

# Epigenome-wide association study reveals CpG sites associated with thyroid function and regulatory effects on *KLF9*

Antoine Weihs<sup>1\*</sup>, Loyal Chaker<sup>2\*</sup>, Tiphaine C. Martin<sup>3,4\*</sup>, Kim V.E. Braun<sup>5</sup>, Purdey J. Campbell<sup>6</sup>, Simon R. Cox<sup>7</sup>, Myriam Fornage<sup>8,9</sup>, Christian Gieger<sup>10,11,12</sup>, Hans J. Grabe<sup>1,13</sup>, Harald Grallert<sup>10,11</sup>, Sarah E. Harris<sup>7</sup>, Brigitte Kühnel<sup>10,11</sup>, Riccardo E. Marioni<sup>14</sup>, Nicholas G. Martin<sup>15</sup>, Daniel L. McCartney<sup>14</sup>, Allan F. McRae<sup>16</sup>, Christa Meisinger<sup>17</sup>, Joyce B.J. van Meurs<sup>18,19</sup>, Jana Nano<sup>11,20</sup>, Matthias Nauck<sup>21,22</sup>, Annette Peters<sup>10,11,12,20</sup>, Holger Prokisch<sup>23,24</sup>, Michael Roden<sup>25,26,27</sup>, Elizabeth Selvin<sup>28,29</sup>, Marian Beekman<sup>30</sup>, Diana van Heemst<sup>31</sup>, Eline P. Slagboom<sup>30</sup>, Brenton R. Swenson<sup>32</sup>, Adrienne Tin<sup>28,33</sup>, Pei-Chien Tsai<sup>4,34,35</sup>, Andre Uitterlinden<sup>18</sup>, W. Edward Visser<sup>36</sup>, Henry Völzke<sup>22,37</sup>, Melanie Waldenberger<sup>10,11,12</sup>, John P. Walsh<sup>6,38</sup>, Anna Köttgen<sup>28,39</sup>, Scott G. Wilson<sup>4,6,40</sup>, Robin P. Peeters<sup>36</sup>, Jordana T. Bell<sup>4#</sup>, Marco Medici<sup>36#</sup>, Alexander Teumer<sup>1,22,37,41#</sup>

1 Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany

2 Erasmus MC Academic Center for Thyroid diseases, Department of Internal Medicine and Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands

3 Department of Oncological Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

4 Department of Twin Research and Genetic Epidemiology, St Thomas' Hospital Campus, King's College London, Westminster Bridge Road, London, UK

5 Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands

6 Department of Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, WA 6009, Australia

7 Lothian Birth Cohorts, Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, UK

8 Brown Foundation Institute of Molecular Medicine, McGovern Medical School, Houston, TX 77030, USA

9 Human Genetics Center, School of Public Health; The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

10 Research Unit Molecular Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764, Neuherberg, Bavaria, Germany

11 Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, D-85764, Neuherberg, Bavaria, Germany

12 German Center for Cardiovascular Research (DZHK), Partner Site Munich Heart Alliance, Munich, Germany

13 German Centre for Neurodegenerative Diseases (DZNE), Site Rostock/Greifswald, Germany

14 Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK

15 QIMR Berghofer Medical Research Institute, Brisbane, Australia

16 Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland, Australia

17 Epidemiology, Medical Faculty, University of Augsburg, Augsburg, Germany

18 Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

19 Department of Orthopaedics & Sports Medicine, Erasmus Medical Center, Rotterdam, The Netherlands

20 Institute for Medical Informatics, Biometrics and Epidemiology, Ludwig-Maximilians-Universität (LMU) Munich, Munich, Germany

21 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany

22 DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany  
23 Institute of Neurogenomics, Computational Health Center, Helmholtz Munich, Neuherberg, Germany  
24 Institute of Human Genetics, School of Medicine, Technical University Munich, Munich, Germany  
25 Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany  
26 Division of Endocrinology and Diabetology, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany  
27 German Center for Diabetes Research (DZD), Munich-Neuherberg, Germany  
28 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, USA  
29 Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins School of Medicine, USA  
30 Section of Molecular Epidemiology, Department of Biomedical Data Sciences, Leiden University Medical Center  
31 Section of Gerontology and Geriatrics, department of Internal Medicine, Leiden University Medical Center  
32 Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA  
33 Department of Medicine, University of Mississippi Medical Center  
34 Department of Biomedical Sciences, Chang Gung University, Taoyuan, Taiwan  
35 Division of Pediatric Infectious Diseases, Department of Pediatrics, Chang Gung Memorial Hospital, Taoyuan City 333, Taiwan  
36 Erasmus MC Academic Center for Thyroid diseases, Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands  
37 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany  
38 Medical School, University of Western Australia, Crawley, WA 6009, Australia  
39 Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany  
40 School of Biomedical Sciences, University of Western Australia, Perth 6009, Australia  
41 Department of Population Medicine and Lifestyle Diseases Prevention, Medical University of Białystok, Białystok, Poland

\* These authors contributed equally.

# These authors jointly supervised this work.

Correspondence to:

Alexander Teumer, PhD  
Department of Psychiatry and Psychotherapy  
University Medicine Greifswald  
Ellernholzstr. 1-2  
17475 Greifswald  
Germany  
email: [ateumer@uni-greifswald.de](mailto:ateumer@uni-greifswald.de)

Final publication is available from Mary Ann Liebert, Inc.: <http://dx.doi.org/10.1089/thy.2022.0373>

## Abstract

Background: Thyroid hormones play a key role in differentiation and metabolism, and are known regulators of gene expression through both genomic and epigenetic processes including DNA methylation. The aim of this study was to examine associations between thyroid hormones and DNA methylation.

Methods: We carried out a fixed-effect meta-analysis of epigenome-wide association study of blood DNA methylation sites from 8 cohorts from the ThyroidOmics Consortium, incorporating up to 7,073 participants of both European and African ancestry, implementing a discovery and replication stage. Statistical analyses were conducted using normalized beta CpG values as dependent and log-transformed thyrotropin (TSH), free thyroxine and free triiodothyronine levels, respectively, as independent variable in a linear model. The replicated findings were correlated with gene expression levels in whole-blood, and tested for causal influence of TSH and free thyroxine by two-sample Mendelian randomization.

Results: Epigenome-wide significant associations ( $p$ -value  $< 1.1E-7$ ) of 3 CpGs for free thyroxine, 5 for free triiodothyronine, and 2 for TSH concentrations were discovered and replicated (combined  $p$ -values =  $1.5E-9$  to  $4.3E-28$ ). The associations included CpG sites annotated to *KLF9* (cg00049440) and *DOT1L* (cg04173586) that overlap with all three traits, consistent with hypothalamic–pituitary–thyroid axis physiology. Significant associations were also found for CpGs in *FKBP5* for free thyroxine, and at *CSNK1D/LINC01970* and *LRR8D* for free triiodothyronine. Mendelian randomization analyses supported a causal effect of thyroid status on DNA methylation of *KLF9*. DNA methylation of cg00049440 in *KLF9* was inversely correlated with *KLF9* gene expression in blood. The CpG at *CSNK1D/LINC01970* overlapped with thyroid hormone receptor alpha binding peaks in liver cells. The total additive heritability of the methylation levels of the six significant CpG sites was between 25% and 57%. Significant methylation QTLs were identified for CpGs at *KLF9*, *FKBP5*, *LRR8D* and *CSNK1D/LINC01970*.

Conclusions: We report novel associations between TSH, thyroid hormones and blood-based DNA methylation. This study advances our understanding of thyroid hormone action particularly related to *KLF9*, and serves as a proof-of-concept that integrations of EWAS with other -omics data can provide a valuable tool for unravelling thyroid hormone signaling in humans by complementing and feeding classical *in-vitro* and animal studies.

## Introduction

Thyroid hormones play a key role in differentiation and metabolism. Thyroid dysfunction, a condition affecting 5-10% of the adult population, is associated with an increased risk of weight changes, cardiovascular diseases, osteoporosis, psychiatric disorders, and mortality<sup>1,2</sup>. Prohormone thyroxine and the biologically active triiodothyronine are the main circulating thyroid hormones and are regulated by thyroid stimulating hormone (TSH). Thyroid hormones are known regulators of gene expression through both genomic and epigenetic processes<sup>3</sup>, however, the underlying exact molecular mechanisms remain unknown. One of these processes include DNA methylation (DNAm), which predominantly occurs at CpG sites, and is a key regulator of gene expression. While there has been little research so far in humans, various animal models provided evidence of thyroid hormones influencing DNAm. This was supported for example in a study by Kyono and colleagues<sup>4</sup>, showing in developing frog tissue that triiodothyronine directly controls DNA methyltransferase 3a (Dnmt3a) expression, responsible for de novo DNAm. A recent study of two cohorts of Australian adolescents further revealed two and six CpG sites that were associated with TSH and free triiodothyronine (FT3) concentrations, respectively, indicating that although DNAm is highly tissue specific, the analysis of DNAm in blood may provide deeper insights into thyroid hormone action and/or regulation<sup>5</sup>. Motivated by the conclusion that larger sample sizes are necessary to replicate the findings and to reveal additional associations, we conducted a large-scale epigenome-wide association study (EWAS) assessing the effects of free thyroxine (FT4), FT3 and TSH on changes of DNAm in whole blood encompassing up to 7,073 individuals of the ThyroidOmics Consortium ([www.thyroidomics.com](http://www.thyroidomics.com)). We furthermore replicated findings in independent samples, improved the generalizability of the results by including individuals of both European and African ancestry, and provided insights into the underlying molecular mechanisms by analyzing gene expression and assessing causality using Mendelian randomization (MR).

## Materials and Methods

### *Study population*

We conducted the EWAS using population-based studies from the ThyroidOmics Consortium. The discovery stage included three cohorts: SHIP-Trend <sup>6</sup>, ARIC <sup>7</sup>, and KORA <sup>8</sup>. Replication of the findings was sought in the Rotterdam Study (RS) <sup>9</sup>, TwinsUK <sup>10</sup>, the Lothian Birth Cohorts of 1921 and 1936 (LBC1921 and LBC1936) <sup>11</sup>, and the Brisbane Systems Genetics Study (BSGS) <sup>12</sup>. All study protocols were conducted in accordance with the Helsinki Declaration and approved by the respective local ethics committees. All subjects provided written informed consent. Detailed information is given in **Tables 1** and **2**, and in the **Supplementary Methods**.

### *Biomarker measurements*

DNAm was measured from whole blood using Illumina Infinium BeadChip arrays. Details regarding the cohorts, as well as the FT4, FT3 and TSH assays applied, are provided in **Table 2**.

### *Statistical analysis*

The EWAS was undertaken in each cohort separately and subsequently meta-analyzed using quantile-quantile normalized beta CpG values as the dependent and log-transformed hormone levels as the independent variable. Each cohort followed the same analysis plan (**Supplementary Methods**), and the results were processed by our in-house quality control pipeline (**Supplementary Figure 1**). The EWAS results were corrected for inflation and bias where applicable using the *Bacon* R-package (**Supplementary Figure 2**) <sup>13</sup>. A detailed overview of the analyses is provided in Figure 1 and in the **Supplementary Methods**.

Findings of the discovery stage passing significance after Bonferroni correction ( $p\text{-value} < 0.05/450,000 = 1.1\text{E-}07$ ) were replicated in additional independent samples. A successfully replicated site was defined by  $p\text{-value} < 0.05$  in the replication stage with consistent effect direction, and a  $p\text{-value} < 1.1\text{E-}07$  in the discovery and replication combined meta-analysis.

### *Correlation with gene expression levels*

The association of replicated CpG sites with gene expression levels of genes within +/-500 kb vicinity was assessed in up to 713 blood samples of the KORA study (**Supplementary Methods**)<sup>14</sup>. Results passing the false discovery rate (FDR) < 0.05 were declared significant.

### *Additional characterization of the findings*

To analyze whether DNAm sites altered by thyroid hormones levels might be influenced by binding of thyroid hormones on thyroid-hormone receptors elements, we extracted the overlap and distance between replicated CpGs and binding sites of the thyroid hormone receptor alpha from CHIPAtlas ([https://chip-atlas.org/peak\\_browser](https://chip-atlas.org/peak_browser))<sup>15</sup>. The data was based human liver<sup>16</sup> and brain cell lines (GEO-ID: GSE129039). Lookup of transcription factor footprint and binding site associations in thyroid tissue was conducted using eFORGE-TF (<https://eforge-tf.altiusinstitute.org/>)<sup>17</sup>. Additive heritability separating the overall genetic variation from environmental influences on selected DNAm sites, was estimated using summary statistics from a previous published dataset<sup>18</sup>. To evaluate whether the DNAm sites are altered by *cis* and *trans* genetic variants, we explored GoDMC (<http://www.godmc.org.uk>)<sup>19</sup>.

### *Mendelian randomization analysis*

To test for a possible causal effect of the thyroid hormones on DNAm levels of significantly associated CpG sites, we conducted a 2-sample MR using the R-package *TwoSampleMR*<sup>20</sup>. As genetic instruments for FT3 were unavailable, we only conducted the MR on FT4- and TSH-associated sites using independent genome-wide significantly associated index SNPs from a large genome-wide association study (GWAS) as instruments for the thyroid hormone levels<sup>21</sup>. The association of these SNPs with DNAm levels (mQTLs) as outcomes, was assessed in 1,662 blood samples of the KORA study. The mQTLs were estimated by regressing the residuals of the methylation beta-values adjusted for sex, age, technical factors and estimated white-blood cell type composition<sup>22</sup> on the SNP allele dosage. Results of the mQTL analyses were available for 31 and 56 instruments for FT4 and TSH, respectively. The inverse-variance weighted fixed effect MR was conducted as primary analysis, and the more robust

methods regarding violations of the validity of the instruments but less powerful weighted median <sup>23</sup> and MR-Egger <sup>24</sup>, as sensitivity analyses. Significance was assessed by the primary analysis p-value < 0.05 divided by the number of tested CpGs per thyroid function trait. The effects can be interpreted as the proportion increase in DNAm per change in one standard deviation of the thyroid hormone level.

## Results

### *Study sample characteristics*

In the discovery stage EWAS based on 3 cohorts, we included 4,085 individuals for TSH, 1,639 for FT3 and 4,081 for FT4. In the replication stage using 5 independent cohorts, 2,988 additional individuals were included in the TSH analysis, 590 for FT3 and 2,448 for FT4 (Table 1 and **Supplementary Methods**).

### *DNAm sites associated with thyroid hormone levels*

We discovered and replicated two novel CpG sites associated with TSH (cg00049440 in *KLF9* and cg04173586 in *DOT1L*), and three novel sites associated with FT4 (cg00049440 in *KLF9*, cg04173586 in *DOT1L* and cg03546163 in *FKBP5*). Of the five FT3-associated CpGs (in *KLF9*, *DOT1L*, *LRRC8D*, and near *CSNK1D/LINC01970*), the two CpGs in *LRRC8D* (cg06983052 and cg20146909) are novel. The replicated results are provided in **Table 3**, the significant results of the discovery stage are shown in **Supplementary Table 1**, and the cohort-specific results are listed in **Supplementary Table 2**. In total, all but one site identified in the discovery analysis were successfully replicated. Detailed annotation of the replicated CpGs is shown in **Supplementary Table 3**.

As indicated by the Chicago plots, all replicated sites were associated with higher methylation for higher TSH (**Figure 2**) and with lower methylation for both higher FT4 (**Figure 3**) and FT3 (**Figure 4**).

In detail, the TSH meta-analysis revealed two associations cg00049440 ( $\beta_{\text{combined}} = 0.13$ ; p-value = 9.57E-16) and cg04173586 ( $\beta_{\text{combined}} = 0.07$ ; p-value = 1.53E-09). The CpG site cg00049440 is located inside a promoter region in intron 1 of the *KLF9* gene (**Supplementary Figure 3**), and cg04173586 is located in intron 1 of *DOT1L* (**Supplementary Figure 4**).

The FT4 analysis identified three sites (cg00049440, cg03546163, and cg04173586), two of which, namely cg00049440 in *KLF9* ( $\beta_{\text{combined}} = -0.75$ ; p-value= 4.44E-20, **Supplementary Figure 5**) and cg04173586 in *DOT1L* ( $\beta_{\text{combined}} = -0.43$ ; p-value= 4.03E-11, **Supplementary Figure 6**) were also identified in the TSH analysis. The additional association is cg03546163 ( $\beta_{\text{combined}} = -0.84$ ; p-value= 1.45E-16), which lies within the second intron of *FKBP5* (**Supplementary Figure 7**).

The EWAS on FT3 identified five sites cg00049440, cg01695994, cg04173586, cg06983052 and cg20146909, two of which, namely cg00049440 ( $\beta_{\text{combined}} = -1.46$ ; p-value= 1.69E-17, **Supplementary Figure 8**) and cg04173586 ( $\beta_{\text{combined}} = -1.93$ ; p-value= 4.33E-28, **Supplementary Figure 9**) were also identified in the FT4 and TSH and one CpG, namely cg06983052 ( $\beta_{\text{combined}} = -1.04$ ; p-value= 5.45E-14, **Supplementary Figure 10**) was identified in the FT4 analysis (**Supplementary Table 1**). The remaining two sites are cg01695994 ( $\beta_{\text{combined}} = -1.37$ ; p-value= 5.84E-17), which is located 1,518bp downstream of *LINC01970* and 14,809bp upstream of *CSNK1D* (**Supplementary Figure 11**), and cg20146909 ( $\beta_{\text{combined}} = -1.04$ ; p-value= 6.51E-14), which is located inside intron 1 of the gene *LRR8D* (**Supplementary Figure 12**).

The DNAm sites we identified for TSH, FT4 and FT3 were not co-located with SNPs or genes previously highlighted through GWAS of thyroid function.

#### *Effects on gene expression in blood*

As DNAm represents an important regulator of gene expression, we tested the association of the methylation levels of our replicated findings with mRNA levels in blood of nearby genes. Of all 146 CpG-mRNA associations tested, only cg00049440 passed the level of significance (FDR <0.05) and was negatively correlated with *KLF9* gene expression ( $\beta = -0.109$ ; p-value= 3.00E-7, **Supplementary Table 4**). The CpG cg00049440 is located in an island shore and in the promotor region of *KLF9*, and was associated with circulating TSH, FT3 and FT4 levels.

### *Overlap of DNAm sites with thyroid hormone receptor binding sites*

Except for cg20146909, all CpGs that were associated with thyroid function fall in promoters and enhancer regions shared between several tissues including blood, liver, and brain (**Supplementary Table 3**). As thyroid hormone receptor footprints are dynamic, capable of chromatin remodeling and dependent on thyroid hormones levels<sup>25,26</sup>, we furthermore studied whether DNAm sites associated with FT4, FT3 and TSH fall in thyroid hormone receptor binding sites in liver and neural cells where the data was available (see Methods). The cg01695994 at *CSNK1D/LINC01970* overlapped with thyroid hormone receptor alpha binding peaks in liver cells. We did not identify any other overlap between thyroid hormone receptor binding site peaks and our six significant DNAm sites (distance between 517 bp and 2,219 bp, **Supplementary Table 3**). This could be either because of tissue-specificity or because there are more complex relationships between thyroid hormone levels and DNAm.

### *Impacts of genetic variation on thyroid-DNAm sites*

The total additive heritability of individual DNAm levels ( $h^2$ ) at our six significant CpG sites are above 25% (cg06983052) and up to 57% (cg06983052) in blood (**Supplementary Table 3**). DNAm heritability estimates at cg00049440 and cg03546163 were explained by 72% and 100% of common genetic variants in the genome ( $h^2_{\text{SNPs}}/h^2$ ), respectively<sup>18</sup>. On the other hand, the proportion of heritability explained by common genetic variants for the five remaining signals was less than 30%, and near 0% for cg04173586.

Using mQTL results identified in GoDMC<sup>19</sup>, several independent *cis* and *trans*-genetic variants impact the blood DNAm levels of all significant CpG sites except *DOT1L* (**Supplementary Table 5**). Interestingly, DNAm levels at cg00049440 (*KLF9*) and cg03546163 (*FKBP5*) were significantly associated with several genetic variants in *trans* located in an intronic region of the *THRB* gene (rs9310736 and rs869785), which encodes the nuclear hormone receptor for triiodothyronine. However, none of these variants were identified in thyroid hormone GWAS<sup>21,27</sup> or as eQTLs in the GTEx database (<https://gtexportal.org>).

### *Causal effects of thyroid status on DNAm*

We conducted a MR analysis to test if variation in TSH and FT4 causally affect the changes in DNAm of the replicated EWAS findings. MR is a framework using SNPs as instrumental variables to assess an unconfounded exposure-outcome association and thus allowing a causal inference<sup>28,29</sup>. To maximize statistical power, we applied a two-sample MR using published large-scale GWAS results for selecting instruments for TSH and FT4 as exposure. Because no instruments for FT3 from former GWAS were available, no causal effects of triiodothyronine on DNAm as outcome could be assessed. The MR results revealed a significant effect of genetically determined TSH levels on cg00049440 at *KLF9* ( $\beta = 0.005$ ; p-value = 0.01). The MR-Egger and the weighted median MR analysis were nominally significant with the same effect directions and similar effect sizes underpinning the robustness of the findings. No heterogeneity (Q-statistics p-value = 0.35) or directional pleiotropy (MR-Egger intercept p-value = 0.33) being both indicators of wrong instruments were detected. All MR results of the second TSH associated CpG site as well as the FT4 associated sites had p-values > 0.05 (**Supplementary Table 6**). As an additional sensitivity analysis addressing potential horizontal pleiotropy through associations with thyroperoxidase (TPO), we removed instruments that were associated (p-value < 0.05) with TPO antibody positivity in a recent GWAS<sup>30</sup>. After excluding five TPO-associated SNPs for TSH, significance changed only for the weighted median MR result for TSH on cg00049440 to p-value = 0.06 with a similar effect estimate. No instruments needed to be excluded for the MR of FT4 (**Supplementary Table 6**).

### **Discussion**

In the current analyses, we revealed three significantly associated CpGs for FT4, five for FT3, and two for TSH. An overview of our findings in relation to known mechanisms is illustrated in **Figure 5**. Additionally to the known associations between FT3 and Krueppel-like factor 9 (*KLF9*) and DOT1 like histone lysine methyltransferase (*DOT1L*) related CpG sites<sup>5</sup>, we found de-novo associations regarding TSH and FT4 with effect directions consistent with hypothalamic–pituitary–thyroid axis physiology (i.e., opposite effects for TSH and FT4/FT3).

The ubiquitously expressed *KLF9* gene is part of the Sp1 C2H2-type zinc finger transcription factor family that binds to GC box elements located in the promotor (NCBI RefSeq database). *KLF9* was found to be a triiodothyronine response gene in mice<sup>26</sup> and frogs<sup>31</sup>, where triiodothyronine regulated gene expression by binding to and activating thyroid hormone receptors, which in turn regulate *KLF9* transcription via the thyroid response element<sup>26</sup>. Furthermore, there is evidence that this mechanism is also present in human cells including hepatocytes<sup>32</sup> and neural cells<sup>33</sup>. Our *KLF9* findings exemplify how thyroid function affects gene expression via DNAm in humans. In particular, we found that higher thyroid hormone levels (with physiologically fitting lower TSH levels) were associated with lower DNAm levels at cg00049440, located in the promotor region of *KLF9*, which in turn was associated with increased *KLF9* expression in blood cells (**Figure 5**). Importantly, the causality in this cascade was supported by MR analyses showing directionally consistent associations with genetically determined TSH levels. Previous studies showed no overlap between variants detected in TSH and FT4 GWAS<sup>21</sup>, while MR analyses showed that well-known thyroid status-related endpoints such as cholesterol levels and atrial fibrillation were associated with TSH but not with FT4 variants<sup>34,35</sup>. This strongly suggests that TSH-associated genetic variants are a more valid representation of thyroid function, which likely explains the absence of associations in our FT4 MR analyses. We hereby show in humans that next to the known pathway via binding to thyroid hormone receptor beta on a response element in the *KLF9* promoter, thyroid hormones can also affect *KLF9* gene expression via *KLF9* DNAm (**Figure 5**). In this context it is noteworthy to mention that in a study in developing *Xenopus* brain, triiodothyronine has been shown to promote DNA demethylation. This is in part by promoting TET3 recruitment to discrete genomic regions, and in part by inducing genes that encode DNA demethylation enzymes<sup>36</sup>.

The second multi-trait association, *DOT1L*, represents a ubiquitously expressed gene coding for a methyltransferase that methylates lysine-79 of histone H3. A link to thyroid hormones was found in *Xenopus laevis* where triiodothyronine induces *Dot1L* expression during metamorphosis<sup>37</sup>. Our study partially translates these results to humans, where the thyroid function-associated CpG cg04173586 is

located in an enhancer region. However, the correlation of DNAm with *DOT1L* mRNA levels in 699 blood tissue samples was not significant.

This points to a limitation of our study as blood represents only one of the possible target tissues for both DNAm and gene expression analyses. Additionally, the sample size might have been too small to achieve the required statistical power, although the effect direction suggests an inverse association (**Supplementary Table 4**). The advantage of correlating both measured DNAm and mRNA is that this approach does not depend on the availability of known genetic factors needed to predict these measurements in alternative approaches like colocalization.

In addition to two new and three known associations with FT3 revealed in our study, the remaining three associations discovered in the study of Lafontaine *et al.*<sup>5</sup> were statistically significant ( $p < 0.05/6$ ) in our EWAS. Considering that our replication cohort for FT3 was also included in Lafontaine *et al.*, we restricted this lookup to the discovery stage results (**Supplementary Table 7**) providing an independent sample set.

Of note, the CpG cg20146909 in *LRRC8D* that was associated with FT3 in our study, was suggestive of an association in Lafontaine *et al.*<sup>5</sup>. Thus, we confirmed this candidate, and revealed a second site in this gene, namely cg06983052, that was associated with both FT3 and FT4.

The two previously reported TSH associations at *FOXK2* and the lncRNA *AC012506.1* were not replicated in our study (**Supplementary Table 7**). This could be because the previous findings were false positive results, or because these associations are specific to adolescents or young adults who accounted for most participants in the former study.

In addition to datasets from European ancestry, our study also included a substantial proportion of DNAm data of African ancestry individuals via the ARIC cohort. However, no ancestry-specific heterogeneity of the effect estimates was observed for our replicated EWAS results (**Supplementary Table 2, Supplementary Figure 13**).

In conclusion, we conducted an EWAS by developing and applying a standardized analyses and quality control workflow in a large sample of both European and African ancestry individuals, replicated the statistically significant associations in independent cohorts, and characterized the findings by integrating multiple types of -omics data. Our study revealed DNAm associated with thyroid function at five distinct loci. In particular, we showed that DNAm is another previously unknown route underlying thyroid hormone-induced *KLF9* expression, which is a well-known thyroid hormone household gene. This study serves as a proof-of-concept that integration of EWAS with other -omics data can provide a valuable tool to unravel thyroid hormone signaling in humans by complementing and feeding classical *in-vitro* and animal studies.

## **Funding**

Acknowledgements and funding sources are listed in the **Supplementary Note**.

## **Author Contributions**

Project Design: A.Te., J.T.B., L.C., M.M., R.P.P., T.C.M. Data collection: A.F.M., C.G., C.M., D.v.H., E.P.S., H.G., H.J.G., H.P., J.P.W., J.T.B., M.F., M.N., M.R., M.W., N.G.M., P.J.C., S.E.H., S.G.W., S.R.C. Cohort Study Management: A.P., A.U., C.M., E.P.S., H.V., J.B.v.M., J.T.B., M.B., M.F., M.N., N.G.M., R.P.P., S.R.C., T.C.M. Subject Recruitment: C.M., H.V., N.G.M., S.R.C. Drafting of manuscript: A.Te., A.W., T.C.M. Interpretation of Results: A.K., A.Te., A.W., E.S., J.N., J.P.W., J.T.B., M.M., S.G.W., T.C.M. Statistical Methods and Analysis: A.F.M., A.Te., A.Ti., A.W., B.K., B.R.S., D.L.M., J.N., K.V.B., L.C., P.J.C., P.T., R.E.M., S.E.H., T.C.M. Critical review of manuscript: A.F.M., A.K., A.P., A.Te., A.Ti., A.U., A.W., B.K., B.R.S., C.G., C.M., D.L.M., D.v.H., E.P.S., E.S., W.E.V., H.G., H.J.G., H.P., H.V., J.B.v.M., J.N., J.P.W., J.T.B., K.V.B., L.C., M.B., M.F., M.M., M.N., M.R., M.W., N.G.M., P.J.C., P.T., R.E.M., R.P.P., S.E.H., S.G.W., S.R.C., T.C.M.

## **Author Disclosure Statement**

H.J.G. has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag as well as research funding from Fresenius Medical Care not related to this work. R.E.M. has received speaker fees from Illumina and is an advisor to the Epigenetic Clock Development Foundation. All other authors declare no conflicts of interest.

## References

1. Cooper DS, Biondi B. Subclinical thyroid disease. *Lancet* 2012;379(9821):1142–1154; doi: 10.1016/S0140-6736(11)60276-6.
2. Wondisford, Fredric E; Radovick S. *Clinical Management of Thyroid Disease*. 1st ed. (Wondisford, Fredric E; Radovick S. ed). Elsevier; 2009.; doi: 10.1016/B978-1-4160-4745-2.X0001-6.
3. Wu Y, Koenig RJ. Gene Regulation by Thyroid Hormone. *Trends Endocrinol Metab* 2000;11(6):207–211; doi: 10.1016/S1043-2760(00)00263-0.
4. Kyono Y, Sachs LM, Bilesimo P, et al. Developmental and Thyroid Hormone Regulation of the DNA Methyltransferase 3a Gene in *Xenopus* Tadpoles. *Endocrinology* 2016;157(12):4961–4972; doi: 10.1210/en.2016-1465.
5. Lafontaine N, Campbell PJ, Castillo-Fernandez JE, et al. Epigenome-Wide Association Study of Thyroid Function Traits Identifies Novel Associations of ft3 With KLF9 and DOT1L. *J Clin Endocrinol Metab* 2021;106(5):e2191–e2202; doi: 10.1210/clinem/dgaa975.
6. Völzke H, Schössow J, Schmidt CO, et al. Cohort Profile Update: The Study of Health in Pomerania (SHIP). *Int J Epidemiol* 2022; doi: 10.1093/ije/dyac034.
7. Wright JD, Folsom AR, Coresh J, et al. The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. *J Am Coll Cardiol* 2021;77(23):2939–2959; doi: 10.1016/j.jacc.2021.04.035.
8. Holle R, Happich M, Löwel H, et al. KORA - A Research Platform for Population Based Health Research. *Das Gesundheitswes* 2005;67(S 01):19–25; doi: 10.1055/s-2005-858235.
9. Hofman A, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015;30(8):661–708; doi: 10.1007/s10654-015-0082-x.
10. Moayyeri A, Hammond CJ, Valdes AM, et al. Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int J Epidemiol* 2013;42(1):76–85; doi: 10.1093/ije/dyr207.
11. Taylor AM, Pattie A, Deary IJ. Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol* 2018;47(4):1042–1042r; doi: 10.1093/ije/dyy022.
12. Powell JE, Henders AK, McRae AF, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One* 2012;7(4):e35430; doi: 10.1371/journal.pone.0035430.
13. van Iterson M, van Zwet EW, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* 2017;18(1):19; doi: 10.1186/s13059-016-1131-9.
14. Schurmann C, Heim K, Schillert A, et al. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. *PLoS One* 2012;7(12):e50938; doi: 10.1371/journal.pone.0050938.
15. Oki S, Ohta T, Shioi G, et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep* 2018;19(12); doi: 10.15252/EMBR.201846255.
16. Partridge EC, Chhetri SB, Prokop JW, et al. Occupancy maps of 208 chromatin-associated proteins in one human cell type. *Nat* 2020 5837818 2020;583(7818):720–728; doi: 10.1038/s41586-020-2023-4.
17. Breeze CE. Cell Type-Specific Signal Analysis in Epigenome-Wide Association Studies. *Methods*

Mol Biol 2022;2432:57–71; doi: 10.1007/978-1-0716-1994-0\_5.

18. Van Dongen J, Nivard MG, Willemsen G, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* 2016 71 2016;7(1):1–13; doi: 10.1038/ncomms11115.
19. Min JL, Hemani G, Hannon E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* 2021;53(9):1311–1321; doi: 10.1038/S41588-021-00923-X.
20. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 2018;7; doi: 10.7554/eLife.34408.
21. Teumer A, Chaker L, Groeneweg S, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. *Nat Commun* 2018;9(1):4455; doi: 10.1038/s41467-018-06356-1.
22. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;13(1):86; doi: 10.1186/1471-2105-13-86.
23. Bowden J, Davey Smith G, Haycock PC, et al. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol* 2016;40(4):304–314; doi: 10.1002/gepi.21965.
24. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;44(2); doi: 10.1093/ije/dyv080.
25. Grøntved L, Waterfall JJ, Kim DW, et al. Transcriptional activation by the thyroid hormone receptor through ligand-dependent receptor recruitment and chromatin remodelling. *Nat Commun* 2015 61 2015;6(1):1–11; doi: 10.1038/ncomms8048.
26. Denver RJ, Williamson KE. Identification of a Thyroid Hormone Response Element in the Mouse Krüppel-Like Factor 9 Gene to Explain Its Postnatal Expression in the Brain. *Endocrinology* 2009;150(8):3935–3943; doi: 10.1210/en.2009-0050.
27. Zhou W, Brumpton B, Kabil O, et al. GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. *Nat Commun* 2020;11(1):3981; doi: 10.1038/s41467-020-17718-z.
28. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet* 2008;123(1):15–33; doi: 10.1007/s00439-007-0448-6.
29. Medici M, Peeters RP, Teumer A, et al. The importance of high-quality mendelian randomisation studies for clinical thyroidology. *lancet Diabetes Endocrinol* 2019;50(5):668–681; doi: 10.1016/S2213-8587(19)30145-7.
30. Medici M, Porcu E, Pistis G, et al. Identification of novel genetic Loci associated with thyroid peroxidase antibodies and clinical thyroid disease. *PLoS Genet* 2014;10(2):e1004123; doi: 10.1371/journal.pgen.1004123.
31. Furlow JD, Kanamori A. The Transcription Factor Basic Transcription Element-Binding Protein 1 Is a Direct Thyroid Hormone Response Gene in the Frog *Xenopus laevis*. *Endocrinology* 2002;143(9):3295–3305; doi: 10.1210/EN.2002-220126.
32. Cvoro A, Devito L, Milton FA, et al. A Thyroid Hormone Receptor/KLF9 Axis in Human Hepatocytes and Pluripotent Stem Cells. *Stem Cells* 2015;33(2):416–428; doi:

10.1002/stem.1875.

33. Morte B, Gil-Ibáñez P, Bernal J. Regulation of Gene Expression by Thyroid Hormone in Primary Astrocytes: Factors Influencing the Genomic Response. *Endocrinology* 2018;159(5):2083–2092; doi: 10.1210/EN.2017-03084.
34. Marouli E, Kus A, Del Greco M F, et al. Thyroid function affects the risk of stroke via atrial fibrillation: a Mendelian Randomization study. *J Clin Endocrinol Metab* 2020; doi: 10.1210/clinem/dgaa239.
35. Kuś A, Marouli E, Del Greco M F, et al. Variation in normal range thyroid function affects serum cholesterol levels, blood pressure and type 2 diabetes risk: A Mendelian randomization study. *Thyroid* 2020;thy.2020.0393; doi: 10.1089/thy.2020.0393.
36. Raj S, Kyono Y, Sifuentes CJ, et al. Thyroid Hormone Induces DNA Demethylation in *Xenopus* Tadpole Brain. *Endocrinology* 2020;161(11); doi: 10.1210/endo/bqaa155.
37. Wen L, Fu L, Shi Y. Histone methyltransferase Dot1L is a coactivator for thyroid hormone receptor during *Xenopus* development. *FASEB J* 2017;31(11):4821–4831; doi: 10.1096/fj.201700131R.
38. Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol* 2015;16(1):37; doi: 10.1186/s13059-015-0600-x.
39. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 2013;29(2):189–96; doi: 10.1093/bioinformatics/bts680.
40. Tsaprouni LG, Yang T, Bell J, et al. Epigenetics Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics* 2014;9(10):1382–1396; doi: 10.4161/15592294.2014.969637.
41. McRae AF, Powell JE, Henders AK, et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol* 2014;15(5):R73; doi: 10.1186/gb-2014-15-5-r73.

## Tables

**Table 1:** Baseline statistics of the cohorts included in the EWAS analysis.

	<b>SHIP-Trend</b>	<b>ARIC</b>	<b>KORA</b>	<b>RS</b>	<b>TWINSUK</b>	<b>LBC1936</b>	<b>LBC1921</b>	<b>BSGS</b>
	Discovery	Discovery	Discovery	Replication	Replication	Replication	Replication	Replication
<b>Ethnicity</b>	European ancestry	African ancestry	European ancestry					
<b>Age, mean (sd)</b>	50.9 (13.52)	56.2 (5.77)	61.01 (8.86)	59.85 (8.16)	57.20 (10.36)	69.54 (0.83)	79.05 (0.58)	21.0 (18.88)
<b>BMI, Mean (sd), kg/m<sup>2</sup></b>	27.09 (4.01)	29.99 (6.24)	28.81 (4.78)	27.47 (4.79)	27.06 (5.23)	27.70 (4.41)	26.13 (4.07)	20.41 (3.63) (33% missing)
<b>Female, N, %</b>	117 (50)	2061 (64)	882 (51.07)	388 (54.11)	573 (95.98)	390 (47.10)	215 (57.95)	284 (48.14)
<b>Smoking status</b>								
<b>Never Smoker, N, %</b>	93 (39.74)	1459 (45.14)	755 (43.77)	209 (29.15)	354 (59.3)	391 (47.22)	165 (44.59)	NA
<b>Former Smoker, N, %</b>	91 (38.89)	942 (29.15)	720 (41.74)	315 (43.93)	178 (29.9)	341 (41.18)	178 (48.11)	NA
<b>Current Smoker, N, %</b>	50 (21.37)	831 (25.71)	250 (14.5)	193(26.92)	65 (10.9)	96 (11.60)	27 (7.30)	NA
<b>FT3, mean (sd) [pmol/L]</b>	4.80 (0.58)	NA	4.86 (0.61)	NA	NA	NA	NA	4.89 (0.66)
<b>FT4, mean (sd) [pmol/L]</b>	13.28 (1.60)	14.29 (2.96)	13.77 (1.91)	15.64 (2.10)	NA	15.16 (2.20)	13.96 (2.29)	12.67 (1.42)
<b>TSH, mean (sd) [mIU/L]</b>	1.31 (0.63)	1.99 (3.62)	1.45 (0.90)	2.26 (1.72)	1.65 (2.23)	2.05 (1.47)	1.77 (1.20)	1.59 (0.92)

Sample sizes per cohort and trait can be found in Figure 1 and in the Supplementary Methods.

**Table 2:** Technical details of the cohorts used in the EWAS analysis.

	<b>SHIP-Trend</b>	<b>ARIC</b>	<b>KORA</b>	<b>RS</b>	<b>TWINSUK</b>	<b>LBC1936</b>	<b>LBC1921</b>	<b>BSGS</b>
<b>Used For</b>	Discovery	Discovery	Discovery	Replication	Replication	Replication	Replication	Replication
<b>Cohort Reference (PMID)</b>	35348705	2646917	16032513	26386597	22253318	22253310, 29546429	22253310, 29546429	22563384
<b>Medium</b>	Whole blood	Whole blood	Whole blood	Whole blood	Whole blood	Whole blood	Whole blood	Whole blood
<b>DNAm platform</b>	EPIC Illumina	450K Illumina	450K Illumina	450K Illumina	450K Illumina	450K Illumina	450K Illumina	450k Illumina
<b>DNAm Array and QC Pipeline</b>	CPACOR <sup>1</sup>	BMIQ <sup>2</sup>	CPACOR <sup>1</sup>	CPACOR <sup>1</sup>	GenomeStudio <sup>3</sup>	BMIQ <sup>2</sup>	BMIQ <sup>2</sup>	inhouse pipeline <sup>4</sup>
<b>Prediction or real proportion of WBC, method if prediction</b>	prediction, Houseman's method	prediction, Houseman's method	prediction, Houseman's method	Houseman's (CD8T and CD4T) and real proportion WBC	prediction, Houseman's method	prediction, Houseman's method	prediction, Houseman's method	prediction, Houseman's method
<b>TSH Assay</b>	Dimension Vista System, Siemens	Elecsys 2010, Roche	Dimension Vista, Siemens	Elecsys 2010, Roche	ARCHITECT, Abbott	ARCHITECT, Abbott	ARCHITECT, Abbott	ARCHITECT, Abbott
<b>FT3 Assay</b>	Dimension Vista, Siemens	Not measured	Dimension Vista, Siemens	Not measured	Not measured	Not measured	Not measured	ARCHITECT Free T3, Abbott
<b>FT4 Assay</b>	Dimension Vista, Siemens	Elecsys 2010, Roche	Dimension Vista, Siemens	Elecsys 2010, Roche	Not measured	ARCHITECT, Abbott	ARCHITECT, Abbott	ARCHITECT, Abbott

<sup>1</sup> see Lehne et al. (2015)<sup>38</sup>; <sup>2</sup> see Teschendorff et al. (2013)<sup>39</sup>; <sup>3</sup> see Tsaprouni et al. (2014)<sup>40</sup>; <sup>4</sup> see McRae et al. (2014)<sup>41</sup>

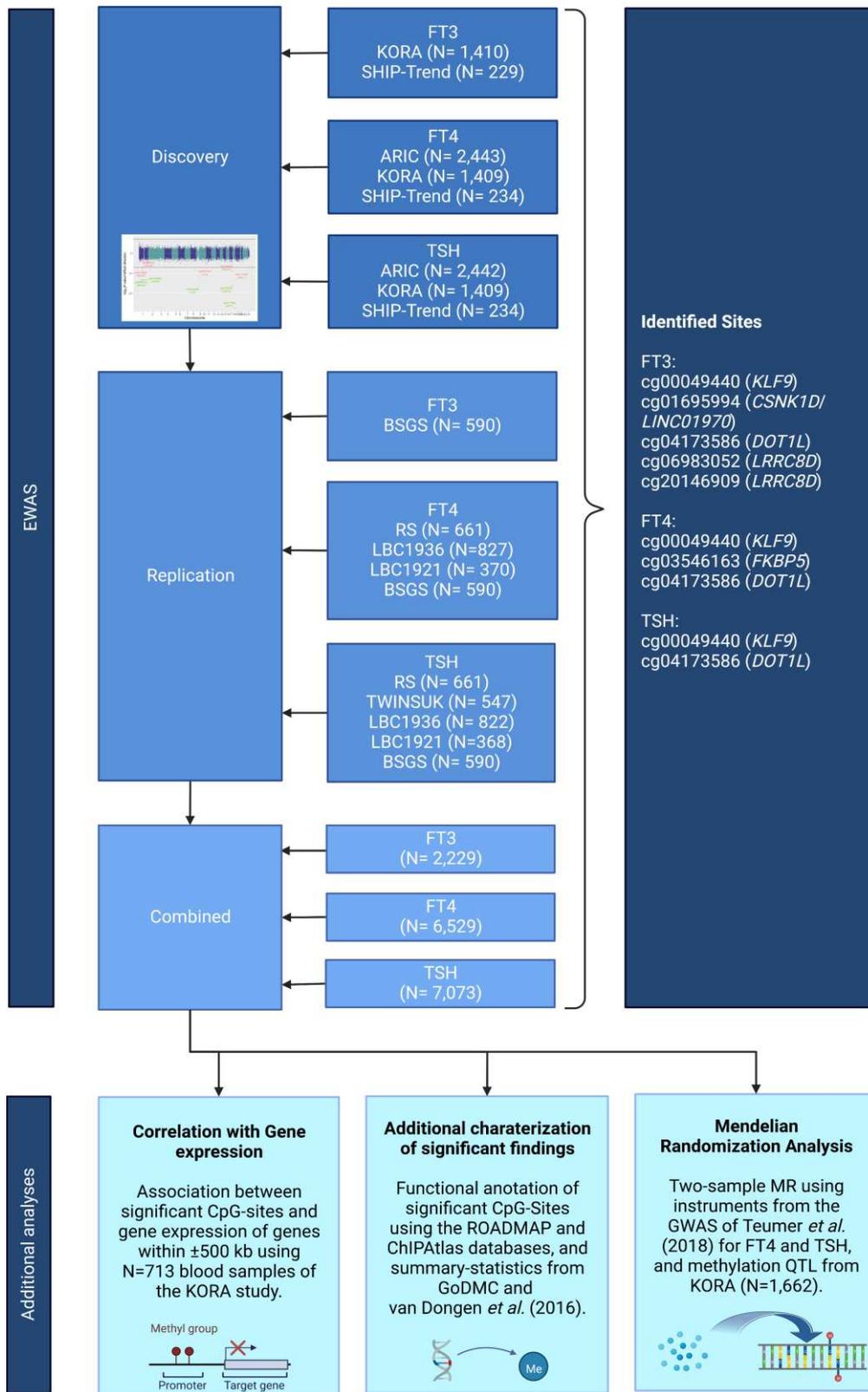
Sample size per cohort and trait can be found in Figure 1 and in the supplementary methods.

**Table 3:** CpG sites significantly associated with thyroid function after the replication stage.

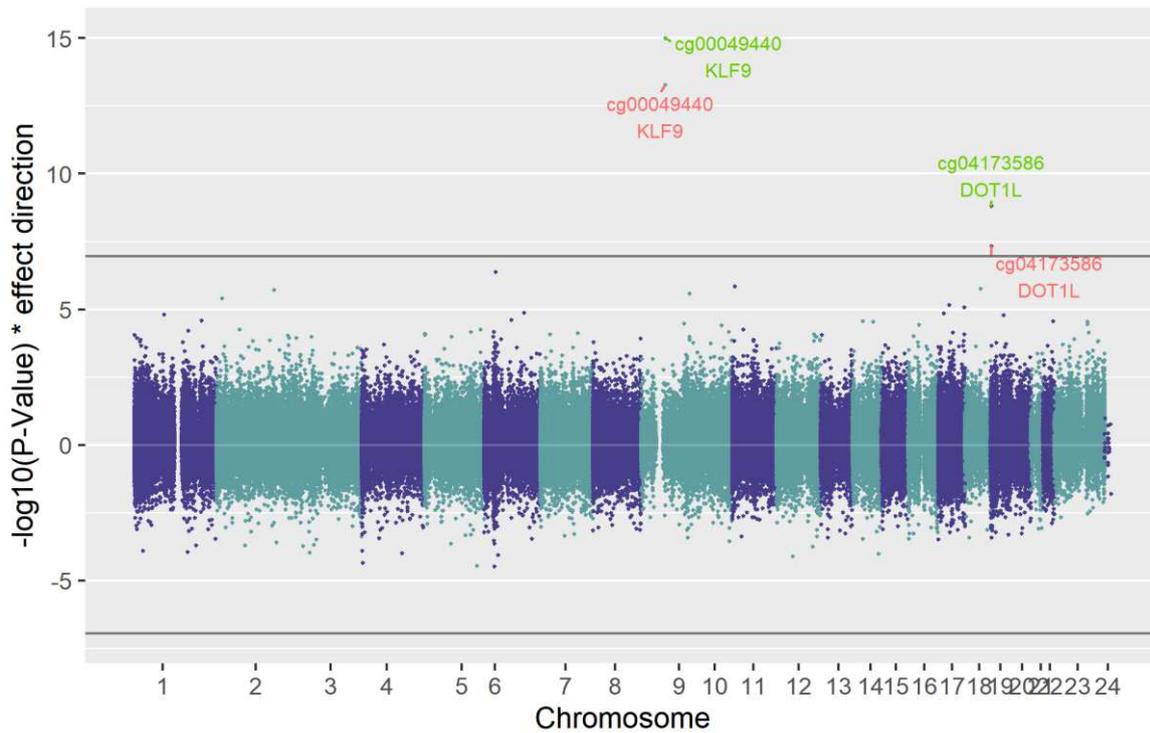
		Discovery			Replication			Combined			
CpG probeID	Chr: Position (build 37)	Estimate	P-value	Sample Size	Estimate	P-value	Sample Size	Estimate	P-value	Sample Size	Nearest Gene
<b><i>FT4</i></b>											
cg00049440	9: 73,026,643	-0.790	4.53E-15	4078	-0.674	1.45E-06	2448	-0.750	<b>4.44E-20</b>	6526	<i>KLF9</i>
cg03546163	6: 35,654,363	-0.849	3.11E-13	3878	-0.825	1.00E-04	1251	-0.843	<b>1.45E-16</b>	5129	<i>FKBP5</i>
cg04173586	19: 2,167,496	-0.445	5.14E-09	4038	-0.380	1.88E-03	2444	-0.427	<b>4.03E-11</b>	6482	<i>DOT1L</i>
<b><i>FT3</i></b>											
cg00049440 <sup>1</sup>	9: 73,026,643	-1.349	2.61E-12	1637	-1.876	1.81E-07	590	-1.463	<b>1.69E-17</b>	2227	<i>KLF9</i>
cg01695994 <sup>1</sup>	17: 80,246,403	-1.112	8.43E-10	1638	-2.497	8.63E-12	590	-1.371	<b>5.84E-17</b>	2228	<i>CSNK1D/LINC01970</i>
cg04173586 <sup>1</sup>	19: 2,167,496	-1.842	7.98E-14	1607	-2.013	1.51E-16	586	-1.935	<b>4.33E-28</b>	2193	<i>DOT1L</i>
cg06983052	1: 90,288,099	-0.875	5.90E-08	1639	-1.474	8.49E-09	590	-1.040	<b>5.45E-14</b>	2229	<i>LRR8D</i>
cg20146909	1: 90,289,611	-0.889	3.38E-09	1639	-1.811	5.67E-08	590	-1.036	<b>6.51E-14</b>	2229	<i>LRR8D</i>
<b><i>TSH</i></b>											
cg00049440	9: 73,026,643	0.143	5.19E-14	4082	0.088	1.11E-03	2988	0.125	<b>9.57E-16</b>	7070	<i>KLF9</i>
cg04173586	19: 2,167,496	0.077	4.52E-08	4041	0.061	8.52E-03	2984	0.072	<b>1.53E-09</b>	7025	<i>DOT1L</i>

<sup>1</sup> known association. The effect estimates and p-values of the discovery and replication combined meta-analysis are provided.

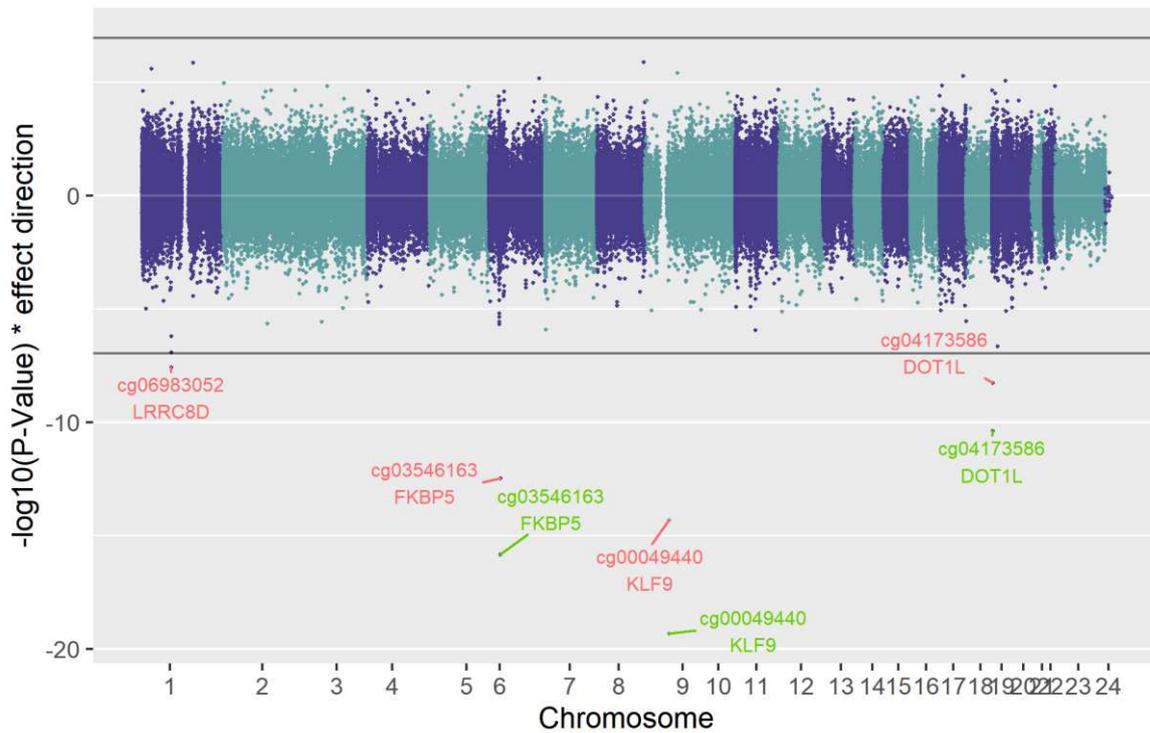
# Figures



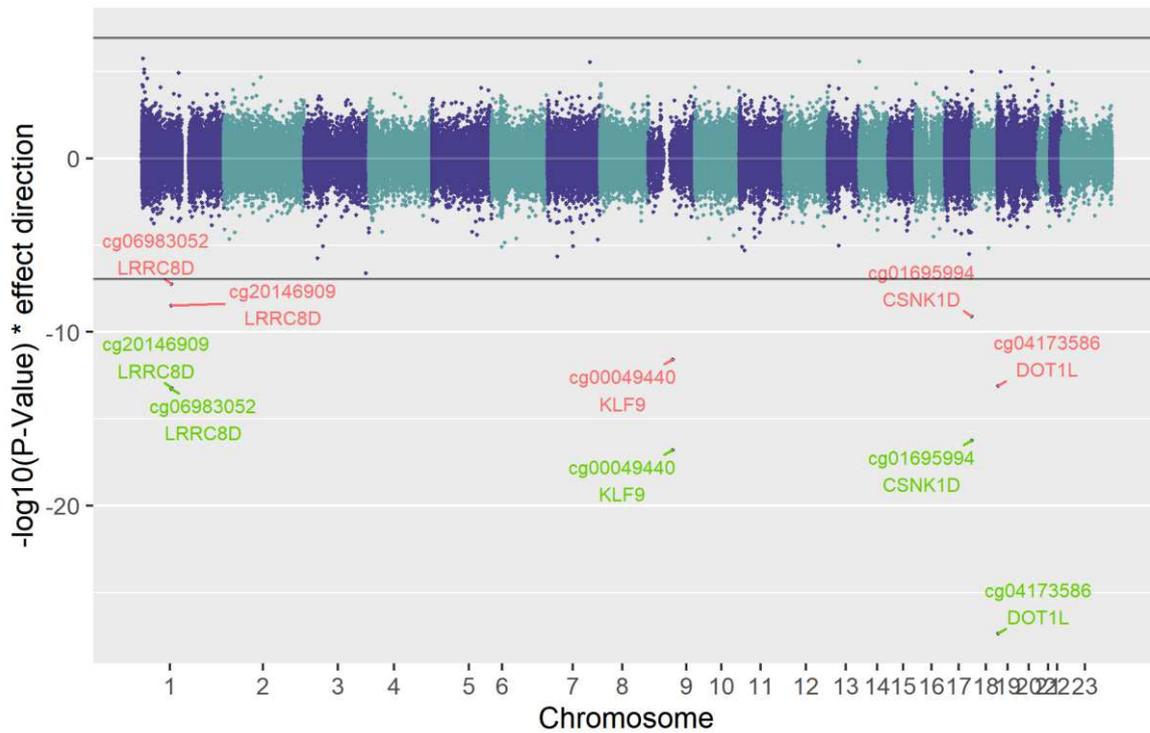
**Figure 1:** Flowchart of the analyses performed in the current study. Image created with BioRender.com



**Figure 2:** Chicago plot showing the results of the TSH discovery analysis. Red labels show the significant sites ( $p$ -value  $< 1.1E-07$ ) of the discovery analysis, while green labels show results of the combined analysis. The black line shows the epigenome wide significance cut-off ( $1.1E-07$ ). Points plotted in the upper panel show a positive correlation with TSH, while those plotted below have a negative association.



**Figure 3:** Chicago plot showing the results of the FT4 discovery analysis. Red labels show the significant sites ( $p\text{-value} < 1.1\text{E-}07$ ) of the discovery analysis, while green labels show results of the combined analysis. The black line shows the epigenome wide significance cut-off ( $1.1\text{E-}07$ ). Points plotted in the upper panel show a positive correlation with FT4, while those plotted below have a negative association.



**Figure 4:** Chicago plot showing the results of the FT3 discovery analysis. Red labels show the significant sites ( $p$ -value  $< 1.1E-07$ ) of the discovery analysis, while green labels show results of the combined analysis. The black line shows the epigenome wide significance cut-off ( $1.1E-07$ ). Points plotted in the upper panel show a positive correlation with FT3, while those plotted below have a negative association.



# Epigenome-wide association study reveals CpG sites associated with thyroid function and regulatory effects on *KLF9*

SUPPLEMENTARY INFORMATION

Supplementary Note: Acknowledgements .....	3
Supplementary Methods .....	6
Study population.....	6
Details on study-specific ethics approvals .....	6
EWAS analyses and meta-analyses .....	7
Correlation with gene expression levels.....	7
EWAS analysis plan provided to the individual cohorts .....	8
Meta-analysis and QC pipeline .....	10
Supplementary Figures .....	12
Supplementary Figure 1: Flow chart of the pre_metaQC() and run_meta() functions .....	12
Supplementary Figure 2: Bacon plots .....	13
Supplementary Figure 3: TSH cg00049440 .....	16
Supplementary Figure 4: TSH cg04173586 .....	18
Supplementary Figure 5: FT4 cg00049440 .....	20
Supplementary Figure 6: FT4 cg04173586 .....	22
Supplementary Figure 7: FT4 cg03546163 .....	24
Supplementary Figure 8: FT3 cg00049440 .....	26
Supplementary Figure 9: FT3 cg04173586 .....	28
Supplementary Figure 10: FT3 cg06983052 .....	30
Supplementary Figure 11: FT3 cg01695994 .....	32
Supplementary Figure 12: FT3 cg20146909 .....	34
Supplementary Figure 13: Forest plots.....	36
References .....	39
Supplementary Tables .....	40

## Supplementary Note: Acknowledgements

CpG probe annotation was supported by software resources provided via the Caché Campus program of the InterSystems Corporation (Cambridge, MA, USA) to Alexander Teumer.

### Cohort acknowledgements and funding sources

#### ARIC

The Atherosclerosis Risk in Communities study has been funded in whole or in part with Federal funds from the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services (contract numbers HHSN268201700001I, HHSN268201700002I, HHSN268201700003I, HHSN268201700004I and HHSN268201700005I), R01HL087641, R01HL059367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. Funding was also supported by 5RC2HL102419 and R01NS087541. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. The work of Anna Köttgen was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 192904750 – CRC 992 Medical Epigenetics and Project-ID 431984000 – CRC 1453 Nephrogenetics. Adrienne Tin is supported by NIAMS grant R01AR073178-01A1, and Dr. Elizabeth Selvin by R01DK089174. Reagents for the thyroid assays were donated by Roche Diagnostics.

#### BSGS

The study samples were collected in the context of the BSGS within the Brisbane Longitudinal Twin Study 1992-2016. This work was supported by the Australian National Health and Medical Research Council (NHMRC) (project grant 1087407, 1031119, 1010374, 496667 and 1046880), the National Institutes of Health (NIH) (grants GM057091 and GM099568), Australian Research Council (A7960034, A79906588, A79801419, DP0212016, DP0343921, DP1093900) and NHMRC Medical Bioinformatics Genomics Proteomics Program (grant 389891) for building and maintaining the adolescent twin family resource, through which samples were collected. Support was also received from The Sir Charles Gairdner Osborne Park Health Care Group Research Advisory Committee (grant 2018–19/015). We thank Anjali Henders, Lisa Bowdler, and Tabatha Goncales for Biobank collection and Kerrie McAloney for collating samples for this study. We also thank Abbott Diagnostics Australia for donating immunoassay reagents. We gratefully acknowledge the participation of the twins and their families. We thank Marlene Grace, Ann Eldridge, and Kerrie McAloney for sample collection and processing;

the staff of the Molecular Epidemiology Laboratory at QIMR for DNA sample processing and preparation; Harry Beeby and David Smyth for IT support; and Dale Nyholt and Scott Gordon for their substantial efforts involving the QC and preparation of the BSGS datasets.

#### KORA F4

The KORA research platform (KORA, Cooperative Health Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München - German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ.

#### Lothian Birth Cohorts of 1921 and 1936

The LBC1921 was supported by the UK's Biotechnology and Biological Sciences Research Council (BBSRC), a Royal Society–Wolfson Research Merit Award to I.J.D., and the Chief Scientist Office (CSO) of the Scottish Government's Health Directorates. The LBC1936 is supported by the BBSRC and Economic and Social Research Council (BB/W008793/1 which support S.E.H.), Age UK (Disconnected Mind project), the Medical Research Council (G0701120, G1001245, MR/M013111/1, MR/R024065/1), and the University of Edinburgh. S.R.C. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (221890/Z/20/Z). Methylation typing in both the LBC1921 and LBC1936 was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland.

#### Rotterdam Study

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the Rotterdam Study participants, the staff involved with the Rotterdam Study and the participating general practitioners and pharmacists.

The generation and management of the Illumina 450K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific

Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). We thank Mr. Michael Verbiest, Ms. Mila Jhamai, Ms. Sarah Higgins, Mr. Marijn Verkerk, and Lisette Stolk for their help in creating the methylation database. We thank Pascal Arp, Mila Jhamai, MarijnVerkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database.

#### SHIP-Trend

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI\_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). DNA methylation data have been supported by the DZHK (grant 81X3400104). The SHIP authors are grateful to Paul S. DeVries for his support with the EWAS pipeline.

#### TwinsUK

TwinsUK is funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR) Clinical Research Network (CRN) and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. This project also received support from the UK Economic and Social Research Council (ESRC ES/N000404/1 to JTB) and JPI-HDHL DIMENSION project (funded by the UK Biotechnology and Biological Sciences Research Council BB/S020845/1 to JTB).

## Supplementary Methods

### *Study population*

The discovery stage EWAS included the following three studies: the Study of Health in Pomerania (SHIP-Trend) whose samples were drawn from the population living in the German Federal State of Mecklenburg-West Pomerania <sup>1</sup>, the Atherosclerosis Risk in Communities (ARIC) study, whose samples were drawn from four US communities, namely Forsyth County, North Carolina, Jackson, Mississippi, suburban Minneapolis, Minnesota and Washington County, Maryland <sup>2</sup>, and the Kooperative Gesundheitsforschung in der Region Augsburg survey F4 (KORA), whose samples were drawn from the population living in and around Augsburg, Germany <sup>3</sup>.

In the replication stage we included the Rotterdam Study (RS) sampled in the Ommoord district, Rotterdam, The Netherlands <sup>4</sup>, TwinsUK consisting of adult twins sampled in the United Kingdom <sup>5</sup>, the Lothian Birth Cohorts of 1921 and 1936 (LBC1921 and LBC1936) consisting of two samples drawn from the Scottish population <sup>6</sup> and the Brisbane Systems Genetics Study (BSGS), which was sampled as part of the Brisbane Longitudinal Twin Study in the region around Brisbane, Australia <sup>7</sup>.

In the discovery stage EWAS, we included 4,085 individuals (ARIC: 2,442, KORA: 1,409, SHIP-Trend: 234) for TSH, 1,639 individuals (KORA: 1,410, SHIP-Trend: 229) for FT3, and 4,081 individuals (ARIC: 2,443, KORA: 1,410, SHIP-Trend: 228) for FT4.

In the replication stage, 2,988 additional individuals (RS: 661, TWINSUK: 547, LBC1936: 822, LBC1921: 368, BSGS: 590) were included in the TSH analysis, 590 (all BSGS) in the FT3 analysis, and 2,448 (RS: 661, LBC1936: 827, LBC1921: 370, BSGS: 590) in the FT4 analysis.

### *Details on study-specific ethics approvals*

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study Personal Registration Data collection is filed with the Erasmus MC Data Protection Officer under registration number EMC1712001. The medical ethics committee of the University of Greifswald approved the study protocol of the SHIP-Trend study (registration number BB 39/08a). The ARIC study was approved by the Institutional Review Board of each ARIC site (IRB No.: 12998/CR517): University of North Carolina at Chapel Hill, Chapel Hill, NC; Wake Forest University, Winston-Salem, NC; Johns Hopkins University, Baltimore, MD; University of Minnesota, Minneapolis, MN; and University of Mississippi Medical Center, Jackson, MS. The QIMR Berghofer HREC ethics committee approved BSGS under the reference number P193. Ethical permission for the Lothian Birth Cohort 1921 study protocol was obtained from the Lothian Research Ethics Committee (LREC/1998/4/183). Ethical permission for the Lothian Birth Cohort 1936 study protocol was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and the Lothian Research Ethics Committee (LREC/2003/2/29). Ethical approval for the TwinsUK study was granted by the National Research Ethics Service London-Westminster, the St Thomas' Hospital Research Ethics Committee (EC04/015 and 07/H0802/84). The KORA cohort ethical approval was granted by the ethics committee of the Bavarian Medical Association (REC

reference number F4: #06068). This covers consent for the use of biological material, including genetics. The KORA data protection procedures were approved by the responsible data protection officer of the Helmholtz Zentrum München.

### *EWAS analyses and meta-analyses*

The EWAS was undertaken by applying a dedicated analysis plan (**Supplementary Methods**), with exclusion criteria defined by missing thyroid hormone concentrations, previously or currently diagnosed thyroid cancer, use of thyroid hormones or antithyroid drugs (Anatomical Therapeutic Chemical [ATC] H03 type drugs, <https://www.whocc.no/atc/>), previously diagnosed thyroid diseases or thyroid surgery. The EWAS were conducted using linear regression with TSH, FT4 and FT3 as the independent variable. The methylation beta-values representing the proportion of methylated DNA at a site were quantile-quantile normalized and included as the outcome in the association analyses. The thyroid hormone concentrations were transformed using the natural logarithm. The models were adjusted for sex, age, age squared, current smoking status and, the percentage of white blood cell subtypes either directly measured or estimated by the Houseman method <sup>8</sup>. Adjustment for technical variables was applied using the principal components calculated from array control probes by the CPACOR pipeline <sup>9</sup> or by array probe location (i.e. sentrix ID and position) where applicable (**Supplementary Table 1**). Family-based studies accounted for the relatedness using a random factor in the association models.

The rule out potential analyses errors and to check the plausibility of the EWAS results, we developed a quality control pipeline (**Supplementary Methods, Supplementary Figure 1**). The results from the cohorts were processed by this pipeline, and then visually inspected for abnormalities. CpG sites with less than 25% of the total sample size were excluded from the subsequent analyses. Additionally, the EWAS results were adjusted for inflation and bias using the R *BACON* package <sup>10</sup>, if the standard deviation of the observed test statistics was bigger than expected under the null (**Supplementary Figure 2**).

Finally, the results were combined using a fixed-effect inverse variance weighted model implemented in the *metafor* R package <sup>11</sup>. Only sites that had DNAm data available in at least 50% of the trait-specific total sample size were included in the final results and subsequent analyses.

### *Correlation with gene expression levels*

The association of the replicated CpG sites with gene expression levels of genes within +/-500 kb vicinity was assessed in up to 713 blood samples of the KORA F4 study using the log2-transformed mRNA levels obtained from the Illumina Human HT-12v3 gene expression array. The gene expression values were regressed on the DNAm beta-values adjusted for sex and age. Prior to the analysis the technical factors as well as the blood cell type proportions were regressed out of the mRNA and DNAm levels, and its residuals were included in the final association model. Annotation and quality control checks of the gene expression probes was based on the table provided in Schurmann *et al.*<sup>12</sup>. All gene expression probes provided in the results passed the annotation-based quality control check.

## EWAS analysis plan provided to the individual cohorts

### A. Phenotypes

We are here interested in 3 numerical phenotypes:

- Continuous blood TSH level
- Continuous blood free T4 level (FT4)
- Continuous blood free T3 level (FT3)

The phenotypes should have been measured at the same sample/time as the DNAm. If this is not the case, please report it. Do not include samples in which the phenotypes and the DNAm were measured more than 5 years apart.

The thyroid hormones could have been measured with different immunoassays/platform, and this can have an influence on the measurements. Please, if your data has been produced using different technologies, keep this information to be used in the models (see covariates section).

We prefer the pre-processing of the array methylation data by CPACOR. Details of the CPACOR pipeline are provided in Lehne *et al.*<sup>9</sup>. Please contact us if you plan to implement CPACOR and need assistance with setting up the pipeline.

### B. Sample Selection

- Patients meeting the following criteria have to be excluded from the analysis:
- Individuals diagnosed with thyroid cancer, previously or after taking blood samples (if information is available)
- Individuals with thyroid medication (H03 type drugs) (if information is available).
- Individuals with previous thyroid disease or operation (if information is available).

### C. Covariates

The following covariates are used in the model:

1. **Age:**  
Age (in years) of the subjects when the blood samples were taken
2. **Age square:**  
Age (in years) squared of the subjects when the blood samples were taken
3. **WBC proportions:**  
When available, the measured total WBC and WBC subtypes (CD8T + CD4T + NK + Bcell + Mono + Gran). Otherwise, the WBC proportions should be estimated using the Houseman method (Houseman et al., 2012) (and use total WBC measured if available).
4. Time between the phenotype and DNA methylation measurements (if the measurements were not performed at the same time, otherwise not applicable)
5. **Immunoassay (if different assays were used for same phenotype within the cohort, otherwise not applicable):**  
If the reference range of the different immunoassay is different and this impacts

the measurements obtained, the immunoassay information should be included as a covariate in all models as a fixed effect.

6. **Smoking:**  
as a factor with 2 levels (0=never and ex/past, 1 = current smoker)
7. **Batch effects (if available):**  
Technical covariates from the methylation array (such as Sentrix\_ID and Sentrix\_Position and batch)
8. **Gender:**  
as factor with 2 levels (0=female and 1=male)
9. Family structure (if data comes from a family-based study)
10. **Principal Components (if available):**  
principal components obtained from the CPACOR pipeline

#### D. Model

Linear or linear mixed models will be used to test for association between each **phenotype (as an independent variable)** and **methylation as the dependent variable (quantile-quantile normalized beta-value)**. The model will be run three times, once for every phenotype (X = TSH, FT4 and FT3), while correcting for all possible confounders.

Batch and family structure variables are defined as random effects, the others as fixed effects. All covariates, unless specified, should be coded as numeric.

The beta value of the CpG sites are **quantile-quantile normalized** and a **natural log transformation** is performed on the thyroid function parameters.

Covariates in grey are cohort-dependent and have to be added to the model only if available or applicable. Control probe specific PCs obtained from the CPACOR pipeline may also be added if appropriate.

#### Model (R-code example):

```
Qqnorm(beta(CpG)) ~ ln(X) + age + age2 + as.factor(sex) + as.factor(current_smoking) +
%WBCs + %Bcell + %NK + %Mono + %Gran + %CD8T + %CD4T+ [diff_time +immunoassay
+as.factor((1|Sentrix_ID))+ as.factor(as.factor((1|Sentrix_Position))) +
as.factor((1|family_structure))] + Principal Components
```

#### E. Data Reporting

Please provide the EWAS results in a CSV file containing the following columns:

Column	Description
Marker name	CpG site (cgXXXXXXXX)
Estimate	beta from phenotype-methylation association, at least 5 decimal places -- "NA" if not available
SE	Standard error of beta estimate, to at least 5 decimal places -- "NA" if not available
Pval	p-value of test statistic, -- "NA" if not available

mean	mean of DNA methylation beta values (raw values, before QQnorm) across all samples
sd	SD of DNA methylation beta values (raw values, before QQnorm) across all samples
n_total	total sample with phenotype and methylation for each CpG site

### Meta-analysis and QC pipeline

To reproducibly conduct the centralized analyses, we created an R package, which automates the quality control (QC) and meta-analysis steps. It can be downloaded from <https://github.com/antoineweihs/MetaPipeline>, and a graphical overview can be seen in Supplementary Figure 1. Overall, the package loads and aggregates the data using a summary file containing the file location and name, the phenotype and gender that should be analysed, as well as the names of the column headers. Then one of two pipelines can be run, the first, *pre\_metaQC()*, performs preliminary quality control steps, that should be run before the meta-analysis. The function is divided into two sections, the first being a visual inspection of the data sets. Since all cohorts followed a similar experimental set up, the distributions of the outcome are expected to look similar. The pipeline therefore creates density plots of the effect, standard error and p-value distribution for each EWAS result file in order to enable a visual comparison of the data and identification of errors. To identify problems concerning genomic coverage, two more plots are created, which show the genomic position of each CpG site and the number of CpG sites per chromosome in relation to how many are present on the Illumina chip.

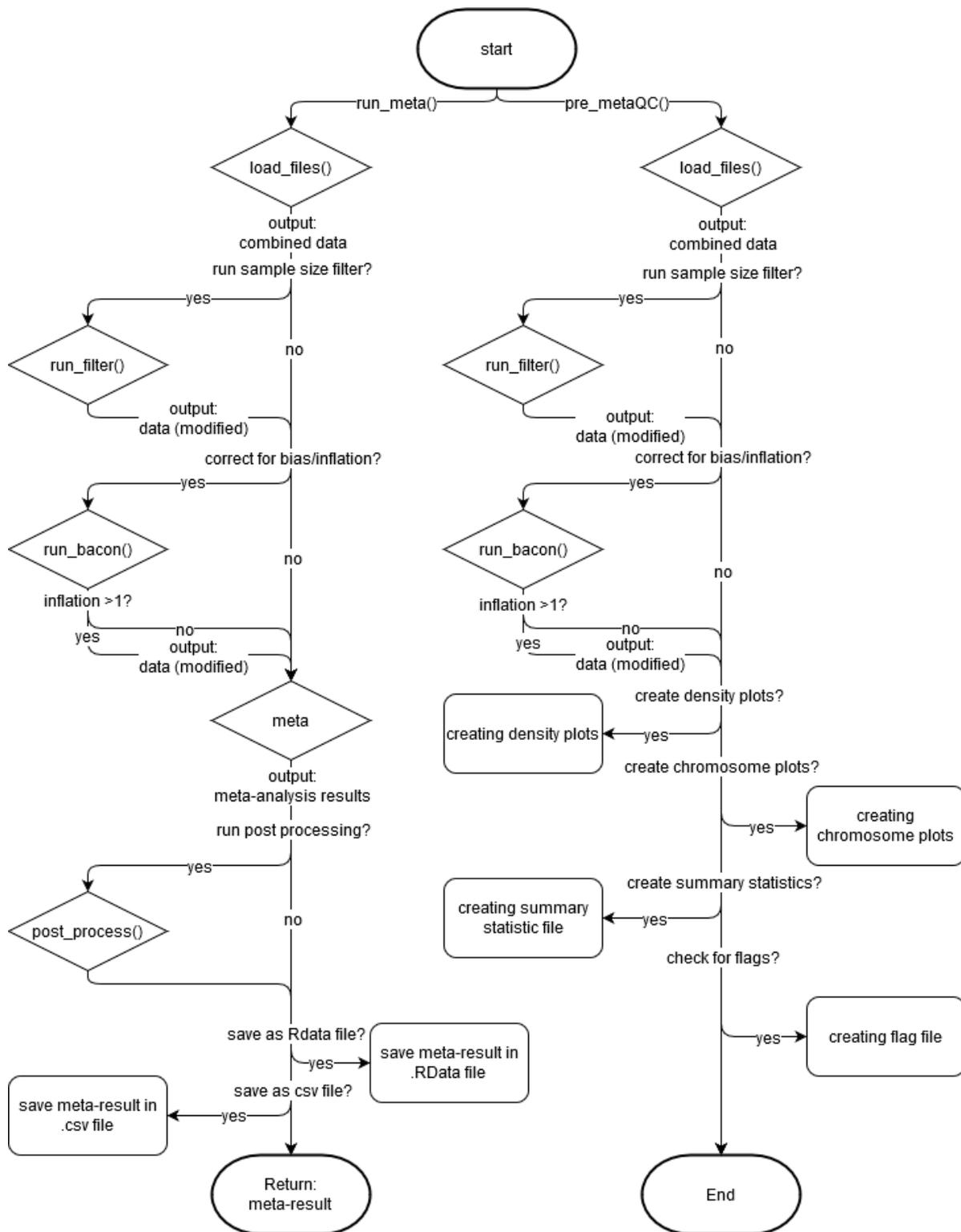
In addition, the quality control method provides descriptive statistics of the EWAS result files. A statistics summary file contains the minimum, maximum, mean, median, 1<sup>st</sup> and 3<sup>rd</sup> quartile values and the number of missings of the effect, standard errors, p-values and sample size columns for each cohort, as well as the *bacon* inflation and bias values<sup>10</sup>. To create the flags file, *preMeta\_QC()* compares various parameters in the data sets to threshold values, and flags instances where the threshold is crossed. Parameters currently considered by the function are extreme effect, standard error and inflation values, different numbers of missing values in the effect, standard error and p-value columns, missing chromosomes, too few decimal places and implausible standard error or p-values.

The second pipeline, *run\_meta()* performs the actual meta-analysis. After loading the EWAS results of the individual cohorts, a pre-processing step is performed by including two adjustments in the analysis pipeline, namely a sample size filter and an inflation/bias adjustment. The sample size filter is used because CpG sites with low sample sizes (i.e. lots of missing values) indicate technical problems occurring at that site often resulting in extreme effect and standard error values. To take the different sample size per individual study into account, a variable study-specific cut-off was applied based on given percentage of the maximal sample size of that study (default: 25%). The inflation/bias adjustment uses the *bacon* package that uses an empirical Bayes approach to estimate the inflation and bias present in each cohort. If the inflation is estimated to be above 1, the effect and standard error values from that cohort will be adjusted, otherwise it will not. After the pre-processing, a fixed-effect inverse variance meta-analysis is performed using the algorithms implemented in the *metaphor* package<sup>11</sup>. Post-processing

functions of our package include double Manhattan plots of the results, forest plots and annotation plots for each significant site, and annotation of the CpG sites with various genetic and epigenetic information such as ROADMAP ChromHMM annotation<sup>13</sup>. Finally, the results can be saved in an RData object or text file.

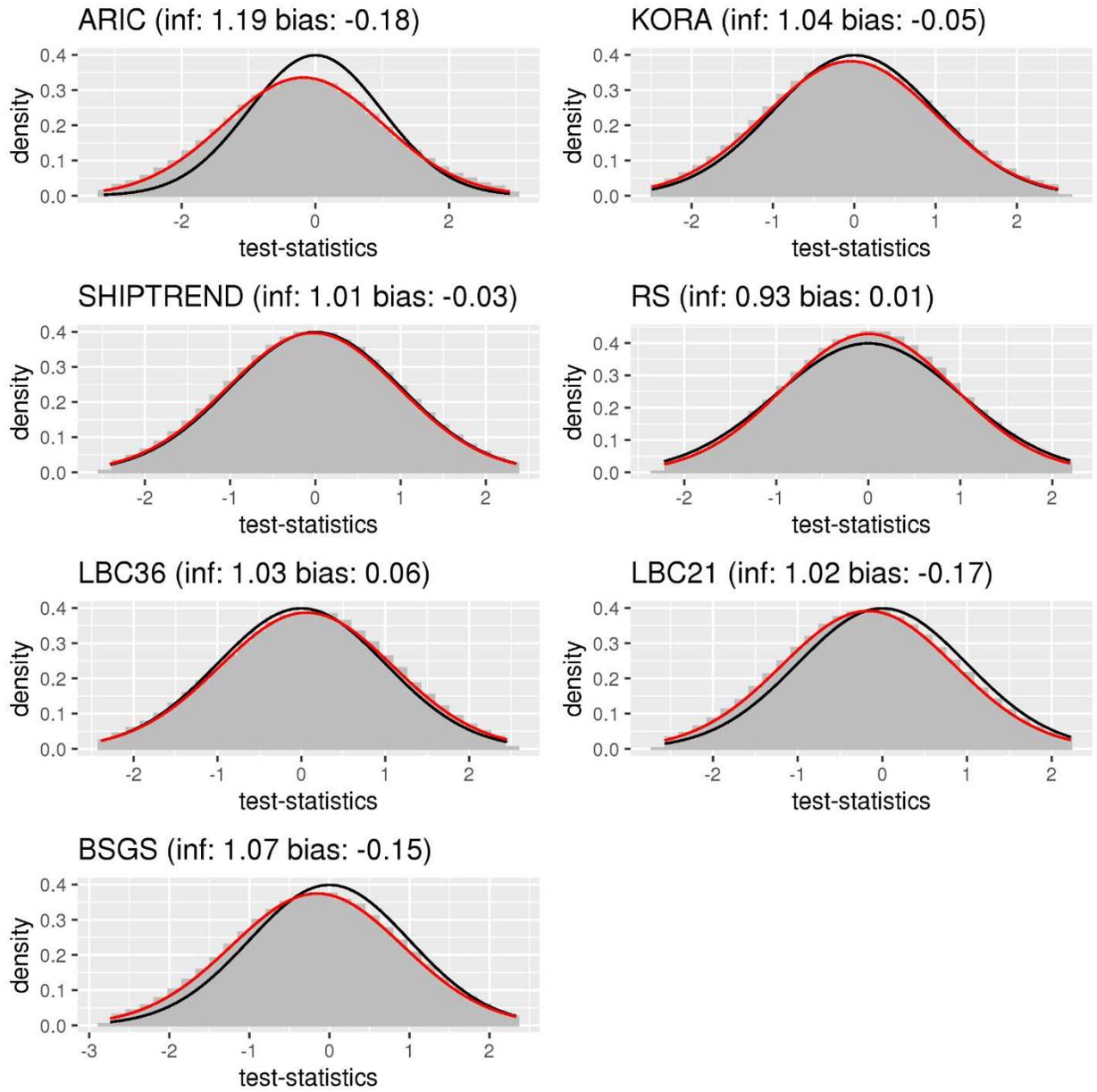
## Supplementary Figures

Supplementary Figure 1: Flow chart of the `pre_metaQC()` and `run_meta()` functions

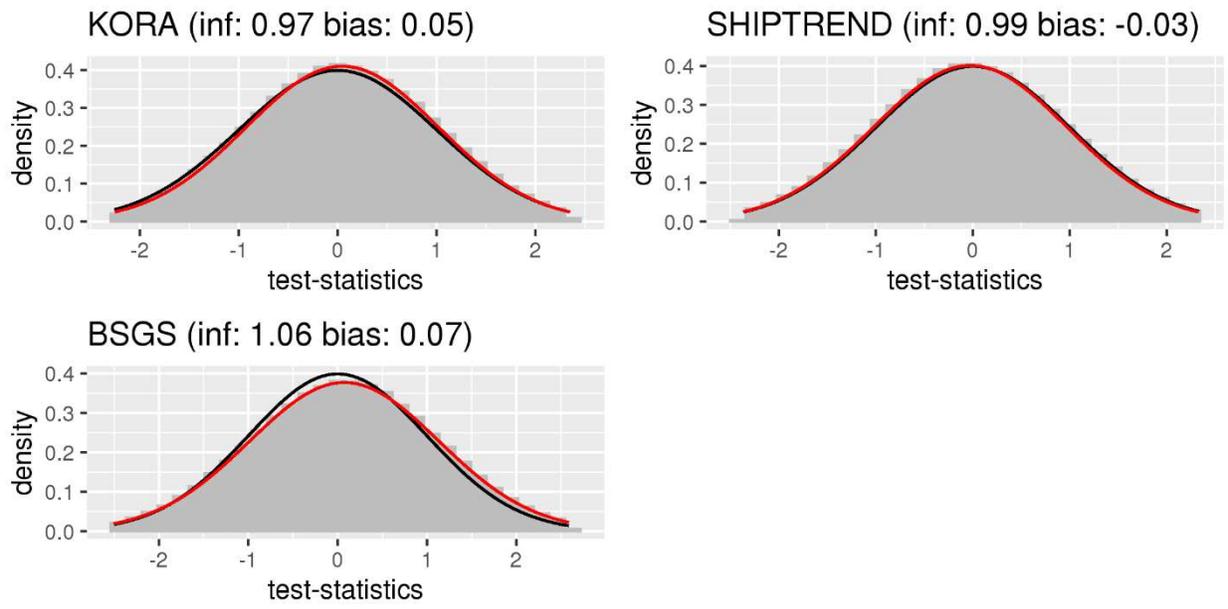


Supplementary Figure 2: Bacon plots

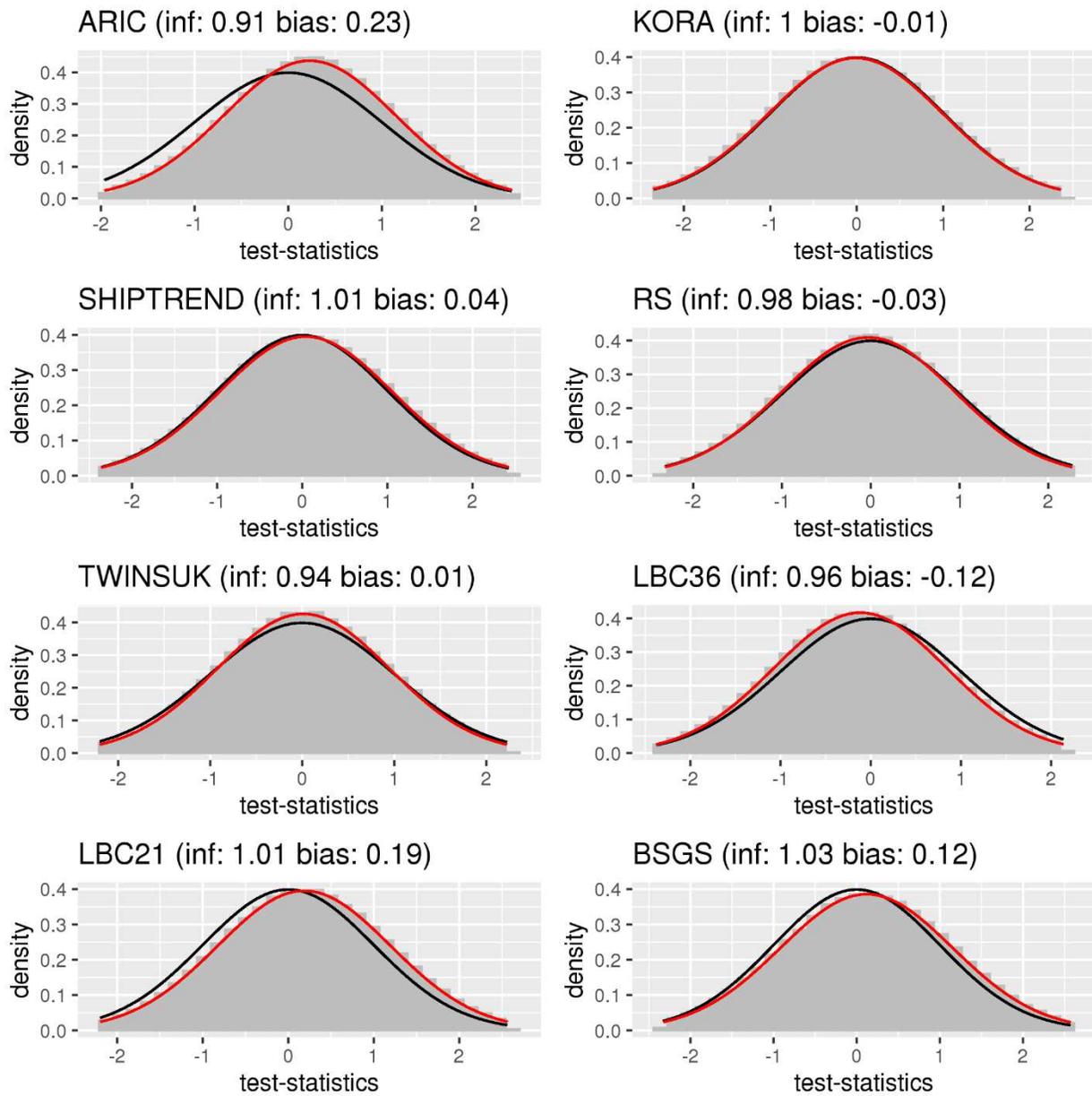
a



b

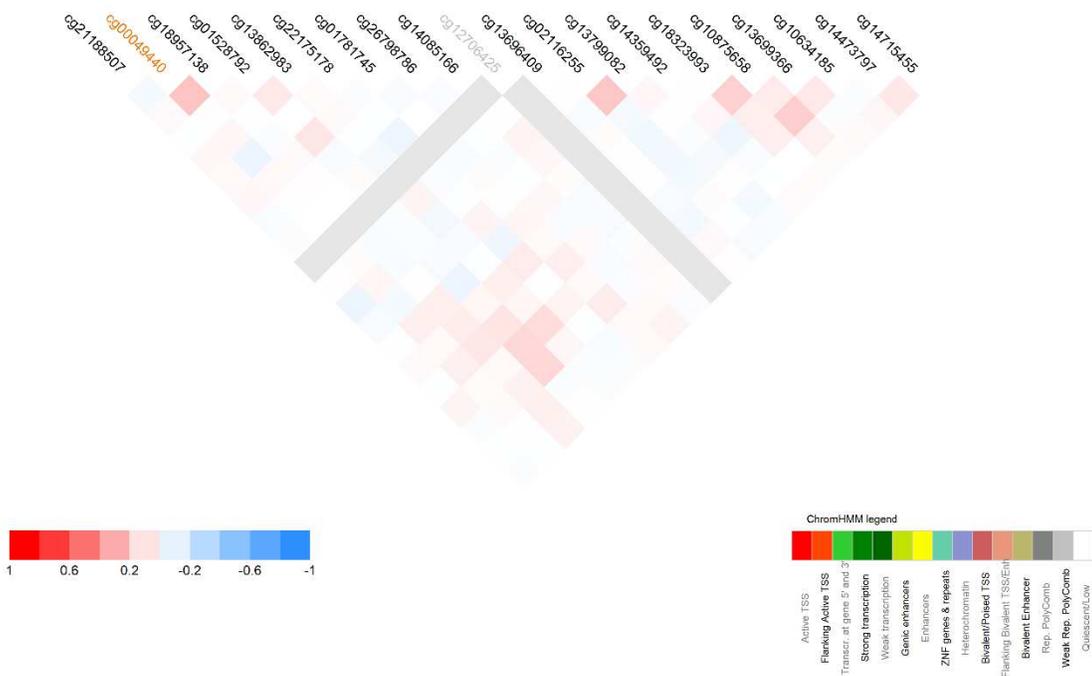
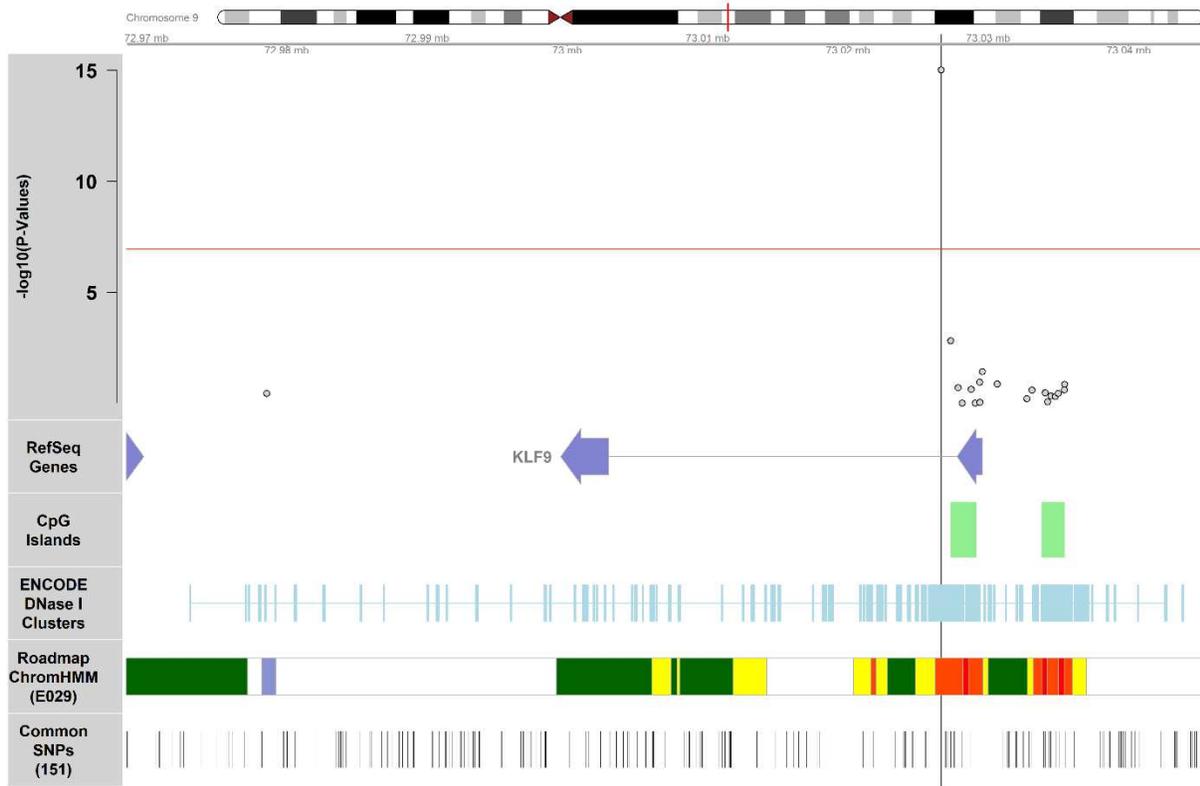


c



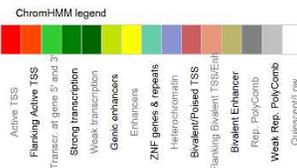
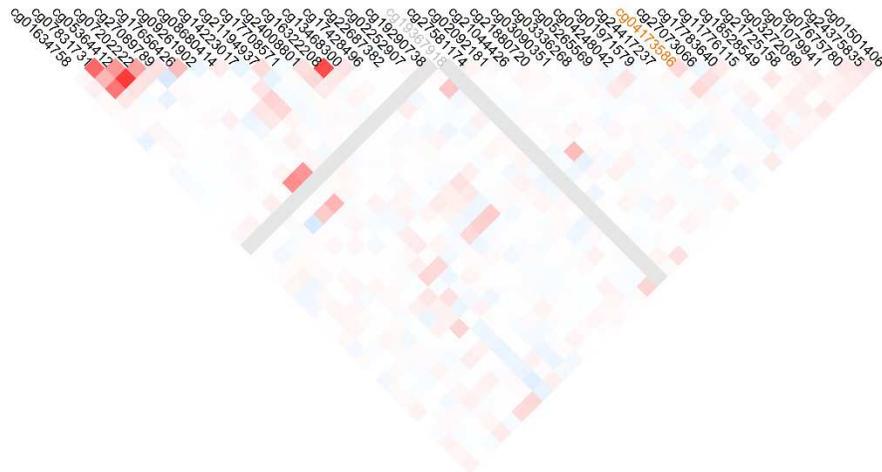
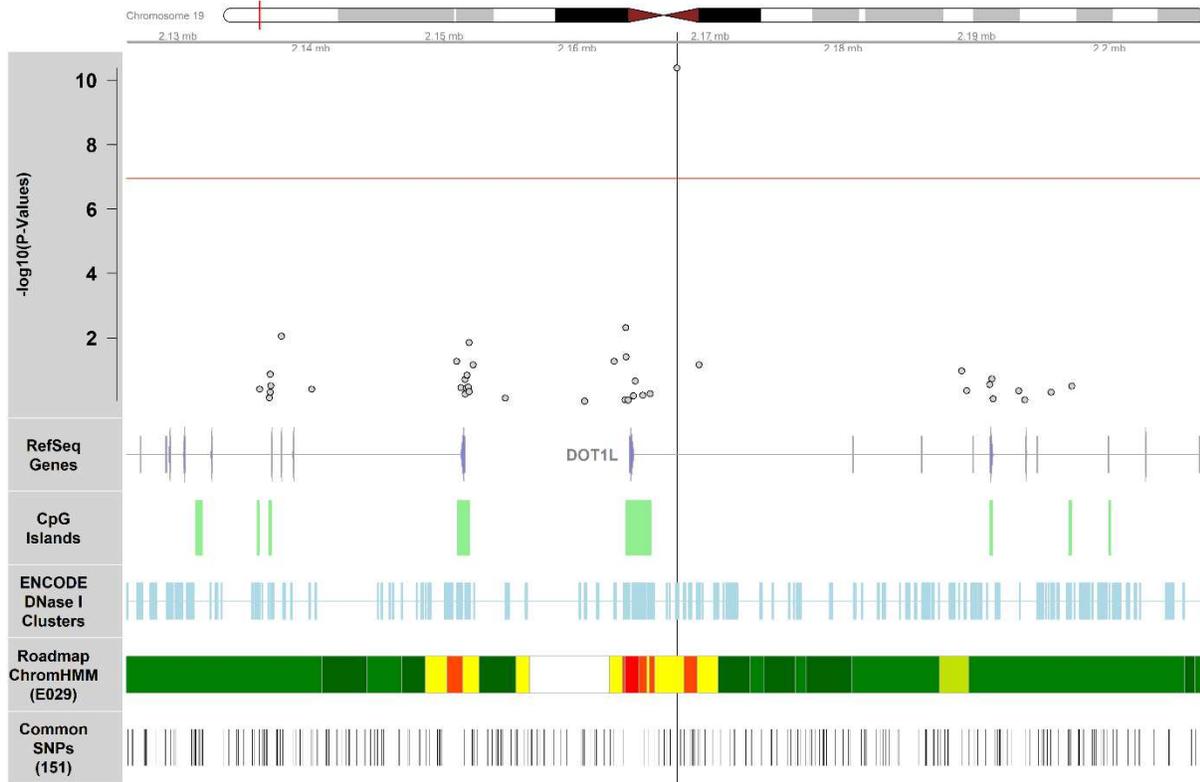
Test statistic distribution of the cohort-specific EWAS results for a) free T4, b) free T3, and c) TSH. The black line represents the expected normal distribution, while the red line represents the actual distribution estimated by bacon <sup>10</sup>.

Supplementary Figure 3: TSH cg00049440



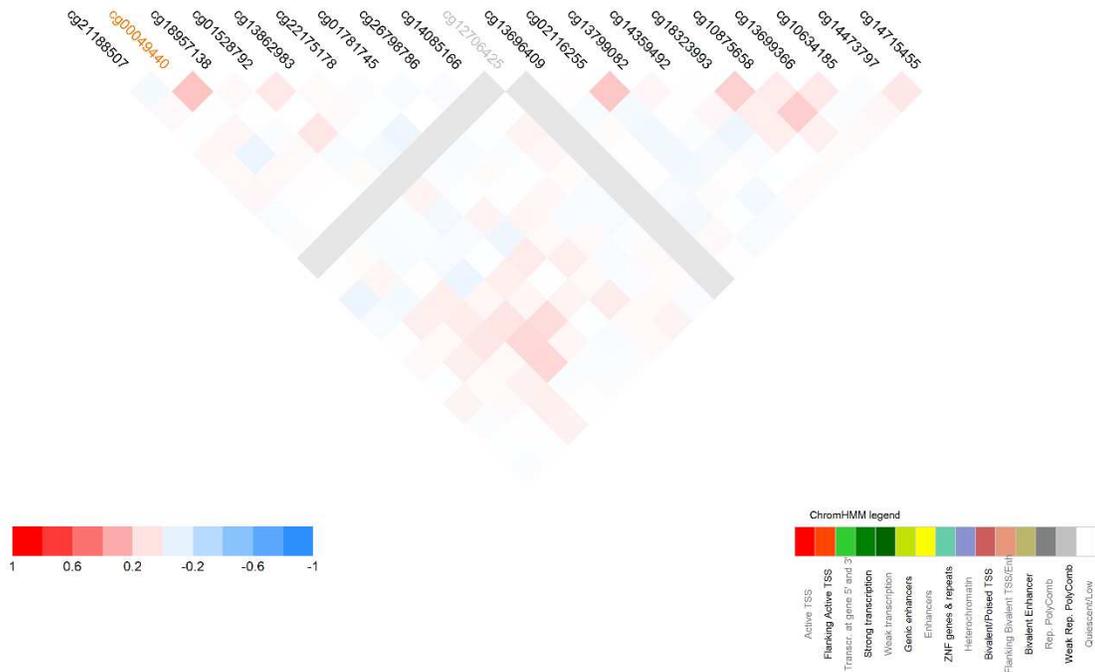
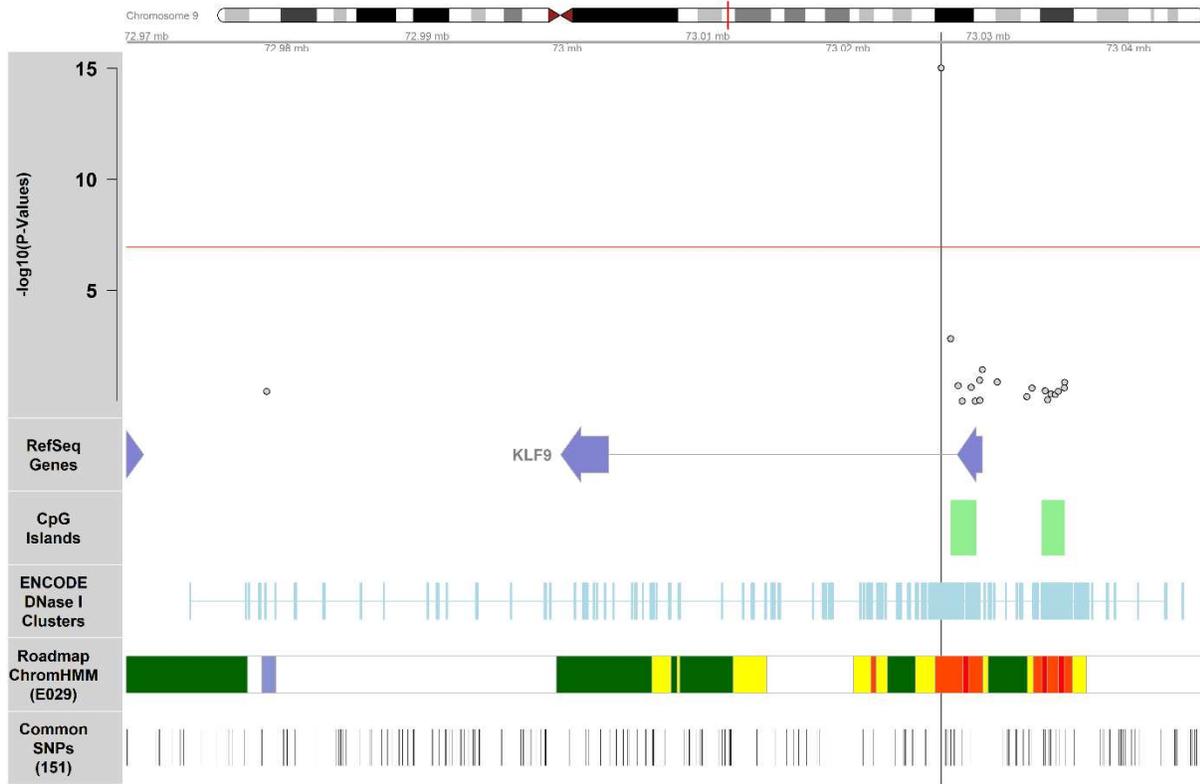
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of  $1.1E-07$ . The correlations were calculated in the KORA cohort.

Supplementary Figure 4: TSH cg04173586



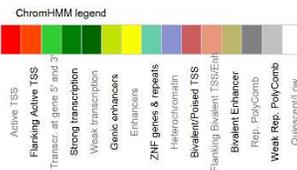
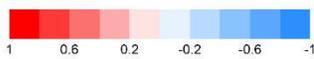
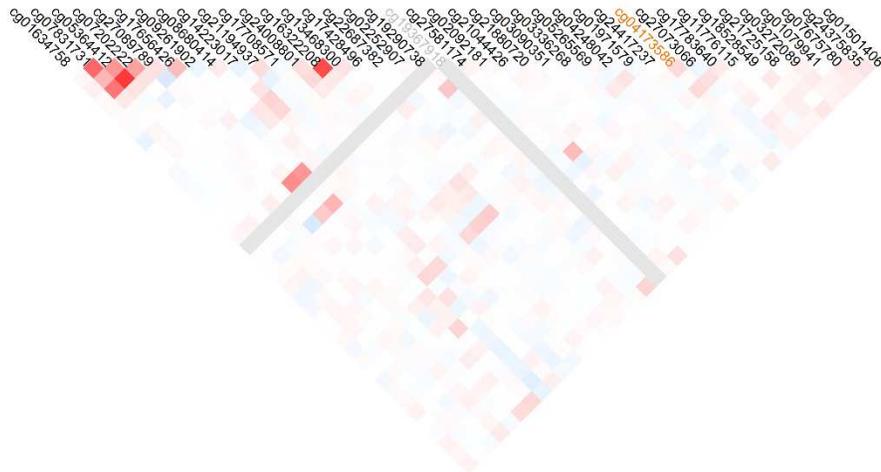
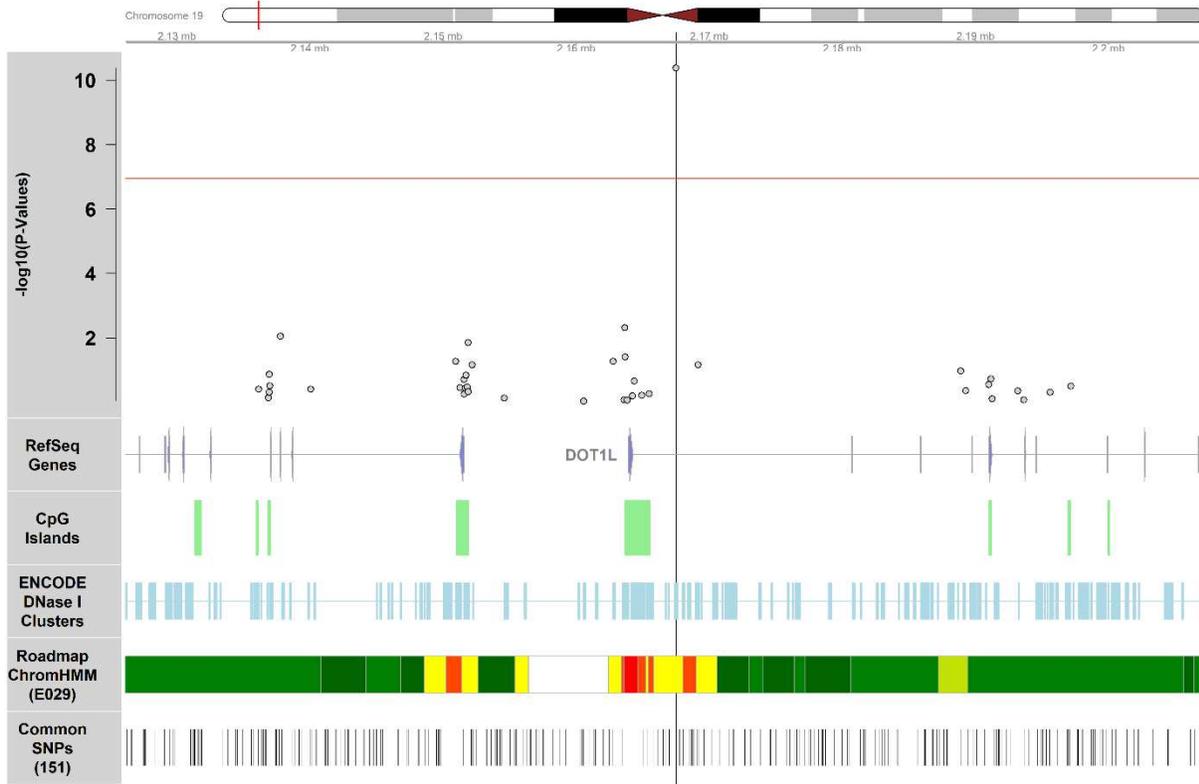
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of 1.1E-07. The correlations were calculated in the KORA cohort.

Supplementary Figure 5: FT4 cg00049440



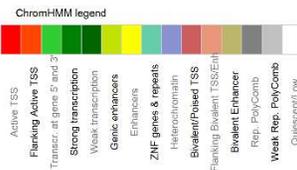
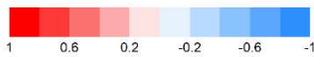
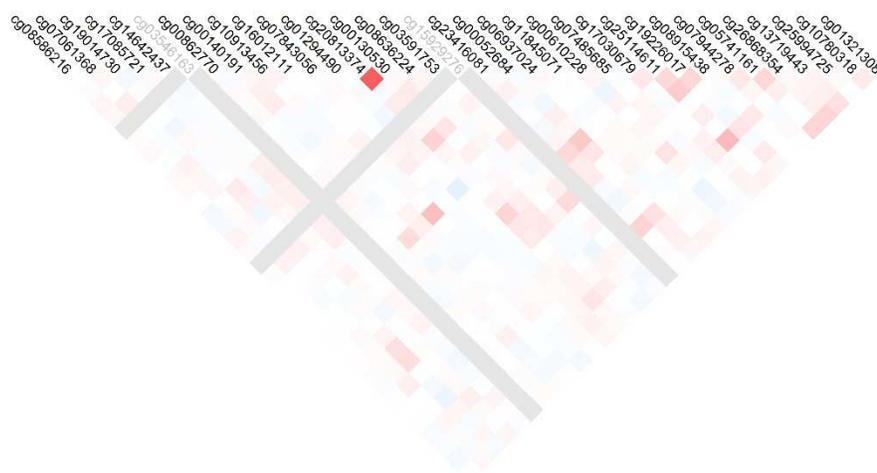
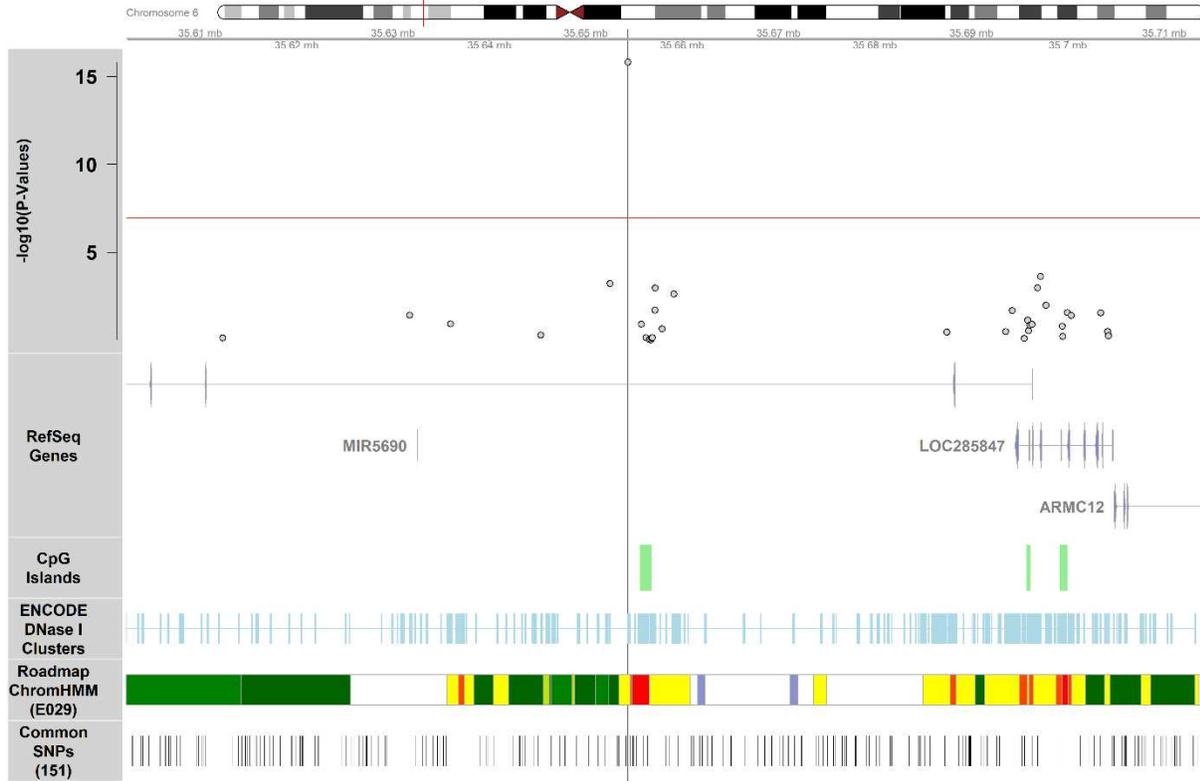
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of 1.1E-07. The correlations were calculated in the KORA cohort.

Supplementary Figure 6: FT4 cg04173586



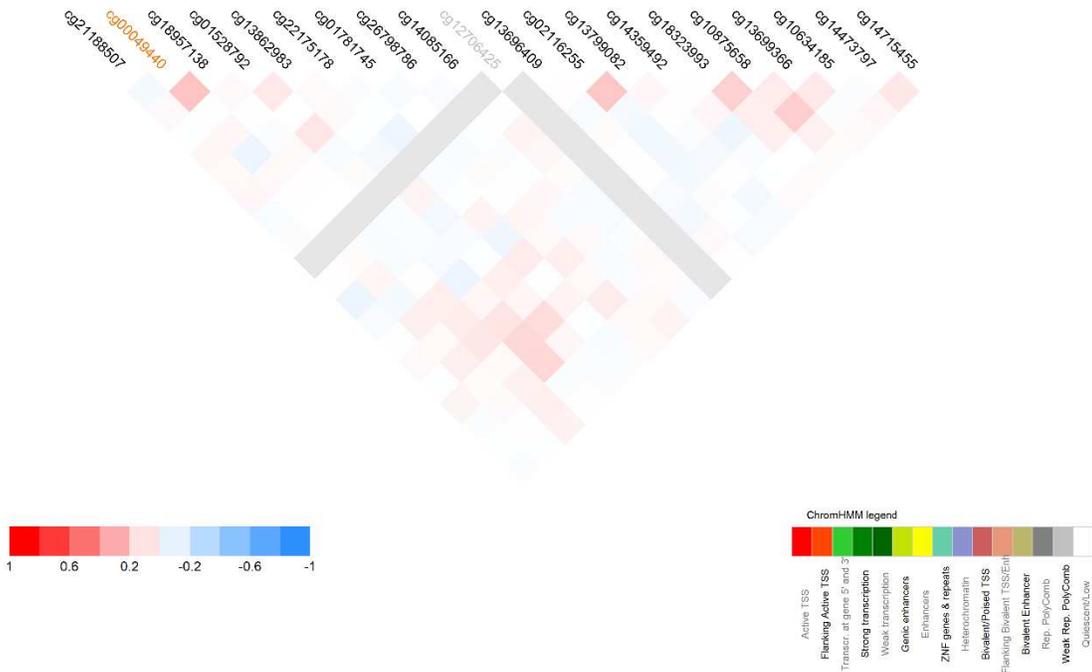
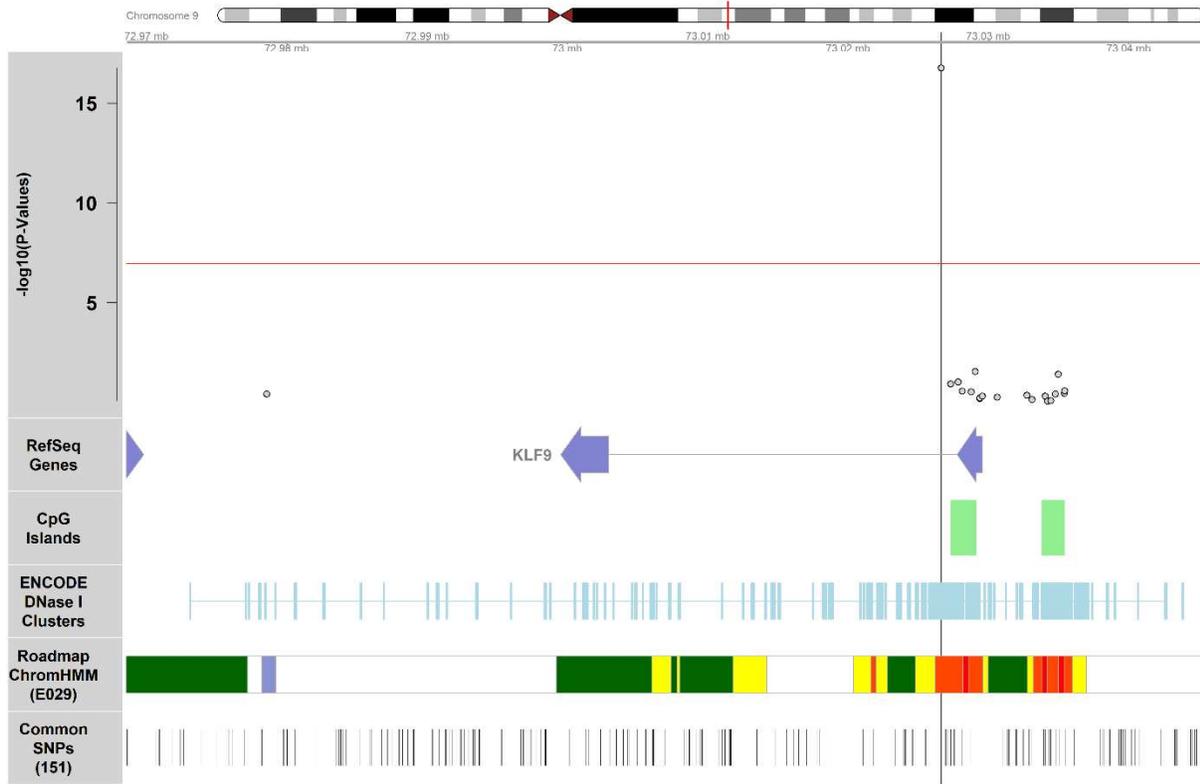
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of 1.1E-07. The correlations were calculated in the KORA cohort.

Supplementary Figure 7: FT4 cg03546163



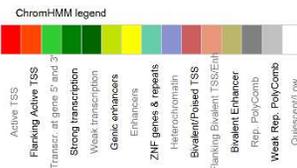
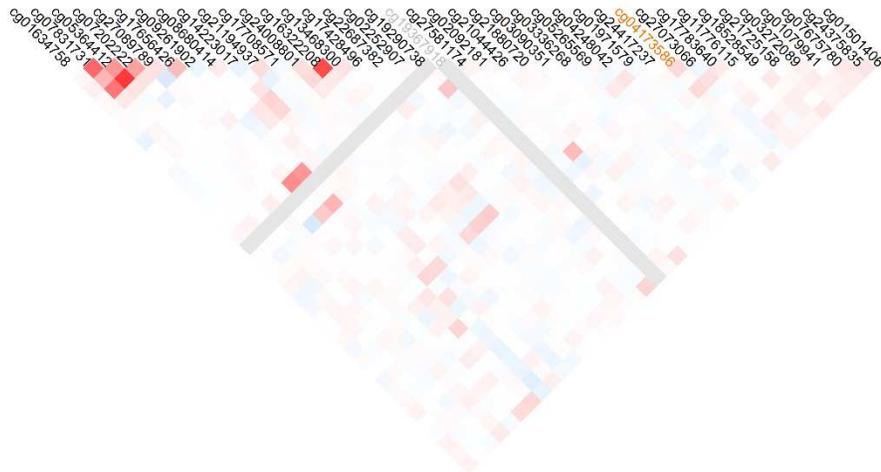
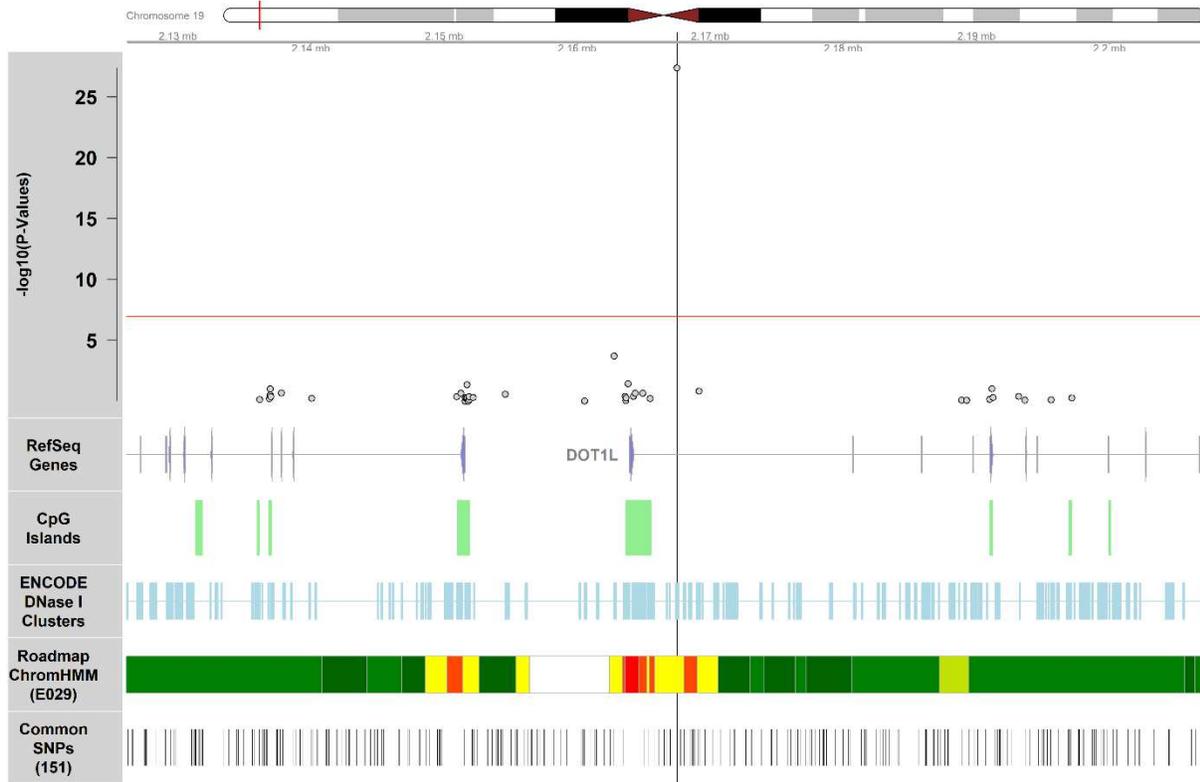
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of 1.1E-07. The correlations were calculated in the KORA cohort.

Supplementary Figure 8: FT3 cg00049440



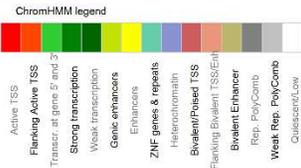
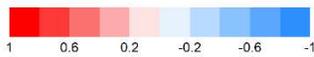
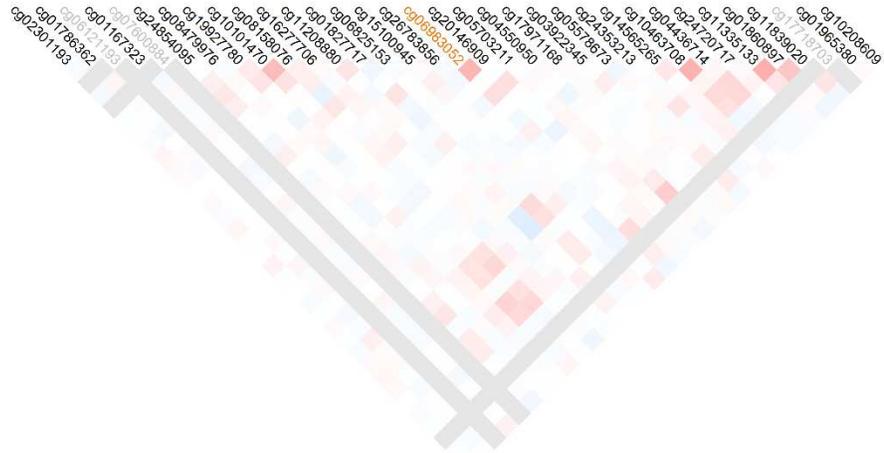
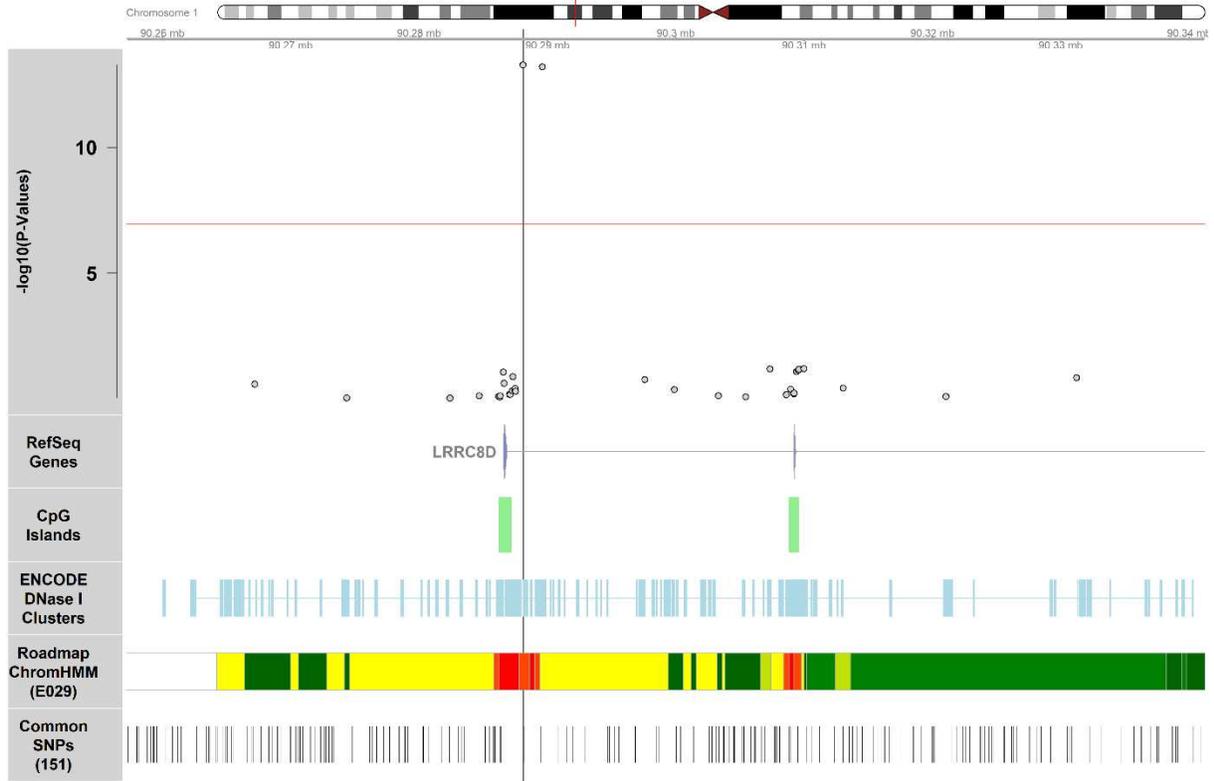
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of 1.1E-07. The correlations were calculated in the KORA cohort.

Supplementary Figure 9: FT3 cg04173586



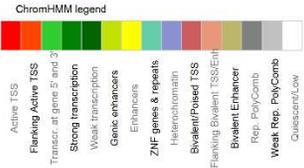
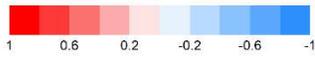
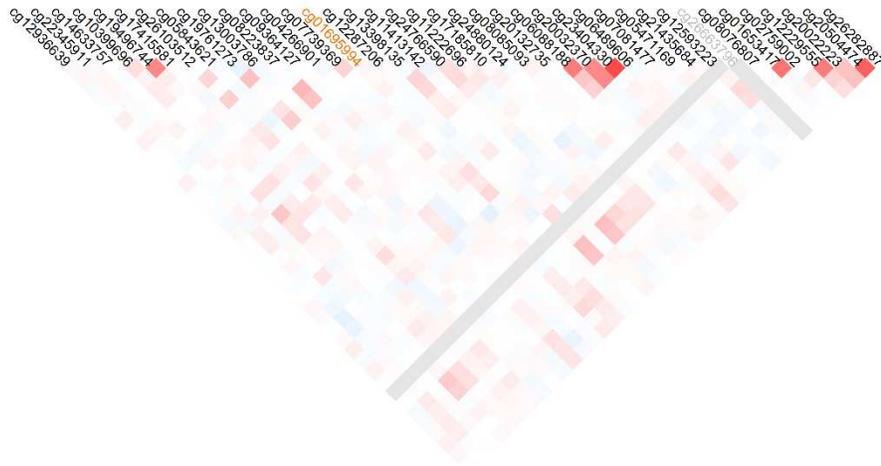
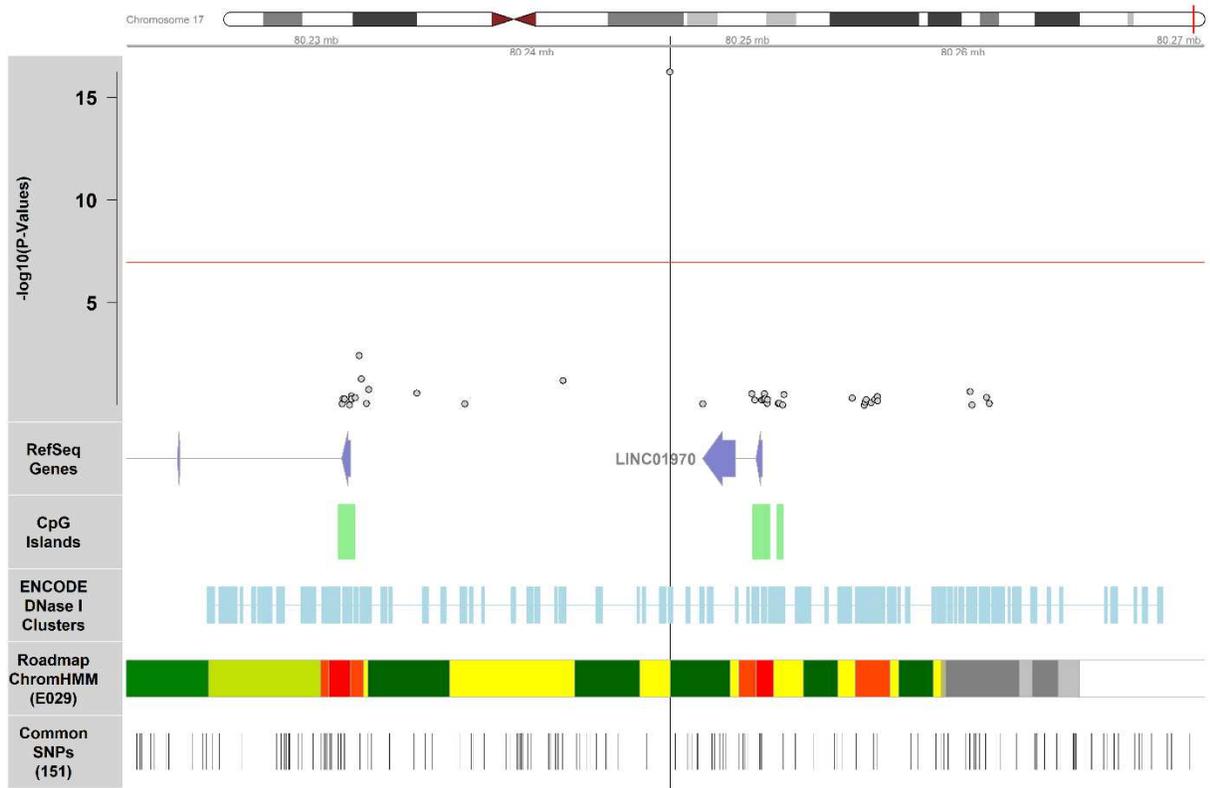
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of  $1.1E-07$ . The correlations were calculated in the KORA cohort.

Supplementary Figure 10: FT3 cg06983052



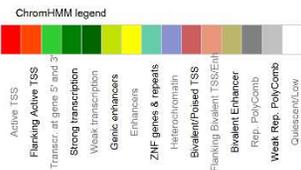
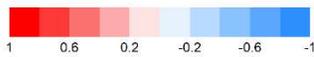
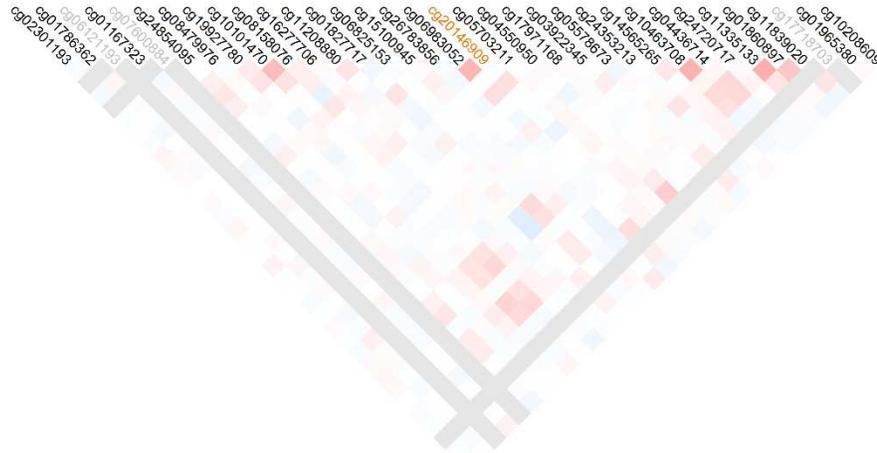
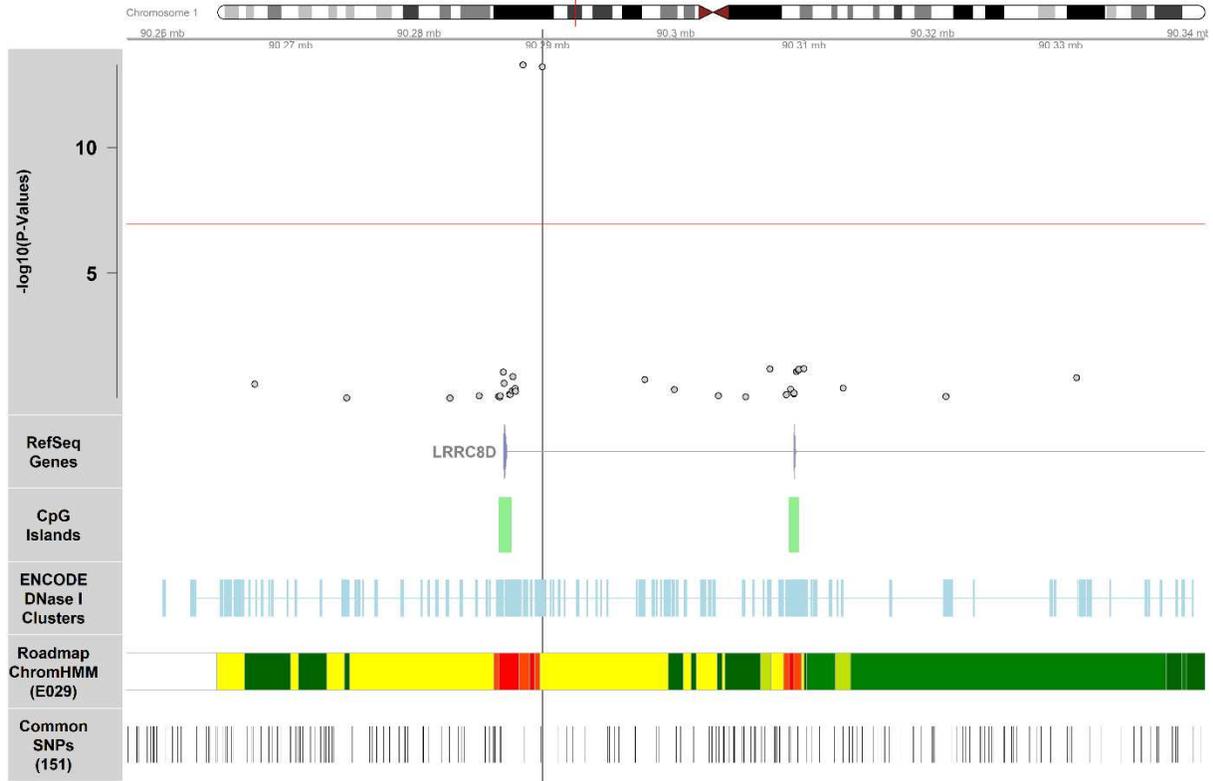
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of 1.1E-07. The correlations were calculated in the KORA cohort.

Supplementary Figure 11: FT3 cg01695994



P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of 1.1E-07. The correlations were calculated in the KORA cohort.

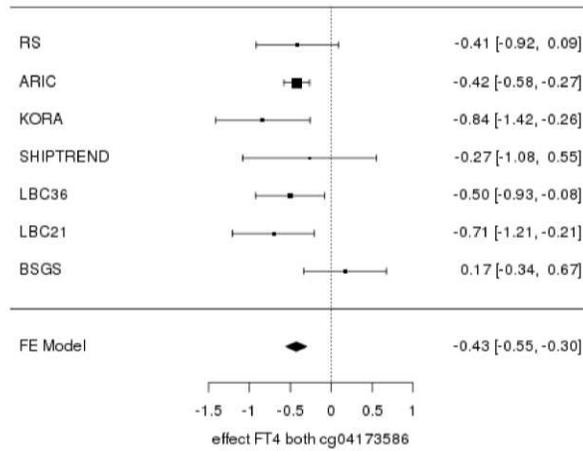
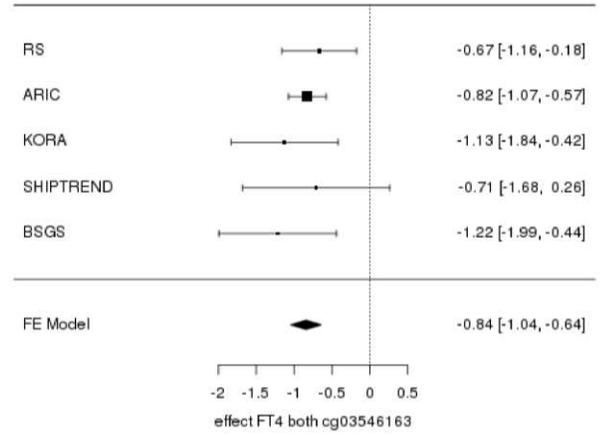
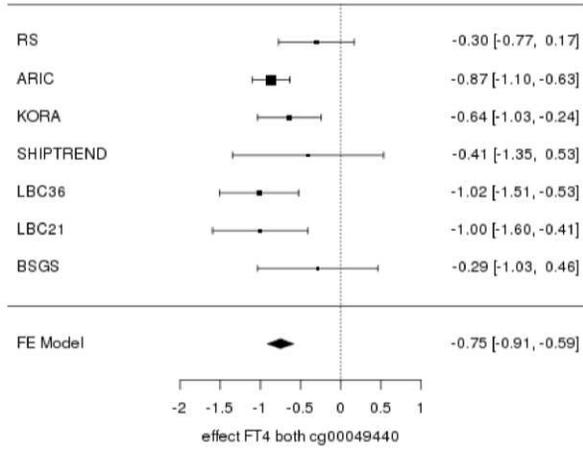
Supplementary Figure 12: FT3 cg20146909



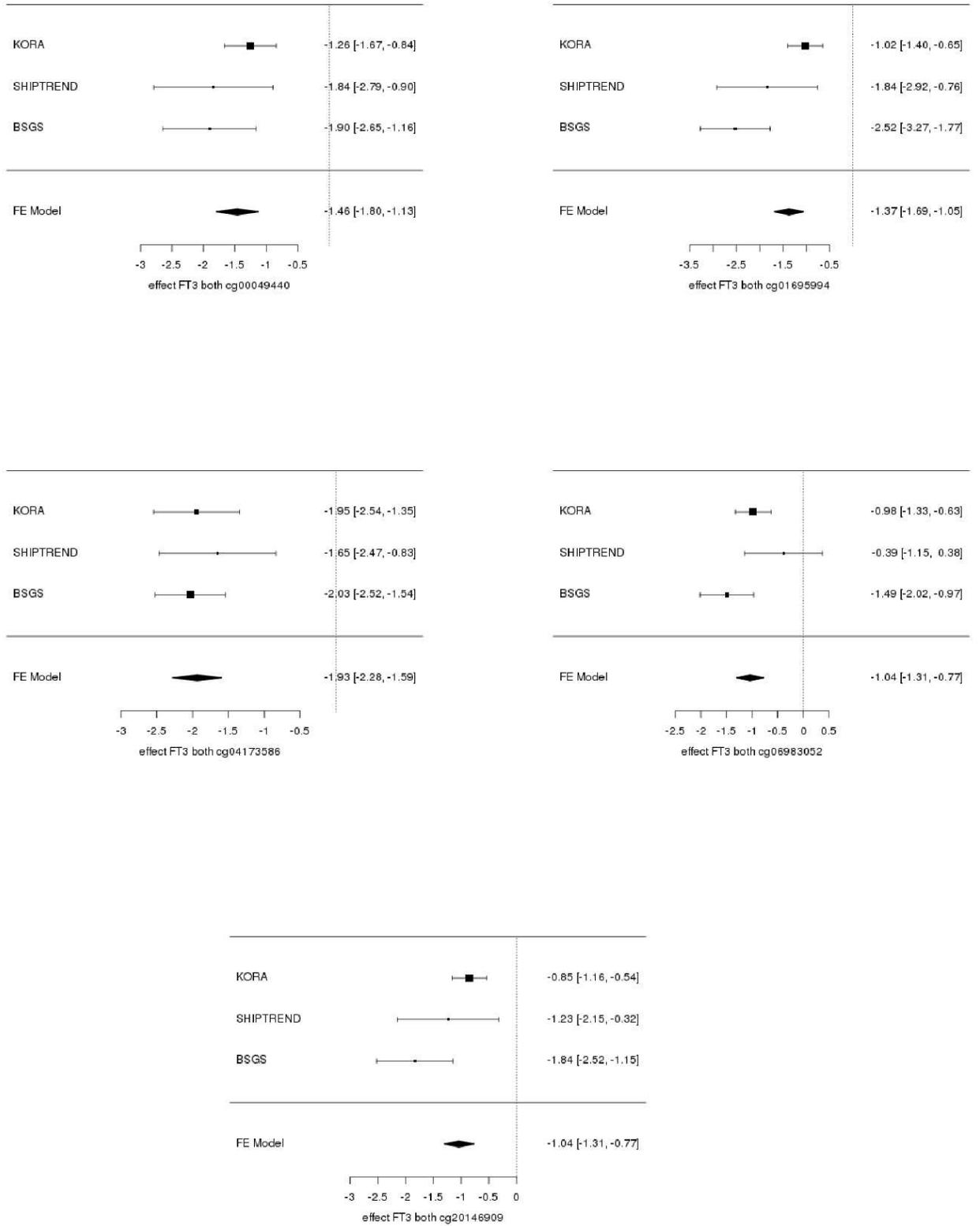
P-Values come from the combined analysis, the UCSC Genes track was extracted from the UCSC RefSeq track, the CpG Islands track was extracted from the UCSC CpG Island track, the Roadmap ChromHMM track was extracted from ROADMAP epigenomics project's core 15-state model using the E029 epigenome (primary monocytes from peripheral blood) and the common SNPs 151 track was extracted from the UCSC dbSNP 151 (common) track <sup>13,14</sup>. The red line indicates the epigenome-wide significance level of 1.1E-07. The correlations were calculated in the KORA cohort.

Supplementary Figure 13: Forest plots

a



b



C



Forest plots of the CpG sites significantly associated with a) FT4, b) FT3, and c) TSH of both the discovery and replicated fixed-effects meta-analysis (FE model).

## References

1. Völzke H, Schössow J, Schmidt CO, et al. Cohort Profile Update: The Study of Health in Pomerania (SHIP). *Int J Epidemiol* 2022; doi: 10.1093/ije/dyac034.
2. Wright JD, Folsom AR, Coresh J, et al. The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. *J Am Coll Cardiol* 2021;77(23):2939–2959; doi: 10.1016/j.jacc.2021.04.035.
3. Holle R, Happich M, Löwel H, et al. KORA - A Research Platform for Population Based Health Research. *Das Gesundheitswes* 2005;67(S 01):19–25; doi: 10.1055/s-2005-858235.
4. Hofman A, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol* 2015;30(8):661–708; doi: 10.1007/s10654-015-0082-x.
5. Moayyeri A, Hammond CJ, Valdes AM, et al. Cohort Profile: TwinsUK and Healthy Ageing Twin Study. *Int J Epidemiol* 2013;42(1):76–85; doi: 10.1093/ije/dyr207.
6. Taylor AM, Pattie A, Deary IJ. Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol* 2018;47(4):1042–1042r; doi: 10.1093/ije/dyy022.
7. Powell JE, Henders AK, McRae AF, et al. The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One* 2012;7(4):e35430; doi: 10.1371/journal.pone.0035430.
8. Houseman EA, Accomando WP, Koestler DC, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 2012;13(1):86; doi: 10.1186/1471-2105-13-86.
9. Lehne B, Drong AW, Loh M, et al. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol* 2015;16(1):37; doi: 10.1186/s13059-015-0600-x.
10. van Iterson M, van Zwet EW, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* 2017;18(1):19; doi: 10.1186/s13059-016-1131-9.
11. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw* 2010;36(3); doi: 10.18637/jss.v036.i03.
12. Schurmann C, Heim K, Schillert A, et al. Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. *PLoS One* 2012;7(12):e50938; doi: 10.1371/journal.pone.0050938.
13. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518(7539):317–330; doi: 10.1038/nature14248.
14. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC. *Genome Res* 2002;12(6):996–1006; doi: 10.1101/gr.229102.

## Supplementary Tables

Supplementary Table 1: Overview of the Meta-Analysis results of significant discovery stage CpGs

Supplementary Table 2: Cohort EWAS results of significant hits

Supplementary Table 3: Annotation of the sites found in the analyses

Supplementary Table 4: Gene expression results in blood cells

Supplementary Table 5: Lookup of meQTL associations in the GoDMC

Supplementary Table 6: Results of the Mendelian randomization analysis

Supplementary Table 7: Lookup of known CpG associations in the discovery stage

**Supplementary Table 1: Overview of the Meta-Analysis results of significant discovery stage CpGs**

Hormone	CpG probeID	Chr	Position (build 37)	EWAS Meta-Analysis Results						Nearest Gene	Located in Island	Located in Promotor region	Roadmap		
				Analysis	Estimate	P-Value	Combined Sample Size	Included Cohorts	I2				Het. P-Value	Located in Enhancer Region	Located in Dyadic Region
FT4	cg00049440	9	73,026,643	Discovery	-0.790	4.53E-15	4078	ARIC, KORA, SHIP-Trend	30	0.2	KLF9	Island Shore	Yes	No	No
				Replication	-0.674	1.45E-06	2448	RS, LBC1936, LBC1921,BSGS							
				Combined	-0.750	<b>4.44E-20</b>	6526	ARIC, KORA, SHIP-Trend, RS, LBC1936, LBC1921,BSGS							
FT4	cg03546163	6	35,654,363	Discovery	-0.849	3.11E-13	3878	ARIC, KORA, SHIP-Trend	0	0.72	FKBP5	Island Shore	Yes	No	No
				Replication	-0.825	1.00E-04	1251	RS, BSGS							
				Combined	-0.843	<b>1.45E-16</b>	5129	ARIC, KORA, SHIP-Trend, RS, BSGS							
FT4	cg04173586	19	2,167,496	Discovery	-0.445	5.14E-09	4038	ARIC, KORA, SHIP-Trend	31	0.19	DOT1L	Island Shore	No	Yes	No
				Replication	-0.380	1.88E-03	2444	RS, LBC1936, LBC1921,BSGS							
				Combined	-0.427	<b>4.03E-11</b>	6482	ARIC, KORA, SHIP-Trend, RS, LBC1936, LBC1921,BSGS							
FT4	cg06983052	1	90,288,099	Discovery	-0.388	2.63E-08	4081	ARIC, KORA, SHIP-Trend	65	0.008	LRRC8D	Island Shore	No	No	Yes
				Replication	-0.032	7.99E-01	2448	RS, LBC1936, LBC1921,BSGS							
				Combined	-0.304	6.11E-07	6529	ARIC, KORA, SHIP-Trend, RS, LBC1936, LBC1921,BSGS							
FT3	cg00049440	9	73,026,643	Discovery	-1.349	2.61E-12	1637	KORA, SHIP-Trend	32	0.23	KLF9	Island Shore	Yes	No	No
				Replication	-1.876	1.81E-07	590	BSGS							
				Combined	-1.463	<b>1.69E-17</b>	2227	KORA, SHIP-Trend, BSGS							
FT3	cg01695994	17	80,246,403	Discovery	-1.112	8.43E-10	1638	KORA, SHIP-Trend	85	0.002	CSNK1D/LINC01970	Island Shelf	No	Yes	No
				Replication	-2.497	8.63E-12	590	BSGS							
				Combined	-1.371	<b>5.84E-17</b>	2228	KORA, SHIP-Trend, BSGS							
FT3	cg04173586	19	2,167,496	Discovery	-1.842	7.98E-14	1607	KORA, SHIP-Trend	0	0.74	DOT1L	Island Shore	No	Yes	No
				Replication	-2.013	1.51E-16	586	BSGS							
				Combined	-1.935	<b>4.33E-28</b>	2193	KORA, SHIP-Trend, BSGS							
FT3	cg06983052	1	90,288,099	Discovery	-0.875	5.90E-08	1639	KORA, SHIP-Trend	66	0.05	LRRC8D	Island Shore	No	No	Yes
				Replication	-1.474	8.49E-09	590	BSGS							
				Combined	-1.040	<b>5.45E-14</b>	2229	KORA, SHIP-Trend, BSGS							
FT3	cg20146909	1	90,289,611	Discovery	-0.889	3.38E-09	1639	KORA, SHIP-Trend	70	0.03	LRRC8D	Island Shelf	No	No	No
				Replication	-1.811	5.67E-08	590	BSGS							
				Combined	-1.036	<b>6.51E-14</b>	2229	KORA, SHIP-Trend, BSGS							
TSH	cg00049440	9	73,026,643	Discovery	0.143	5.19E-14	4082	ARIC, KORA, SHIP-Trend	0	0.46	KLF9	Island Shore	Yes	No	No
				Replication	0.088	1.11E-03	2988	RS, TWINSUK, LBC1936, LBC1921, BSGS							
				Combined	0.125	<b>9.57E-16</b>	7070	ARIC, KORA, SHIP-Trend, RS, TWINSUK, LBC1936, LBC1921, BSGS							
TSH	cg04173586	19	2167496	Discovery	0.07656	4.52E-08	4041	ARIC, KORA, SHIP-Trend	44	0.08	DOT1L	Island Shore	No	Yes	No
				Replication	0.06083	0.008519	2984	RS, TWINSUK, LBC1936, LBC1921, BSGS							
				Combined	0.07234	1.53E-09	7025	ARIC, KORA, SHIP-Trend, RS, TWINSUK, LBC1936, LBC1921, BSGS							

**Supplementary Table 2: Cohort EWAS results of significant hits**

Hormone	CpG probeID	ARIC			KORA			SHIP-Trend		
		Estimate	P-Value	Sample Size	Estimate	P-Value	Sample Size	Estimate	P-Value	Sample Size
FT4	cg00049440	-0.89	3.48E-18	2442	-0.65	7.90E-04	1408	-0.42	3.80E-01	228
FT4	cg03546163	-0.84	4.86E-15	2434	-1.14	9.72E-04	1216	-0.72	1.45E-01	228
FT4	cg04173586	-0.43	1.58E-10	2432	-0.85	2.59E-03	1378	-0.28	5.05E-01	228
FT3	cg00049440				-1.26	2.96E-09	1408	-1.84	1.80E-04	229
FT3	cg01695994				-1.02	1.11E-07	1409	-1.84	9.97E-04	229
FT3	cg04173586				-1.95	2.76E-10	1378	-1.65	1.02E-04	229
FT3	cg06983052				-0.98	4.30E-08	1410	-0.39	3.23E-01	229
FT3	cg20146909				-0.85	1.05E-07	1410	-1.23	8.88E-03	229
TSH	cg00049440	0.14	7.40E-11	2441	0.15	1.60E-04	1407	0.05	6.80E-01	234
TSH	cg04173586	0.07	1.59E-06	2430	0.18	1.89E-03	1377	0.09	4.16E-01	234

Hormone	CpG probeID	BSGS			LBC1921			LBC1936		
		Estimate	P-Value	Sample Size	Estimate	P-Value	Sample Size	Estimate	P-Value	Sample Size
FT4	cg00049440	-0.34	3.35E-01	590	-1.05	4.38E-04	370	-1.00	4.03E-05	827
FT4	cg03546163	-1.27	6.17E-04	590						
FT4	cg04173586	0.13	5.86E-01	586	-0.75	2.95E-03	370	-0.49	1.93E-02	827
FT3	cg00049440	-1.88	1.81E-07	590						
FT3	cg01695994	-2.50	8.63E-12	590						
FT3	cg04173586	-2.01	1.51E-16	586						
FT3	cg06983052	-1.47	8.49E-09	590						
FT3	cg20146909	-1.81	5.67E-08	590						
TSH	cg00049440	0.06	4.88E-01	590	0.17	1.42E-02	368	0.14	7.59E-03	822
TSH	cg04173586	-0.01	8.72E-01	586	0.15	9.25E-03	368	-0.01	7.82E-01	822

Hormone	CpG probeID	RS			TWINSUK		
		Estimate	P-Value	Sample Size	Estimate	P-Value	Sample Size
FT4	cg00049440	-0.30	2.05E-01	661			
FT4	cg03546163	-0.67	8.01E-03	661			
FT4	cg04173586	-0.41	1.07E-01	661			
FT3	cg00049440						
FT3	cg01695994						
FT3	cg04173586						
FT3	cg06983052						
FT3	cg20146909						
TSH	cg00049440	0.04	3.36E-01	661	0.05	4.73E-01	547
TSH	cg04173586	0.13	6.25E-03	661	0.08	1.44E-01	547

Supplementary Table 3: Annotation of the sites found in the analyses

CpG probeID	Associated Traits	Chromosome	Position (hg19)	Nearest Gene (relative location)	CpG Island Association (identification)	ROADMAP Annotation (Cell Type)	Distance in bp to THRA peak (liver cell) *	Total heritability of CpG (h <sup>2</sup> ) **	CpG variation explained by common genetic variants (h <sup>2</sup> <sub>SNPs</sub> ) **	CpG heritability attributed to common genetic variants
cg00049440	TSH, FT4, FT3	9	73026643	KLF9 (2,930bp inside)	cpg shore (chr9:73027331-73029184)	in promotor anatomy (ESC; ESC_DERIVED; LUNG; IPSC; FAT; STROMAL_CONNECTIVE; BREAST; BLOOD; MUSCLE; BRAIN; SKIN; VASCULAR; LIVER; GI_COLON; GI_DUODENUM; GI_ESOPHAGUS; GI_INTESTINE; KIDNEY; PANCREAS; PLACENTA; GI_STOMACH; HEART; OVARY; GI_RECTUM; THYMUS; SPLEEN; CERVIX; BONE)	2282	0.57	0.41	72%
cg04173586	TSH, FT4, FT3	19	2167496	DOT1L (3,348bp inside)	cpg shore (chr19:2163633-2165603)	in enhancer anatomy (ESC; ESC_DERIVED; LUNG; IPSC; STROMAL_CONNECTIVE; BREAST; BLOOD; BRAIN; SKIN; VASCULAR; LIVER; GI_COLON; GI_DUODENUM; GI_ESOPHAGUS; ADRENAL; HEART; GI_INTESTINE; MUSCLE; MUSCLE_LEG; PLACENTA; GI_STOMACH; THYMUS; OVARY; PANCREAS; GI_RECTUM; SPLEEN; CERVIX)	828	0.43	5.70E-10	0%
cg03546163	FT4	6	35654363	FKBP5 (2,217bp inside)	cpg shore (chr6:35655608-35656856)	in promotor anatomy (ESC; ESC_DERIVED; IPSC; FAT; BREAST; BLOOD; STROMAL_CONNECTIVE; MUSCLE; SKIN; VASCULAR; LIVER; BRAIN; GI_COLON; GI_DUODENUM; ADRENAL; HEART; PANCREAS; THYMUS; OVARY; GI_RECTUM; GI_STOMACH)	2219	0.40	0.40	100%
cg06983052	FT3	1	90288099	LRRC8D (619bp inside)	cpg shore (chr1:90286190-90287222)	in dyadic anatomy (ESC; ESC_DERIVED; LUNG; IPSC; FAT; STROMAL_CONNECTIVE; BREAST; BLOOD; MUSCLE; BRAIN; SKIN; LIVER; GI_COLON; GI_DUODENUM; ADRENAL; HEART; GI_INTESTINE; KIDNEY; PANCREAS; PLACENTA; GI_STOMACH; THYMUS; GI_RECTUM; SPLEEN; CERVIX)	517	0.25	0.02	6%
cg01695994	FT3	17	80246403	CSNK1D (14,809bp upstream)	cpg shelf (chr17:80250227-80251057)	in enhancer anatomy (ESC_DERIVED; ESC; LUNG; FAT; BREAST; BLOOD; SKIN; GI_COLON; GI_DUODENUM; GI_ESOPHAGUS; GI_INTESTINE; MUSCLE; MUSCLE_LEG; PLACENTA; GI_STOMACH; HEART; OVARY; PANCREAS; GI_RECTUM; SPLEEN; LIVER; VASCULAR; BRAIN; BONE)	0	0.53	0.09	18%
cg20146909	FT3	1	90289611	LRRC8D (2,131bp inside)	cpg shelf (chr1:90286190-90287222)	none	-995	0.41	0.12	29%

\* source: [https://chip-atlas.org/peak\\_browser](https://chip-atlas.org/peak_browser)

\*\* source: van Dongen et al. (2016)

CpG is in a transcription factor binding site in thyroid tissue

CpG is associated with transcription factor footprint (determined by DNase footprinting) in thyroid tissue

**Supplementary Table 4: Gene expression results in blood cells**

CpG probeID	expression probeID	Chromosome	expression probe start position	expression probe stop position	expression (mRNA) gene	expression beta	expression SE	expression pvalue	FDR (Benjamini-Hochberg)	Sample size
cg00049440	ILMN_1679391	9	72,658,496	72,841,888	MAMDC2	0.050	0.040	0.208	0.94	712
cg00049440	ILMN_1740622	9	73,149,965	73,736,514	TRPM3	-0.012	0.038	0.763	0.95	712
cg00049440	ILMN_1757132	9	72,873,877	72,969,789	SMCS	-0.002	0.040	0.951	1.00	712
<b>cg00049440</b>	<b>ILMN_1778523</b>	<b>9</b>	<b>72,999,512</b>	<b>73,029,573</b>	<b>KLF9</b>	<b>-0.109</b>	<b>0.021</b>	<b>3.00E-07</b>	<b>4.38E-05</b>	<b>712</b>
cg03546163	ILMN_1656849	6	35,911,290	35,992,413	SLC26A8	-0.086	0.077	0.265	0.94	619
cg03546163	ILMN_1659359	6	35,182,189	35,218,609	SCUBE3	-0.153	0.126	0.224	0.94	619
cg03546163	ILMN_1672575	6	35,911,290	35,992,413	SLC26A8	-0.114	0.082	0.167	0.94	619
cg03546163	ILMN_1674282	6	35,310,334	35,395,968	PPARD	0.116	0.085	0.171	0.94	619
cg03546163	ILMN_1712616	6	35,704,858	35,716,688	ARMC12	-0.089	0.108	0.411	0.94	619
cg03546163	ILMN_1720656	6	35,995,453	36,070,535	MAPK14	0.000	0.083	0.995	1.00	619
cg03546163	ILMN_1737627	6	35,995,453	36,079,013	MAPK14	-0.093	0.060	0.121	0.94	619
cg03546163	ILMN_1749327	6	36,098,260	36,112,301	MAPK13	-0.059	0.067	0.378	0.94	619
cg03546163	ILMN_1755843	6	35,911,290	35,992,413	SLC26A8	-0.036	0.036	0.318	0.94	619
cg03546163	ILMN_1778444	6	35,541,361	35,696,397	FKBP5	-0.044	0.059	0.460	0.94	619
cg03546163	ILMN_1779794	6	35,310,334	35,393,173	PPARD	0.053	0.113	0.636	0.94	619
cg03546163	ILMN_1788291	6	35,773,070	35,791,852	LHFPL5	0.071	0.123	0.565	0.94	619
cg03546163	ILMN_1789600	6	35,748,830	35,755,841	CLPSL1	0.293	0.101	0.004	0.28	619
cg03546163	ILMN_1798804	6	35,800,810	35,888,957	SRPK1	-0.119	0.071	0.094	0.94	619
cg03546163	ILMN_1804263	6	35,227,509	35,263,760	ZNF76	0.074	0.075	0.319	0.94	619
cg03546163	ILMN_1808041	6	35,436,177	35,438,558	RPL10A	0.028	0.052	0.594	0.94	619
cg03546163	ILMN_1811973	6	35,310,334	35,395,968	PPARD	-0.013	0.124	0.914	1.00	619
cg03546163	ILMN_1814002	6	35,441,373	35,464,861	TEAD3	0.034	0.094	0.720	0.94	619
cg03546163	ILMN_2043452	6	35,420,137	35,434,881	FANCE	0.027	0.074	0.712	0.94	619
cg03546163	ILMN_2054121	6	35,744,391	35,747,329	CLPSL2	-0.107	0.108	0.320	0.94	619
cg03546163	ILMN_2131811	6	35,762,758	35,765,121	CLPS	-0.010	0.110	0.925	1.00	619
cg03546163	ILMN_2154566	6	35,436,177	35,438,558	RPL10A	-0.081	0.100	0.418	0.94	619
cg03546163	ILMN_2163547	6	35,465,650	35,480,647	TULP1	0.224	0.099	0.024	0.89	619
cg03546163	ILMN_2394210	6	35,911,290	35,992,413	SLC26A8	-0.039	0.048	0.419	0.94	619
cg04173586	ILMN_1653372	19	1,782,073	1,812,275	ATP8B3	0.083	0.103	0.424	0.94	699
cg04173586	ILMN_1672461	19	2,328,628	2,355,100	SPPL2B	0.047	0.061	0.440	0.94	699
cg04173586	ILMN_1674926	19	2,274,630	2,282,181	C19ORF35	-0.015	0.033	0.649	0.94	699
cg04173586	ILMN_1678165	19	2,321,519	2,328,614	LSM7	-0.013	0.050	0.794	0.96	699
cg04173586	ILMN_1679363	19	1,876,974	1,885,518	ABHD17A	0.003	0.090	0.972	1.00	699
cg04173586	ILMN_1686557	19	2,425,621	2,427,875	TIMM13	-0.027	0.060	0.647	0.94	699
cg04173586	ILMN_1693559	19	2,164,147	2,232,577	DOT1L	-0.055	0.100	0.585	0.94	699
cg04173586	ILMN_1697701	19	2,233,154	2,236,951	PLEKHJ1	-0.113	0.065	0.082	0.94	699
cg04173586	ILMN_1706521	19	1,941,160	1,981,336	CSNK1G2	0.011	0.026	0.678	0.94	699
cg04173586	ILMN_1708101	19	2,428,162	2,456,966	LMNB2	0.056	0.075	0.455	0.94	699
cg04173586	ILMN_1718977	19	2,476,122	2,478,257	GADD45B	0.072	0.057	0.205	0.94	699
cg04173586	ILMN_1721344	19	2,071,034	2,096,269	MOB3A	-0.003	0.051	0.961	1.00	699
cg04173586	ILMN_1728107	19	2,511,217	2,702,746	GNG7	0.029	0.051	0.574	0.94	699
cg04173586	ILMN_1734346	19	2,096,867	2,099,583	IZUMO4	0.043	0.088	0.624	0.94	699
cg04173586	ILMN_1753008	19	1,815,244	1,848,452	REXO1	-0.010	0.059	0.871	0.98	699
cg04173586	ILMN_1754220	19	2,236,815	2,248,678	SF3A2	-0.040	0.063	0.525	0.94	699
cg04173586	ILMN_1764945	19	2,100,986	2,151,556	AP3D1	-0.085	0.088	0.330	0.94	699
cg04173586	ILMN_1770387	19	2,389,783	2,426,086	TMPRSS9	0.012	0.032	0.719	0.94	699
cg04173586	ILMN_1770451	19	1,852,397	1,863,564	KLF16	-0.042	0.086	0.626	0.94	699
cg04173586	ILMN_1773080	19	2,269,519	2,273,487	OAZ1	-0.189	0.181	0.297	0.94	699
cg04173586	ILMN_1773163	19	1,952,525	1,954,548	CSNK1G2-AS1	0.142	0.088	0.106	0.94	699
cg04173586	ILMN_1788337	19	2,096,867	2,099,583	IZUMO4	-0.107	0.097	0.268	0.94	699
cg04173586	ILMN_1800843	19	1,905,370	1,926,012	SCAMP4	0.053	0.077	0.493	0.94	699
cg04173586	ILMN_1802315	19	2,328,628	2,355,100	SPPL2B	0.051	0.083	0.539	0.94	699
cg04173586	ILMN_1806275	19	1,905,370	1,913,446	ADAT3	0.033	0.069	0.630	0.94	699
cg04173586	ILMN_2067444	19	1,753,661	1,775,444	ONECUT3	-0.147	0.102	0.152	0.94	699
cg04173586	ILMN_2140623	19	1,876,974	1,885,518	ABHD17A	-0.277	0.107	0.010	0.47	699
cg04173586	ILMN_2239772	19	1,876,974	1,885,518	ABHD17A	-0.028	0.054	0.611	0.94	699
cg04173586	ILMN_2284667	19	2,100,986	2,129,162	AP3D1	-0.037	0.100	0.709	0.94	699
cg04173586	ILMN_2307978	19	1,876,974	1,885,518	ABHD17A	-0.017	0.053	0.747	0.95	699
cg04173586	ILMN_2390114	19	2,100,986	2,151,556	AP3D1	0.016	0.083	0.852	0.96	699
cg04173586	ILMN_2400546	19	2,096,867	2,099,583	IZUMO4	-0.050	0.105	0.633	0.94	699
cg04173586	ILMN_2408771	19	2,328,628	2,355,100	SPPL2B	-0.058	0.096	0.544	0.94	699
cg06983052	ILMN_1662021	1	90,460,677	90,494,094	ZNF326	0.070	0.035	0.043	0.89	713
cg06983052	ILMN_1669440	1	90,460,677	90,494,094	ZNF326	0.029	0.034	0.401	0.94	713
cg06983052	ILMN_1712128	1	89,990,396	90,063,420	LRRC8B	-0.002	0.034	0.945	1.00	713
cg06983052	ILMN_1720004	1	90,090,407	90,098,453	FLJ27354	0.014	0.029	0.636	0.94	713
cg06983052	ILMN_1753950	1	90,460,677	90,494,094	ZNF326	-0.017	0.028	0.548	0.94	713
cg06983052	ILMN_1756953	1	89,829,435	89,853,719	GBP6	-0.009	0.024	0.696	0.94	713
cg06983052	ILMN_1838313	1	89,990,396	90,063,420	LRRC8B	0.027	0.020	0.186	0.94	713

cg06983052	ILMN_2094396	1	90,098,643	90,185,094	<i>LRR8C</i>	0.005	0.017	0.748	0.95	713
cg06983052	ILMN_2159471	1	90,458,823	90,460,525	<i>GEMIN8P4</i>	0.028	0.035	0.423	0.94	713
cg01695994	ILMN_1652277	17	80,278,899	80,291,921	<i>SECTM1</i>	0.000	0.021	0.982	1.00	713
cg01695994	ILMN_1652333	17	80,674,581	80,685,893	<i>FN3KRP</i>	-0.026	0.015	0.085	0.94	713
cg01695994	ILMN_1652846	17	79,860,776	79,869,340	<i>PCYT2</i>	0.031	0.023	0.189	0.94	713
cg01695994	ILMN_1653342	17	80,015,747	80,023,697	<i>DUS1L</i>	-0.013	0.018	0.476	0.94	713
cg01695994	ILMN_1660749	17	79,935,425	79,975,282	<i>ASPSCR1</i>	-0.011	0.022	0.611	0.94	713
cg01695994	ILMN_1661427	17	80,005,777	80,009,650	<i>RFNG</i>	0.035	0.023	0.126	0.94	713
cg01695994	ILMN_1665289	17	80,347,085	80,376,513	<i>OGFOD3</i>	-0.006	0.023	0.789	0.96	713
cg01695994	ILMN_1673376	17	80,709,939	80,901,062	<i>TBCD</i>	-0.021	0.028	0.448	0.94	713
cg01695994	ILMN_1681437	17	79,993,756	79,995,573	<i>DCXR</i>	-0.016	0.021	0.460	0.94	713
cg01695994	ILMN_1685763	17	80,572,437	80,606,411	<i>WDR45B</i>	-0.002	0.022	0.943	1.00	713
cg01695994	ILMN_1687437	17	80,111,485	80,170,689	<i>CCDC57</i>	0.000	0.028	0.996	1.00	713
cg01695994	ILMN_1692260	17	79,876,144	79,885,587	<i>MAFG</i>	-0.013	0.023	0.563	0.94	713
cg01695994	ILMN_1693340	17	79,989,531	79,992,080	<i>RAC3</i>	-0.019	0.026	0.460	0.94	713
cg01695994	ILMN_1715178	17	79,890,268	79,895,172	<i>PYCR1</i>	-0.029	0.023	0.207	0.94	713
cg01695994	ILMN_1716546	17	80,069,865	80,109,489	<i>CCDC57</i>	-0.047	0.024	0.045	0.89	713
cg01695994	ILMN_1719303	17	79,801,033	79,818,544	<i>P4HB</i>	-0.014	0.033	0.672	0.94	713
cg01695994	ILMN_1720708	17	80,200,536	80,231,594	<i>CSNK1D</i>	-0.033	0.023	0.154	0.94	713
cg01695994	ILMN_1722102	17	79,849,598	79,858,363	<i>ANAPC11</i>	-0.042	0.021	0.049	0.89	713
cg01695994	ILMN_1725311	17	79,762,009	79,771,889	<i>GCGR</i>	-0.024	0.032	0.464	0.94	713
cg01695994	ILMN_1726281	17	79,849,598	79,858,363	<i>ANAPC11</i>	-0.012	0.032	0.693	0.94	713
cg01695994	ILMN_1730351	17	80,317,122	80,321,652	<i>TEX19</i>	0.031	0.036	0.389	0.94	713
cg01695994	ILMN_1733799	17	79,780,292	79,791,167	<i>FAM195B</i>	0.003	0.017	0.836	0.96	713
cg01695994	ILMN_1734742	17	79,825,596	79,829,628	<i>ARHGDI1A</i>	0.001	0.021	0.968	1.00	713
cg01695994	ILMN_1741180	17	80,376,251	80,400,516	<i>HEXDC</i>	0.008	0.025	0.730	0.94	713
cg01695994	ILMN_1742507	17	79,981,279	79,989,027	<i>LRR8C45</i>	-0.007	0.028	0.808	0.96	713
cg01695994	ILMN_1744455	17	79,910,382	79,919,057	<i>NOTUM</i>	-0.010	0.028	0.729	0.94	713
cg01695994	ILMN_1747962	17	80,332,200	80,333,370	<i>UTS2R</i>	0.029	0.033	0.376	0.94	713
cg01695994	ILMN_1750401	17	80,400,462	80,408,707	<i>C17ORF62</i>	0.005	0.023	0.816	0.96	713
cg01695994	ILMN_1750518	17	79,845,710	79,849,462	<i>ALYREF</i>	-0.028	0.024	0.254	0.94	713
cg01695994	ILMN_1753773	17	79,849,598	79,858,363	<i>ANAPC11</i>	-0.029	0.030	0.324	0.94	713
cg01695994	ILMN_1754827	17	79,981,279	79,989,027	<i>LRR8C45</i>	-0.007	0.029	0.802	0.96	713
cg01695994	ILMN_1757253	17	79,791,367	79,792,926	<i>PPP1R27</i>	-0.006	0.033	0.846	0.96	713
cg01695994	ILMN_1759732	17	80,005,777	80,009,650	<i>RFNG</i>	-0.009	0.025	0.731	0.94	713
cg01695994	ILMN_1765923	17	80,347,085	80,376,513	<i>OGFOD3</i>	-0.050	0.027	0.062	0.94	713
cg01695994	ILMN_1769103	17	79,876,144	79,881,444	<i>MAFG</i>	-0.010	0.027	0.708	0.94	713
cg01695994	ILMN_1769634	17	79,976,578	79,980,785	<i>STRA13</i>	-0.013	0.025	0.611	0.94	713
cg01695994	ILMN_1769960	17	80,009,762	80,015,346	<i>GPS1</i>	-0.005	0.021	0.802	0.96	713
cg01695994	ILMN_1777453	17	79,860,071	79,860,781	<i>NPB</i>	-0.024	0.034	0.475	0.94	713
cg01695994	ILMN_1784341	17	80,551,297	80,562,483	<i>FOKK2</i>	-0.031	0.031	0.323	0.94	713
cg01695994	ILMN_1784871	17	80,036,213	80,056,106	<i>FASN</i>	-0.004	0.015	0.768	0.95	713
cg01695994	ILMN_1792538	17	80,272,745	80,275,480	<i>CD7</i>	-0.006	0.020	0.752	0.95	713
cg01695994	ILMN_1795400	17	80,709,939	80,901,062	<i>TBCD</i>	0.004	0.019	0.834	0.96	713
cg01695994	ILMN_1795876	17	80,009,762	80,015,346	<i>GPS1</i>	-0.020	0.029	0.500	0.94	713
cg01695994	ILMN_1801588	17	80,693,451	80,709,073	<i>FN3K</i>	0.017	0.031	0.585	0.94	713
cg01695994	ILMN_1802603	17	80,005,777	80,009,650	<i>RFNG</i>	-0.037	0.025	0.138	0.94	713
cg01695994	ILMN_1811991	17	80,347,085	80,376,513	<i>OGFOD3</i>	-0.024	0.028	0.401	0.94	713
cg01695994	ILMN_2077858	17	79,869,814	79,876,058	<i>SIRT7</i>	0.015	0.022	0.488	0.94	713
cg01695994	ILMN_2140342	17	80,059,345	80,170,689	<i>CCDC57</i>	-0.001	0.021	0.972	1.00	713
cg01695994	ILMN_2162298	17	79,791,367	79,792,926	<i>PPP1R27</i>	0.037	0.032	0.251	0.94	713
cg01695994	ILMN_2166275	17	79,910,382	79,919,057	<i>NOTUM</i>	-0.023	0.032	0.463	0.94	713
cg01695994	ILMN_2202930	17	79,860,776	79,869,340	<i>PCYT2</i>	0.017	0.023	0.460	0.94	713
cg01695994	ILMN_2230566	17	80,614,942	80,656,598	<i>RAB40B</i>	0.018	0.021	0.385	0.94	713
cg01695994	ILMN_2241317	17	80,477,593	80,562,483	<i>FOKK2</i>	-0.040	0.030	0.177	0.94	713
cg01695994	ILMN_2275803	17	79,981,279	79,989,027	<i>LRR8C45</i>	0.012	0.019	0.544	0.94	713
cg01695994	ILMN_2286870	17	80,200,536	80,231,594	<i>CSNK1D</i>	-0.010	0.020	0.630	0.94	713
cg01695994	ILMN_2304996	17	80,005,777	80,009,650	<i>RFNG</i>	-0.024	0.029	0.403	0.94	713
cg01695994	ILMN_2309228	17	80,009,762	80,015,346	<i>GPS1</i>	0.041	0.022	0.065	0.94	713
cg01695994	ILMN_2310685	17	80,477,593	80,562,483	<i>FOKK2</i>	0.046	0.026	0.080	0.94	713
cg01695994	ILMN_2334693	17	80,416,059	80,446,143	<i>NARF</i>	0.002	0.018	0.906	1.00	713
cg01695994	ILMN_2343563	17	79,849,598	79,858,363	<i>ANAPC11</i>	0.009	0.026	0.721	0.94	713
cg01695994	ILMN_2364022	17	80,186,281	80,197,375	<i>SLC16A3</i>	-0.013	0.022	0.553	0.94	713
cg01695994	ILMN_2364062	17	79,845,710	79,849,462	<i>ALYREF</i>	0.013	0.016	0.423	0.94	713
cg01695994	ILMN_2388539	17	80,347,085	80,376,513	<i>OGFOD3</i>	0.004	0.020	0.827	0.96	713
cg01695994	ILMN_2395092	17	79,890,266	79,894,968	<i>PYCR1</i>	-0.014	0.030	0.637	0.94	713
cg01695994	ILMN_2412761	17	79,876,144	79,885,587	<i>MAFG</i>	-0.013	0.025	0.607	0.94	713
cg20146909	ILMN_1662021	1	90,460,677	90,494,094	<i>ZNF326</i>	0.017	0.029	0.566	0.94	713
cg20146909	ILMN_1669440	1	90,460,677	90,494,094	<i>ZNF326</i>	-0.018	0.028	0.523	0.94	713
cg20146909	ILMN_1712128	1	89,990,396	90,063,420	<i>LRR8C8B</i>	0.029	0.029	0.314	0.94	713
cg20146909	ILMN_1720004	1	90,090,407	90,098,453	<i>FLJ27354</i>	-0.014	0.024	0.570	0.94	713
cg20146909	ILMN_1753950	1	90,460,677	90,494,094	<i>ZNF326</i>	-0.019	0.023	0.409	0.94	713
cg20146909	ILMN_1756953	1	89,829,435	89,853,719	<i>GBP6</i>	0.017	0.020	0.402	0.94	713

cg20146909	ILMN_1838313	1	89,990,396	90,063,420	<i>LRRC8B</i>	0.033	0.017	0.048	0.89	713
cg20146909	ILMN_2094396	1	90,098,643	90,185,094	<i>LRRC8C</i>	-0.012	0.014	0.390	0.94	713
cg20146909	ILMN_2159471	1	90,458,823	90,460,525	<i>GEMIN8P4</i>	0.000	0.029	0.991	1.00	713
cg01695994	ILMN_1652277	17	80,278,899	80,291,921	<i>SECTM1</i>	0.000	0.021	0.982	1.00	713
cg01695994	ILMN_1734742	17	79,825,596	79,829,282	<i>ARHGDI1A</i>	0.001	0.021	0.968	1.00	713

### Supplementary Table 5: Lookup of meQTL associations in the GoDMC

cpG	rsid	a1	a2	SNP name (chr:position:type)	beta	a1	se	samplesize	pvalue	cis/trans	het $r^2$	closest meQTL gene name	closest meQTL gene distance	DNAm locus
cg00049440	rs10868849	C	G	chr9:73136309:SNP	0.087	0.015	9565	1.22E-08	cis	80.9	<i>TRPM3</i>	13656	<i>KLF9</i>	
cg00049440	rs11142385	A	G	chr9:72992666:SNP	0.337	0.009	25133	0	cis	70.5	<i>KLF9</i>	6847	<i>KLF9</i>	
cg00049440	rs9310736*	A	G	chr3:24350811:SNP	-0.191	0.008	27594	1.82E-113	trans	22.2	<i>THRB</i>	0	<i>KLF9</i>	
cg03546163	rs869785*	T	C	chr3:24347800:SNP	-0.145	0.009	25104	3.57E-55	trans	54	<i>THRB</i>	0	<i>FKBP5</i>	
cg03546163	rs9470079	A	G	chr6:35643063:SNP	0.226	0.014	17650	2.66E-57	cis	48	<i>FKBP5</i>	0	<i>FKBP5</i>	
cg06983052	rs6673007**	G	A	chr1:90272970:SNP	-0.165	0.012	27730	8.97E-45	cis	0	<i>LRR8D</i>	13602	<i>LRR8D</i>	
cg01695994	rs1278601	T	C	chr12:133499481:SNP	-0.212	0.009	22192	9.80E-114	trans	56.6	<i>ZNF605</i>	0	<i>CSNK1D</i>	
cg01695994	rs9913095	T	C	chr17:80304189:SNP	0.143	0.011	21954	1.27E-38	cis	15.5	<i>SECTM1</i>	12268	<i>CSNK1D</i>	
cg20146909	rs6693717**	T	C	chr1:90132480:SNP	-0.126	0.009	27744	0	cis	0	<i>LRR8C</i>	0	<i>LRR8D</i>	

\*LD  $r^2 = 0.98$

\*\*LD  $r^2 = 0.21$

SNP positions are provided according to build 37

**Supplementary Table 6: Results of the Mendelian randomization analysis**

Hormone (exposure)	CpG Site (outcome)	Method	Main analysis					After excluding significant (p<0.05) TPOAb SNPs			
			# SNPs	Effect	SE	p-Value	Heterogeneity p-value (Q-stat/Egger intercept)	# SNPs	Effect	SE	p-Value
FT4	cg00049440 <i>KLF9</i>	inverse-variance weighted fixed effect	31	0.001	0.002	0.738	0.446	No SNPs were removed			
		weighted median	31	0.001	0.004	0.835	-				
		MR-Egger (bootstrapped)	31	0.002	0.005	0.362	0.439				
FT4	cg03546163 <i>FKBP5</i>	inverse-variance weighted fixed effect	31	0.009	0.017	0.573	0.931				
		weighted median	31	0.005	0.026	0.844	-				
		MR-Egger (bootstrapped)	31	0.029	0.034	0.193	0.928				
FT4	cg04173586 <i>DOT1L</i>	inverse-variance weighted fixed effect	31	0.014	0.012	0.244	0.156				
		weighted median	31	0.008	0.019	0.679	-				
		MR-Egger (bootstrapped)	31	0.019	0.025	0.213	0.141				
TSH	cg00049440 <i>KLF9</i>	inverse-variance weighted fixed effect	56	0.005	0.002	<b>0.010</b>	0.352	51	0.005	0.002	<b>0.007</b>
		weighted median	56	0.006	0.003	<b>0.036</b>	-	51	0.006	0.003	0.060
		MR-Egger (bootstrapped)	56	0.007	0.004	<b>0.026</b>	0.329	51	0.008	0.004	<b>0.018</b>
TSH	cg04173586 <i>DOT1L</i>	inverse-variance weighted fixed effect	56	-0.007	0.009	0.443	0.557	51	-0.004	0.009	0.649
		weighted median	56	-0.015	0.013	0.245	-	51	-0.001	0.015	0.967
		MR-Egger (bootstrapped)	56	0.012	0.019	0.268	0.640	51	0.016	0.019	0.198

**Supplementary Table 7: Lookup of known CpG associations in the discovery stage**

Hormone	CpG probeID	Estimate	Standard Error	P-Value	Sample Size	Chr	Position (build 37)	Nearest Gene
FT3	cg00024471	-0.900	0.17	2.55E-07	1635	3	188692547	<i>TPRG1</i>
FT3	cg00049440	-1.349	0.19	2.61E-12	1637	9	73026643	<i>KLF9</i>
FT3	cg01695994	-1.112	0.18	8.43E-10	1638	17	80246403	<i>CSNK1D/LINC01970</i>
FT3	cg02183564	-0.736	0.17	8.97E-06	1636	7	76874892	<i>CCDC146</i>
FT3	cg04173586	-1.842	0.25	7.98E-14	1607	19	2167496	<i>DOT1L</i>
FT3	cg19837174	-0.553	0.20	6.50E-03	1634	10	6389707	<i>LOC399715</i>
TSH	cg03445151	-0.002	0.02	9.32E-01	4053	2	23516881	<i>KLHL29/AC012506.1</i>
TSH	cg20065905	0.003	0.02	8.72E-01	4068	17	80560980	<i>FOXK2</i>