2012) were also downloaded from the UCSC genome browser (Kent et al., 2002).

### Enrichment of Genomic Locations

To test whether specific genomic locations showed a stronger association between DNA methylation and WB, eight categories were tested for being enriched using the EWAS test statistics for all genome-wide methylation sites. The locations tested were: (1) gene body, (2) proximal promoter, (3) distal promoter, (4) downstream region, (5) CGI, (6) CGI shore, (7) CGI shelf, and (8) DHS. To account for differences in variability between methylation sites we also included the mean and standard deviation of DNA methylation level in the model as covariates. For a detailed description of the method used for this analysis, see also Van Dongen et al. (2015), published in this issue.

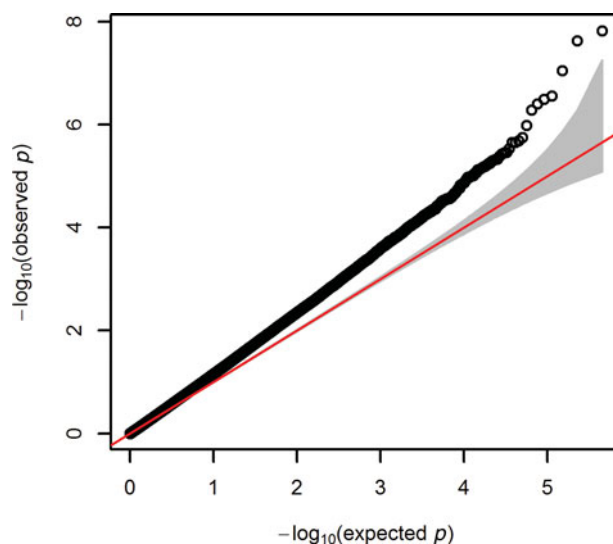
### Enrichment of the Gene Ontology Terms

Methylation sites with a stronger association with WB were tested for enrichment of GO terms. To do so, all methylation sites that were tested were ranked by EWAS  $p$  value and the resulting ranked gene list was supplied to the online software tool GOrilla (Eden et al., 2009). GOrilla performs GO enrichment analyses based on gene rank, and therefore no  $p$

**TABLE 2**  
Characteristics of the Wellbeing Data

Data	N	Wellbeing					Age at survey			Age at blood sampling		
		Mean	Median	SD	Min	Max	Mean	Median	SD	Mean	Median	SD
Survey 6	10087	26.62	28	5.26	5	35	39.44	34.00	14.35			
Survey 8	19746	27.22	29	5.32	5	35	40.41	39.94	16.32			
Survey 10	11604	26.80	28	5.24	5	35	44.42	47.00	17.49			
EWAS <sup>a</sup>	2519	27.02	29	5.46	5	35	38.30	35.33	13.53	36.79	33.10	13.01

Note: <sup>a</sup>Includes individuals with 450k methylation data and data on white blood cell counts.

**FIGURE 1**

(Colour online) Quantile–quantile (QQ) plot from the EWAS of wellbeing.

Note: The observed  $p$  values (y-axis) are plotted against the  $p$  values expected under the null hypothesis (x-axis). The straight diagonal line denotes the pattern that is expected under the null hypothesis, with 95% confidence intervals indicated by the shaded grey area.

value cut-off for defining the input list of genes is required. The background in this analysis consisted of all genes for which methylation sites were analyzed in the EWAS. In all analyses, we accounted for multiple testing by controlling the false discovery rate (FDR). An FDR  $q$  value  $< 0.05$  was considered statistically significant.

## Results

### Characteristics of the Study Sample

Data on WB were available for 1,747 individuals who filled out survey 6, 2,056 individuals who filled out survey 8, and 1,059 individuals who filled out survey 10. The EWAS was performed on 2,519 blood samples from 2,456 subjects (mean age at blood sampling = 36.8 years,  $SD = 13\%$  male = 31.1), for which the WB score closest to the moment of blood draw was selected: For 1,799 samples, WB was assessed after blood draw (mean 3.1 years), and for 720 samples, WB was assessed before blood drawn (mean 2.5 years). Table 2 summarizes the characteristics of the WB data and EWAS study sample. The average WB score of the EWAS study sample was comparable to the averages of the different survey waves (mean EWAS study = 27.0, mean

survey 6 = 26.8, mean survey 8 = 27.5, mean survey 10 = 27.0), and was also comparable to the mean of the overall NTR survey database SAT score (26.9,  $n = 38,740$ ). (For a distribution of the WB data, see Supplementary Figure S1).

### EWAS

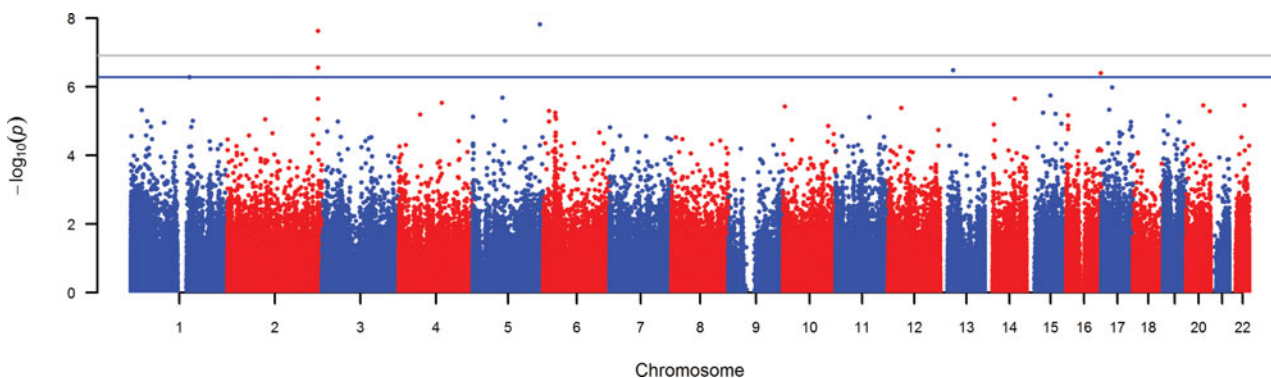
After QC, we tested 411,169 autosomal sites in the genome for their association between DNA methylation and WB score, while correcting for white blood cells counts, age at blood sampling, age squared, sex, array row and three PCs from the methylation data. Figure 1 shows the Quantile–Quantile (QQ) plot. The genomic inflation factor ( $\lambda$ ) was 1.227.

Two of the methylation sites reached the genome wide significant threshold of  $p = 1.22 \times 10^{-7}$  when Bonferroni corrected and six of the methylation sites reached genome significance when using a threshold of FDR  $q$  value  $< 0.05$  ( $p = 5.23 \times 10^{-7}$ ; Figure 2). The highest ranking methylation site was cg10845147 (UCSC Genome Browser hg19 assembly;  $p = 1.51 \times 10^{-8}$ ), located on chromosome 5: 172149624, which was negatively associated with WB (Figure 3a).

**TABLE 3**  
Top-Ranking CpG Sites From the EWAS of Wellbeing

CpG site	Chr	Position (hg19)	(Nearest) gene name	Distance gene from each site (bps)	Mean methylation level <sup>a</sup>	SD methylation level <sup>a</sup>	Estimate <sup>b</sup>	Robust SE <sup>c</sup>	p value
cg10845147	5	172149624	<i>DKFZP761M1511</i>	491642	0.69	0.04	-0.00072	1.27 e <sup>-04</sup>	1.51 e <sup>-08</sup>
cg01940273	2	233284934	<i>ALPPL2</i>	874220	0.66	0.05	0.00101	1.80 e <sup>-04</sup>	2.33 e <sup>-08</sup>
cg03329539	2	233283329	<i>ALPPL2</i>	874220	0.45	0.05	0.00065	1.27 e <sup>-04</sup>	2.76 e <sup>-07</sup>
cg09716613	13	33000534	<i>CG018</i>	572356	0.27	0.04	-0.00052	1.02 e <sup>-04</sup>	3.23 e <sup>-07</sup>
cg04387347	16	88537187	<i>ZFPM1</i>	171	0.22	0.05	-0.00075	1.48 e <sup>-04</sup>	3.94 e <sup>-07</sup>
cg02290168	1	151255971	<i>ZNF687</i>	25647	0.12	0.02	0.00030	5.98 e <sup>-05</sup>	5.23 e <sup>-07</sup>

Note: Top hits from the EWAS for the association between methylation and wellbeing. <sup>a</sup>Mean and standard deviation of the methylation proportion ( $\beta$ -value) in the entire 450K cohort. <sup>b</sup>Estimate from the regression of methylation proportion on wellbeing score. <sup>c</sup>Robust standard error of the estimate (accounting for the clustering of observations within families).

**FIGURE 2**

(Colour online) Manhattan plot showing the  $p$  values for the association between wellbeing and DNA methylation level at genome-wide autosomal sites.

Note: The horizontal grey line represents the Bonferroni-adjusted  $p$  value threshold. The blue horizontal line represents the FDR  $q$  value  $< 0.05$ .

The nearest gene associated with this site is the *DKFZP761M1511* gene (synonym is *NEURL1B*). This gene spans 50,274 bps of chromosome 5 and ranges from 172641266 to 172691540. The other site reaching Bonferroni genome-wide significance is cg01940273 ( $p = 2.34 \times 10^{-8}$ ), located on chromosome 2:233284934, which showed a positive relationship between DNA methylation and WB (Figure 3b). The gene closest located to this site is the *ALPPL2* gene, which ranges from chromosome 2:232406843 to 2: 232410714 (3871 bps) (Supplementary Figure S2). The four additional CpG sites that were genome-wide significant using FDR  $q$  value  $< 0.05$  are cg03329539 ( $p = 2.76 \times 10^{-7}$ , chromosome 2), cg09716613 ( $p = 3.23 \times 10^{-7}$ , chromosome 13), cg04387347 ( $p = 3.95 \times 10^{-7}$ , chromosome 16), and cg02290168 ( $p = 5.23 \times 10^{-7}$ , chromosome 1). The significant CpG sites located on chromosomes 1 and 2 were positively associated with WB, whereas the CpG sites located on chromosomes 13 and 16 were negative associated with WB. Characteristics of genome-wide significant methylation sites as well as the location of the nearest genes are provided in Table 3.

Next, we looked at the association with WB for all CpGs in relatively close proximity (within 10 kb) of each significant CpG site (for an overview of all CpG sites located within

this window and their  $p$  value and regression coefficient, see Supplementary Table S1). For the highest ranked CpG site (cg10845147, chromosome 5), CpG site cg07853407 was located closest at 2563 bp. This site showed no association with WB ( $p = 8.17 \times 10^{-1}$ ,  $\beta = -1.22 \times 10^{-5}$ ). The two genome-wide significant CpG sites on chromosome 2 were located within 1,606 bp from each other. Within this window, five additional probes were measured. All of these probes showed a change in methylation in the same direction of association with WB (see Supplementary Table S1). On chromosome 13, CpG site cg12054869 was located closest to the significant CpG site (cg09716613) at 716 bp, while for chromosome 16, the CpG site located most closely to the leading CpG site cg04387347 was located 73 bp away. Finally, the CpG site located closest to the leading CpG site at chromosome 1 was cg01062937, 668 bp away. For each of these probes, the regression coefficient for WB indicated a similar direction of effect as the significant probe in that specific region.

#### Enrichment of Genomic Locations

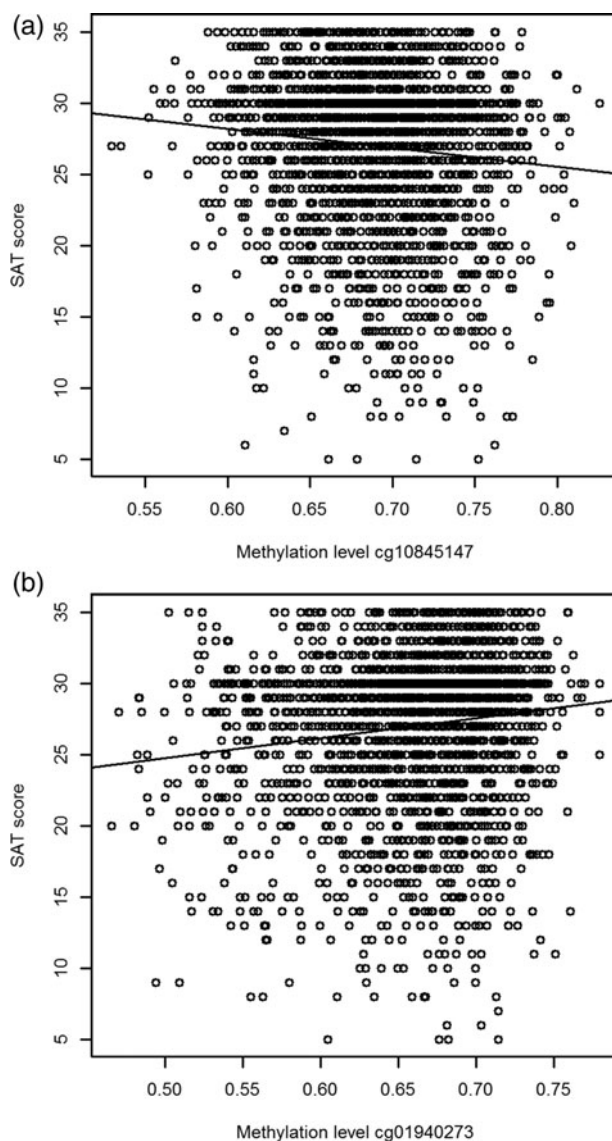
Table 4 shows the results of the regression of the EWAS test statistics on annotation categories across all genome-wide sites. Enrichment of signal was seen in the gene body ( $p =$

**TABLE 4**

Results From the Regression of EWAS Test Statistics on Genomic Annotation Categories

Regression parameter	Estimate	SE	t value	p value
Intercept	1.110	0.012	85.59	0
Gene body	0.035	0.008	4.35	$1.34 \times 10^{-5}$
Proximal promotor	0.077	0.009	8.63	$6.01 \times 10^{-18}$
Distal promotor	0.009	0.014	0.60	.54
Downstream region	0.022	0.021	1.05	.29
DNase I hypersensitive site (DHS)	0.048	0.006	7.57	$3.68 \times 10^{-14}$
CGI shore	0.048	0.008	6.12	$9.26 \times 10^{-10}$
CGI shelf	0.002	0.010	0.23	.82
CpG island	0.001	0.008	0.06	.94
Mean methylation level <sup>a</sup>	-0.086	0.011	-7.75	$9.13 \times 10^{-15}$
SD methylation level <sup>b</sup>	3.57	0.182	19.64	$6.63 \times 10^{-36}$

Note: <sup>a</sup>Mean methylation level of the site. <sup>b</sup>Standard deviation of the methylation level. Results based on jackknife are presented in Supplementary Table 2.

**FIGURE 3**

Scatterplots for the two top CpGs based on the entire NTR study sample. (a) The relationship between WB and methylation level of CpG site cg10845147. (b) The relationship between WB and methylation level of CpG site cg01940273.

Note: Wellbeing scores are plotted against methylation level.

$1.34 \times 10^{-5}$ ), proximal promoters ( $p = 6.01 \times 10^{-18}$ ), CGI shores ( $p = 9.26 \times 10^{-10}$ ), and DHS ( $p = 3.68 \times 10^{-14}$ ). CpG sites with a lower mean methylation level, showed, on average, a stronger association with WB ( $p = 9.13 \times 10^{-15}$ ). To account for the fact that the errors in this regression are not normally distributed, jackknife standard errors were computed, but this analysis led to the same conclusions as the normal linear regression standard errors (see Supplementary Table S2). These findings indicate that methylation sites associated with WB are enriched in gene bodies promoter areas and other regions of regulatory active DNA.

### Gene Ontology Analysis

GO enrichment analysis based on EWAS  $p$  value rank identified a large number of GO terms that were significantly enriched among higher ranked methylation sites. The strongest enriched GO term were positive regulation of biological processes (GO:0048518,  $p = 5.38 \times 10^{-21}$ ), positive regulation of cellular processes (GO:0048522,  $p = 1.34 \times 10^{-16}$ ) and developmental processes (GO:0032502,  $p = 2.38 \times 10^{-16}$ ). Also, many brain and central nervous system processes, such as regulation of neurogenesis (GO:0050767,  $p = 3.72 \times 10^{-12}$ ), neuron projector guidance (GO:0097485,  $p = 5.77 \times 10^{-10}$ ), neurotrophic signaling pathway (GO:0038179,  $p = 2.53 \times 10^{-8}$ ), as well as regulation of neuron differentiation (GO:0045664,  $p = 1.81 \times 10^{-8}$ ), were significantly enriched among higher ranking methylation sites (see Supplementary Table S3 for a complete list of significant GO terms).

### Discussion

By performing an epigenome-wide methylation analysis, the aim of the present study was to identify genomic locations at which differences in DNA methylation level are associated with differences in WB in a population-based sample of adults. Six genome-wide significant hits were identified after correction for multiple testing (FDR  $q$  value  $< 0.05$ ), while two hits remained significant after applying the stricter Bonferroni correction. Annotation analysis showed that enrichment of signal was seen in the gene body, proximal promoters, CpG shores and DHS. GO analysis, which tests categories of genes instead of single methylation sites, revealed that genes involved in regulation of cellular processes and central nervous system processes were enriched among higher-ranking genes from our EWAS. Here, we describe the six CpGs that were genome-wide significant using an FDR  $q$  value  $< 0.05$ . The gene located closest (at  $\sim 500$  kb distance) to our top-ranked CpG site cg10845147 (genomic location: chr5:172149624) is *DKFZP761M1511* (synonym is *NEURL1B*). *NEURL1B* (Neutralized E3 Ubiquitin Protein Ligase 1B) is a ligase that is involved in the regulation of the Notch pathway by influencing the stability and activity of several Notch ligands. Notch pathways acts as a regulator of cell survival and cell proliferation (Artavanis-tsakonas

et al., 1999; Miele & Osborne, 1999) and are suggested to play a role in human mammary development (Dontu et al., 2004). ENCODE data on transcription factor binding sites (TFBS) and DHS were accessed through the UCSC genome browser and showed that our top CpG, cg10845147, does not overlap with a TFBS but is located within a DHS peak in several cell types, suggesting that it is located within a regulatory region (ENCODE TFBS ChIP-seq data Mar 2012 Freeze).

The second- and third-ranked CpG sites (cg01940273, location: chr2:233,284,934 and cg03329539, location: chr:233283329) are approximately 875 kb located from the nearest gene called alkaline phosphatase, placental-like 2 (*ALLPL2*). Alkaline phosphatases (ALPs) are responsible for the dephosphorylation of various molecules, including proteins, nucleotides or alkaloids. Circulating ALP concentration is associated with premature birth (Meyer et al., 1995), and low birth weight (Brock & Barron, 1988). ALLP2 enzyme levels are increased up to 10-fold in 80% of cigarette smokers (Koshida et al., 1990) and elevated in patients with a number of cancers (Koshida et al., 1996). Both CpGs found in our study have been associated with smoking in several studies (Kleinschmidt et al., 2013; Philibert et al., 2015; Tsaprouni et al., 2014). In those studies, it was shown that methylation at multiple CpGs, including our two CpGs, was decreased among heavy smokers, but slowly increased among former smokers. Because of their association with smoking, we tested whether adding smoking as a covariate would alter the significance level of the two CpGs. For both sites, the  $p$  values did not reach the genome-wide significance threshold when adjusting for smoking (cg01940273,  $p = 1.21 \times 10^{-5}$  and cg03329539,  $p = 7.89 \times 10^{-5}$ ). Although no longer significant, the association with WB was reduced rather than fully attenuated when correcting for smoking. The regression coefficient remained in the same direction as before, suggesting a positive relationship between WB and an increase in methylation. A growing field of research has been focusing on the effects of smoking cessation on WB. The general findings of these studies are all pointing in the same direction: smoking cessation improves WB. For instance, a study by Wilson et al. (1999) found that light, moderate, and heavy smokers scored significantly lower than never-smokers as well as ex-smokers on the health-related quality-of-life scale (HR-QoL), with the strongest difference between heavy smokers and never-smokers. A similar finding was found by Piper et al. (2012) and Shahab & West (2012), who found that successful quitters reported improved subjective WB, in contrast to continuing smokers, after one to three years. Finally, the genes located nearby the other significant CpG sites, were *CGO18/N4BP2L1* (chromosome 13), *ZFPM1* (chromosome 16), and *ZNF687* (chromosome 1) and have not been previously linked to WB or related phenotypes.

Additionally, we looked up the top three probes associated with depression as reported by Dempster et al. (2014),

and the genes located most closely to these probes (*DPYSL4*, *STK32C*, *KIF13B*, *DUSP4*, *PQLC3*, and *KCNF1*) to investigate whether these genes are also associated with WB. The six genes were located in close proximity to 341 probes in our dataset, but none of these probes reach genome-wide significance, with  $p$  values ranging from  $p = .001$  to  $p = .99$ . Also, alterations in DNA methylation of the *BDNF* pathway have been associated with depression (Dalton et al., 2014). However, the 80 probes lying in close proximity in our dataset did not show an association with WB with  $p$  values ranging from .003 to .99 (for a complete overview see Supplementary Table S4).

A limitation of this study is that we did not have access to a validation data set. Therefore, future studies are warranted, especially for follow-up of the CpG sites that reached genome-wide significance. Additionally, we limited this study to WB and did not consider other aspects of behavior such as different personality traits or psychiatric symptoms. Since WB is strongly associated with a wide range of mental health diseases like depression or neuroticism, DNA methylation levels associated with WB as measured in this study may be informative for associated phenotypes. The ideal EWAS approach would therefore encompass WB and related phenotypes (e.g., different aspects of personality and depression). Such an approach would give insight into which methylation sites are common for WB and its related phenotypes and which sites are specific for WB.

In conclusion, this study provides the first genome-wide methylation association study of WB. We found six genome-wide significant DNA methylation sites of which two remained significant after the more stringent Bonferroni correction. Once genetic variants have been identified for WB, future studies that integrate both genetic and epigenetic information are warranted to investigate the intriguing interplay between genetic and environmental mechanisms in a complex trait like WB.

## Acknowledgments

This study was supported by the BBRMI-NL -financed BIOS Consortium (NWO 184.021.007), the European Research Council: ERC-230374: Genetics of Mental Illness, and ERC-284167: Beyond the genetics of Addiction, and the Netherlands Organization for Scientific Research [NWO 904-61-193: Resolving cause and effect in the association between regular exercise and psychological wellbeing and ZONMW 31160008: Genetic determinants of risk behavior in relation to alcohol use and alcohol use disorder]. MN is supported by Royal Netherlands Academy of Science Professor Award (PAH/6635) to DIB. M. Bartels is and was financially supported by a senior fellowship of the (EMGO+) Institute for Health and Care and a VU University Research Chair position.

## Supplementary Material

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/thg.2015.85>

## References

- Allione, A., Marcon, F., Fiorito, G., & Guarrera, S. (2015). Novel epigenetic changes unveiled by monozygotic twins discordant for smoking habits. *PLoS One*, *10*, e0128265.
- Artavanis-tsakonas, S., Rand, M. D., & Lake, R. J. (1999). Notch signaling: Cell fate control and signal integration in development. *Science*, *284*, 770–777.
- Bartels, M. (2015). Genetics of happiness, satisfaction with life, and wellbeing: A review and meta-analysis of heritability studies. *Behavior Genetics*, *45*, 137–156.
- Bartels, M., & Boomsma, D. I. (2009). Born to be happy? The etiology of subjective well-being. *Behavior Genetics*, *39*, 605–615.
- Bartels, M., Cacioppo, J. T., van Beijsterveldt, T. C., & Boomsma, D. I. (2013). Exploring the association between well-being and psychopathology in adolescents. *Behavior Genetics*, *43*, 177–190.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.
- Bernstein, B. E., Meissner, A., & Lander, E. S. (2007). The mammalian epigenome. *Cell*, *128*, 669–681.
- Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., ... Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*, 288–295.
- Blanchflower, D. G., & Oswald, A. J. (2008). Is well-being U-shaped over the life cycle? *Social Science & Medicine*, *66*, 1733–1749.
- Brock, D. J., & Barron, L. (1988). Measurement of placental alkaline phosphatase in maternal plasma as an indicator of subsequent low birthweight outcome. *Journal of Obstetrics & Gynaecology*, *95*, 79–83.
- Chen, Y. A., Lemire, M., Choufani, S., Butcher, D. T., Grafodatskaya, D., Zanke, B. W., ... Weksberg, R. (2013). Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, *8*, 203–209.
- Dalton, V. S., Kolshus, E., & McLoughlin, D. M. (2014). Epigenetics and depression: Return of the repressed. *Journal of Affective Disorders*, *155*, 1–12.
- Dempster, E. L., Wong, C. C. Y., Lester, K. J., Burrage, J., Gregory, A. M., Mill, J., & Eley, T. C. (2014). Genome-wide methylomic analysis of monozygotic twins discordant for adolescent depression. *Biological Psychiatry*, *76*, 977–983.
- Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*, 71–75.
- Dominguez-salas, P., Moore, S. E., Baker, M. S., Bergen, A. W., Cox, S. E., Dyer, R. A., ... Hennig, B. J. (2014). Maternal nutrition at conception modulates DNA methylation of



- human metastable epialleles. *Nature Communications*, 5, 1–7.
- Dontu, G., Jackson, K. W., McNicholas, E., Kawamura, M. J., Abdallah, W. M., & Wicha, M. S. (2004). Research article role of notch signaling in cell-fate determination of human mammary stem/progenitor cells. *Breast Cancer Research*, 6, 605–615.
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, 10, 48.
- Fortin, J. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biology*, 15, 503.
- Fujita, F., & Diener, E. D. (2005). Life satisfaction set point: Stability and change. *Journal of Personality and Social Psychology*, 88, 158–164.
- Genome of the Netherlands Consortium. (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics*, 46, 818–825.
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12, 996–1006.
- Kleinschmidt, A., Klopp, N., Zeilinger, S., Ku, B., Gieger, C., Weidinger, S., . . . Illig, T. (2013). Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*, 8, e63812.
- Klengel, T., Pape, J., & Binder, E. B. (2014). The role of DNA methylation in stress-related psychiatric disorders. *Neuropharmacology*, 80, 115–132.
- Koshida, K., Stigbrand, T., Hisazumi, H., & Wahren, B. (1990). Analysis of serum placental alkaline phosphatase activity in testicular cancer and cigarette smokers. *Urological Research*, 18, 169–173.
- Koshida, K., Uchibayashi, T., Yamamoto, H., & Hirano, K. (1996). Significance of placental alkaline phosphatase (PLAP) in the monitoring of patients with seminoma. *British Journal of Urology*, 77, 138–142.
- Layard, R. (2010). Measuring subjective well-being. *Science*, 327, 534–535.
- Lee, K. W. K., Richmond, R., Hu, P., French, L., Shin, J., Bourdon, C., & Reischl, E. (2015). Prenatal exposure to maternal cigarette smoking and dna methylation: Epigenome-wide association in a discovery sample of adolescents and replication in an independent cohort at birth through 17 years of age. *Environmental Health Perspectives*, 123, 193–199.
- Meyer, R. E., Thompson, S. J., Addy, C. L., Garrison, C. Z., & Best, R. G. (1995). Maternal serum placental alkaline phosphatase level and risk for preterm delivery. *American Journal of Obstetrics & Gynecology*, 173, 181–186.
- Miele, L., & Osborne, B. (1999). Arbiter of differentiation and death: Notch signaling meets apoptosis. *Journal of Cellular Physiology*, 409, 393–409.
- Oh, G., Wang, S. C., Pal, M., Chen, Z. F., Khare, T., Tochigi, M., . . . Petronis, A. (2015). DNA modification study of major depressive disorder: Beyond locus-by-locus comparisons. *Biological Psychiatry*, 77, 246–255.
- Philibert, R., Hollenbeck, N., Andersen, E., Osborn, T., Gerrard, M., Gibbons, F. X., & Wang, K. (2015). A quantitative epigenetic approach for the assessment of cigarette consumption. *Frontiers in Psychology*, 6, 1–8.
- Piper, M. E., Kenford, S., Fiore, M. C., & Baker, T. B. (2012). Smoking cessation and quality of life: Changes in life satisfaction over 3 years following a quit attempt. *Annals of Behavioral Medicine*, 43, 262–270.
- Rakyan, V. K., Down, T. A., Balding, D. J., & Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nature Reviews. Genetics*, 12, 529–541.
- Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, 447, 425–432.
- Romens, S. E., McDonald, J., Svaren, J., & Pollak, S. D. (2015). Associations between early life stress and gene methylation in children. *Child Development*, 86, 303–309.
- Seaford, C. (2011). Policy: Time to legislate for the good life. *Nature*, 477, 532–533.
- Shahab, L., & West, R. (2012). Differences in happiness between smokers, ex-smokers and never smokers: Cross-sectional findings from a national household survey. *Drug and Alcohol Dependence*, 121, 38–44.
- Sliker, R. C., Bos, S. D., Goeman, J. J., Bovée, J. V., Talens, R. P., van der Breggen, R., . . . Heijmans, B. T. (2013). Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics & Chromatin*, 6, 26.
- Stephoe, A., Deaton, A., & Stone, A. A. (2014). Subjective wellbeing, health, and ageing. *The Lancet*, 385, 640–648.
- Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). *Report by the commission on the measurement of economic performance and social progress*. Issaquah, WA: Community Indicators Consortium. Retrieved from <http://www.citymaking.com/wp-content/uploads/2010/01/19784660-Happiness-and-Measuring-Economic-Progress-by-Joseph-Stiglitz.pdf>
- Stubbe, J. H., Posthuma, D., Boomsma, D. I., & De Geus, E. J. (2005). Heritability of life satisfaction in adults: A twin-family study. *Psychological Medicine*, 35, 1581–1588.
- The ENCODE project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57–74.
- Tsaprouni, L. G., Yang, T. P., Bell, J., Dick, K. J., Kanoni, S., Nisbet, J., & Deloukas, P. (2014). Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics*, 9, 1382–1396.
- Van Beijsterveldt, C. E. M., Groen-Blokhuis, M., Hottenga, J. J., Franić, S., Hudziak, J. J., Lamb, D., . . . Boomsma, D. I. (2013). The young Netherlands twin register (YNTR):

- Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics*, 16, 252–267.
- Van Dongen, J., Nivard, M. G., Baselmans, B. M. L., Zilhão, N. R., Ligthart, L., Bios Consortium, . . . Boomsma, D. I. (2015). Epigenome-wide association study of aggressive behavior. *Twin Research and Human Genetics*, 18, 1–13.
- Van Dongen, J., Nivard, M. G., Willemsen, A. H. M., Hottenga, J. J., Helmer, Q., Dolan, C. V., . . . Boomsma, D. I. (2015). Genetic and environmental influences interact with age and sex in shaping the human methylome. Manuscript submitted for publication.
- Van Iterson, M., Tobi, E. W., Sliker, R. C., den Hollander, W., Luijk, R., Slagboom, P. E., & Heijmans, B. T. (2014). MethylAid: Visual and interactive quality control of large Illumina 450k datasets, *Bioinformatics*, 30, 3435–3437.
- Willemsen, G., de Geus, E. J. C., Bartels, M., van Beijsterveldt, C. E. M. T., Brooks, A. I., Estourgie-van Burk, G. F., . . . Boomsma, D. I. (2010). The Netherlands Twin Register biobank: A resource for genetic epidemiological studies. *Twin Research and Human Genetics*, 13, 231–245.
- Willemsen, G., Vink, J. M., Abdellaoui, A., den Braber, A., van Beek, J. H., Draisma, H. H., . . . Boomsma, D. I. (2013). The adult Netherlands twin register: Twenty-five years of survey and biological data collection. *Twin Research and Human Genetics*, 16, 271–281.
- Wilson, D., Parsons, J., & Wakefield, M. (1999). The health-related quality-of-life of never smokers, ex-smokers, and light, moderate, and heavy smokers. *Preventive Medicine*, 29, 139–144.
-