THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

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

Citation for published version:
Weihs, A, Chaker, L, Martin, TC, Braun, KVE, Campbell, PJ, Cox, SR, Fornage, M, Gieger, C, Grabe, HJ, Grallert, H, Harris, S, Kuhnel, B, Marioni, RE, Martin, N, McCartney, DL, Mcrae, AF, Meisinger, C, van Meurs, JBJ, Nano, J, Nauck, M, Peters, A, Prokisch, H, Roden, M, Selvin, E, Beekman, M, van Heemst, D, Slagboom, PE, Swenson, BR, Tin, A, Tsai, P-C, Uitterlinden, AG, Visser, WE, Völzke, H, Waldenberger, M, Walsh, JP, Köttgen, A, Wilson, SG, Peeters, RP, Bell, JT, Medici, M & Teumer, A 2023, 'Epigenome-wide association study reveals CpG sites associated with thyroid function and regulatory effects on KLF9', *Thyroid Journal Program*, vol. 33, no. 3, pp. 301–311. https://doi.org/10.1089/thy.2022.0373

Digital Object Identifier (DOI):
10.1089/thy.2022.0373

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Thyroid Journal Program

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 19. May. 2023

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

Running title: EWAS on Thyroid Function

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

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 Orthopeadics & 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 Human Genetics, Helmholtz Zentrum München German Research Center for Environmental Health, 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
29 Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins School of Medicine
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 Bialystok, Bialystok, Poland

* These authors contributed equally.
# These authors jointly supervised this work.

Correspondence to:

Alexander Teumer, PhD
Institute for Community Medicine
University Medicine Greifswald, Germany
ateumer@uni-greifswald.de
phone: +49 (0) 3834-86-19579

keywords: thyroid function, DNA methylation, KLF9, Mendelian randomization, gene expression

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 an epigenome-wide association study of blood DNA methylation sites in up to 7,073 participants from 8 cohorts of both European and African ancestry individuals from the ThyroidOmics Consortium. Significant associations from the discovery stage were replicated in independent samples. Subsequently, the validated findings were correlated with gene expression levels and genetic variants. Causal influence of thyroid hormones on the DNA methylation levels was assessed by 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 thyroid stimulating hormone (TSH) concentrations were discovered and replicated in independent cohorts (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/LINCO1970* and *LRRC8D* 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*, *LRRC8D* 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 (1, 2). The 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 (3). However, the exact molecular mechanisms underlying these processes are still 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 (4), showing in developing frog tissue that triiodothyronine directly controls the expression of the DNA methyltransferase 3a (Dnmt3a), 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, demonstrating that although DNAm is highly tissue specific, the analysis of DNAm in blood may provide deeper insights into thyroid hormone action and/or regulation (5). Motivated by the conclusion of that study, namely 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 patterns in whole blood encompassing up to 7,073 individuals of the ThyroidOmics Consortium (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 (6), ARIC (7), and KORA (8). To validate our EWAS findings, we sought additional cohorts for replication. In the replication stage we included the Rotterdam Study (RS) (9), TwinsUK (10), the Lothian Birth Cohorts of 1921 and 1936 (LBC1921 and LBC1936) (11), and the Brisbane Systems Genetics Study (BSGS) (12). Detailed information is provided in the **Supplementary Methods** and in **Supplementary Table 1**. All study protocols were approved by the respective local ethics committees.

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 **Supplementary Table 1**.

Statistical analysis

The EWAS was undertaken in each cohort separately and subsequently meta-analysed. 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 (**Supplementary Figure 2**). Details of the analyses workflow are provided in the **Supplementary Methods**.

The findings of the discovery stage that passed significance after Bonferroni correction for the number of sites tested (p-value < 0.05/450,000 = 1.1E-07) were taken forward for replication in additional independent samples. A successfully replicated site was defined by p-value < 0.05 in the replication stage with consistent effect direction, and a p-value < 1.1E-07 in the discovery and replication combined meta-analysis.

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 study (**Supplementary Methods**)(13). Results passing the false discovery rate (FDR) < 0.05 were declared significant.

Additional characterization of the findings

We assessed the genetic impact on DNAm, and the overlap of the thyroid-associated CpGs with thyroid hormone receptor binding sites using published datasets (**Supplementary Methods**).

Mendelian randomization analysis

To test for a possible causal effect of the thyroid hormones on the DNAm levels of the significantly associated CpG sites, we conducted a 2-sample MR using the R package *TwoSampleMR* (14). As no genetic instruments for FT3 were available, we only conducted the MR on the FT4- and TSH-associated sites using the independent genome-wide significantly associated index SNPs from a large genome-wide association study (GWAS) as instruments for the thyroid hormone levels (15). 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 (16) 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 (17) and MR-Egger (18), 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 cohorts, 2,988 additional individuals were included in the TSH analysis, 590 for FT3 and 2,448 for FT4 (**Supplementary Methods**).

DNAm sites associated with thyroid hormone levels

We discovered and replicated in independent cohorts 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 CpGs associated with FT3 (in *KLF9*, *DOT1L*, *LRRC8D*, and near *CSNK1D/LINC01970*), the two CpGs in *LRRC8D* (cg06983052 and cg20146909), also represent novel associations. The replicated results are provided in **Table 1**, the significant results of the discovery stage are shown in **Supplementary Table 2**, and the cohort-specific results are listed in **Supplementary Table 3**. In total, all but one site identified in the discovery analysis were successfully replicated (**Supplementary Table 2**). Detailed annotation of the replicated CpGs is shown in **Supplementary Table 4**.

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 the two associations cg00049440 ($\beta_{combined}$ = 0.13; p-value = 9.58E-16) and cg04173586 ($\beta_{combined}$ = 0.07; p-value = 1.53E-09). The CpG site cg00049440 is located inside 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* ($\beta_{combined}$ = -0.75; p-value = 4.76E-20, **Supplementary Figure 5**) and cg04173586 in *DOT1L* ($\beta_{combined}$ = -0.43; p-value = 4.18E-11, **Supplementary Figure 6**) were also

identified in the TSH analysis. The additional association is cg03546163 ($\beta_{combined}$ = -0.84; p-value = 1.54E-16), which lies within the second intron of *FKBP5* (**Supplementary Figure 7**).

The EWAS on FT3 identified the five sites cg00049440, cg01695994, cg04173586, cg06983052 and cg20146909, of which two of the sites, namely cg00049440 ($\beta_{combined}$ = -1.46; p-value = 1.68E-17, **Supplementary Figure 8**) and cg04173586 ($\beta_{combined}$ = -1.93; p-value = 4.22E-28, **Supplementary Figure 9**) were also identified in the FT4 and TSH and one CpG, namely cg06983052 ($\beta_{combined}$ = -1.04; p-value = 5.40E-14, **Supplementary Figure 10**) was identified in the FT4 analysis. The remaining two novel sites are cg01695994 ($\beta_{combined}$ = -1.37; p-value = 5.76E-17), which is located 1,518bp downstream of *LINC01970* and 14,809bp upstream of *CSNK1D* (**Supplementary Figure 11**), and cg20146909 ($\beta_{combined}$ = -1.036; p-value = 6.51E-14), which is located inside intron 1 of the gene *LRRC8D* (**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 the cg00049440 passed the level of significance (FDR <0.05) and was negatively correlated with the gene expression of *KLF9* (β = -0.109; p-value = 3.00E-7, **Supplementary Table 5**). 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 4**). As thyroid hormone receptor footprints are dynamic, capable of chromatin remodeling and dependent on the level of thyroid hormones (19, 20), 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 in liver and brain tissues (distance between 517 bp and 2,219 bp, **Supplementary Table 4**). 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 4**). DNAm heritability estimates at cg00049440 and cg03546163 were explained by 72% and 100% of common genetic variants in the genome (h^2_{SNPs}/h^2), respectively (21). 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 the GoDMC (http://www.godmc.org.uk)(22), several genetic variants, locally (*cis*) and distally (*trans*), impact the blood DNAm levels of all significant CpG sites except *DOT1L* (**Supplementary Table 6**). Interestingly, the 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(15, 23) 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 (24, 25). 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* (β = 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 7**). 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 (26). After excluding the 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 7**).

Discussion

In our EWAS of DNAm associated with thyroid function, 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 4**. Although the associations of the CpG sites at Krueppel-like factor 9 (*KLF9*) and DOT1 like histone lysine methyltransferase (*DOT1L*) with FT3 were known (5), we additionally discovered associations with 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 gene *KLF9* is part of the Sp1 C2H2-type zinc finger family of transcription factors that binds to GC box elements located in the promotor (NCBI RefSeq database). It has been shown that *KLF9* is a triiodothyronine response gene in mice (20) and frogs (27), where triiodothyronine regulates gene expression by binding to and activating thyroid hormone receptors that regulate the gene

transcription of *KLF9* by eliciting an effect via a thyroid response element (20). In addition, there is evidence that this mechanism is also present in human cells including hepatocytes (28) and neural cells (29). Our *KLF9* findings serve as an example of thyroid function affecting gene expression via DNAm in humans. In particular, our results indicate 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* leading to increased *KLF9* expression in blood cells (**Figure 4**). Importantly, causality in this cascade was furthermore 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 (15), 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 (30, 31). This strongly suggests that the TSH associated genetic variants are a more valid representation of thyroid function, which likely explains the absence of associations in our FT4 variant 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 4**). 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 (32).

The second multi-trait association, *DOT1L*, represents a ubiquitously expressed gene that codes 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 (33). 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. In addition, the sample size might have been too small to

achieve the required statistical power for the *DOT1L* gene expression analysis, although the effect direction suggests an inverse association (**Supplementary Table 5**). 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 the two new and three known associations with FT3 revealed in our study, the remaining three associations with FT3 discovered in the study of Lafontaine *et al.* (5) were statistically significant ($p < 0.05/6$) in our EWAS. Taking into account that our replication sample for FT3 was included also in Lafontaine *et al.*, we restricted this lookup to the discovery stage results (**Supplementary Table 8**) 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.* (5). 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 lnRNA *AC012506.1* were not replicated in our study (**Supplementary Table 8**). 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 3, 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 additional 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 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.

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 2012 Subclinical thyroid disease. Lancet **379**:1142–1154.
2. Wondisford, Fredric E; Radovick S 2009 Clinical Management of Thyroid Disease, 1st ed. Elsevier.
3. Wu Y, Koenig RJ 2000 Gene Regulation by Thyroid Hormone. Trends Endocrinol Metab **11**:207–211.
4. Kyono Y, Sachs LM, Bilesimo P, et al. 2016 Developmental and Thyroid Hormone Regulation of the DNA Methyltransferase 3a Gene in Xenopus Tadpoles. Endocrinology **157**:4961–4972.
5. Lafontaine N, Campbell PJ, Castillo-Fernandez JE, et al. 2021 Epigenome-Wide Association Study of Thyroid Function Traits Identifies Novel Associations of fT3 With KLF9 and DOT1L. J Clin Endocrinol Metab **106**:e2191–e2202.
6. Völzke H, Schössow J, Schmidt CO, et al. 2022 Cohort Profile Update: The Study of Health in Pomerania (SHIP). Int J Epidemiol https://doi.org/10.1093/ije/dyac034.
7. Wright JD, Folsom AR, Coresh J, et al. 2021 The ARIC (Atherosclerosis Risk In Communities) Study: JACC Focus Seminar 3/8. J Am Coll Cardiol **77**:2939–2959.
8. Holle R, Happich M, Löwel H, et al. 2005 KORA - A Research Platform for Population Based Health Research. Das Gesundheitswes **67**:19–25.
9. Hofman A, Brusselle GGO, Murad SD, et al. 2015 The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol **30**:661–708.
10. Moayyeri A, Hammond CJ, Valdes AM, et al. 2013 Cohort Profile: TwinsUK and Healthy Ageing Twin Study. Int J Epidemiol **42**:76–85.
11. Taylor AM, Pattie A, Deary IJ 2018 Cohort Profile Update: The Lothian Birth Cohorts of 1921 and 1936. Int J Epidemiol **47**:1042-1042r.
12. Powell JE, Henders AK, McRae AF, et al. 2012 The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. PLoS One **7**:e35430.
13. Schurmann C, Heim K, Schillert A, et al. 2012 Analyzing illumina gene expression microarray data from different tissues: methodological aspects of data analysis in the metaxpress consortium. PLoS One **7**:e50938.
14. Hemani G, Zheng J, Elsworth B, et al. 2018 The MR-Base platform supports systematic causal inference across the human phenome. Elife **7**.
15. Teumer A, Chaker L, Groeneweg S, et al. 2018 Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nat Commun **9**:4455.
16. Houseman EA, Accomando WP, Koestler DC, et al. 2012 DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics **13**:86.
17. Bowden J, Davey Smith G, Haycock PC, et al. 2016 Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol **40**:304–314.
18. Bowden J, Smith GD, Burgess S 2015 Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. Int J Epidemiol **44**.
19. Grøntved L, Waterfall JJ, Kim DW, et al. 2015 Transcriptional activation by the thyroid hormone

receptor through ligand-dependent receptor recruitment and chromatin remodelling. Nat Commun 2015 61 **6**:1–11.

20. Denver RJ, Williamson KE 2009 Identification of a Thyroid Hormone Response Element in the Mouse Krüppel-Like Factor 9 Gene to Explain Its Postnatal Expression in the Brain. Endocrinology **150**:3935–3943.

21. Van Dongen J, Nivard MG, Willemsen G, et al. 2016 Genetic and environmental influences interact with age and sex in shaping the human methylome. Nat Commun 2016 71 **7**:1–13.

22. Min JL, Hemani G, Hannon E, et al. 2021 Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat Genet **53**:1311–1321.

23. Zhou W, Brumpton B, Kabil O, et al. 2020 GWAS of thyroid stimulating hormone highlights pleiotropic effects and inverse association with thyroid cancer. Nat Commun **11**:3981.

24. Ebrahim S, Davey Smith G 2008 Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? Hum Genet **123**:15–33.

25. Medici M, Peeters RP, Teumer A, et al. 2019 The importance of high-quality mendelian randomisation studies for clinical thyroidology. lancet Diabetes Endocrinol **50**:668–681.

26. Medici M, Porcu E, Pistis G, et al. 2014 Identification of novel genetic Loci associated with thyroid peroxidase antibodies and clinical thyroid disease. PLoS Genet **10**:e1004123.

27. Furlow JD, Kanamori A 2002 The Transcription Factor Basic Transcription Element-Binding Protein 1 Is a Direct Thyroid Hormone Response Gene in the Frog Xenopus laevis. Endocrinology **143**:3295–3305.

28. Cvoro A, Devito L, Milton FA, et al. 2015 A Thyroid Hormone Receptor/KLF9 Axis in Human Hepatocytes and Pluripotent Stem Cells. Stem Cells **33**:416–428.

29. Morte B, Gil-Ibáñez P, Bernal J 2018 Regulation of Gene Expression by Thyroid Hormone in Primary Astrocytes: Factors Influencing the Genomic Response. Endocrinology **159**:2083–2092.

30. Marouli E, Kus A, Del Greco M F, et al. 2020 Thyroid function affects the risk of stroke via atrial fibrillation: a Mendelian Randomization study. J Clin Endocrinol Metab https://doi.org/10.1210/clinem/dgaa239.

31. Kuś A, Marouli E, Del Greco M F, et al. 2020 Variation in normal range thyroid function affects serum cholesterol levels, blood pressure and type 2 diabetes risk: A Mendelian randomization study. Thyroid thy.2020.0393.

32. Raj S, Kyono Y, Sifuentes CJ, et al. 2020 Thyroid Hormone Induces DNA Demethylation in Xenopus Tadpole Brain. Endocrinology **161**.

33. Wen L, Fu L, Shi Y 2017 Histone methyltransferase Dot1L is a coactivator for thyroid hormone receptor during Xenopus development. FASEB J **31**:4821–4831.

Tables

Table 1: CpG sites significantly associated with thyroid function after the replication stage.

CpG probeID	Chr	Position (Build 37)	Estimate	Standard Error	P-value	Sample Size	Nearest Gene
FT4							
cg00049440	9	73,026,643	-0.750	0.082	**4.44E-20**	6526	*KLF9*
cg03546163	6	35,654,363	-0.843	0.102	**1.45E-16**	5129	*FKBP5*
cg04173586	19	2,167,496	-0.427	0.065	**4.03E-11**	6482	*DOT1L*
FT3							
cg00049440[1]	9	73,026,643	-1.463	0.172	**1.69E-17**	2227	*KLF9*
cg01695994[1]	17	80,246,403	-1.371	0.164	**5.84E-17**	2228	*CSNK1D/LINC01970*
cg04173586[1]	19	2,167,496	-1.935	0.176	**4.33E-28**	2193	*DOT1L*
cg06983052	1	90,288,099	-1.040	0.138	**5.45E-14**	2229	*LRRC8D*
cg20146909	1	90,289,611	-1.036	0.138	**6.51E-14**	2229	*LRRC8D*
TSH							
cg00049440	9	73,026,643	0.125	0.016	**9.57E-16**	7070	*KLF9*
cg04173586	19	2,167,496	0.072	0.012	**1.53E-09**	7025	*DOT1L*

[1] known association. The effect estimates and p-values of the discovery and replication combined meta-analysis are provided.

Figure legends

Figure 1: 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 genome wide significance cut-off (1.1E-07). Points plotted in the upper panel show a positive correlation with TSH, while those plotted blow have a negative association.

Figure 2: Chicago plot showing the results of the FT4 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 genome wide significance cut-off (1.1E-07). Points plotted in the upper panel show a positive correlation with FT4, while those plotted blow have a negative association.

Figure 3: 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 genome wide significance cut-off (1.1E-07). Points plotted in the upper panel show a positive correlation with FT3, while those plotted blow have a negative association.

Figure 4: Overview of the study analyses and findings. Known mechanisms and associations are colored in black, new findings revealed by this study are marked in blue color. Image created with BioRender.com