Beyond the CpG: an integrative approach to decoding DNA methylation in immunometabolic health

Sinke, L.J.

BEYOND THE CPG

an integrative approach to decoding DNA methylation in immunometabolic health

Lucy Sinke

Stellingen behorend bij het proefschrift getiteld:

BEYOND THE CPG

an integrative approach to decoding DNA methylation in immunometabolic health

1. Meta-analysis of large-scale blood-based datasets can reveal biologically meaningful DNA methylation, even for traits non-hematopoietic in origin (**this thesis; Hannon *et al.* eLife, 2021**).
2. Only by triangulating evidence from complementary techniques can we extract the most promising loci from high-dimensional results (**this thesis; Gutierrez *et al.*, Eur J Epidemiol, 2025**).
3. CpGs should not be interpreted independently of the regulatory proteins they encounter, their broader chromatin environments, or the cells they were discovered in (**this thesis; Machado *et al.* Brief Funct Genomics, 2014; Schübeler, Nature, 2015; Yin *et al.*, Science, 2021**).
4. With an unprecedented wealth of public data at our fingertips, progress now depends on our ability to translate associations into clinically relevant results (**this thesis; Jiang *et al.* Nat Rev Cancer, 2022; Singh *et al.* Adv Protein Chem Struct Biol, 2024**).
5. Following robust identification of disease-associated CpGs across tissues and studies, our focus must shift from discovery to mechanistic interpretation (**Kaluscha *et al.*, Nat Genet, 2022; Hawe *et al.*, Nat Genet, 2022**).
6. Novelty accelerates science as long as fundamental questions remain at the steering wheel (**Cohen, eLife, 2017; Teplitskiy *et al.*, Proc Natl Acad Sci, 2022**).
7. Advances in infrastructure and accessible tools now support collaborative discovery across sensitive data without compromising privacy (**Teo *et al.*, Cell Rep Med, 2024; The Galaxy Community, Nucleic Acids Res, 2024**).
8. Even the most sophisticated algorithms cannot rescue a flawed experimental design (**Leek *et al.*, Nat Rev Genet, 2013; Goh *et al.*, Trends Biotechnol, 2018**).
9. While discovery charts our scientific paths, it is the people who nurture our creativity and curiosity that illuminate them.
10. Humans are emerging as an unparalleled model organism for studying human biology.

BEYOND THE CPG

an integrative approach to
decoding DNA methylation in immunometabolic
health

Lucy Sinke

Beyond the CpG:
an integrative approach to decoding DNA methylation in immunometabolic health
L. Sinke, M.Sc.
ISBN: 978-94-6534-235-1

The research described in this thesis was performed at the Department of Biomedical Data Sciences of the Leiden University Medical Centre in The Netherlands, and was supported by the Joint Programming Initiative "a Healthy Diet for a Healthy Life" (JPI-HDHL) DIMENSION project (ZonMw grant number: 529051021).

Financial support for publication of this thesis from Leiden University Medical Centre, Cantoni Therapeutics, and Biopartner Leiden is gratefully acknowledged.

The cover illustration depicts the scientific detective work performed within this thesis. By tracing multiple lines of evidence, researchers uncover *wanted* loci where DNA methylation is intimately connected to immunometabolic health.

Printed by ProefschriftMaken.nl
Layout and design by Lucy and Mark Sinke

BEYOND THE CPG

an integrative approach to
decoding DNA methylation in immunometabolic health

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. S. de Rijcke,
volgens besluit van het college voor promoties
te verdedigen op donderdag 7 mei 2026
klokke 14:30 uur

door
Lucy Sinke
geboren te Londen, Verenigd Koninkrijk
in 1990

Promotores:

Prof. dr. B. T. Heijmans

Prof. dr. P. E. Slagboom

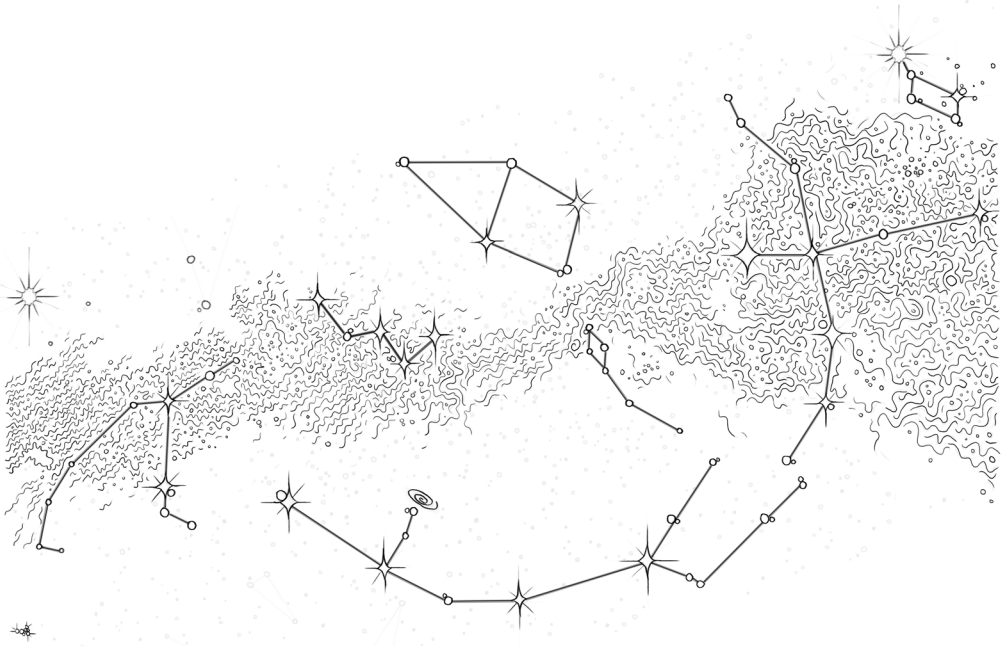
Leden promotiecommissie:

Prof. dr. G. W. E. Santen
Secretaris

Dr. ir. N. van Geloven

Prof. dr. ir. C. P. G. M. de Groot
Wageningen University & Research, Wageningen

Prof. dr. J. B. J. van Meurs
Erasmus University Medical Center, Rotterdam



for Mark

looking at the same sky together

Table of contents

CHAPTER ONE

Introduction

Healthy Ageing

Age is the leading risk factor for most common, chronic diseases, but the ageing process itself is highly heterogenous[1,2]. While some individuals suffer from decades of multimorbidity and functional decline, others enjoy a comparatively healthy older age with lower levels of frailty and limited disease development. Modifiable lifestyle factors, such as diet and physical activity, can influence these trajectories and targeted interventions have successfully improved health in older adults[3,4]. Yet, key challenges remain to identify the molecular determinants of age-related phenotypes and uncover the biological roots driving intervention effects[5].

Markers of health offer valuable insights into the processes and pathways underlying ageing. They encompass a broad spectrum of measures, from functional tests and anthropometry to clinical chemistry and cellular biomarkers, each with distinct strengths and limitations[6]. While physiological metrics are typically easier to observe and interpret in the clinic, they emerge following a series of subtler shifts in health[7]. Early detection of such inter-individual variation is where biochemical and molecular markers shine[8]. Interleukin-6 (IL-6) levels can predict cardiovascular disease and all-cause mortality years prior to clinical onset[9,10], nutrient-sensing genes are robustly linked to longevity and represent promising therapeutic targets[11–13], and epigenetic clocks integrating DNA methylation (DNAm) at sensitive sites can accurately estimate chronological age[14,15]. These examples offer a glimpse into the promise of molecular biology in ageing research, but the links between markers and the mechanics of their regulation await full elucidation.

Immunometabolism

Profound shifts in both immunity and metabolism accompany ageing, and these systems are intimately engaged in a bidirectional crosstalk termed *immunometabolism*[16,17]. This dynamic communication shapes physiological function and dysfunction throughout the lifecourse, providing a compelling framework for studying the biology of ageing, particularly in tissues, such as blood, adipose, and muscle, which sit at the crossroads between immune surveillance and metabolic regulation[18,19].

Among these interconnected systems, age-related changes in the immune compartment have been especially well characterized[20,21]. Leukocyte populations undergo extensive remodelling with age, culminating in a distinct immunological landscape known as *immunosenescence*[22]. This state is marked by a decline in naïve T-cell populations relative to terminally differentiated memory cells, inversion of CD4+/CD8+ T-cell ratios, and reduced T-cell receptor diversity[23,24]. Such qualitative and quantitative immunological alterations have consequences for the health of older adults, contributing to impaired tumour surveillance, increased susceptibility to autoimmunity, and

higher risk of chronic, non-communicable disease (**Fig. 1a**)[25–27].

Paradoxically, even as adaptive immunity declines, a persistent, low-grade inflammation referred to as *inflammaging* emerges[28]. Sustained in part by chronic stimulation from misplaced self-molecules and cellular debris, affected leukocytes remain in a state of prolonged activation[29]. These exhausted immune cells, no longer finely attuned to specific threats, mount maladaptive responses and exhibit a heightened propensity to differentiate into pro-inflammatory subtypes, such as Th17 effectors (**Fig. 1b**)[30,31]. Inflammaging is characterized by elevated serum levels of molecular mediators, including IL-6 and TNF-α[32,33]. The physiological relevance of these changes is underscored by studies in centenarians, where genetic variants in the *IL6* locus strongly correlate with longer lifespans[34], and by growing evidence that dampening inflammation may be central to attaining exceptional longevity[35,36].

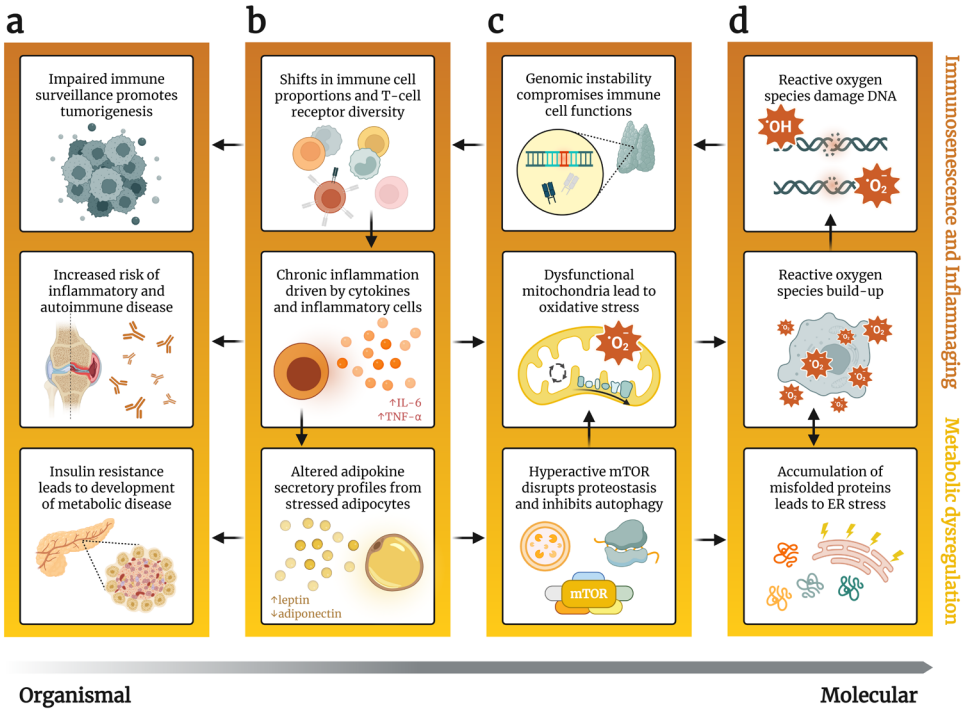


Figure 1 | Immune, inflammatory, and metabolic components of ageing across biological scales. a) Organismal-level alterations increase incidence of cancer, autoimmune, and metabolic diseases, including type 2 diabetes. **b)** Loss of naïve T-cells impairs adaptive immunity, cytokine production promotes chronic inflammation, and stressed adipocytes adjust their secretory profiles. **c)** Cellular effects include genomic disruptions of immune cell function, mitochondrial dysfunction, and hyperactive mTOR signalling. **d)** Molecular changes propagate endoplasmic reticulum (ER) and oxidative stress, fuelling DNA damage and genomic instability. *Figure created in BioRender.*

At the cellular level, age-associated changes in immune function both rely on and shape shifts in metabolism[37]. Pro-inflammatory cytokines, such as IL-6, trigger stress pathways that impair mitochondrial function, disrupt proteostasis, and contribute to genomic instability (**Fig. 1c,d**)[38,39]. Additionally, immune cells depend on metabolic reprogramming to meet the energetic and biosynthetic demands of activation[40]. While oxidative phosphorylation is largely sufficient to support the modest energy requirements of naïve T-cells, activating cells rapidly shift to glycolysis for ATP generation. Indeed, glucose deprivation alone can suppress T-cell expansion and cytokine production even when alternative metabolic substrates are available, highlighting the critical role of metabolic flexibility in ageing and the immune response[41,42].

This interplay between immunity and metabolism extends beyond the circulation as well. Chronic inflammation infiltrates other tissues, including adipose and muscle, where activation of cytokine receptors on resident cells both impairs glucose homeostasis and promotes insulin resistance[43]. Yet, adipocytes and myocytes are not mere passive targets of inflammaging; they also secrete pro- and anti-inflammatory mediators, including IL-6, adiponectin, and leptin, thereby modulating immune responses and influencing organismal energy balance[44,45]. This bidirectional crosstalk systemically propagates age-related dysfunction, underpinning a constellation of immune, inflammatory, and metabolic alterations[46]. Taken together, these processes demonstrate the pivotal role of distinct cellular phenotypes and their regulation for shaping health trajectories in later life, and position immunometabolism as a unifying framework for understanding ageing and health in the population at large.

Gene regulation and epigenetics

Cellular phenotypes are fundamentally determined by gene expression and protein production patterns[47]. This is exemplified by effector T-cell subtypes, whose distinct transcriptional and translational signatures enable their transformation from naïve cells into specialized lineages[48]. The classical markers that distinguish these subsets underscores the vital role of genomic regulation in shaping their fates (**Fig. 2a**). Lineage-defining *transcription factors* (TFs) such as T-bet drive Th1 identity by promoting expression of IFN-γ, *receptors* including CD25 detect IL-2 in the extracellular milieu and support regulatory T-cell (Treg) survival and immunosuppressive functions, and IL-17A *secretions* enables Th17 subtypes to amplify local inflammation and recruit neutrophils to sites of tissue damage[49].

The execution of these differential programs depends on a complex and tightly regulated network of molecular interactions. TFs act in concert with co-factors and the surrounding chromatin landscape to activate or repress target genes[50,51]. Epigenetic alterations, including DNAm and histone modifications, modulate chromatin accessibility, thereby determining which genomic regions are transcriptionally permissive, and which are

not[52]. Through the integration of these regulatory layers of control, cells establish the expression profiles that define their phenotypes and ultimately enable their precise functional specialization[53].

DNAm is one such epigenetic modification, involving the addition of a methyl group to the 5' position of cytosine residues, primarily at CpG dinucleotides (**Fig. 2b**). By influencing local chromatin structure and TF binding potential, DNAm can modulate expression without altering underlying genetics[54]. While traditionally regarded as a repressive mark, particularly when located in promoters or enhancers, the regulatory effects of DNAm are now recognized as strongly context dependent (**Fig. 2c**)[55,56].

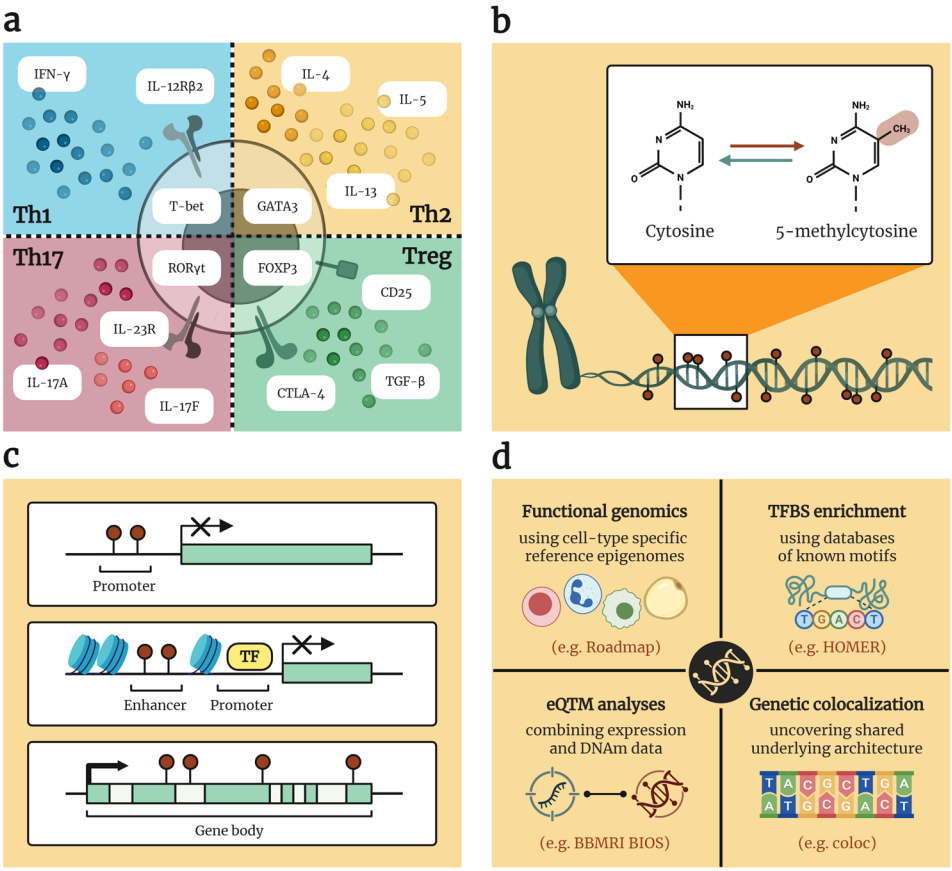


Figure 2 | Gene regulation involves co-operation between transcription factors (TFs) and the chromatin landscape. a) T-cell subtypes are defined by lineage-defining TFs, receptors, and secretions. **b)** DNA methylation (DNAm) attaches a methyl group (red) to cytosine residues. **c)** DNAm is context-dependent: it may block the binding of TFs at promoter and enhancer regions but is also linked to transcriptional activation when located in gene bodies. **d)** Large-scale integrative analyses can improve CpG-gene annotations including functional genomics, TF binding site (TFBS) enrichment, eQTM analyses, and genetic colocalization. *Figure created in BioRender.*

Methylation within gene bodies, for example, is frequently associated with active transcription, and intergenic DNAm can help preserve the boundaries of regulatory domains. Furthermore, expression quantitative trait methylation (eQTM) studies have demonstrated that DNAm correlates with expression in both directions; although associations are predominantly negative, a substantial proportion are positive, underscoring a nuanced role for DNAm in transcriptional regulation[57].

This context-dependent regulatory nature has complicated the development of reliable maps linking DNAm to target genes[58]. While CpGs are frequently annotated to the nearest gene body, this pairing is crude: it overlooks tissue-specificity, ignores local chromatin structure, and provides no insight into the regulatory factors involved. Emerging evidence supports more distal CpG-gene associations and reveals that TFs differ in their sensitivity to DNAm, highlighting the limitations of simple proximity based approaches[59,60]. Refining CpG-gene annotations will require large-scale, longitudinal, and tissue-specific analyses that combine available epigenetic, transcriptomic, and genetic data, such as those generated by the Roadmap Epigenomics and BBMRI BIOS consortia[61]. These resources offer valuable opportunities to uncover the shared regulatory architecture of molecular traits. Yet, systematic integration of evidence-based mapping is not yet routine in epigenomic studies, and functional and mechanistic interpretations continue to lag behind their scientific potential (**Fig. 2d**).

Epigenome-wide association studies

DNAm is a stable and accessible epigenetic mark that can be profiled using minimally invasive sampling and cost-effective, microarray-based technologies[62]. These practical advantages have fuelled a wave of large-scale epidemiological investigations examining DNAm variation across populations. A core method in this field is the epigenome-wide association study (EWAS), which systematically associates DNAm with a trait of interest[63,64]. EWAS have become a cornerstone of population epigenetics, revealing DNAm signatures for a broad range of traits, including age, lifestyle factors, and immunometabolic health[14,65–67].

Methylation changes across the lifecourse have been particularly well-characterised[68]. Genome-wide analyses and twin studies uncovered a hallmark signature of global hypomethylation, particularly in intergenic and repetitive elements, accompanied by site-specific hypermethylation at CpG-rich promoters[69,70]. A subset of these predictable modifications forms the foundation of epigenetic clocks, which estimate calendar age using highly informative CpGs[14,15]. Beyond chronological prediction, these clocks also capture biological changes, including shifts in immune cell composition, mortality risk, and functional decline, with reorganization of the epigenetic landscape now established as a characteristic component of human ageing[8,71,72].

EWAS have also revealed robust associations between DNAm and immunometabolic health, linking CpG sites at key regulatory genes to body mass index (BMI), circulating lipid levels, and type 2 diabetes across multiple studies[73–75]. Notable examples include *ABCG1*, which encodes a cholesterol transporter central to lipid homeostasis and inflammation, *SREBF1*, a transcriptional regulator of lipogenesis, and *CPT1A*, a rate-limiting enzyme in mitochondrial fatty acid oxidation[65,76]. These findings reveal a central role for methylation of metabolic loci in health and disease, illustrating the power of EWAS to uncover compelling CpGs with clear biological relevance.

Yet, despite these successes, the epigenetic architecture of many clinically relevant traits and exposures remains poorly characterised. While certain risk factors, such as adiposity and smoking, have been extensively studied, many others lie unexplored[64]. As a result, gaps persist in our understanding of how DNAm contributes and responds to early shifts in immunometabolic health, and EWAS efforts will need to be expanded to capture a broader spectrum of inter- and intra-individual epigenetic variation.

Technical considerations

Although EWAS primarily yield correlational findings, their interpretation requires careful consideration of plausible underlying causal structures[77]. By directly altering the mechanical properties of DNA, CpG methylation can modulate chromatin accessibility and alter TF binding[54,60]. However, this regulatory *potential* by no means guarantees that CpGs identified in EWAS are a product of direct, functional effects within studied cells (**Fig. 3a**). Observed DNAm-trait associations may also arise from technical or biological confounding, reverse causation, or shared upstream factors (**Fig. 3b-f**)[62,78]. Distinguishing between these scenarios requires a combination of thoughtful study design, high-quality data, and analytical strategies capable of separating biological signals from technical artifacts.

Among these scenarios, technical confounding (**Fig. 3b**) is arguably the only one that is wholly undesirable. Arising from differences in sample handling, array position, or environmental conditions during processing, these effects represent variation introduced only after sample collection and therefore do not reflect underlying biology in the organism or test system sampled. As DNAm datasets expand, they become increasingly vulnerable to these confounders which, if not properly addressed, can obscure signals or even generate misleading results[79,80]. While *post hoc* adjustment strategies, such as ComBat and other normalisation techniques, are widely employed, they risk inadvertently masking genuine biological variation[81]. Consequently, proactive strategies to minimise batch effects through experimental design and randomisation are preferred. Despite this, practical frameworks and accessible tools to support researchers at the design stage remain limited, and there is a disconnect between best-practice recommendations and their routine implementation in high-throughput studies.

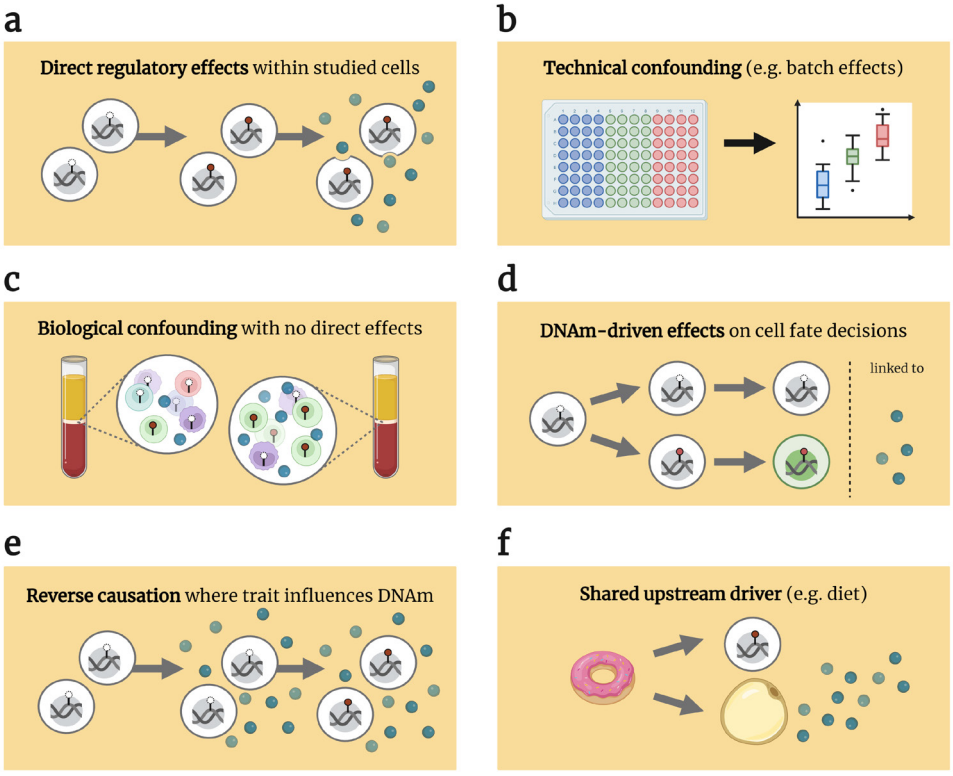


Figure 3 | Six potential causal structures. a) Direct regulatory DNAm effects within studied cells alters expression and thereby influences the trait. **b)** Associations between the trait and technical factors induce systematic DNAm differences. **c)** Associations between cell composition and the trait lead to apparent DNAm differences at CpGs that mark cell types but are not linked to their differentiation. **d)** DNAm directed cell fate decisions lead to changes in cell-type composition associated with the trait. **e)** The trait influences DNAm in studied cells. **f)** Shared upstream factors like diet cause independent effects on DNAm in studied cells and the trait, including through DNAm effects in a non-studied cell type. *Figure created in BioRender.*

Since DNAm patterns are tightly linked to cellular identity, biological confounding poses another major consideration for EWAS[82,83]. Blood, the most accessible and widely studied tissue in epigenetic epidemiology, offers rich opportunities for large-scale analysis, but its cellular heterogeneity complicates interpretation. DNAm-trait associations may arise from shifts in immune cell proportions (**Fig. 3c**) rather than within-cell differential methylation, and this is further complicated by methylation's ability to direct cell fate decisions (**Fig. 3d**)[84,85]. Although routine clinical tests measure only a small number of immune cell categories, recent developments in algorithmic deconvolution now allow estimation of a much wider spectrum of cellular subtypes across multiple tissues[86–88]. This has greatly improved the available resolution for disentangling drivers of observed results. Crucially, however, changes in sample composition are also biologically meaningful readouts, such as the shifts in immune cell populations that occur

with age[24]. As such, EWAS interpretation must carefully distinguish between cellular variation that introduces bias and that which reveals key aspects of immune or physiological function.

Processing and analysing DNAm data remains a rapidly evolving science, driven by novel platforms and technological innovations. Bioconductor packages, such as *methylAid*[89], *omicsPrint*[90], and *bacon*[91] provide powerful, interactive frameworks for monitoring data quality, verifying sample identify, and reducing bias in analysis. Yet, computational pipelines and packages require regular updates, and optimal analytical choices vary with study aims and designs[92,93]. Flexible, scalable, and reproducible workflows, supported by clear and accessible documentation, are essential to empower researchers to keep pace with advances while ensuring high-quality analysis. Even so, modular and transparent resources remain scarce and the steep learning curve of DNAm data processing and analysis continues to limit broad participation.

Yet, despite these ongoing methodological challenges, the expanding availability of large-scale DNAm data offers opportunities to aggregate EWAS findings across populations. In this context, meta-analysis has become an indispensable tool, enhancing the statistical power available to detect modest but consistent DNAm-trait associations. Scalable software such as *METAL* facilitates the integration of summary statistics while quantifying cohort-specific heterogeneity, allowing researchers to detect robust signals that may not reach significance in individual studies[94]. To fully capitalise on EWAS meta-analysis, careful consideration of plausible causal structures will be required alongside collaborative efforts and consortia aimed at elucidating how the DNAm landscape varies with complex traits.

Functional interpretation

Most EWAS to date remain largely descriptive and there is a pressing need to move beyond site discovery towards in-depth functional interpretation[77]. While genic annotation of CpGs, particularly when strengthened by tissue-specific reference epigenomes and TF binding site information, can offer preliminary clues about regulatory potential, this alone is insufficient to contextualize high-dimensional results. Achieving meaningful biological interpretation requires a more systematic and integrative approach. Yet, despite growing enthusiasm and the increasing availability of relevant resources, such follow-up remains inconsistently applied. This limits the field's capacity to generate mechanistic insights and leaves the full potential of EWAS unrealized.

One promising avenue to address this gap is to use curated databases containing gene sets and prior EWAS findings (**Fig. 4a,b**)[95,96]. These resources can be systematically interrogated through enrichment analyses to determine whether differentially methylated CpGs and their putative target genes converge on shared biological

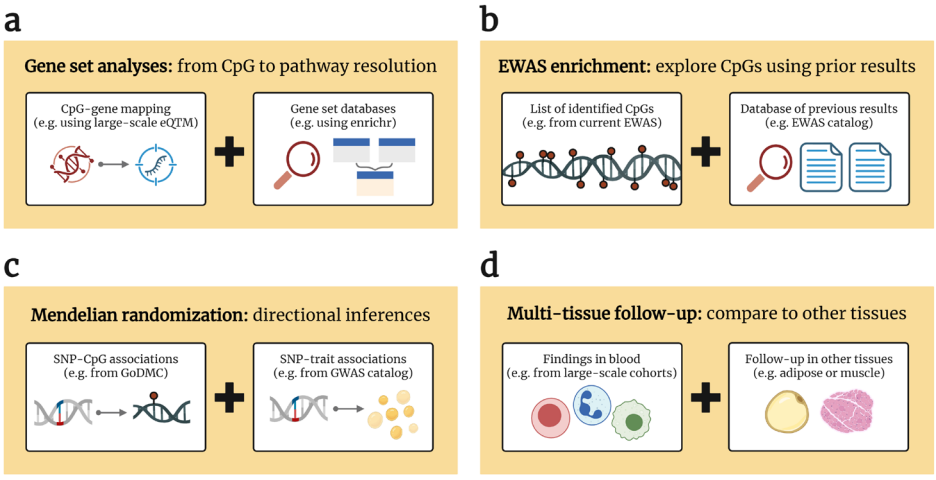


Figure 4 | Follow-up analyses available for epigenome-wide studies including a) gene-set analyses, **b)** EWAS enrichment analysis, **c)** bidirectional two-sample Mendelian randomization, and **d)** multi-tissue investigations and interventions. *Figure created in BioRender.*

processes or pathways. This can illuminate underlying mechanisms, reveal points of regulatory control, and help prioritize loci for functional follow-up. By embedding biological and clinical context into epigenomic workflows, researchers can shift from the isolated nucleotide resolution of CpGs back up to interpretable, systems-level biology.

Yet even biologically plausible associations do not imply causality. A central challenge in epigenetic epidemiology is determining whether differential methylation plays a direct role in shaping cellular phenotypes, or if it instead reflects a different underlying causal structure such as reverse causation (**Fig. 3e**) or shared upstream drivers (**Fig. 3f**). Robust study designs and causal inference approaches can assist in disentangling these possibilities: bidirectional two-sample Mendelian randomization provides a powerful statistical framework to assess directionality of associations by leveraging genetic instruments as proxies for exposure and outcome (**Fig. 4c**)[82], tissue- or cell-type specific follow-up can assess whether blood-based signals mirror changes in other relevant compartments (**Fig. 4d**)[78], and intervention studies, in which exposure precedes DNAm by design, offer a valuable window into causality by revealing if epigenetic control is responsive to established modulators of immunometabolic health.

When applied in concert and thoughtfully interpreted, these integrative approaches will allow researchers to move beyond CpG associations, highlighting focal points for mechanistic investigations and generating testable hypotheses about how epigenetics and immunometabolic health interact. In doing so, such frameworks have potential to transform the field of epigenetic epidemiology from a cataloguing exercise into a mature, mechanistically grounded discipline that advances translational research.

Outline of this thesis

A central feature of ageing is a progressive decline in immunometabolic health, amplifying the risk of certain conditions in older adults. This thesis aims to examine the epigenetic component of such immunometabolic health variation through cohort and intervention studies, using in-depth functional epigenomics to gain fresh insights into the mechanisms that limit or lead to the pathogenesis of age-related disease.

In *Chapters 2 and 3*, we explore DNAm signatures of circulating adipocytokines, addressing key knowledge gaps in the epigenetics of immunometabolic risk. **Chapter 2** focuses on IL-6, a central inflammatory mediator that increases with age and is predictive of cardiovascular events and all-cause mortality[9,10]. By combining blood-based data from over 4,000 individuals, we identify CpG sites associated with circulating IL-6 levels and assess their regulatory potential. Furthermore, we uncover the shared molecular architecture of DNAm, expression, and IL-6 through integration with transcriptomic and genetic data, including large-scale quantitative trait loci databases and genome-wide association studies. In this manner, the resulting epigenetic signature is connected to immunometabolic risk and disease, with three complementary causal inference approaches placing CpGs along the aetiological pathway from elevated inflammation to IL-6 related conditions, such as increased BMI and inflammatory bowel disease.

In **Chapter 3** we conduct a meta-analysis of five European cohorts to explore links between DNAm and serum adiponectin and leptin, two circulating adipokines with established regulatory roles in ageing and immunometabolic disease. Using comprehensive functional annotation, integration with BIOS consortium data, and follow-up in adipocytes, we carefully consider the directionality, tissue-specificity, and broader relevance of our findings for immunometabolic health, including by exploring causal effects on lipid levels and type 2 diabetes.

Recognising the common challenges in epigenome-wide studies, *Chapter 4 and 5* focus on improving analytical workflows and data quality for use in a novel experimental setting. In **Chapter 4** we present statistical software designed to proactively minimise batch effects in high-throughput genomic studies with special attention for accessible implementation and paired designs. **Chapter 5** introduces *DNAmArray*, a reproducible and modular workflow with extensive documentation that streamlines quality control, preprocessing, and analysis in EWAS using existing packages, including *methylAid*[89], *omicsPrint*[90], and *bacon*[91], ensuring high-quality DNAm data for downstream analyses.

In contrast to observational cohort study designs, interventions can examine alteration to immunometabolic health and molecular traits in parallel. Therefore, in **Chapter 6** we evaluate multi-tissue epigenomic responses to a combined lifestyle intervention, where older adults improved their health through a combination of calorie restriction and increased physical activity. We comprehensively characterise epigenetic effects in

fasted blood, subcutaneous adipose tissue, and skeletal muscle, linking them to changes in immunometabolic health markers, including BMI, body fat percentage, grip strength, and serum adipocytokine levels. Furthermore, we incorporate epigenetic clocks to assess the efficacy of DNAm-based algorithms for capturing intervention-driven effects, demonstrating value for combining experimental and computational methods in epigenomic studies.

The general discussion in **Chapter 7** synthesizes our main findings within the broader context of immunosenescence and inflammaging. We discuss the importance of integrative approaches in large-scale EWAS, emphasizing the role of functional annotation and causal inference in moving beyond CpG associations. By comparing findings across *Chapters 2, 3, and 6*, we identify shared epigenetic indicators of health and focal points for follow-up investigations. In closing, we offer fresh perspectives for future directions and outline next steps on the path towards a more mechanistic and translational epigenetic science.

References

1. Niccoli, T. *and* Partridge, L. Ageing as a risk factor for disease. *Curr Biol* **22** (17): R741-752 (2012).
2. Partridge, L., Deelen, J. *and* Slagboom, P. E. Facing up to the global challenges of ageing. *Nature* **561** (7721): 45–56 (2018).
3. Reiner, M. *et al.* Long-term health benefits of physical activity: A systematic review of longitudinal studies. *BMC Public Health* **13**: 813 (2013).
4. Green, C. L., Lamming, D. W. *and* Fontana, L. Molecular mechanisms of dietary restriction promoting health and longevity. *Nat Rev Mol Cell Biol* **23** (1): 56-73 (2022).
5. Campisi, J. *et al.* From discoveries in ageing research to therapeutics for healthy ageing. *Nature* **571** (7764): 183–192 (2019).
6. Wagner, K. H. *et al.* Biomarkers of Aging: From Function to Molecular Biology. *Nutrients* **8** (6): 338 (2016).
7. Khan, S. S., Singer, B. D. *and* Vaughan, D. E. Molecular and physiological manifestations and measurement of aging in humans. *Aging Cell* **16** (4): 624–633 (2017).
8. López-Otín, C. *et al.* Hallmarks of aging: An expanding universe. *Cell* **186** (2): 243–278 (2023).
9. Hirata, T. *et al.* Associations of cardiovascular biomarkers and plasma albumin with exceptional survival to the highest ages. *Nat Commun* **11** (1): 3820 (2020).
10. Mossmann, M. *et al.* Increased serum IL-6 is predictive of long-term cardiovascular events in high-risk patients submitted to coronary angiography: an observational study. *Diabetol Metab Syndr* **14** (1): 125 (2022).
11. Mannick, J. B. *and* Lamming, D. W. Targeting the biology of aging with mTOR inhibitors. *Nat Aging* **3** (6): 642-660 (2023).
12. Johnson, S. C., Rabinovitch, P. S. *and* Kaeberlein, M. MTOR is a key modulator of ageing and age-related disease. *Nature* **493** (7432): 338-345 (2013).
13. Passtoors, W. M. *et al.* Gene expression analysis of mTOR pathway: Association with human longevity. *Aging Cell* **12** (1): 24-31 (2013).
14. Bernabeu, E. *et al.* Refining epigenetic prediction of chronological and biological age. *Genome Med* **15** (1): 12 (2023).
15. Horvath, S. *and* Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat Rev Genet* **19** (6): 371-384 (2018).
16. Hotamisligil, G. S. Inflammation, metaflammation and immunometabolic disorders. *Nature* **542** (7640): 177-185 (2017).
17. Zhu, X. *et al.* Inflammation, epigenetics, and metabolism converge to cell senescence and ageing: the regulation and intervention. *Signal Transduct Target Ther* **6** (1): 245 (2021).
18. Mraz, M. *and* Haluzik, M. The role of adipose tissue immune cells in obesity and low-grade inflammation. *J Endocrinol* **222** (3): R113-127 (2014).
19. Juban, G. *and* Chazaud, B. Metabolic regulation of macrophages during tissue repair: insights from skeletal muscle regeneration. *FEBS Lett* **591** (19): 3007-3021 (2017).
20. Montecino-Rodriguez, E., Berent-Maoz, B. *and* Dorshkind, K. Causes, consequences, and reversal of immune system aging. *J Clin Invest* **123** (3): 958-965 (2013).
21. Liu, Z. *et al.* Immunosenescence: molecular mechanisms and diseases. *Signal Transduct Target Ther* **8** (1): 200 (2023).
22. Walford, R. L. The immunologic theory of aging. *Gerontologist* **4**: 195–197 (1964).
23. Nikolich-Žugich, J. The twilight of immunity: Emerging concepts in aging of the immune system. *Nat Immunol* **19** (1): 10-19 (2018).
24. Mittelbrunn, M. *and* Kroemer, G. Hallmarks of T cell aging. *Nat Immunol* **22** (6): 687-698 (2021).
25. Weyand, C. M. *and* Goronzy, J. J. Aging of the immune system: Mechanisms and therapeutic targets. *Ann Am Thorac Soc* **13**: S422-S428 (2016).
26. Barbé-Tuana, F. *et al.* The interplay between immunosenescence and age-related diseases. *Semin Immunopathol* **42** (5): 545–557 (2020).
27. Lian, J. *et al.* Immunosenescence: a key player in cancer development. *J Hematol Oncol* **13** (1): 151 (2020).
28. Fulop, T. *et al.* Immunosenescence and inflammaging as two sides of the same coin: Friends or Foes? *Front Immunol* **8**: 1960 (2018).
29. Franceschi, C. *et al.* Inflammaging and 'Garb-aging'. *Trends Endocrinol Metab* **28** (3): 199-212 (2017).
30. Haynes, L. *and* Maue, A. C. Effects of aging on T cell function. *Curr Opin Immunol* **21** (4): 414-417 (2009).
31. Schmitt, V., Rink, L. *and* Uciechowski, P. The Th17/Treg balance is disturbed during aging. *Exp Gerontol* **48** (12): 1379-1386 (2013).
32. Ferrucci, L. *and* Fabbri, E. Inflammageing: chronic inflammation in ageing, cardiovascular disease, and frailty. *Nat Rev Cardiol* **15** (9): 505-522 (2018).
33. Michaud, M. *et al.* Proinflammatory cytokines, aging, and age-related diseases. *J Am Med Dir Assoc* **14** (12): 887-882 (2013).
34. Zeng, Y. *et al.* Novel loci and pathways significantly associated with longevity. *Sci Rep* **6:** 21243 (2016).

35. Zhou, L. *et al.* Centenarians Alleviate Inflammaging by Changing the Ratio and Secretory Phenotypes of T Helper 17 and Regulatory T Cells. *Front Pharmacol* **13**: 877709 (2022).
36. Minciullo, P. L. *et al.* Inflammaging and Anti-Inflammaging: The Role of Cytokines in Extreme Longevity. *Arch Immunol Ther Exp (Warsz)* **64** (2): 111-126 (2016).
37. Martin, D. E. *et al.* Targeting Aging: Lessons Learned From Immunometabolism and Cellular Senescence. *Front Immunol* **12**: 714742 (2021).
38. Desdín-Micó, G. *et al.* T cells with dysfunctional mitochondria induce multimorbidity and premature senescence. *Science* **368** (6497): 1371-1376 (2020).
39. Hipp, M. S., Kasturi, P. *and* Hartl, F. U. The proteostasis network and its decline in ageing. *Nat Rev Mol Cell Biol* **20** (7): 421-435 (2019).
40. Soto-Heredero, G. *et al.* Glycolysis: a key player in the inflammatory response. *FEBS J* **287** (16): 3350-3369 (2020).
41. Macintyre, A. N. *et al.* The glucose transporter Glut1 is selectively essential for CD4 T cell activation and effector function. *Cell Metab* **20** (1): 61-72 (2014).
42. Cham, C. M. *et al.* Glucose Deprivation Inhibits Multiple Key Gene Expression Events and Effector Functions in CD8+ T Cells. *Eur J Immunol* **38** (9): 2438-2450 (2008).
43. Brestoff, J. R. *and* Artis, D. Immune regulation of metabolic homeostasis in health and disease. *Cell* **161** (1): 146-160 (2015).
44. Maury, E. *and* Brichard, S. M. Adipokine dysregulation, adipose tissue inflammation and metabolic syndrome. *Mol Cell Endocrinol* **314** (1): 1-16 (2010).
45. Leal, L. G., Lopes, M. A. *and* Batista, M. L. Physical exercise-induced myokines and muscle-adipose tissue crosstalk: A review of current knowledge and the implications for health and metabolic diseases. *Front Physiol* **9**: 1307 (2018).
46. De Carvalho, F. G. *et al.* Adipose tissue quality in aging: How structural and functional aspects of adipose tissue impact skeletal muscle quality. *Nutrients* **11** (11): 2553 (2019).
47. Klemm, S. L., Shipony, Z. *and* Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* **20** (4): 207-220 (2019).
48. Zhou, L., Chong, M. M. W. *and* Littman, D. R. Plasticity of CD4+ T Cell Lineage Differentiation. *Immunity* **30** (5): 646-655 (2009).
49. Saravia, J., Chapman, N. M. *and* Chi, H. Helper T cell differentiation. *Cell Mol Immunol* **16** (7): 634-643 (2019).
50. Wilson, C. B., Rowell, E. *and* Sekimata, M. Epigenetic control of T-helper-cell differentiation. *Nat Rev Immunol* **9** (2): 91-105 (2009).
51. Henning, A. N., Roychoudhuri, R. *and* Restifo, N. P. Epigenetic control of CD8+ T cell differentiation. *Nat Rev Immunol* **18** (5): 340-356 (2018).
52. Allis, C. D. *and* Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17** (8): 487-500 (2016).
53. Smith, Z. D. *and* Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14** (3): 204-220 (2013).
54. Moore, L. D., Le, T. *and* Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **38** (1): 23-38 (2013).
55. Greenberg, M. V. C. *and* Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20** (10): 590-607 (2019).
56. Jones, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13** (7): 484-492 (2012).
57. Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49** (1): 131-138 (2017).
58. Hannon, E. *et al.* Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. *Am J Hum Genet* **103** (5): 654-665 (2018).
59. Calo, E. *and* Wysocka, J. Modification of enhancer chromatin: what, how and why? *Mol Cell* **49** (5): 825-837 (2013).
60. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356** (6337): eaaj2239 (2017).
61. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518** (7539): 317-330 (2015).
62. Bock, C. Analysing and interpreting DNA methylation data. *Nat Rev Genet* **13** (10): 705–719 (2012).
63. Rakyan, V. K. *et al.* Epigenome-wide association studies for common human diseases. *Nat Rev Genet* **12** (8): 529-541 (2011).
64. Wei, S. *et al.* Ten Years of EWAS. *Adv Sci (Weinh)* **8** (20): e2100727 (2021).
65. Ling, C. *and* Rönn, T. Epigenetics in Human Obesity and Type 2 Diabetes. *Cell Metab* **29** (5): 1028-1044 (2019).
66. Szarc Vel Szic, K. *et al.* From inflammaging to healthy aging by dietary lifestyle choices: Is epigenetics the key to personalized nutrition? *Clin Epigenetics* **7** (1): 33 (2015).

67. Fragoso-Bargas, N. *et al.* Epigenome-wide association study of objectively measured physical activity in peripheral blood leukocytes. *BMC Genomics* **26** (1): 62 (2025).
68. Jones, M. J., Goodman, S. J. *and* Kobor, M. S. DNA methylation and healthy human aging. *Aging Cell* **14** (6): 924-932 (2015).
69. Teschendorff, A. E., West, J. *and* Beck, S. Age-associated epigenetic drift: implications, and a case of epigenetic thrift? *Hum Mol Genet* **22** (R1): R7-R15 (2013).
70. Talens, R. P. *et al.* Epigenetic variation during the adult lifespan: Cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* **11** (4): 694-703 (2012).
71. Jonkman, T. H. *et al.* Functional genomics analysis identifies T and NK cell activation as a driver of epigenetic clock progression. *Genome Biol* **23** (1): 24 (2022).
72. Kuiper, L. M. *et al.* Epigenetic and Metabolomic Biomarkers for Biological Age: A Comparative Analysis of Mortality and Frailty Risk. *J Gerontol A Biol Sci Med Sci* **78** (10): 1753-1762 (2023).
73. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541** (7635): 81-86 (2017).
74. Pfeiffer, L. *et al.* DNA methylation of lipid-related genes affects blood lipid levels. *Circ Cardiovasc Genet* **8** (2): 334-342 (2015).
75. Walaszczyk, E. *et al.* DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA1c levels: a systematic review and replication in a case–control sample of the Lifelines study. *Diabetologia* **61** (2): 354-368 (2018).
76. van Dijk, S. J. *et al.* Recent developments on the role of epigenetics in obesity and metabolic disease. *Clin Epigenetics* **7**: 66 (2015).
77. Mill, J. *and* Heijmans, B. T. From promises to practical strategies in epigenetic epidemiology. *Nat Rev Genet* **14** (8): 585-594 (2013).
78. Birney, E., Smith, G. D. *and* Greally, J. M. Epigenome-wide Association Studies and the Interpretation of Disease -Omics. *PLoS Genet* **12** (6): e1006105 (2016).
79. Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* **11** (10): 733-739 (2010).
80. Harper, K. N., Peters, B. A. *and* Gamble, M. V. Batch effects and pathway analysis: Two potential perils in cancer studies involving DNA methylation array analysis. *Cancer Epidemiol Biomarkers Prev* **22** (6): 1052-1060 (2013).
81. Yang, H. *et al.* Randomization in laboratory procedure is key to obtaining reproducible microarray results. *PLoS One* **3** (11): e3724 (2008).
82. Teschendorff, A. E. *and* Relton, C. L. Statistical and integrative system-level analysis of DNA methylation data. *Nat Rev Genet* **19** (3): 129-147 (2018).
83. Loyfer, N. *et al.* A DNA methylation atlas of normal human cell types. *Nature* **613** (7943): 355-364 (2023).
84. Dekkers, K. F. *et al.* Human monocyte-to-macrophage differentiation involves highly localized gain and loss of DNA methylation at transcription factor binding sites. *Epigenetics Chromatin* **12** (1): 34 (2019).
85. Sun, D. *et al.* Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* **14** (5): 678-688 (2014).
86. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**: 86 (2012).
87. Salas, L. A. *et al.* Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat Commun* **13** (1):761 (2022).
88. Wang, X. *et al.* Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat Commun* **10** (1): 380 (2019).
89. Van Iterson, M. *et al.* MethylAid: visual and interactive quality control of large Illumina 450k datasets. *Bioinformatics* **30** (23): 3435-3437 (2014).
90. Van Iterson, M. *et al.* omicsPrint: detection of data linkage errors in multiple omics studies. *Bioinformatics* **34** (12): 2142-2143 (2018).
91. van Iterson, M. *et al.* Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18** (1): 19 (2017).
92. Michels, K. B. *et al.* Recommendations for the design and analysis of epigenome-wide association studies. *Nat Methods* **10** (10): 949-955 (2013).
93. Wilhelm-Benartzi, C. S. *et al.* Review of processing and analysis methods for DNA methylation array data. *Br J Cancer* **109** (6): 1394-1402 (2013).
94. Willer, C. J., Li, Y. *and* Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26** (17): 2190-2191 (2010).
95. Battram, T. *et al.* The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res* **7**: 41 (2022).
96. Li, M. *et al.* EWAS Atlas: A curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res* **47** (D1): D983-D988 (2019).

CHAPTER TWO

Interleukin-6

Epigenome-wide association study of circulating interleukin-6 connects DNA methylation to immunometabolic and inflammatory health

Lucy Sinke[1], Jenny van Dongen[2,3], Thomas Delerue[4], Rory Wilson[4], Yujing Xia[5], Marian Beekman[1], Gonneke Willemsen[2,3], Christian Gieger[4,6], Christian Herder[7,8], Wolfgang Koenig[6,9,10], Annette Peters[4,6,11], Eco J.C. de Geus[2,3], Jose M. Ordovas[12], Jordana T. Bell[5], Melanie Waldenberger[4,6], Dorret I. Boomsma[2,3], P. Eline Slagboom[1], and Bastiaan T. Heijmans[1]

[1] Leiden University Medical Centre, Leiden, The Netherlands
[2] Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[3] Amsterdam University Medical Centre, Amsterdam, The Netherlands
[4] Helmholtz Munich, Neuherberg, Germany
[5] King's College London, London, United Kingdom
[6] German Centre for Cardiovascular Research (DZHK), Munich, Germany
[7] Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[8] German Centre for Diabetes Research (DZD), Neuherberg, Germany
[9] Technical University of Munich, Munich, Germany
[10] Ulm University, Ulm, Germany
[11] Ludwig Maximilian University of Munich, Munich, Germany
[12] Tufts University, Boston, Massachusetts, United States of America

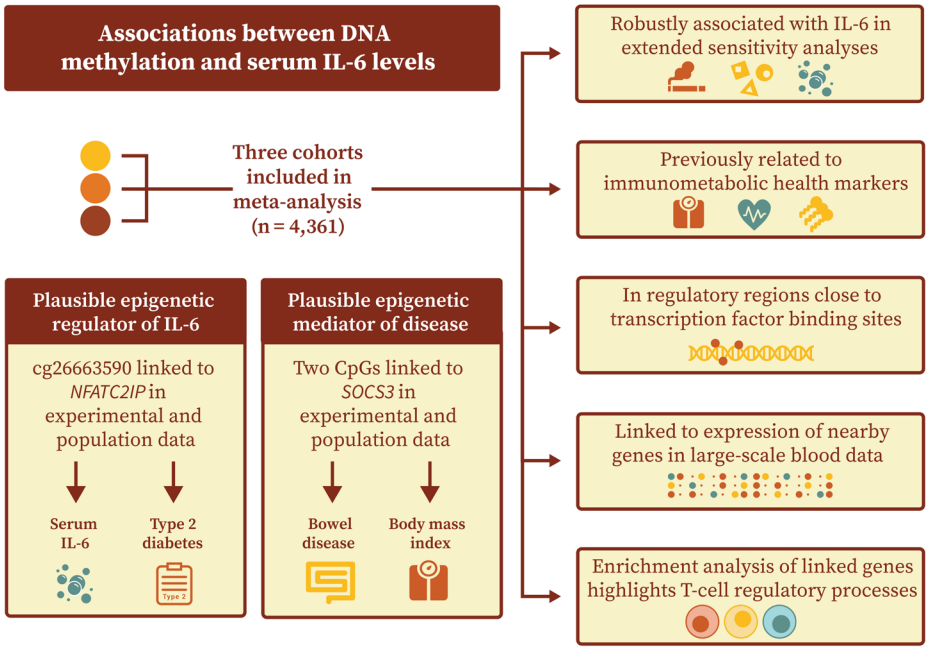
Published in Communications Biology 9, 242 (2026)
DOI: 10.1038/s42003-026-09520-2

Abstract

Interleukin-6 (IL-6) drives metabolic and inflammatory processes central to disease. Current knowledge implicates epigenetic mechanisms in the regulation of these pathways, including through the methylation of CpG sites. This blood-based meta-analysis of three cohorts (*n* = 4,361) identifies 401 IL-6–associated CpGs enriched in regulatory regions and linked to key immunometabolic genes, including *AIM2*, *MTOR*, and *IL6R*.

Three complementary causal inference approaches support most sites as responding to IL-6, with *SOCS3* (Suppressor of Cytokine Signalling 3) methylation statistically mediating inflammatory bowel disease risk. Notably, one CpG connected to *NFATC2IP* (Nuclear Factor of Activated T-cells 2 Interacting Protein) plausibly influences both IL-6 production and multiple immunometabolic conditions, including body mass index and type 2 diabetes. Collectively, our results map the DNA methylation landscape surrounding circulating IL-6 levels and unveil directional effects and distinct functional relationships between epigenetics and inflammation.

Graphical Abstract



Highlights

- 401 CpGs are robustly associated with circulating IL-6 levels in blood.
- These CpGs disproportionately reside in active chromatin close to regulatory binding sites for cellular stress and immune responses.
- Triangulation supports IL-6 as driving DNA methylation at the majority of the identified CpGs.
- In line with this, IL-6 associated DNA methylation correlates with genes that promote pro-inflammatory immune cell fates in response to cytokine signalling.
- Public data reveals a subset of these genes as responsive to IL-6 stimulation of isolated CD4+ T-cells *in vitro*.
- Mediation analysis connects the DNA methylation response at *SOCS3* to specific adverse health outcomes, including increased BMI and risk of inflammatory bowel disease.
- Bidirectional two-sample Mendelian randomisation identifies a single CpG, correlated with *NFATC2IP* expression in both population and experimental data, as driving IL-6 production and immunometabolic health.

Keywords: *DNA methylation; cytokines; inflammation; immunometabolic health; diabetes; functional genomics; EWAS; colocalisation; triangulation; Mendelian randomisation*

Background

Interleukin-6 (IL-6) is a multifunctional cytokine and a central player in immunity, metabolism and disease[1,2]. A low-grade chronic inflammation, characterised by elevated levels of circulating IL-6, is observed in both ageing and multiple health conditions, including type 2 diabetes (T2D), rheumatoid arthritis (RA), and inflammatory bowel disease (IBD)[3–5]. Evidence from preclinical models, genome-wide association studies (GWAS), and Mendelian randomisation directly implicates IL-6 in the pathogenesis and progression of these diseases[6,7], and inhibitors of IL-6 signalling are successfully used as therapeutics[8–10]. Despite this, the molecular mechanisms leading to, and effects of elevated IL-6 levels, remain incompletely understood.

Inflammatory cascades are complex and highly context-dependent[1,11]. IL-6 activation of JAK/STAT signalling, for example, can both amplify *IL6* expression in a positive feedback loop and also restrain IL-6 activity via *SOCS3*. Furthermore, IL-6 production is governed by a multi-layered regulatory network, integrating transcription factors (TFs) with post-transcriptional control to ensure rapid yet reversible induction[1,2]. Such specific and tightly regulated processes can be orchestrated in part by epigenetic mechanisms, including DNA methylation (DNAm), which modifies genomic accessibility and TF binding, leading to altered expression of target genes[12]. Yet, while epigenome-wide association studies (EWAS) have robustly defined how DNAm associates with IL-6 related phenotypes such as body mass index (BMI), IL-6 itself has been comparatively underexplored[13–15].

Here, we perform a meta-analysis of epigenome-wide associations with circulating IL-6 levels from over 4,000 participants, exploring sensitivity of effects to smoking, twelve distinct immune cell types, and C-reactive protein (CRP). We comprehensively characterise the resulting DNAm signature through functional genomics, large-scale integrative analyses, and colocalisation. Finally, we apply three complementary causal inference techniques: triangulation, mediation analysis, and two-sample Mendelian randomisation (2SMR), to directionally connect CpGs to both IL-6 and a broad range of immunometabolic traits and diseases.

Results

Circulating IL-6 levels are associated with DNA methylation in blood

We conducted a meta-analysis of epigenome-wide associations with circulating IL-6 levels in a combined sample of 4,361 individuals from three cohorts (**Table 1**). Mean age ranged from 37.6 years in the NTR to 68.8 years in KORA, and the population was predominantly female (60.9%). Analyses were performed on DNAm profiled from whole blood samples; representing an accessible, metabolic tissue widely used in clinical diagnostics and comprised of a mixture of immune cell-types that produce and respond to circulating cytokines[16,17]. The top 20 CpGs are presented (**Supplementary Table 1**) alongside a study design overview with key findings (**Fig. 1**).

Characteristic	KORA F4	LLS	NTR
Sample size	799	668	2894
Age, years	68.8 ± 4.4	58.8 ± 6.7	37.6 ± 12.7
Sex, female	390 (48.8)	346 (51.8)	1921 (66.4)
Smoking, current	69 (8.6)	85 (12.7)	608 (21.0)
Smoking, never	381 (47.7)	205 (30.7)	1603 (55.4)
hsCRP, mg/L	1.47 (2.25)	2.00 (2.33)	1.41 (2.80)
IL-6, pg/mL	1.51 (1.26)	0.63 (0.43)	1.00 (0.90)

Table 1 | Characteristics of the three cohorts included in this IL-6 EWAS meta-analysis. Values are shown as mean ± standard deviation for age (in years), as median (IQR) for IL-6 levels (in pg/mL) and high-sensitivity C-reactive protein (hsCRP; in mg/L), and as total (%) for sex and smoking status.

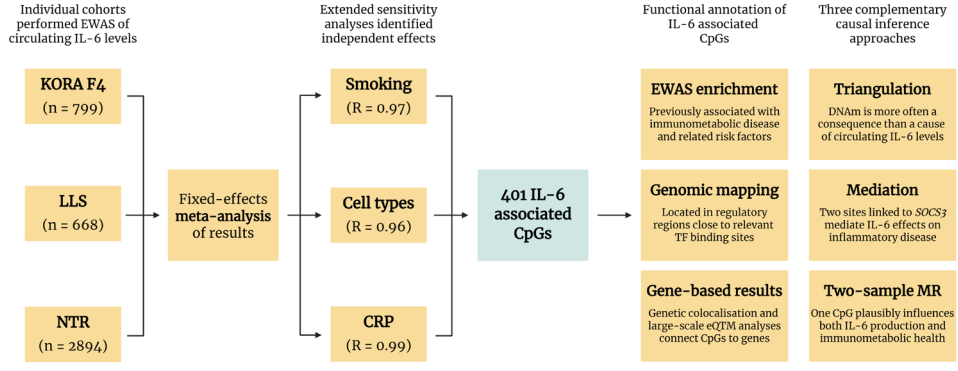


Figure 1 | Flowchart summarizing the study design. This study included meta-analysis, extended sensitivity analyses, functional annotation, and multiple causal inference approaches. *Figure created in BioRender.*

After adjusting for age, sex, technical covariates, and six immune cell-types predicted from DNAm data, IL-6 was associated with 531 CpGs ($p_{FDR} < 0.05$). These associations were obtained following extensive quality control and correction for minor inflation ($0.94 \leq \lambda \leq 1.25$) and bias ($|\mu| \leq 0.25$) in the test statistics. Six CpGs were excluded due to between-cohort heterogeneity ($I^2 > 90\%$), comprising three positively and three inversely associated with IL-6. These removed CpGs, as well as the top 100 IL-6 associated CpGs, showed consistent directions of effect across cohorts, indicating that overall directions were comparable despite estimate variability (**Supplementary Fig. 1**).

Since data preprocessing steps, cell-type proportions, and smoking can influence EWAS findings, we evaluated the stability of IL-6 associations using sensitivity analyses. For DNAm data, outlying values were removed in line with published guidelines and prior work[18,19]. A sensitivity analysis in LLS excluding this processing step confirmed that it did not unduly influence effect sizes in this cohort ($R = 0.95$, **Supplementary Fig. 2a**). For circulating IL-6 levels, undetectable values were left-censored resulting in removal of 1.19% of total samples ($n = 52$). A sensitivity analysis in the cohort with the largest number of undetectable values (LLS) showed a strong correlation between effect sizes using left censoring or limit of detection capping, indicating that findings were also similar across IL-6 data preprocessing strategies ($R = 0.98$, **Supplementary Fig. 2b**).

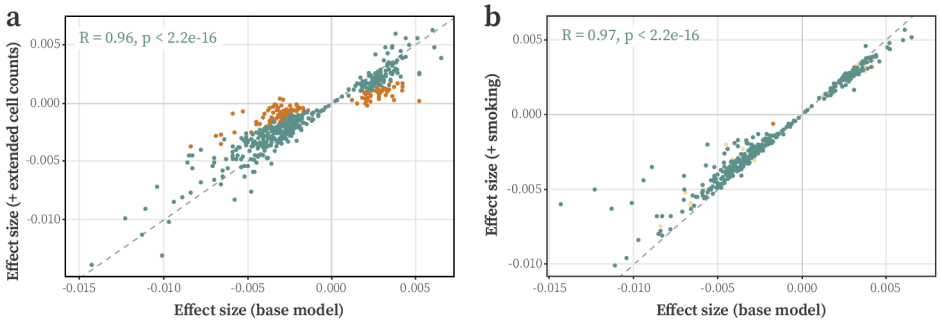


Figure 2 | Sensitivity analysis of IL-6 associated CpG sites. a) Scatter plot comparing effect sizes from the base model (unadjusted) versus a model adjusted for twelve predicted blood cell-type proportions. CpG sites with insufficient evidence for an independent IL-6 effect ($p_{FDR} \geq 0.05$) after adjustment are highlighted in orange ($n = 130$). **b)** Scatter plot comparing base model effect sizes versus a model adjusted for smoking status. CpGs previously removed due to cell-type confounding are shown in yellow and one CpG removed due to smoking effects is highlighted in orange.

Furthermore, all three cohorts extended base models by additionally adjusting for smoking status or the predicted proportions of twelve immune cell-types[20]. Effect size estimates from these adjusted models were highly correlated with those from the base model ($R_{CELL} = 0.96$ and $R_{SMOKE} = 0.97$; **Fig. 2a,b**), suggesting that these common confounders minimally impacted IL-6 effects at the majority of identified CpGs.

For 130 CpGs, IL-6 associations were no longer significant after correcting for multiple testing in one or more sensitivity analyses ($p_{FDR} \geq 0.05$), indicating that signals at these sites were partly attributable to variation in cell-type proportions or smoking. These CpGs were therefore excluded from the final results to ensure focus on IL-6 specific effects in downstream analyses. In summary, our analyses uncovered 401 IL-6 associated CpGs mapped to 384 distinct genomic loci. Effect sizes ranged from -0.0143 to 0.0065, with most associations representing an inverse relationship between IL-6 and DNA methylation (n = 282, 70.3%; **Fig. 3a-c**).

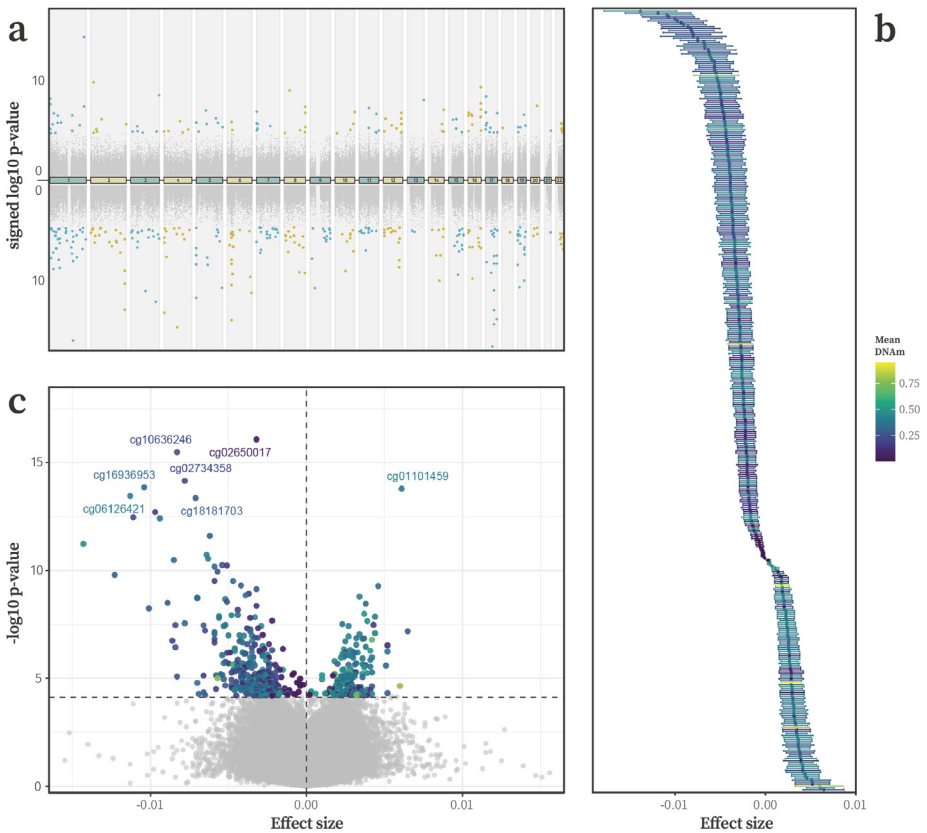


Figure 3 | DNAm associations with circulating IL-6 levels. a) Directional Manhattan plot showing the genomic position of all 412,226 tested CpGs (x-axis) and their signed -$\log_{10}$ p-values (y-axis). CpGs positively associated with IL-6 are displayed in the upper half, and inversely correlated CpGs in the lower half. 401 CpGs associated with IL-6 after multiple testing correction and sensitivity analyses are highlighted in blue (odd chromosomes) and yellow (even chromosomes). Grey dots represent all other tested CpGs. **b)** Effect sizes and 95% confidence intervals (CIs) for the 401 IL-6 associated CpGs, with points coloured by mean DNAm level across the three cohorts. **c)** Volcano plot of IL-6 EWAS effect sizes (x-axis) against -$\log_{10}$ p-values (y-axis) for all tested CpGs. Significant CpGs are coloured by mean DNAm level and the top hits are labelled. All other tested CpGs are shown in grey.

IL-6 associated CpGs are relevant for immunometabolic health and disease

To assess the broader relevance of the IL-6 associated CpGs to other phenotypes, we cross-referenced our findings against publicly available EWAS databases. Of the 401 CpGs associated with IL-6, 329 (82.0%) had been previously connected to at least one other trait. As anticipated, the largest overlap was with CRP (241 CpGs, 60.1%), an acute-phase protein whose hepatic production is directly upregulated by IL-6, and this represented a strong enrichment (OR = 438.9, p_{FDR} < 2.2e-308)[21].

To investigate the bidirectional relationship between epigenetic signatures of these two inflammatory biomarkers further, we first assessed the impact of adjusting for hsCRP on IL-6 associations. Remarkably, following meta-analysis, all 401 IL-6 associated CpGs retained their association with IL-6 (p_{FDR} < 0.05) and effect sizes were only minimally attenuated (R = 0.99, **Fig. 4a**). To explore the reciprocal direction, we examined 1,649 CpGs reported as associated with CRP in previous studies and conducted an hsCRP EWAS meta-analysis in the three cohorts both with and without adjustment for IL-6[22–24]. Correction for IL-6 here resulted in visual attenuation of the hsCRP-DNAm effect sizes (**Fig. 4b**). Nevertheless, the correlation between effects before and after adjustment remained high (R = 0.90). These results suggested that, despite considerable overlap between IL-6 and CRP associated loci alongside their biological similarity, these methylation signatures of inflammation may represent partially distinct and independent signals.

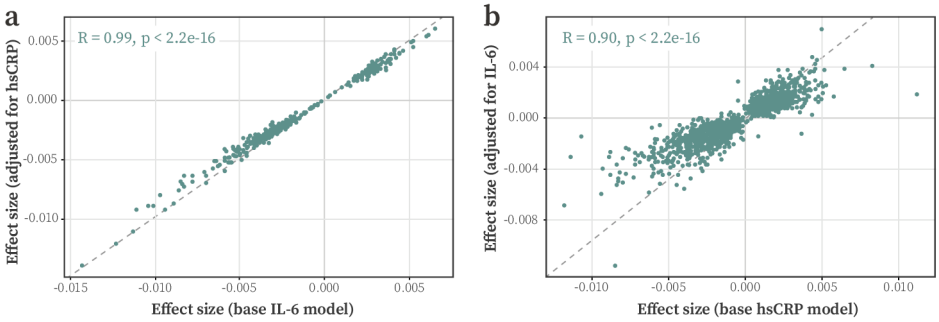


Figure 4 | Sensitivity analyses for the relationship between IL-6 and hsCRP DNAm signals. a) Scatter plot comparing IL-6 associated DNAm effect sizes before and after adjustment for hsCRP levels. **b)** Scatter plot comparing hsCRP-associated DNAm effect sizes before and after adjustment for IL-6 levels. The Pearson correlation coefficient and associated *p*-value are shown in blue for both plots.

Overall, our analyses identified 39 traits enriched among the IL-6 associated CpGs (**Fig. 5; Supplementary Table 2**). Some enrichments, such as those for education level and aggression, may have reflected interactions with environmental, social, and biological factors. For instance, education is frequently correlated with lifestyle factors, including

diet, which can influence systemic inflammation, while aggression may arise as a consequence of neuroinflammatory conditions[25,26]. Additionally, although models were adjusted for age and sex, we still observed enrichment for CpGs linked to these variables. This suggested that some CpGs may have represented signatures shared between age, sex, and IL-6.

Notably, IL-6 associated CpGs were also enriched for diseases with inflammatory components, including T2D (OR = 68.0, p_{FDR} = 7.4e-23) and post-traumatic stress disorder (OR = 89.7, p_{FDR} = 3.1e-56)[3,25], as well as key immunometabolic risk factors, such as BMI (OR = 148.6, p_{FDR} < 2.2e-308), lipid levels (triglyceride OR = 128.3, p_{FDR} = 8.2e-33 and HDL cholesterol OR = 109.1, p_{FDR} = 2.8e-72), systolic blood pressure (OR = 71.8, p_{FDR} = 3.1e-48), and fasting insulin (OR = 62.8, p_{FDR} = 3.3e-87). Collectively, these enrichments pointed to a shared DNAm signature underlying IL-6 and a spectrum of inflammatory comorbidities, supporting the utility of IL-6–associated CpGs as potential biomarkers of health[27].

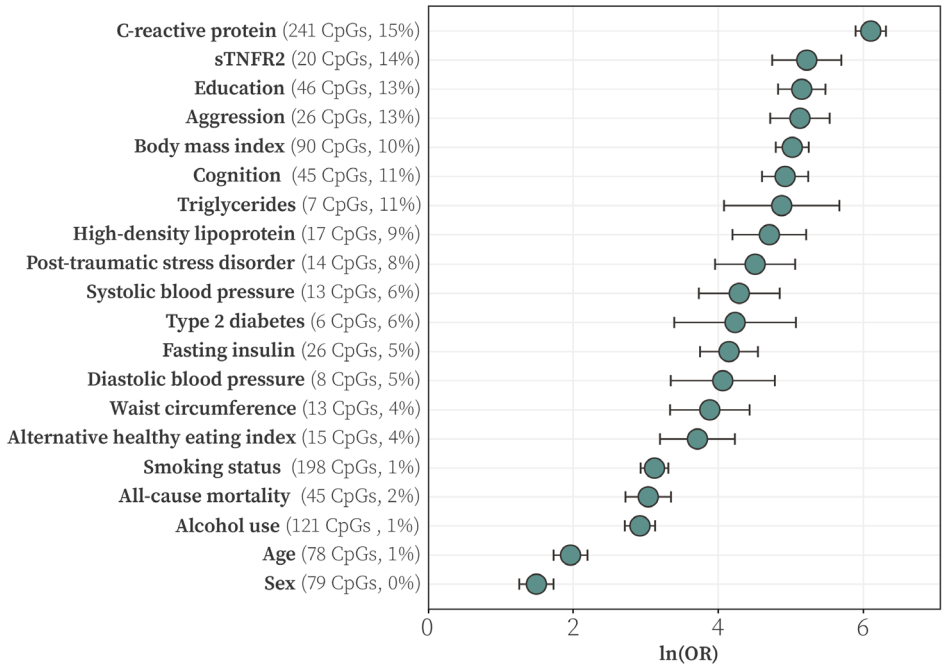


Figure 5 | Trait enrichment among IL-6 associated CpGs. Forest plot showing natural log-transformed odds ratios (ORs) with their corresponding 95% confidence intervals for the twenty most strongly enriched traits. For each trait, the count and percentage of overlapping CpGs are displayed alongside the plot.

Functional genomics highlights regulatory potential for identified CpGs

To assess their regulatory context, we mapped the 401 IL-6 associated CpGs to 15 chromatin states, defined by the Roadmap Epigenomics Consortium PBMC reference epigenome (E062)[28]. These encompassed eight active and seven repressive configurations, each characterised by distinct patterns of DNAm, chromatin accessibility, and regulator binding. Our analysis revealed that IL-6 associated CpGs were enriched for four active annotations (**Fig. 6a; Supplementary Table 3**), including regions flanking active transcription (OR = 6.5, p_{FDR} = 7.0e-6), weakly transcribed regions (OR = 1.6, p_{FDR} = 5.8e-3), genic enhancers (OR = 3.2, p_{FDR} = 3.2e-4), and enhancers (OR = 3.9, p_{FDR} = 1.1e-18). Conversely, three repressive chromatin states were depleted, including polycomb repressed regions (OR = 0.48, p_{FDR} = 5.8e-3) and bivalent transcription start sites (OR = 0.26, p_{FDR} = 3.5e-2). These patterns indicated that IL-6 associated CpGs were preferentially located in regulatory and transcribed chromatin, rather than in transcriptionally silenced regions.

Indeed, 12.7% (n = 51) of the IL-6 associated CpGs were annotated to enhancers in the PBMC reference, compared with 3.6% of all tested CpGs (n = 14,767). This indicated a high probability of colocalisation with markers of open chromatin, specifically H3K4me1[29]. To evaluate cell-type specificity, we extended this analysis using reference epigenomes for 22 other immune cell types (E029-48, E050-51). In all tested cell types, the genomic locations of the IL-6 CpGs were enriched for enhancers (p_{FDR} < 0.05) demonstrating robust regulatory potential independent of immune cell identity.

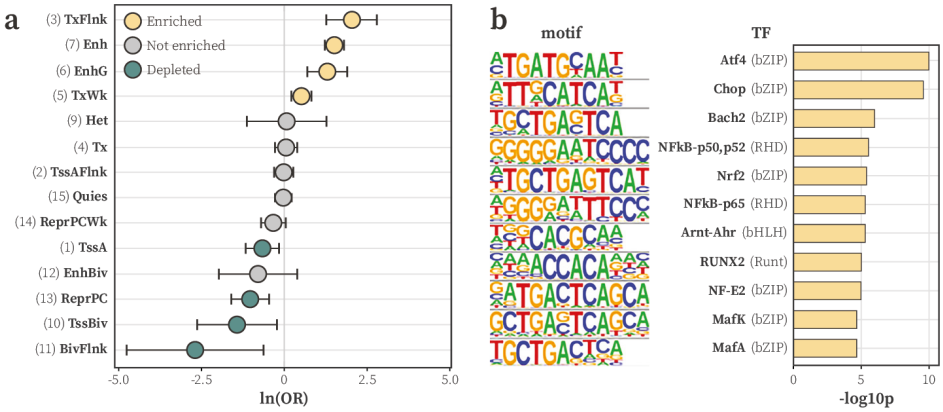


Figure 6 | Functional enrichment of IL-6 associated CpGs. a) Forest plot of ORs and 95% CIs with enrichment across 14 chromatin states, calculated using the Roadmap PBMC reference epigenome (E062) with all tested CpGs as the background set. States are sorted by OR, with yellow denoting enrichment, blue indicating depletion, and grey for non-significant results. One non-significant term (ZNF/Rpts) not shown due to an extremely wide 95% CI. **b)** Bar plot of -$\log_{10}$ p-values from TFBS enrichment analysis, performed using HOMER on sequences within 50bp of IL-6 associated CpGs, tested against a GC-matched random genomic background.

Since DNAm predominantly influences gene expression through TF binding modulation[30], we examined whether sequences within 50bp of IL-6 associated CpGs were enriched for TF binding sites (TFBS; **Fig. 6b, Supplementary Table 4**). Consistent with their regulatory potential, the tested regions were enriched for eleven TFBS. Notably, this included direct regulators of inflammatory gene expression and celluar stress responses, such as Atf4 (9 CpGs), Chop (8 CpGs), Nrf2 (3 CpGs), and subunits of the NF-κB protein complex (p50/p52: 5 CpGs; p65: 10 CpGs)[31–34]. Taken together, these analyses indicated that IL-6 associated CpGs were preferentially located in active regulatory regions across immune cell-types and supported a role for them in modulating inflammatory and cellular stress pathways.

Integrative analyses connect IL-6 associated CpGs to immunometabolic genes

To gain insight into the plausible functional consequences of this epigenetic regulation, we examined correlations between DNAm at IL-6 associated CpGs and expression of genes in *cis* (±100kb) using blood-based data from the BIOS consortium (n = 3,152; **Supplementary Table 5**). Among the 1,156 CpG-gene pairs evaluated, 320 (29.4%) were correlated after correction for multiple testing ($p_{FDR} < 0.05$), with most displaying an inverse relationship (n = 255, 75.0%). In total, 200 unique CpGs were linked to 295 distinct genes, with evidence that 29 of these genes were connected to multiple IL-6 associated CpGs. Notable examples included *SOCS3*, a negative regulator of IL-6 signalling, and core inflammasome components, *IFI16* and *AIM2*[35,36].

Additionally, we leveraged large-scale quantitative trait loci (QTL) datasets to perform a colocalisation analysis[37,38]. This framework evaluated whether IL-6 associated CpGs and nearby genes shared underlying genetic architecture. Unlike direct expression-methylation correlations, colocalisation is more robust to short-term fluctuations in expression and can reveal stable genetically driven relationships, even when transcriptional potential is not realised at time of sampling. Methylation QTL (mQTL) and expression QTL (eQTL) data were available for 914 CpG-gene pairs, of which 214 (23.4%) were flagged as colocalised, representing 130 unique CpGs and 197 unique genes.

In total, 421 genes were linked to IL-6 associated DNAm by one or both integrative analyses. Despite capturing distinct regulatory mechanisms, there was substantial overlap in CpG-gene pairs identified by both approaches (n = 76), indicating convergence on common biological patterns (χ^2 = 15.2, p = 9.9e-5). Notably, several non-nearest gene relationships were both biologically plausible and highly relevant to IL-6 biology, including the IL-6 receptor (*IL6R*) and polycomb group ring finger 6 (*PCGF6*), which transcriptionally represses *IL6* in quiescent dendritic cells (DCs)[39]. Combined with the established regulatory capacity of identified CpGs, these findings supported functional relevance for our results and provided a framework to confidently map IL-6 associated CpGs to plausible target genes in downstream analyses.

Cytokine-specific pathways are linked to IL-6 associated DNA methylation

To systematically characterise the biological roles of the 421 genes connected to IL-6 associated DNAm, we performed an over-representation analysis. Of the 16,037 gene sets tested, 265 were enriched in our results ($p_{FDR} < 0.05$; **Fig. 7a; Supplementary Table 6**) with 331 of the input genes mapping to at least one pathway (78.6%). The most strongly enriched gene set was Immune System (65 genes, $p_{FDR} = 1.0e\text{-}7$), which included several key mediators of IL-6 signalling, such as *IL6R* and *SOCS3*[35,40]. Notably, although the top five terms included many large and generic sets (Immune System, Disease, and Metabolism), 25 of the enriched pathways were cytokine-specific (9.4%), including TNFα signalling via NF-κB (14 genes, p_{FDR}=7.2e-4), IL-2/STAT5 signalling (16 genes, $p_{FDR} = 3.9e\text{-}5$), and interleukin-6 family signalling (4 genes, $p_{FDR} = 2.0e\text{-}2$).

We next performed TF enrichment analysis using the DoRothEA database to assess whether genes linked to IL-6 associated CpGs were over-represented among known TF regulons (**Fig. 7b; Supplementary Table 7**). One of the top ten enriched TFs was STAT3, the canonical downstream effector of IL-6 signalling, indicating that many genes connected to IL-6 associated DNAm were direct IL-6 targets mediated through STAT3 activation[21]. Other highly enriched TFs included ETS1 and ELF1, both key regulators of immune cell differentiation[41,42], and TCF3, a central factor in B-cell commitment and early T-cell development[43]. Collectively, these results suggested that IL-6 associated DNAm was concentrated within regulatory networks that govern leukocyte differentiation and cytokine signalling, whilst also highlighting a distinction between the TFs whose regulons are enriched in genes connected to IL-6 associated DNAm and those with binding sites located near the CpGs themselves.

To clarify how the enriched pathways converged on immune-cell and IL-6 biology, we assembled a mechanistic map tracing terms from T-cell receptor (TCR) and co-stimulatory inputs through to mTORC1-driven glycolysis and expansion of pro-inflammatory effectors. Within this axis, Th17 cells represented a conduit linking IL-6 signalling to the IL-17 dominated inflammation characteristic of many chronic inflammatory diseases[44]. We positioned eleven enriched terms along this aetiological trajectory and highlighted 17 genes with specific relevance for this process (**Fig. 7c**). These encompassed four co-ordinated cues critical for Th17 activation: TCR engagement (**Fig. 7ci**), co-stimulatory inputs (**Fig. 7cii**), and cytokine signals, including IL-1β (**Fig. 7ciii**) and IL-6 (**Fig. 7civ**)[45]. The latter of these triggers JAK/STAT3 activation, upregulating *RORC*, the lineage-defining TF of Th17 cells[46]. These inputs converge on mTORC1, promoting glycolysis as activated cells metabolically shift to rapid ATP generation (**Fig. 7cv**)[47]. Consistent with this model, our gene set included canonical glycolytic genes (*GLUT1, HK2, TPI1, GAPDH,* and the lactate exporter *SLC16A3*), which support Th17 persistence in inflamed, often hypoxic, tissues, and genes relevant for four arms of T-cell activation[48].

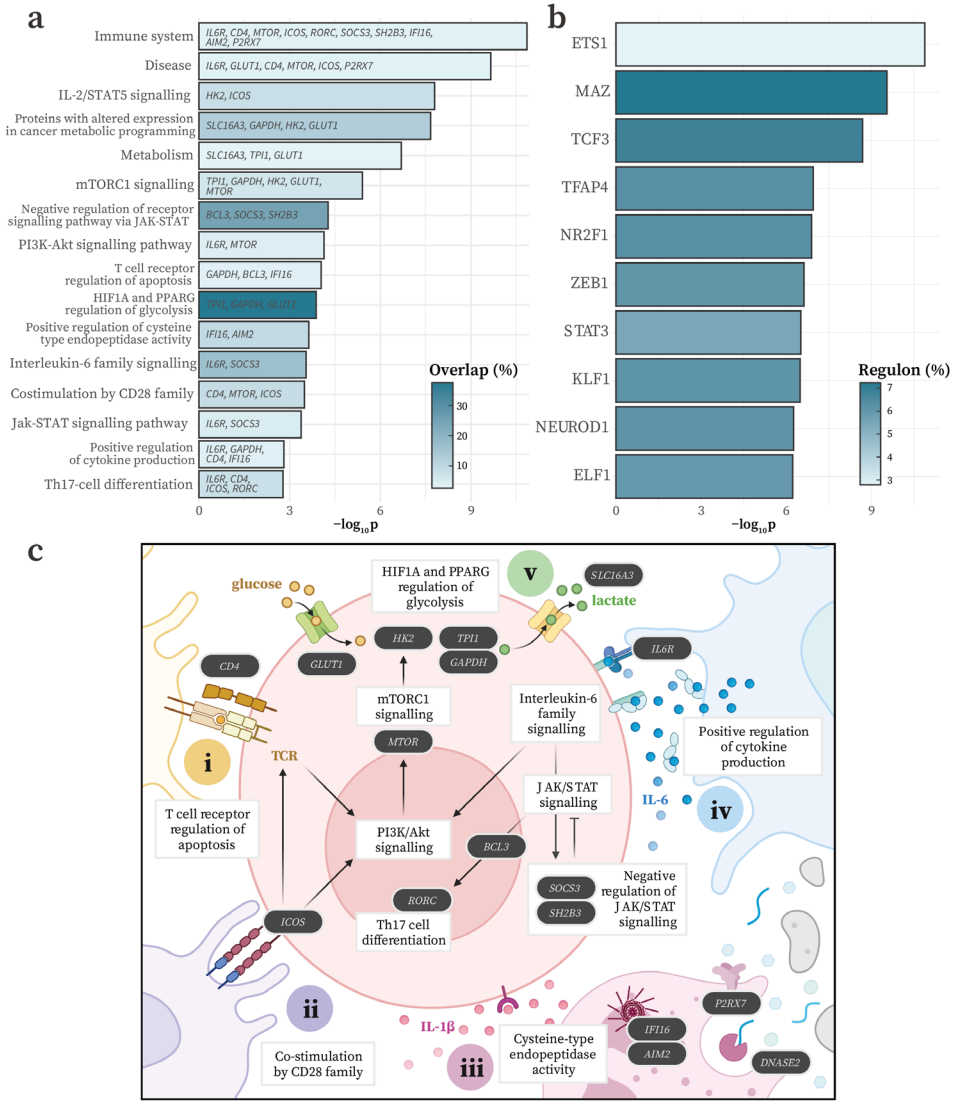


Figure 7 | Enrichment analyses of genes linked to IL-6–associated DNAm. a) Top five enriched gene sets, plus eleven additional terms with specific relevance to Th17 activation and IL-6 signalling. Bars are colored by the percentage of overlapping genes. Genes mapped onto the aetiological cascade from T-cell activation to glycolysis-fuelled Th17 proliferation are indicated. **b)** Top ten TFs whose regulons are enriched in the 421 genes linked to IL-6 associated DNAm based on DoRothEA TF target annotations. **c)** Selected enriched terms (white boxes) and mapped genes (black boxes) arranged along the pathway towards Th17 differentiation. Four activation signals: **(i)** TCR activation, **(ii)** co-stimulation, **(iii)** IL-1B stimulation, **(iv)** IL-6/JAK/STAT3 signalling, culminate in **(v)** mTORC1-driven glycolysis, which fuels rapid proliferation and effector functions. *Figure created in BioRender.*

Triangulation uncovers IL-6 as a driver of DNAm at the majority of CpGs

Overall, the genes and TFs linked to IL-6 associated methylation played a vital role in immune and inflammatory processes. However, ascertaining the directionality of these relationships was far from straightforward. Although we were able to map genes and pathways on an aetiological cascade predominantly downstream of IL-6, immune and inflammatory responses are inherently complex. Cytokines such as IL-6 can initiate and amplify the processes that activated them, creating positive feedback loops. As a result, many pathways such as inflammasome-mediated release of IL-1β and PI3K/Akt/mTOR signalling, can occur both upstream and downstream of IL-6 production[49–52].

To disentangle these interactions we employed triangulation analysis, using genetic variants as proxies for IL-6 levels and DNAm. This framework assumes that if genetically-determined outcome levels (*observed effects*) are predominantly driven by an exposure, then they should be predictable from a combination of genetically-determined exposures and exposure-outcome associations (*predicted effects*)[13,19,24]. The correlation between observed and predicted effects then quantifies the combined support for each direction across CpGs, even if there is insufficient power at the individual sites. By performing this analysis bidirectionally, we comparatively inferred whether the data more strongly supported IL-6 as a cause or consequence of DNAm[53].

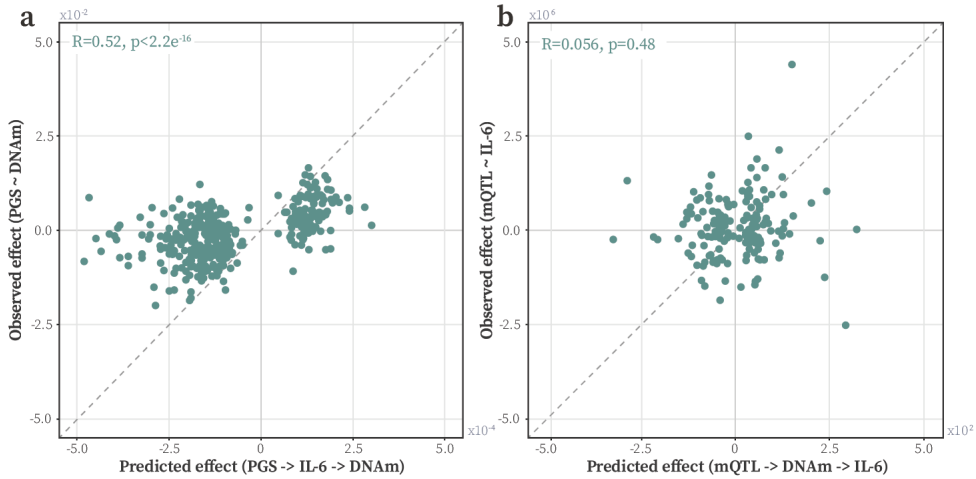


Figure 8 | Triangulation analysis of the directional relationship between IL-6 and DNAm. a) Scatter plot comparing the observed effect (PGS-DNAm associations) with the predicted effect (combining PGS effect on IL-6 and IL-6 CpG associations). Correlation coefficients and associated *p*-values are shown in blue. **b)** Scatter plot comparing the observed (mQTL IL-6 associations) and predicted effects (combined effect of the top mQTL SNP on DNAm, and DNAm association with IL-6). Each point represents one of the 401 IL-6 associated CpGs.

Building on conclusions from previous EWAS, where blood-based DNAm was frequently proposed as a downstream marker of complex traits, and informed by our gene set mapping, we first examined whether IL-6 might be driving DNAm[13,24,54]. Correlations between predicted (through IL-6) and observed effects supported this direction (R = 0.52, p < 2.2e-16, **Fig. 8a**). In contrast, testing the reverse pathway yielded only a weak and non-significant correlation (R = 0.06, p = 0.48, **Fig. 8b**), further indicating that IL-6 most plausibly drives DNA methylation at the majority of identified CpGs.

Causal inference supports DNAm at SOCS3 as mediating IL-6 effects on health

Considering the finding that the majority of IL-6 associated CpGs likely reflected a response to IL-6 levels, we next sought to identify candidate loci where this effect was most strongly supported. We leveraged publicly available transcriptomic data from isolated CD4+ T-cells stimulated with IL-6 *in vitro*, evaluating which of the 421 genes responded in an experimental setting[55].

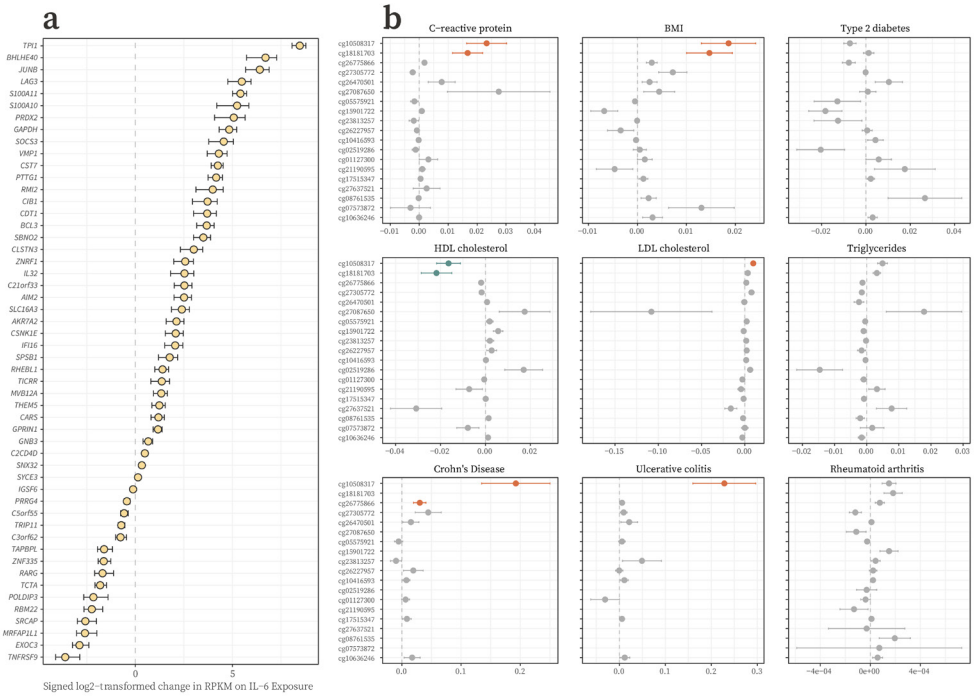


Figure 9: Downstream effects of IL-6. a) Forest plot showing the 52 genes linked to IL-6 associated DNAm that responded to IL-6 stimulation of isolated immune cells *in vitro*. **b)** Forest plots of CpGs ranked in the top ten for at least one tested trait in the mediation analysis. Bars are coloured by the direction and significance of the mediated effect: orange for significant positive mediation, blue for significant negative mediation, and grey for non-significant findings after FDR correction.

This analysis uncovered 52 loci whose responsiveness to IL-6 was identified in both population and experimental data, including core inflammasome components *AIM2* ($\Delta = 4.70$, $p_{FDR} = 0.017$) and *IFI16* ($\Delta = 3.15$, $p_{FDR} = 0.017$), as well as *SOCS3* ($\Delta = 22.44$, $p_{FDR} = 0.020$; **Fig. 9a; Supplementary Table 8**).

To investigate whether DNAm at these IL-6 responsive genes mediated IL-6 effects on nine traits and diseases, we performed a mediation analysis. The tested phenotypes covered major immunometabolic risk factors, such as CRP, BMI, and lipid levels (triglyceride, HDL and LDL cholesterol levels), as well as a spectrum of IL-6 associated diseases, including T2D, RA, and two forms of IBD (Crohn's disease and ulcerative colitis). Two CpGs connected to *SOCS3* expression plausibly propagated IL-6 increases in CRP and BMI, as well as decreases in HDL cholesterol. One of these sites also showed evidence for mediating IL-6 increases in LDL cholesterol and risk of both forms of IBD ($p_{FDR} < 0.05$). Collectively, this experimental and causal evidence not only reinforced conclusions from triangulation that IL-6 associated CpGs are responsive to IL-6, but also highlighted *SOCS3* as a critical regulatory locus that may mediate IL-6 effects on inflammatory health and disease (**Fig. 9b; Supplementary Table 9**).

Mendelian randomisation reveals a CpG plausibly influencing IL-6 production

Although triangulation supported DNAm as predominantly responding to IL-6, we also evaluated causal effects at the individual CpG level. To this end, we applied bidirectional 2SMR, an instrumental variable approach that uses genetic proxies to assess if DNAm at specific CpGs might be regulating or responding to IL-6. Such analysis are inherently vulnerable to low power, which can bias estimates towards the null. This was also true here, as most CpGs were instrumented by only a few independent mQTLs and IL-6 GWAS have only identified a small number of independent loci. Despite this limitation, 2SMR identified a CpG (cg26663590) where DNAm may causally influence IL-6 production ($\beta = 0.11$, $p_{FDR} = 3.6e\text{-}2$), and this represented a notable exception to the overall pattern observed through triangulation and mediation. This CpG was instrumented by two independent SNPs (rs7185941 and rs7186573), with the causal signal primarily driven by rs7186573 ($p_{rs7186573} = 7.5e\text{-}4$ vs. $p_{rs7185941} = 1.2e\text{-}1$).

DNAm at this CpG was correlated with Nuclear Factor of Activated T cells 2 Interacting Protein (*NFATC2IP*) expression in BIOS but shared underlying genetic architecture with *ATP2A1*. *NFATC2IP* is an established regulator of NFAT-driven transcription and has been linked to *IL6* expression specifically[56–58]. While *ATP2A1* is best characterised in muscle, its role in Ca^{2+} transport is relevant to NFAT TFs, which are critically sensitive to intracellular calcium dynamics[59,60]. Therefore, these findings suggested that this putatively causal CpG may influence IL-6 through modulation of NFAT TF activity[61].

To validate the connections between this CpG and expression, we examined publicly

available data containing matched methylation and expression from B-cells at various stages of development (GSM1104678; **Supplementary Fig. 3a,b**), and CD4+ T-cells (GSM4364456; **Supplementary Fig. 3c,d**)[62,63]. Across both datasets, this CpG was associated with *NFATC2IP* and *ATP2A1*. For *NFATC2IP*, the direction was consistent across BIOS and both immune cell types, with higher DNAm correlated with lower expression (BIOS: β = -0.005, p_{FDR} = 9.3e-4; CD4+ T-cells: R = -0.44, p = 0.085; B-cells: R = -0.64, p = 0.002). In contrast, the relationship between *ATP2A1* and DNAm differed between cell-types, showing a positive correlation in CD4+ T-cells (R = 0.65, p = 0.007) but a negative one in B-cells (R = -0.49, p = 0.023). Despite differences, results illustrated that this CpG plausibly regulated nearby transcription in immune cells and could therefore act upstream of IL-6 production.

cg26663590 may influence immunometabolic risk as well as IL-6 production

In addition to its potential role as a driver of IL-6 in this study, DNAm at cg26663590 has also been causally implicated in BMI, incident T2D, and Crohn's disease[13,64–66]. Coupled with biological and functional evidence linking cg26663590 to altered expression of NFAT TF regulators, this motivated investigation into the clinical significance of this CpG[67]. To this end, we evaluated directional relationships between cg26663590 and the nine traits assessed in our prior mediation analysis, this time using a bidirectional 2SMR framework. For all but one of the tested traits (88.9%), including T2D and both forms of IBD, evidence supported cg26663590 as a causal ($p_{FDR} < 0.05$; **Supplementary Table 10**) rather than a consequential factor ($p_{FDR} \geq 0.05$; **Supplementary Table 11**). The predicted effect was negative for HDL cholesterol (β = -0.027, p_{FDR}= 2.1e-3) but positive for CRP, BMI, and all tested diseases, in line with a protective effect from decreasing DNAm at this CpG. This conclusion was further supported by the positive association between cg26663590 and IL-6 (β = 0.004, p_{FDR} = 1.5e-2). In summary, this CpG represented a site worth investigating for plausibly driving immunometabolic disease risk through IL-6 dependent mechanisms.

Discussion

In this study, we performed an epigenome-wide analysis of the relationship between circulating IL-6 levels and DNAm in blood from 4,361 individuals across three cohorts. We identified 401 IL-6 associated CpGs, enriched for open chromatin across multiple immune cell lineages. These sites were located near TFBS relevant to IL-6 biology, including the primary transcriptional regulator of *IL6*, NF-κB[31], as well as mediators of unfolded protein (Atf4, CHOP) and oxidative stress responses (Nrf2)[32–34]. Integrative analyses linked these CpGs to inflammatory genes central to four interconnected arms of Th17-cell differentiation: TCR signalling (*CD4, LAG3, TAPBPL*), co-stimulation (*ICOS, TNFRSF8, TNFRSF9*), inflammasome activation (*P2RX7, AIM2, IFI16*), and the IL-6/JAK/STAT pathway (*IL6R, BCL3, SOCS3*)[45]. In addition, CpGs were connected to *RORC*, a gene encoding the Th17 lineage defining TF RORγt[46], as well as canonical genes for the mTORC1-driven glycolytic shift required for proliferative and pro-inflammatory effector functions (*HK2, GLUT1, MTOR*)[47,68,69]. Over-representation analysis supported these findings, identifying highly specific pathways including mTORC1 signalling and Th17 cell differentiation. Collectively, our findings highlight epigenetic mechanisms in the regulation of critical inflammatory and metabolic processes, which are either marked, stabilised, or controlled by DNAm[70].

To investigate whether DNAm was the cause or consequence of circulating IL-6, we applied three complementary causal inference approaches. Bidirectional triangulation, which takes a holistic view of causality and jointly evaluates evidence for both directions across CpGs, indicated that DNAm was predominantly responding to IL-6. This pattern mirrored findings from previous EWAS of BMI, Crohn's disease, and CRP[13,24,66], as well as prior 2SMR and experimental studies[6,54]. Consistent with this relationship, 52 genes linked to identified CpGs had previously responded to IL-6 stimulation of isolated immune cells *in vitro*[55]. To assess whether DNAm at these loci might propagate IL-6 effects on health, we performed a mediation analysis. This detected two CpGs, both annotated to *SOCS3*, that plausibly mediate IL-6 driven increases in BMI and CRP and decreases in HDL cholesterol. One of these CpGs additionally connected IL-6 to increases in LDL cholesterol and both forms of IBD (Crohn's disease and ulcerative colitis). Taken together, these analyses highlight epigenetic control of *SOCS3* expression as plausibly propagating the adverse health effects of elevated IL-6.

By contrast, bidirectional 2SMR revealed a putatively causal association in the reverse direction: DNAm at a CpG previously implicated in BMI and T2D, may also influence IL-6 production. In both large-scale blood-based cohorts and *in vitro* data from isolated B- and CD4+ T-cells, methylation at this site was inversely correlated with *NFATC2IP*, which regulates NFAT-driven cytokine transcription[56-58]. Therefore, we propose that this CpG affects IL-6 through modulation of NFAT-dependent signalling. Further analyses revealed that this site may have broader effects, potentially driving not only IL-6 but also a spectrum of immunometabolic risk factors and diseases including BMI, T2D, and

two forms of IBD. These findings support epigenetic regulation at this locus as playing a role in the multi-layered network that controls IL-6 production in immune cells.

This study comprehensively characterised IL-6-associated DNAm, exploring sensitivity to confounding from smoking and twelve immune cell subtypes. Furthermore, by performing EWAS for both IL-6 and CRP, we compared the stability of associations under mutual adjustment. This revealed that CRP-associated methylation was more sensitive to IL-6 than vice versa. This asymmetric attenuation pattern is consistent with a conditional dependence structure where IL-6 accounts for a portion of the DNAm signal shared with CRP. Consistent with this, mediation analysis positioned DNAm on the pathway from IL-6 to CRP, suggestive that epigenetics could contribute to IL-6 driven CRP production during the acute phase response[21].

Previous studies have also investigated links between DNAm and IL-6, providing important context for our findings. In a trial of breast cancer patients, chemotherapy triggered hypomethylation at eight CpGs in parallel with increased IL-6 levels, pointing to a role for DNAm in chemotherapy-driven inflammation[71]. We replicated six of these CpGs (75%), although one was excluded from our final set due to sensitivity to immune cell-type composition. While this manuscript was under review, a separate IL-6 EWAS also replicated four of these chemotherapy-associated CpGs, meaning that these sites that have now been independently observed in three studies[72]. This same EWAS reported 178 IL-6–associated CpGs, of which we replicated 66 (37.1%). However, 18 of these 66 sites (27.3%) were excluded in our final results due to confounding from extended cell counts, underscoring the critical importance of accurately modelling leukocyte heterogeneity in blood-based EWAS. In our study, IL-6 associated CpGs were further characterised through systematic follow-up analyses, which linked them to core inflammatory processes and assessed their directions and disease relevance. Moreover, a prior study developed an IL-6 predictor comprising 35 CpGs[27], of which two were identified here, while the remaining 33 have yet to be replicated in any independent study.

Notably, STAT3, which directly responds to IL-6 signalling, did not have binding sites enriched within 50bp of IL-6 associated CpGs. Nevertheless, genes linked to these CpGs were enriched in the STAT3 regulon. This could reflect IL-6 modulation of STAT3 targets through multiple mechanisms, both directly through STAT3 phosphorylation, but also indirectly through DNAm modulation of other TFs, such as NF-κB, which may cooperate with STAT3 to regulate target genes. This interpretation would align with the CpG-poor nature of STAT3 motifs[73]. Alternatively, IL-6 could influence DNAm via STAT3-independent TFs that share downstream targets with STAT3, or via modulation of methylation enzymes, a mechanism supported by prior experimental evidence[6].

There were several key limitations to this study. Firstly, the immune response is governed by a complex network of inflammatory proteins, interacting through positive and

negative feedback loops. Our analysis, however, was restricted to only two mediators: IL-6, the primary focus, and CRP, included as a sensitivity analysis. Secondly, although IL-6 is produced by and acts upon circulating cells, many of its effects occur in other tissues, such as the liver or pancreas. Because our study relied on blood-based DNAm, distinct directional effects could be observed in other IL-6 responsive tissues. Blood is also highly heterogenous. While we adjusted for twelve predicted immune cell-types, incorporated cell-type specific reference epigenomes, and validated findings using data from isolated immune cells, these approaches were insufficient to pinpoint the specific subtypes driving associations. Adjusting for extended cell-type proportions slightly reduced the number of IL-6 associated CpGs in this study, but strong concordance in effect estimates before and after this adjustment suggested that our findings were largely robust to this finer-grained modelling of immune heterogeneity. Nevertheless, some initial associations may have been masked by the lower resolution adjustment in base models and could be revealed by further initial deconvolution. Our extended characterisation of genes linked to IL-6 associated CpGs also identified several cell-type specific processes, including cytokine production by DCs, inflammasome activation in neutrophils, and metabolic reprogramming of T-cells, underscoring the critical importance of considering isolated cell-types for IL-6 biology.

Furthermore, our causal inference was constrained by weak instruments composed of only a small number of SNPs. As a result, rigorous testing of the relevance assumption was not feasible and we cannot rule out horizontal pleiotropy as an explanation for our directional findings[74]. The three cohorts in this study also differed in terms of population demographics, fasting status, and assay sensitivity. These factors likely contributed to higher effect size heterogeneity, although observed directions were largely consistent across cohorts, and fewer than 1% of IL-6 associated CpGs showed discordance (n = 4). Therefore, we adopted a higher heterogeneity threshold, retaining more CpGs while acknowledging that these may come at the expense of precision. As resulting CpGs were also thoroughly characterised using external datasets, multiple causal inference approaches, and experimental data, our conclusions were not solely dependent on discovery cohorts, somewhat mitigating this limitation. Lastly, this study provided only a cross-sectional snapshot of DNAm and IL-6, and therefore could not distinguish between acute and chronic inflammation. To confirm and extend on the links between DNAm and IL-6, future work could incorporate intervention studies, experimental approaches, or longitudinal designs.

In summary, our findings illustrate the value of integrative analytical frameworks and bidirectional causal inference for dissecting the regulatory interface between epigenetics and inflammation. IL-6 represents a critical circulating cytokine with broad implications for the pathogenesis and progression of disease, and here we have positioned DNAm as a potential contributor to both its regulation and downstream effects on human health.

Methods

Cohort descriptions

Leiden Longevity Study (LLS) The LLS is a cohort of long-lived, Dutch, Caucasian siblings (*n* = 944), who were recruited with their offspring (*n* = 1,671) and their offspring's partners (*n* = 744)[75]. Between 2002 and 2006, research nurses collected non-fasted blood samples from living subjects for isolation of DNA, RNA, serum, and plasma, and these were stored at -80˚C until analysis. From these samples, IL-6 levels were assessed in plasma using a PeliKine Compact Human IL-6 ELISA kit (Sanquin Reagents, Amsterdam, The Netherlands) and plasma hsCRP was assayed using an automated enzymatic colorimetric method with a Modular P analyzer (Roche, Almere, The Netherlands).

DNAm data of whole blood samples was generated for 821 unrelated participants from the offspring-partner generation of the LLS by the Human Genotyping facility (HuGe-F, Erasmus MC, Rotterdam, The Netherlands) within the Biobank-Based Integrative Omics Studies (BIOS) consortium, funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007). Genomic DNA (500ng) was bisulfite converted using the Zymo EZ-96 DNA methylation kit (Zymo Research Corp, Irvine, CA, USA). 4 µl was then hybridized on the Infinium HumanMethylation450 BeadChip array (Illumina Inc, San Diego, CA, USA) according to the manufacturer's protocol. Data preprocessing and quality control followed the *DNAmArray* workflow[76].

Kooperative Gesundheitsforschung in der Region Augsburg (KORA) The KORA F4 study (2006-2008) followed up individuals (*n* = 3,080) living in the region of Augsburg, Southern Germany, aged between 32 and 81 years old[77]. Blood samples were drawn after an 8 hour fast and frozen at -80˚C until analysis. Serum concentrations of IL-6 were measured with the Human IL-6 Quantikine HS ELISA kit (R&D Systems, Wiesbaden, Germany), and hsCRP was determined by a high-sensitivity latex-enhanced nephelometric assay on a BN II analyzer (Dade Behring)[78].

Following isolation according to standard procedures, genomic DNA (750 ng) from 1,707 whole blood samples was bisulfite converted with the Zymo EZ-96 DNA Methylation Kit (Zymo Research Corp, Irvine, CA, USA). 4 µl from each sample underwent amplification, enzymatic fragmentation, and application to the Infinium HumanMethylation450 BeadChip array (Illumina Inc, San Diego, CA, USA) according to the manufacturer's protocol. The methylation module of GenomeStudio was used to extract and process the raw image data, and initial quality assessment was conducted using the Control Dashboard. For data preprocessing, a common pipeline was used as previously described[79].

Netherlands Twin Register (NTR) The NTR, a national register established in 1987, recruited Dutch twins and their families, collecting fasted whole blood samples during a home visit[80,81]. EDTA plasma was harvested and aliquoted (0.5ml), snap-frozen in dry ice, and stored at -30˚C. IL-6 levels were measured with an UltraSensitive ELISA (R&D systems, Minneapolis, USA, Quantikine HS HSTA00C) and hsCRP with the Immulite 1000 CRP assay (Diagnostic Product Corporation, USA).

DNAm data of whole blood samples was generated by the Human Genotyping facility (HuGe-F, Erasmus MC, Rotterdam, The Netherlands) within the BIOS consortium[74]. Quality control and normalization have been described in detail previously[82]. The methylation data were normalized with functional normalization, expressed as β-values, and any data points further than 3 interquartile ranges (IQR) from the nearest quartile were excluded[83].

Cohort-based analyses

Base models Cohorts were contacted to participate in the study and those with required data types were recruited. All cohorts followed a common analysis plan and all samples analysed were taken from distinct individuals (i.e., there were no repeat measurements included in the analysis). Any IL-6 measurements below the limit of detection and outlying values for both DNAm and IL-6 (more than three IQR from the nearest quartile) were removed prior to analysis rather than imputed to avoid introducing bias or heterogeneity due to the differences in assay detection limits or types between cohorts.

For each of j CpGs measured in i individuals, a linear regression model (see *Equation 1* for general specification) was fitted of DNAm β-values on IL-6 values, measured in pg/mL and natural log-transformed to improve Normality. β-values were used to improve downstream interpretation and harmonization with previous analyses. All models were adjusted for age (in years), sex, cell-type proportions predicted from DNAm data using the IDOL algorithm (monocytes, CD8+ T, CD4+ T, natural killer, and B cells)[20], and technical covariates (left to the analyst's discretion).

$$DNAm_{\beta_{ij}} = \beta_0 + \beta_1 \cdot \ln(IL6)_i + \beta_2 \cdot age_i + \beta_3 \cdot sexFemale_i + \beta_4 \cdot CD8T_i + \beta_5 \cdot CD4T_i + \beta_6 \cdot NK_i + \beta_7 \cdot Mono_i + \beta_8 \cdot technical\_factors_i \quad [1]$$

Sensitivity analyses Each cohort ran additional analyses, adding trichotomous smoking to the cohort-specific base models as a categorical fixed effect. In some cases, this resulted in a reduction of the sample size as there were missing values in the smoking data.

To investigate the effect of adjusting for additional cell-types whose proportions could be estimated from DNAm data using the IDOL extended algorithm, basophils, memory B-cells, naïve B-cells, CD4+ memory T-cells, CD4+ naïve T-cells, CD8+ memory T-cells, CD8+ naïve T-cells, eosinophils, monocytes, natural killer cells, and regulatory T-cells (Treg) were added to the base model for all cohorts. Neutrophils were excluded to avoid collinearity (as proportions for all cells sum to 1, meaning a model without neutrophils already contains information about neutrophil proportions and their addition would introduce multicollinearity)[84]. Plots comparing effect sizes from the base model with each of the sensitivity analyses were inspected and tests of correlation were performed. IL-6 associated CpGs no longer associated after adjustment were removed prior to downstream analysis ($p_{\text{FDR}} \geq 0.05$).

Additional phenotypes To distinguish inflammatory signals, cohorts additionally adjusted the base IL-6 model for hsCRP (measured in mg/L). All cohorts additionally ran an hsCRP EWAS with the same specification as the IL-6 base model (see *Equation 2* for specification in LLS) both with and without adjustment for IL-6 levels.

$$DNAm_{\beta_j} = \beta_0 + \beta_1 \cdot \ln(CRP)_i + \beta_2 \cdot age_i + \beta_3 \cdot sexFemale_i + \beta_4 \cdot CD8T_i + \beta_5 \cdot CD4T_i + \beta_6 \cdot NK_i + \beta_7 \cdot Mono_i + \beta_8 \cdot plate1_i + \cdots + \beta_{17} \cdot plate10_i + \beta_{18} \cdot arrayRow_i \quad [2]$$

Meta-analyses

Differentially methylated probes Cohorts provided summary statistics for each CpG, including the mean and SE of DNAm levels, size of association with IL-6 along with the SE, nominal *p*-value, and number of individuals used to derive the IL-6 association. Results from each cohort were inspected and CpGs were removed if they were estimated from fewer than 50 participants. Probes located on sex chromosomes, in ENCODE Blacklist regions, that contained common genetic variants, or which were ambiguously mapped were also removed[85,86]. To ensure good quality data, we inspected QQ, volcano, and Manhattan plots, alongside boxplots of the effect sizes and SE distributions across cohorts. Following these steps, data was available on 412,226 CpGs for the base IL-6 model.

The Bioconductor package *bacon* was used to inspect and adjust for bias and inflation of the test statistics, using default priors ($\alpha = 1.28$, $\beta = 0.36$)[87]. After running bacon, inflation and bias were re-estimated for all models to ensure that these had been adequately corrected for. Bacon-adjusted effect sizes and SEs were used as input in a fixed-effects meta-analysis in *METAL*[88]. Separate analyses were performed for the base model, each extended model (adjusted for smoking, extended cell counts, and hsCRP), and hsCRP models (base and IL-6 adjusted). Any CpGs where there was evidence of high heterogeneity of effect sizes between cohorts ($I^2 > 90\%$) were removed.

Differentially methylated regions To assess the number of distinct genomic loci associated with circulating IL-6 levels, differentially methylated regions (DMRs) were identified using the *DMRfinder* algorithm as implemented in the *DNAmArray* workflow[76,89]. DMRs were defined as regions with at least three differentially methylated positions (DMPs) with an inter-CpG distance of less than 1kb, allowing a maximum of three non-DMPs across a DMR. Then, the number of distinct loci were calculated as the total number of DMPs minus the number of DMPs in DMRs plus the number of DMRs called by *DMRfinder*.

Functional annotation

Phenotype enrichment Using summary data from the EWAS catalog and EWAS atlas, we investigated whether IL-6 associated CpGs had been previously associated with other phenotypes[90,91]. Following loading of data into R, we filtered results to include only large-scale, peer-reviewed, and relevant findings (with a PubMed ID, sample size above 500, finding over 100 CpGs at the threshold of inclusion into the respective database, performed in adults, reporting nominal *p*-values, and using whole blood or leukocyte samples). We also recoded traits to ensure consistency between names, for example by combining EWAS of "BMI" and "body mass index". Finally, we combined data with summary statistics from a recent hsCRP EWAS, which was not included in either catalogue[22]. This resulted in a list of 57 traits, for which we tested enrichment of associations with our CpGs using logistic regression.

Chromatin state enrichment IL-6 associated CpGs were annotated to chromatin state using the Roadmap reference epigenome for peripheral blood mononuclear cells (PBMCs; E062) which includes histone modification ChIP peaks (H3K4me1, H3K4me3, and H3K27ac), marks of open chromatin, and DNase I hypersensitivity sites[28]. We mapped each probe tested on the Illumina 450k array background to eight active and seven repressive chromatin states, recording overlap at each probe as a binary variable. To determine whether enrichment occurred more often than expected by chance, we performed logistic regression using the *glm* function in R to calculate and test odds ratios (ORs) of significance for each state. Nominal *p*-values were adjusted for multiple testing using the false discovery rate (FDR) and enrichments or depletions were assessed using a 5% significance threshold.

Transcription factor binding site (TFBS) enrichment A 50 bp window around IL-6 associated CpGs was scanned using *findMotifsGenome.pl* from HOMER to evaluate enrichment of known motifs compared to a random background matched for GC content[92]. ENCODE TFBS annotation for TFs and CpGs on the 450k array was used to further investigate the size of binding sites and distance from CpG to summit[86]. TFs associated with enriched TFBS were examined for links with inflammation, immune cell functioning, and IL-6 related pathways and interactions.

eQTM analyses Blood-based RNA-sequencing and DNAm data from matched samples were available from the BIOS consortium (n = 3,152), comprising six Dutch biobanks: namely the Cohort on Diabetes and Atherosclerosis Maastricht[93], LifeLines[94], LLS[75], NTR[80,81], Rotterdam Study[95], and Prospective ALS Study Netherlands[96]. After filtering out non-autosomal and lowly expressed genes, count data was transformed into $\log_2$ counts per million ($\log_2$CPM), and values for each gene were rank-inverse normal (RIN) transformed prior to analysis. Genomic locations of human transcripts, exons, coding sequences, and genes were imported from the Ensembl database using *makeTxDbFromEnsembl* from the GenomicFeatures Bioconductor package[97]. These were used to identify all genes within 100kb of each IL-6 associated CpG. To examine correlations between DNAm and gene expression we carried out linear regression with RIN-transformed gene expression values as the response variable and methylation β-values as the independent variable, adjusting for the effects of age, sex, technical covariates (row, plate, and flowcell), and twelve predicted blood cell-types predicted from DNAm data using IDOLext[20].

Colocalisation analysis To investigate if DNAm and local gene expression shared underlying causal variances we performed colocalisation analysis[98]. Since the assumption of single causal variants in traditional colocalisation is a spurious one, we used the Sum of Single Effects (SuSiE) regression framework in combination with *coloc*[99]. SuSiE improves fine-mapping and colocalisation by modelling multiple causal variants within a region simultaneously, rather than assuming a single causal signal. It separates the statistical evidence supporting each variant by conditioning on other signals, thereby reducing confounding between correlated variants and yielding more reliable colocalisation inferences[100].

To perform this analysis, we downloaded mQTL data from the Genetics of DNA Methylation Consortium (GoDMC), saving associations between DNAm at the IL-6 associated CpGs and all single nucleotide polymorphisms (SNPs) within ±1000kb regardless of significance[38]. For each CpG-gene pair, we used the full eQTL summary statistics from eQTLGen and stored associations between the identified SNPs and expression of nearby genes alongside minor allele frequencies (MAF)[37]. Linkage disequilibrium (LD) matrices based on reference haplotype data from the publicly released Phase 3 (Version 5) 1000 Genomes Project population were provided by the *LDlinkR* package, which allows access to the LDlink API from the R console[101]. Variances in methylation and expression at each IL-6 associated CpG and linked gene were calculated using available DNAm and RNA-seq data from the BIOS consortium. Finally, the *coloc.susie* function from the *coloc* R package was used to run the analysis with default priors for the probability a SNP was associated with DNAm (p_1 = 1e-4), the probability a SNP was associated with expression (p_2 = 1e-4), and the probability a SNP was associated with both traits (p_{12} = 5e-6). Colocalisation of DNAm and expression signals was considered confirmed if the posterior probability of H4, the hypothesis where both gene expression and DNAm share causal variants, was above 0.9.

Over-representation analysis Following eQTM and colocalisation analyses, we generated a list of plausible target genes for each IL-6 associated CpG. The associated gene names were then used as input for over-representation analysis, using eleven recent (updated in the last 6 years) databases relating to human health and disease downloaded from *Enrichr* (BioPlanet 2019, Elsevier Pathway Collection, GeDiPNet 2023, GO Biological Process 2023, KEGG Human 2021, MsigDB Hallmark 2020, OMIM, PhenGenI Association 2021, PheWeb 2019, Reactome 2022, and WikiPathway Human 2021). These were imported into R and analyses were performed using the *enrichr* function from *clusterProfiler*[102]. *p*-values were adjusted for multiple testing using FDR and significance was assessed at the 5% level.

TF regulon enrichment To identify TFs potentially regulating the set of genes connected to IL-6 associated CpGs, we performed TF enrichment analysis using the *DoRothEA* human regulon database. *DoRothEA* provides curated TF-target interactions with confidence levels ranging from A (highest) to E (lowest). We restricted this analysis to high or medium confidence interactions (levels A to D) and considered three nested subsets: A_B, A_B_C, and A_B_C_D. The input gene list consisted of unique HGNC gene symbols from our dataset, intersected with the *DoRothEA* target universe to define the background. For each TF, we calculated the statistical significance of the overlap between its regulon and the input gene set using a one-sided hypergeometric test, with the universe defined as all genes present in *DoRothEA*. *p*-values were adjusted for multiple testing using FDR and significance was assessed at the 5% level.

Causal inference

Triangulation analyses To perform triangulation analyses, we interrogated the correlation between the observed effect of an instrumental variable on an outcome (i.e., methylation quantitative trait locus (mQTL)-IL-6 or polygenic score (PGS)-DNAm associations) and the predicted effect via the exposure. This analysis assumed that if the effect of an exposure on an outcome is causal, then it would be possible to predict the effect of a genetic instrument on that outcome via the combination of the instrument's effect on the exposure and the exposure's association with the outcome.

In detail, when looking at the effect of DNAm on IL-6 levels (consequential analysis), the *observed effect* is the association between the top mQTL and ln(IL-6), extracted from full GWAS summary data[103]. The *predicted effect* can then be calculated by combining the mQTL effect extracted from GoDMC data[36] (the association effect of +1 effect allele with DNAm, β_{mQTL}) and the EWAS effect (the association effect of +1 ln(IL-6) with DNAm, β_{EWAS}). Since we wanted to predict the effect of +1 effect allele on IL-6, the calculation for the predicted effect was then $\beta_{mQTL}/\beta_{EWAS}$ in this instance. When looking in the reverse direction (i.e., IL-6 as a cause of DNAm), the *observed effect* is a PGS, where the effect of two IL-6 associated SNPs on DNAm were weighted by their effect allele frequencies (EAF). The predicted effect used equivalent EAF weighting and was calculated by multiplying the effect of the PGS on IL-6 (β_{PGS}) with the EWAS effect (β_{EWAS}). Correlation between the *observed* and *predicted effects* in both directions were assessed using Pearson correlation coefficients.

Transcriptional responses to IL-6 To identify CpGs that were plausibly regulating transcriptional responses to IL-6, we used publicly available gene expression data from an experimental study where isolated CD4+ T-cells were stimulated with IL-6 *in vitro* (GSE65621)[55]. Data was provided as reads per kilobase of transcript, per million reads mapped (RKPM) for IL-6 exposed ($n = 9$) and control ($n = 9$) samples, following 3 days of culturing. Differential expression of genes annotated to IL-6 associated CpGs was calculated using linear models fitted in *limma*, adjusting for patient status and using FDR to correct for multiple testing.

Mediation analysis For identified CpGs connected to IL-6 responsive genes we performed a mediation analysis. This tested if DNAm at these sites mediated the causal effect of IL-6 on downstream traits. The tested phenotypes covered a number of key risk factors for diseases, such as high density lipoprotein (HDL) cholesterol, body mass index (BMI), and CRP levels, as well as a diverse range of IL-6 associated diseases, including T2D, RA, and two types of inflammatory bowel disease (IBD; Crohn's disease and ulcerative colitis).

For each CpG and trait, we estimated (a) the effect of IL-6 on DNAm at that CpG, (b) the effect of DNAm at that CpG on the trait, and (c) the total effect of IL-6 on the trait. Each of these effects were estimated using genetic proxies for the exposure in an instrumental variable framework, using either the Wald ratio (for single SNP instruments) or IVW methods (for multiple, independent SNPs). Instrumental SNPs were obtained from recent large-scale GWAS, including for IL-6 (ebi-a-GCST90012005), CRP (ebi-a-GCST90029070), BMI (ieu-b-40), T2D (ebi-a-GCST006867), HDL cholesterol (ieu-b-109), LDL cholesterol (ieu-b-110), triglycerides (ieu-b-111), RA (ukb-d-M13_RHEUMA), Crohn's disease (ieu-a-12), and ulcerative colitis (ieu-a-970).

The total effect (c) was decomposed into an indirect (mediated) effect (ab) with variance estimated using the delta method, and a direct (non-mediated) effect (c – ab). For each trait, CpG-specific mediated *p*-values were corrected for multiple testing using FDR, and significance was assessed at the 5% level.

Two-sample Mendelian Randomization To assess site-specific directional effects between IL-6 and DNAm at IL-6 associated CpGs, we used the *TwoSampleMR* package to perform bidirectional 2SMR. This instrumental variable based method uses GWAS summary statistics to infer that a risk factor causally influences an outcome. 2SMR relies on several key assumptions, namely that instruments are relevant, independent, and that there is no horizontal pleiotropy. To interrogate the effects of DNAm at our CpGs on IL-6, we extracted SNP-based *cis*-mQTL data from GoDMC and combined these with summary statistics from a large-scale GWAS of IL-6[38,103]. For some CpGs (n = 169, 42.14%), there was insufficient data available to interrogate the effects of DNAm at that CpG. For the remaining CpGs, between one and five independent SNPs had data on both their association with DNAm and with IL-6. Following harmonisation of the data and clumping of these instrumental SNPs, they were combined using the Wald ratio or IVW methods.

To interrogate the influence of IL-6 on DNAm at identified CpGs, independent GWAS variants from a large-scale analysis was used. Of the two variants that could instrument IL-6, there was *trans*-mQTL data in *GoDMC* available for both of them (rs4537545, rs6734238). The GWAS summary statistics and mQTL effects were then combined using the IVW method and the *TwoSampleMR* package in R. To interrogate directional effects between DNAm and inflammatory traits and ensure harmonisation across this study, the GWAS from the mediation analysis were also used for 2SMR. For all analyses, *p*-values were adjusted for multiple testing using the FDR method and potential causal effects were assessed at the 5% significance threshold.

The following cohorts were used to derive both mQTL effects in GoDMC and IL-6 GWAS effects and therefore had overlapping individuals in both the exposure and outcome datasets for the 2SMR analysis: Lothian Birth Cohorts 1921 (GoDMC: 435 samples; IL-6 GWAS: 166 samples), Lothian Birth Cohorts 1936 (GoDMC: 905 samples; IL-6 GWAS: 759 samples), LLS (GoDMC: 718 samples; IL-6 GWAS: 1798 samples), NTR (GoDMC: 2757 samples; IL-6 GWAS: 3668 samples), Rotterdam Study (GoDMC: 1472 samples; IL-6 GWAS: 599 samples), and TwinsUK (GoDMC: 843 samples; IL-6 GWAS: 1103 samples). Considering that both meta-analyses incorporated data from over 26 cohorts, this overlap was acceptably low.

eQTM validation In order to validate eQTM effects at a CpG identified by 2SMR as causally influencing IL-6, we used publicly available DNAm and expression data from isolated CD4+ T- and B-cells[62,63]. Counts were normalised to $\log_2$CPM and DNAm was represented by β-values. Correlations between genes identified by large-scale association and colocalisation analyses and the CpG were calculated using the Pearson correlation coefficient and significance was assessed at the 5% level.

Ethics statement

Leiden Longevity Study (LLS) Written informed consent for DNA collection and its use for genetic analyses was obtained from all participants prior to their enrolment into the study. Good clinical practice guidelines were maintained, and the study protocol was approved by the local Medical Ethical Committee of the Leiden University Medical Center. All ethical regulations relevant to human research participants were followed.

Kooperative Gesundheitsforschung in der Region Augsburg (KORA) In accordance with the Declaration of Helsinki, written informed consent was obtained from all participants prior to their enrolment into the study. Good clinical practice guidelines were maintained, and the study protocol was approved by the Ethics Committee of the Bavarian Medical Association. All ethical regulations relevant to human research participants were followed.

Netherlands Twin Register (NTR) Informed consent was obtained from all participants. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance FWA00017598; IRB/institute codes, NTR 03-180). All ethical regulations relevant to human research participants were followed.

Author contributions

L.S. conceived and designed the study, performed data preprocessing and quality control, undertook cohort-specific and main analyses, interpreted results, and drafted the manuscript; B.T.H. conceived and designed the study, supervised conduct of the study, iteratively revised the manuscript, and provided critical intellectual contributions and interpretations of results; J.v.D, T.D., and R.W. performed cohort-specific analyses, provided summary statistics for the meta-analyses, provided critical intellectual contributions and interpretations of results; Y.X., J.O., and J.T.B. provided critical intellectual contributions and interpretations of results; M.B., G.W., C.G., C.H., W.K., A.P., E.C.D.G. M.W., D.I.B., and P.E.S. provided and verified data, phenotype acquisition, and harmonization for these analyses, provided critical intellectual contributions and interpretation of results. All authors read and approved the manuscript prior to submission.

Funding

The work of L.S. was supported by the Joint Programming Initiative 'a Healthy Diet for a Healthy Life' (JPI-HDHL) DIMENSION project [ZonMW project number: 529051021]. This paper makes use of data generated by the BIOS consortium. Funding for the BIOS consortium was provided by the Netherlands Organization for Scientific Research (NWO 184.021.007), made available as a Rainbow Project of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). This paper makes use of data generated by the CHARGE Inflammation Working Group.

NTR This research was funded by Genotype/phenotype database for behaviour genetic and genetic epidemiological studies (ZonMW Middelgroot 911-09-032); Netherlands Twin Registry Repository: researching the interplay between genome and environment (NWO-Groot 480-15-001/674); the Avera Institute, Sioux Falls (USA) and the National Institutes of Health (NIH R01 HD042157-01A1, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995); NWO, project 184.034.019, and by the BBRMI-NL-financed BIOS Consortium (NWO 184.021.007); epigenetic data were generated at the Human Genomics Facility (HUGE-F) at EUR. J.v.D is supported by NWO Large Scale infrastructures, X-Omics (184.034.019). D.I.B. acknowledges the Royal Netherlands Academy of Science Professor Award (PAH/6635).

KORA F4 The KORA research platform (KORA, Cooperative Health Research in the Region of Augsburg) was initiated and financed by the Helmholtz Munich German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research 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. The work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the EU Joint Programming Initiative 'A Healthy Diet for a Healthy Life' (DIMENSION; grant number 01EA1902A). The German Diabetes Center is supported by the Ministry of Culture and Science of the state of North Rhine-Westphalia (Düsseldorf, Germany) and the German Federal Ministry of Health (Berlin, Germany). This study was supported in part by a grant from the German Federal Ministry of Education and Research to the German Center for Diabetes Research (DZD).

LLS The LLS was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE01014 and IGE05007), the Centre for Medical Systems Biology, and the National Institute for Healthy Ageing (Grant 05040202 and 05060810), in the framework of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research.

Declarations of interest

The authors declare no competing interests.

Data & code availability

Individual-level data The informed consents given by KORA study participants does not cover data posting in public databases. However, data are available upon request from KORA Project Application Self-Service Tool. Requests can be submitted online and are subject to approval by the KORA Board. The HumanMethylation450 BeadChip data from the NTR and LLS are available as part of the BIOS Consortium in the European Genome-phenome Archive (EGA), under the accession code EGAD00010000887. Additional -omic and phenotype data are available upon request via the BBMRI-NL BIOS consortium. All data can be requested by researchers from the respective cohorts.

All other data used in this study is publicly available EWAS summary statistics can be downloaded from the EWAS Catalog[90] and EWAS atlas[91], reference epigenome data is available from ROADMAP[28], TFBS data is available within the HOMER software[92], full mQTL summary statistics can be requested from GoDMC[38], eQTL data can be downloaded from eQTLgen[37], IL-6 GWAS summary statistics are available upon request from the CHARGE Inflammation Working Group[103], LD matrices can be accessed using LDlink[101], variances in methylation and expression were calculated from data generated by the BIOS consortium, and libraries for GSEA were downloaded directly from Enrichr (https://maayanlab.cloud/Enrichr/).

Instrumental SNPs were obtained from recent large-scale GWAS, including for IL-6 (ebi-a-GCST90012005), CRP (ebi-a-GCST90029070), BMI (ieu-b-40), T2D (ebi-a-GCST006867), HDL cholesterol (ieu-b-109), LDL cholesterol (ieu-b-110), triglycerides (ieu-b-111), RA (ukb-d-M13_RHEUMA), Crohn's disease (ieu-a-12), and ulcerative colitis (ieu-a-970). Data from isolated CD4+ T-cells following IL-6 exposure can be accessed via the Gene Expression Omnibus (GEO; GSE65621). Data containing matched methylation and expression from B-cells at various stages of development (GSM1104678), and CD4+ T-cells (GSM4364456) is also accessible freely from GEO.

Custom code availability R scripts are deposited on GitHub at `nebulyra/il6_ewas` and archived on Zenodo[104].

All software used is open source and freely available Unless stated otherwise, all calculations were performed using R version 4.2.2. For all meta-analyses, METAL, version 2011-03-25 was used. TFBS enrichment analyses were performed using HOMER v3.1.

Acknowledgements

The authors thank the staff and participants of all cohorts involved in this study.

References

1. Tanaka, T., Narazaki, M. *and* Kishimoto, T. IL-6 in inflammation, Immunity, And disease. *Cold Spring Harb Perspect Biol* **6** (10): a016295 (2014).
2. Aliyu, M. *et al.* Interleukin-6 cytokine: An overview of the immune regulation, immune dysregulation, and therapeutic approach. *Int Immunopharmacol* **111**: 109130 (2022).
3. Akbari, M. *and* Hassan-Zadeh, V. IL-6 signalling pathways and the development of type 2 diabetes. *Inflammopharmacology* **26** (3): 685-698 (2018).
4. Hirata, T. *et al.* Associations of cardiovascular biomarkers and plasma albumin with exceptional survival to the highest ages. *Nat Commun* **11** (1): 3820 (2020).
5. Mossmann, M. *et al.* Increased serum IL-6 is predictive of long-term cardiovascular events in high-risk patients submitted to coronary angiography: an observational study. *Diabetol Metab Syndr* **14** (1): 125 (2022).
6. Balakrishnan, A. *et al.* Interleukin-6 determines protein stabilization of DNA methyltransferases and alters DNA promoter methylation of genes associated with insulin signaling and angiogenesis. *Lab Invest* **98** (9): 1143-1158 (2018).
7. Ridker, P. M. *and* Rane, M. Interleukin-6 Signaling and Anti-Interleukin-6 Therapeutics in Cardiovascular Disease. *Circ Res* **128** (11): 1728-1746 (2021).
8. Osman, E. E. A. *and* Neamati, N. Ironing Out the Mechanism of gp130 Signaling. *Pharmacol Rev* **76** (6): 1399-1443 (2024).
9. Feng, Y. *et al.* The Role of Interleukin-6 Family Members in Cardiovascular Diseases. *Front Cardiovasc Med* **9**: 818890 (2022).
10. Kishimoto, T. *and* Kang, S. IL-6 Revisited: From Rheumatoid Arthritis to CAR T Cell Therapy and COVID-19. *Annu Rev Immunol* **40**: 323-348 (2022).
11. Taniguchi, K. *and* Karin, M. IL-6 and related cytokines as the critical lynchpins between inflammation and cancer. *Semin Immunol* **26** (1): 54-74 (2014).
12. Yin, Y. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356** (6337): eaaj2239 (2017).
13. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541** (7635): 81-86 (2017).
14. Demerath, E. W. *et al.* Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. *Hum Mol Genet* **24** (15): 4464-4479 (2015).
15. Mendelson, M. M. *et al.* Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. *PLoS Med* **14** (1): e1002215 (2017).
16. Korn, T. *and* Hiltensperger, M. Role of IL-6 in the commitment of T cell subsets. *Cytokine* **146**: 155654 (2021).
17. Xu, Y. D. *et al.* Role of IL-6 in dendritic cell functions. *J Leukoc Biol* **111** (3): 695-709 (2022).
18. Van Rooij, J. *et al.* Evaluation of commonly used analysis strategies for epigenome- and transcriptome-wide association studies through replication of large-scale population studies. *Genome Biol* **20** (1): 235 (2019).
19. Sinke, L. *et al.* DNA methylation of genes involved in lipid metabolism drives adiponectin levels and metabolic disease. *Diabetologia* **69** (1): 127-145 (2026).
20. Salas, L. A. *et al.* Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat Commun* **13** (1): 761 (2022).
21. Ngwa, D. N., Pathak, A. *and* Agrawal, A. IL-6 regulates induction of C-reactive protein gene expression by activating STAT3 isoforms. *Mol Immunol* **146**: 50-56 (2022).
22. Wielscher, M. *et al.* DNA methylation signature of chronic low-grade inflammation and its role in cardio-respiratory diseases. *Nat Commun* **13** (1): 2408 (2022).
23. Marzi, C. *et al.* Epigenetic signatures at AQP3 and SOCS3 engage in low-grade inflammation across different tissues. *PLoS One* **11** (11): e0166015 (2016).
24. Ligthart, S. *et al.* DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol* **17** (1): 255 (2016).
25. Coccaro, E. F. *et al.* Plasma and cerebrospinal fluid inflammatory markers and human aggression. *Neuropsychopharmacology* **48** (7): 1060-1066 (2023).
26. Fard, N. A. *et al.* On the interplay between educational attainment and nutrition: a spatially-aware perspective. *EPJ Data Sci* **10**: 18 (2021).
27. Stevenson, A. J. *et al.* Creating and Validating DNA Methylation-Based Proxy for Interleukin-6. *J Gerontol A Biol Sci Med Sci* **76** (12): 2284-2292 (2021).
28. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518** (7539): 317-330 (2015).
29. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39** (3): 311-318 (2007).

30. Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49** (1): 131-138 (2017).

31. Liu, T. *et al.*. NF-κB signaling in inflammation. *Signal Transduct Target Ther* **2**: 17023 (2017).

32. Iwasaki, Y. *et al.* Activating transcription factor 4 links metabolic stress to interleukin-6 expression in macrophages. *Diabetes* **63** (1): 152-161 (2014).

33. Hattori, T. *et al.* C/EBP homologous protein (CHOP) up-regulates IL-6 transcription by trapping negative regulating NF-IL6 isoform. *FEBS Lett* **541** (1-3): 33-39 (2003).

34. Wruck, C. J. *et al.* Nrf2 induces interleukin-6 (IL-6) expression via an antioxidant response element within the IL-6 promoter. *J Biol Chem* **286** (6): 4493-4499 (2011).

35. Croker, B. A. *et al.* SOCS3 negatively regulates IL-6 signaling in vivo. *Nat Immunol* **4** (6): 540-545 (2003).

36. Fan, Z. *et al.* Effects of AIM2 and IFI16 on Infectious Diseases and Inflammation. *Viral Immunol* **36** (7): 438-448 (2023).

37. Võsa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* **53** (9): 1300-1310 (2021).

38. Min, J. L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* **53** (9): 1311-1321 (2021).

39. Boukhaled, G. M. *et al.* The Transcriptional Repressor Polycomb Group Factor 6, PCGF6, Negatively Regulates Dendritic Cell Activation and Promotes Quiescence. *Cell Rep* **16** (7): 1829-1837 (2016).

40. Salmond, R. J. *and* Alexander, D. R. SHP2 forecast for the immune system: fog gradually clearing. *Trends Immunol* **27** (3): 154-160 (2006).

41. Luchtel, R. A. ETS1 Function in Leukemia and Lymphoma. *Adv Exp Med Biol* **1459**: 359-378 (2024).

42. Varghese, P., Garcia-Cuellar, M.-P. *and* Slany, R. K. The myeloid transcription factor Elf1 regulates genes with function in innate immunity. *Exp Hematol* **151**: 104864 (2025).

43. Somasundaram, R. *et al.* Transcription factor networks in B-cell differentiation link development to acute lymphoid leukemia. *Blood* **126** (2): 144-152 (2015).

44. Zhou, L., Chong, M. M. W. *and* Littman, D. R. Plasticity of CD4+ T Cell Lineage Differentiation. *Immunity* **30** (5): 646-655 (2009).

45. Tai, Y. *et al.* Molecular mechanisms of T cells activation by dendritic cells in autoimmune diseases. *Front Pharmacol* **9**: 642 (2018).

46. Saravia, J., Chapman, N. M. *and* Chi, H. Helper T cell differentiation. *Cell Mol Immunol* **16** (7): 634-643 (2019).

47. Hwang, J. R. *et al.* Recent insights of T cell receptor-mediated signaling pathways for T cell activation and development. *Exp Mol Med* **52** (5): 750 -761 (2020).

48. Shi, L. Z. *et al.* HIF1alpha-dependent glycolytic pathway orchestrates a metabolic checkpoint for the differentiation of TH17 and Treg cells. *J Exp Med* **208** (7): 1367-1376 (2011).

49. Brasier, A. R. The nuclear factor-B-interleukin-6 signalling pathway mediating vascular inflammation. *Cardiovasc Res* **86** (2): 211–218 (2010).

50. Wegiel, B. *et al.* Interleukin-6 activates PI3K/Akt pathway and regulates cyclin A1 to promote prostate cancer cell survival. *Int J Cancer* **122** (7): 1521-1529 (2008).

51. Cahill, C. M. *and* Rogers, J. T. Interleukin (IL) 1β induction of IL-6 is mediated by a novel phosphatidylinositol 3-kinase-dependent AKT/IκB kinase α pathway targeting activator protein-1. *J Biol Chem* **283** (38): 25900-25912 (2008).

52. Ruwanpura, S. M. *et al.* Cross-talk between IL-6 trans-signaling and AIM2 inflammasome/IL-1β axes bridge innate immunity and epithelial apoptosis to promote emphysema. *Proc Natl Acad Sci U S A* **119** (36): e2201494119 (2022).

53. Hemani, G., Tilling, K. *and* Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet* **13** (11): e1007081 (2017).

54. Dekkers, K. F. *et al.* Blood lipids influence DNA methylation in circulating cells. *Genome Biol* **17** (1): 138 (2016).

55. Hirahara, K. *et al.* Asymmetric Action of STAT Transcription Factors Drives Transcriptional Outputs and Cytokine Specificity. *Immunity* **42** (5): 877-889 (2015).

56. Huang, T. *et al.* miR-301b-5p and its target gene NFATC2IP regulate inflammatory responses in the liver of rainbow trout (Oncorhynchus mykiss) under high temperature stress. *Ecotoxicol Environ Saf* **242**: 113915 (2022).

57. Mowen, K. A. *et al.* Arginine methylation of NIP45 modulates cytokine gene expression in effector T lymphocytes. *Mol Cell* **15** (4): 559-571 (2004).

58. Hodge, M. R. *et al.* NF-AT-driven interleukin-4 transcription potentiated by NIP45. *Science* **274** (5294): 1903-1905 (1996).

59. Jin, S. *et al.* Ebf factors and MyoD cooperate to regulate muscle relaxation via Atp2a1. *Nat Commun* **5**: 3793 (2014).

60. Park, Y. J. *et al.* The Role of Calcium–Calcineurin–NFAT Signaling Pathway in Health and Autoimmune Diseases. *Front Immunol* **11**: 195 (2020).

61. Nilsson, L. M. *et al.* Novel blocker of NFAT activation inhibits IL-6 production in human myometrial arteries and reduces vascular smooth muscle cell proliferation. *Am J Physiol Cell Physiol* **292** (3): C1167-C1178 (2007).
62. Li, T. *et al.* Epigenomics and transcriptomics of systemic sclerosis CD4+ T cells reveal long-range dysregulation of key inflammatory pathways mediated by disease-associated susceptibility loci. *Genome Med* **12** (1): 81 (2020).
63. Lee, S. T. *et al.* A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res* **40** (22): 11339-11351 (2012).
64. Hillary, R. F. *et al.* Blood-based epigenome-wide analyses of 19 common disease states: A longitudinal, population-based linked cohort study of 18,413 Scottish individuals. *PLoS Med* **20** (7): e1004247 (2023).
65. Nones, K. *et al.* Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. *Int J Cancer* **135** (5): 1110-1118 (2014).
66. Kalla, R., Adams, A. T., *and* Satsangi, J. Blood-based DNA methylation in Crohn's disease and severity of intestinal inflammation. *Transl Gastroenterol Hepatol* **4**: 76 (2019).
67. Xu, X. *et al.* Calcium channel TRPV6 promotes breast cancer metastasis by NFATC2IP. *Cancer Lett* **519**: 150-160 (2021).
68. Macintyre, A. N. *et al.* The glucose transporter Glut1 is selectively essential for CD4 T cell activation and effector function. *Cell Metab* **20** (1): 61-72 (2014).
69. Cham, C. M. *et al.* Glucose Deprivation Inhibits Multiple Key Gene Expression Events and Effector Functions in CD8+ T Cells. *Eur J Immunol* **38** (9): 2438-2450 (2008).
70. Schübeler, D. Function and information content of DNA methylation. *Nature* **517** (7534): 321-326 (2015).
71. Smith, A. K. *et al.* Epigenetic changes associated with inflammation in breast cancer patients treated with chemotherapy. *Brain Behav Immun* **38**: 227-236 (2014).
72. Lundin, J. I. *et al.* Epigenetic mechanisms underlying variation of IL-6, a well-established inflammation biomarker and risk factor for cardiovascular disease. *Atherosclerosis* **407**: 120219 (2025).
73. Zhang, S. C. *et al.* Reversible promoter methylation determines fluctuating expression of acute phase proteins. *Elife* **9:** e51317 (2020).
74. Davies, N. M., Holmes, M. V. *and* Davey Smith, G. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ* **362**: k601 (2018).
75. Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: The Leiden Longevity Study. *Eur J Hum Genet* **14** (1): 79-84 (2006).
76. Sinke, L. *et al.* DNAmArray: Streamlined workflow for the quality control, normalization, and analysis of Illumina methylation array data (2.1) *Zenodo* (2019).
77. Meisinger, C. *et al.* Prevalence of undiagnosed diabetes and impaired glucose regulation in 35-59-year-old individuals in Southern Germany: The KORA F4 study. *Diabet Med* **27** (3): 360-362 (2010).
78. Herder, C. *et al.* Association of subclinical inflammation with polyneuropathy in the older population: KORA F4 study. *Diabetes Care* **36** (11): 3663-3670 (2013).
79. Touleimat, N. *and* Tost, J. Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* **4** (3): 325-341 (2012).
80. Willemsen, G. *et al.* The Netherlands twin register biobank: A resource for genetic epidemiological studies. *Twin Res Hum Genet* **13** (3): 231-245 (2010).
81. Van Dongen, J. *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* **7**: 11115 (2016).
82. Francioli, L. C. *et al.* Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* **46** (8): 818-825 (2014).
83. Fortin, J. P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* **15** (12): 503 (2014).
84. Jonkman, T. H. *et al.* Probing epigenetic clocks as a rational markers of biological age using blood cell counts. *medRxiv* (2025).
85. Amemiya, H. M., Kundaje, A. *and* Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9** (1): 9354 (2019).
86. Zhou, W., Laird, P. W. *and* Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45** (4): e22 (2017).
87. van Iterson, M. *et al.* Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18** (1): 19 (2017).

88. Willer, C. J., Li, Y. *and* Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26** (17): 2190-2191 (2010).

89. Slieker, R. C. *et al.* Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* **6** (1): 26 (2013).

90. Battram, T. *et al.* The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res* **7**: 41 (2022).

91. Li, M. *et al.* EWAS Atlas: A curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res* **47** (D1): D983-D988 (2019).

92. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38** (4): 576-589 (2010).

93. van Greevenbroek, M. M. J. *et al.* The cross-sectional association between insulin resistance and circulating complement C3 is partly explained by plasma alanine aminotransferase, independent of central obesity and general inflammation (the CODAM study). *Eur J Clin Invest* **41** (4): 372-379 (2011).

94. Tigchelaar, E. F. *et al.* Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* **5** (8): e006772 (2015).

95. Hofman, A. *et al.* The Rotterdam Study: 2014 objectives and design update. *Eur J Epidemiol* **28** (11): 889-926 (2013).

96. Huisman, M. H. B. *et al.* Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. *J Neurol Neurosurg Psychiatry* **82** (10): 1165-1170 (2011).

97. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9** (8): e1003118 (2013).

98. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet* **10** (5): e1004383 (2014).

99. Wang, G. *et al.* A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. *J R Stat Soc Series B Stat Methodol* **82** (5): 1273-1300 (2020).

100. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet* **17** (9): e1009440 (2021).

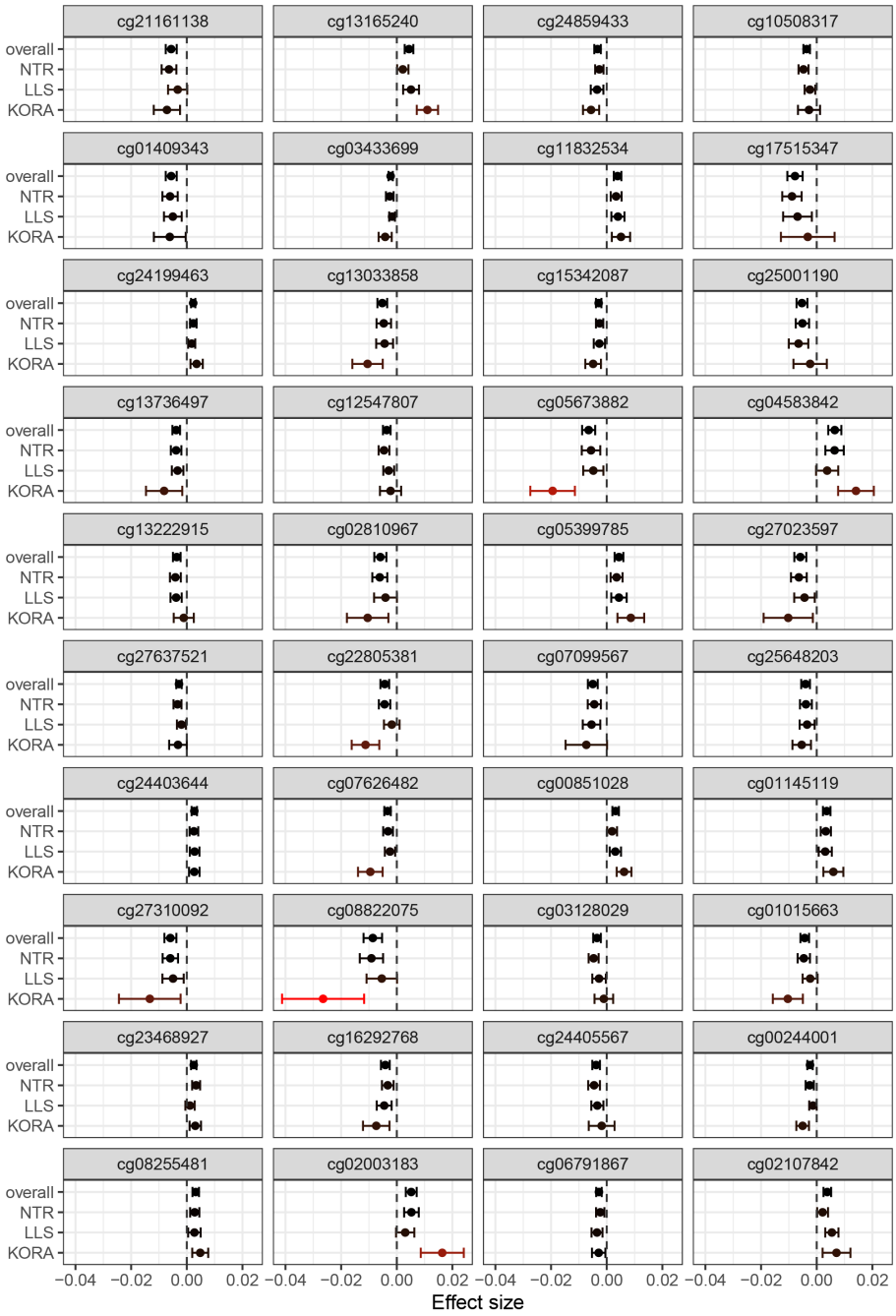
101. Machiela, M. J. *and* Chanock, S. J. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31** (21): 3555-3557 (2015).

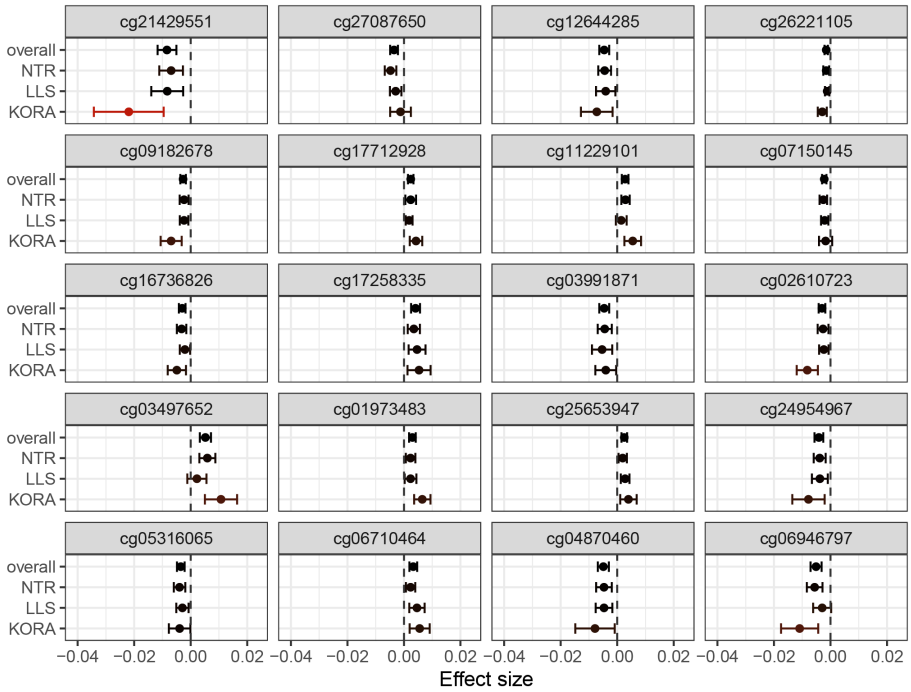
102. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2** (3): 100141 (2021).

103. Ahluwalia, T. S. *et al.* Genome-wide association study of circulating interleukin 6 levels identifies novel loci. *Hum Mol Genet* **30** (5): 393-409 (2021).

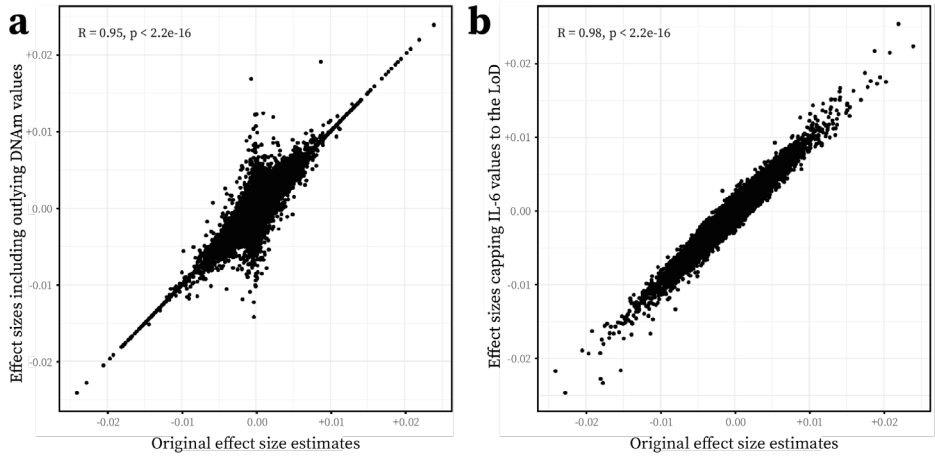
104. Sinke, L. nebulyra/il6_ewas: IL-6 EWAS meta-analysis. *Zenodo* (2025).

Supplementary Material

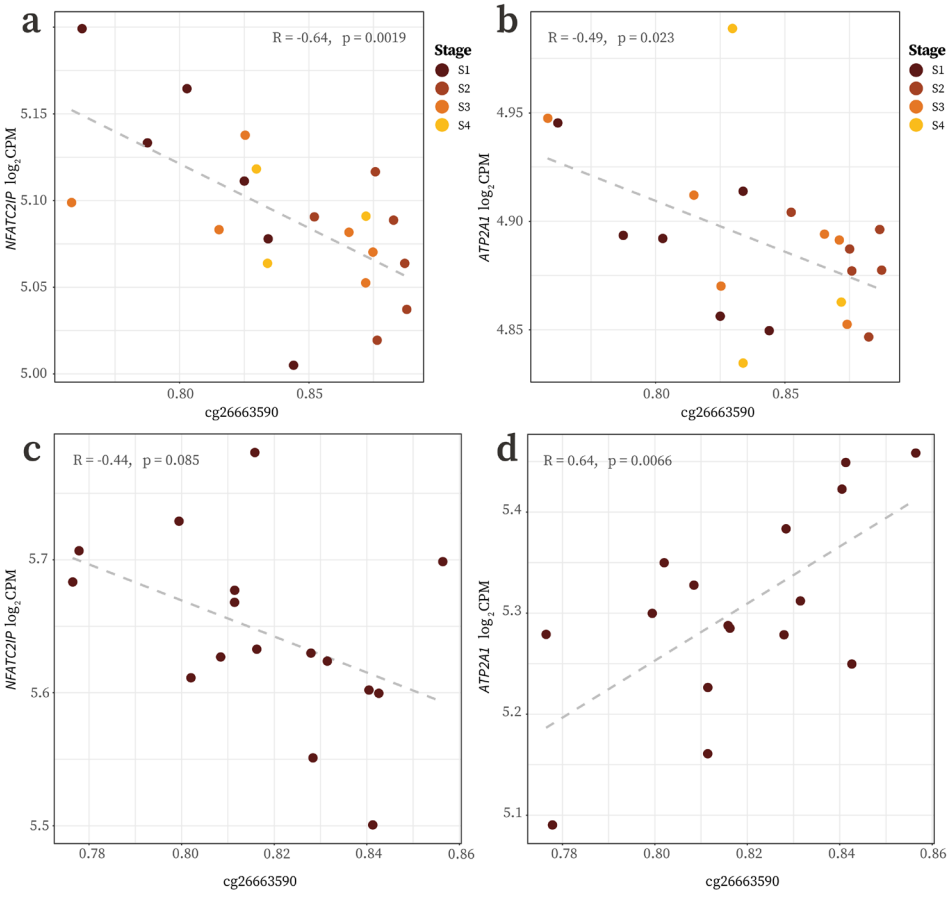

cg21161138
cg13165240
cg24859433
cg10508317
cg01409343
cg03433699
cg11832534
cg17515347
cg24199463
cg13033858
cg15342087
cg25001190
cg13736497
cg12547807
cg05673882
cg04583842
cg13222915
cg02810967
cg05399785
cg27023597
cg27637521
cg22805381
cg07099567
cg25648203
cg24403644
cg07626482
cg00851028
cg01145119
cg27310092
cg08822075
cg03128029
cg01015663
cg23468927
cg16292768
cg24405567
cg00244001
cg08255481
cg02003183
cg06791867
cg02107842
overall
NTR
LLS
KORA
−0.04
−0.02
0.00
0.02
Effect size



Supplementary Figure 1 | Cohort-specific effects for the top 100 CpGs associated with IL-6, ordered by p-value. The effects measured in each cohort as well as the overall meta-analysed effect estimate are shown alongside their 95% confidence intervals, coloured by the absolute difference between the cohort-specific estimate and overall estimate. A null effect is indicated by a dashed grey line.



Supplementary Figure 2 | Data processing sensitivity analysis results in LLS. a) Scatter plot comparing IL-6 EWAS effect sizes with (x-axis) and without (y-axis) removal of outlying DNA methylation (DNAm) values. **b)** Scatter plot comparing IL-6 EWAS effect sizes when removing IL-6 values below the limit of detection (LoD, x-axis) or capping them at the LoD (y-axis). Pearson correlation coefficients and respective *p*-values are shown in the top left corner of each plot.



Supplementary Figure 3 | eQTM validation in isolated immune cells. a) Scatter plot of *NFATC2IP* $\log_2$CPM against DNAm at cg26663590 in isolated B-cells at four stages of development. **b)** Scatter plot of *ATP2A1* $\log_2$CPM against DNAm in the same cells. **c)** Scatter plot of *NFATC2IP* $\log_2$CPM against DNAm in CD4+ T-cells, **d)** Scatter plot of *ATP2A1* $\log_2$CPM against DNAm in the same cells. Pearson correlation coefficients and associated *p*-values are shown, as well as a line of best fit as a dashed grey line.

CpG	Position	*n*	beta	SE	*p*	p_{FDR}
cg02650017	chr17:47301614	4341	-0.0032	0.0004	8.38E-17	3.45E-11
cg10636246	chr1:159046973	4318	-0.0083	0.0010	3.26E-16	6.72E-11
cg02734358	chr4:90227074	4350	-0.0078	0.0010	6.96E-15	9.56E-10
cg01101459	chr1:234871477	4350	0.0061	0.0008	1.62E-14	1.33E-09
cg16936953	chr17:57915665	4346	-0.0104	0.0014	1.39E-14	1.33E-09
cg06126421	chr6:30720080	4347	-0.0113	0.0015	3.50E-14	2.41E-09
cg18181703	chr17:76354621	4349	-0.0071	0.0009	4.39E-14	2.58E-09
cg07094298	chr4:2748026	4350	-0.0097	0.0013	1.95E-13	1.00E-08
cg01940273	chr2:233284934	4351	-0.0094	0.0013	3.81E-13	1.57E-08
cg12054453	chr17:57915717	4345	-0.0111	0.0015	3.45E-13	1.57E-08
cg01059398	chr3:172235808	4350	-0.0062	0.0009	2.47E-12	9.25E-08
cg05575921	chr5:373378	4197	-0.0143	0.0021	5.87E-12	2.02E-07
cg17501210	chr6:166970252	4349	-0.0064	0.0010	1.87E-11	5.94E-07
cg12992827	chr3:101901234	4346	-0.0063	0.0009	2.83E-11	8.32E-07
cg18942579	chr17:57915773	4344	-0.0085	0.0013	3.30E-11	9.06E-07
cg02716826	chr9:33447032	4344	-0.0054	0.0008	5.59E-11	1.43E-06
cg15901722	chr5:175974973	4351	-0.0051	0.0008	5.88E-11	1.43E-06
cg19821297	chr19:12890029	4347	-0.0059	0.0009	6.66E-11	1.52E-06
cg14753356	chr6:30720108	4351	-0.0057	0.0009	1.14E-10	2.47E-06
cg21566642	chr2:233284661	4336	-0.0123	0.0019	1.61E-10	3.31E-06

Supplementary Table 1 | Top 20 CpGs associated with circulating IL-6 levels in blood. CpG identifier, genomic position, sample size (n), and association with IL-6 (beta) is shown, alongside standard error (SE) and significance (nominal and FDR-adjusted *p*-value).

Trait	OR	lnOR	SE	*p*	p_{FDR}	Overlap	%
C-reactive protein	438.9	6.08	0.11	2.2E-308	2.2E-308	241	14.61
Soluble TNFR2	181.6	5.20	0.25	2.05E-98	1.46E-97	20	14.39
Education level	169.2	5.13	0.17	1.81E-208	2.58E-207	46	12.74
Aggression	166.9	5.12	0.22	3.12E-123	2.54E-122	26	13.20
Body mass index	148.6	5.00	0.12	2.2E-308	2.2E-308	90	10.11
Cognition	135.7	4.91	0.17	9.93E-192	1.13E-190	45	10.51
Triglycerides	128.3	4.85	0.40	2.59E-33	8.20E-33	7	10.94
HDL-cholesterol	109.1	4.69	0.26	5.36E-73	2.78E-72	17	9.24
PTSD	89.7	4.50	0.28	7.10E-57	3.11E-56	14	7.78
Systolic blood pressure	71.8	4.27	0.29	8.08E-49	3.07E-48	13	6.34
Type 2 diabetes	68.0	4.22	0.42	2.73E-23	7.41E-23	6	6.12
Fasting insulin	62.8	4.14	0.21	5.18E-88	3.28E-87	26	5.42
Diastolic blood pressure	57.4	4.05	0.37	2.28E-28	6.50E-28	8	5.19
Waist circumference	48.4	3.88	0.29	2.60E-41	8.72E-41	13	4.36
AHEI	40.7	3.71	0.27	1.72E-43	6.13E-43	15	3.68
Smoking	22.3	3.11	0.10	5.03E-211	9.56E-210	198	1.14
All-cause mortality	20.8	3.03	0.16	1.24E-80	7.07E-80	45	1.77
Alcohol use	18.5	2.92	0.11	9.06E-157	8.61E-156	121	1.27
Age	7.1	1.96	0.13	2.48E-54	1.01E-53	78	0.57
Sex	4.5	1.50	0.13	1.01E-32	3.03E-32	79	0.37

Supplementary Table 2 | Top 20 traits from EWAS enrichment analysis. Odds ratio of enrichment in IL-6 associated CpGs (OR and log-transformed lnOR) is shown, alongside their SE and significance (nominal and FDR-adjusted *p*-value). The total number of tested CpGs associated with the trait in previous EWAS and the overlap with IL-6 associated CpGs is also presented. (HDL: high-density lipoprotein, PTSD: post-traumatic stress disorder, AHEI: alternative healthy eating index)

Chromatin State	OR	lnOR	SE	p	p_{FDR}	Full Roadmap name
(1) TssA	0.56	-0.57	0.14	1.71E-02	3.45E-02	Active TSS
(2) TssAFlnk	0.97	-0.03	0.13	8.38E-01	9.32E-01	Flanking active TSS
(3) TxFlnk	6.53	1.88	2.74	9.36E-07	7.02E-06	Transcr. at gene 5' and 3'
(4) Tx	1.03	0.03	0.18	8.43E-01	9.32E-01	Strong transcription
(5) TxWk	1.57	0.45	0.23	1.92E-03	5.77E-03	Weak transcription
(6) EnhG	3.23	1.17	1.00	6.44E-05	3.22E-04	Genic enhancers
(7) Enh	3.93	1.37	0.60	7.47E-20	1.12E-18	Enhancers
(8) ZNF/Rpts	0.00	-10.21	9.30E+96	9.32E-01	9.32E-01	ZNF genes + repeats
(9) Het	1.05	0.05	0.75	9.29E-01	9.32E-01	Heterochromatin
(10) TssBiv	0.26	-1.35	0.18	1.84E-02	3.45E-02	Bivalent/poised TSS
(11) BivFlnk	0.08	-2.52	0.14	1.10E-02	2.75E-02	Flanking bivalent TSS/Enh
(12) EnhBiv	0.63	-0.46	0.37	3.58E-01	5.37E-01	Bivalent enhancer
(13) ReprPC	0.48	-0.74	0.12	1.55E-03	5.77E-03	Repressed polycomb
(14) ReprPCWk	0.73	-0.31	0.13	7.99E-02	1.33E-01	Weak repressed polycomb
(15) Quies	0.96	-0.04	0.12	7.64E-01	9.32E-01	Quiescent/low

Supplementary Table 3 | Chromatin state enrichment analysis results using the peripheral blood mononuclear cell reference epigenome (PBMC; E062). The odds ratio of enrichment in IL-6 associated CpGs (OR and natural log-transformed lnOR) is shown, alongside its standard error (SE) and significance (nominal and FDR-adjusted *p*-value). The abbreviated and full Roadmap names are also displayed. (TSS: transcription start site, ZNF: zinc finger)

TF	Type	p	p_{FDR}	CpG n	CpG %	Background %
Atf4	bZIP	6.76E-11	7.44E-10	9	2.24	0.40
Chop	bZIP	1.73E-10	9.52E-10	8	2.00	0.32
Bach2	bZIP	7.83E-07	2.87E-06	6	1.50	0.33
NFκB-p50,p52	RHD	2.19E-06	5.68E-06	5	1.25	0.25
Nrf2	bZIP	3.10E-06	5.68E-06	3	0.75	0.08
NFκB-p65	RHD	2.88E-06	5.68E-06	10	2.49	0.92
Arnt: Ahr	bHLH	3.92E-06	6.16E-06	18	4.49	2.23
RUNX2	Runt	7.71E-06	9.83E-06	13	3.24	1.44
NF-E2	bZIP	8.04E-06	9.83E-06	3	0.75	0.09
MafK	bZIP	1.71E-05	1.79E-05	6	1.50	0.43
MafA	bZIP	1.79E-05	1.79E-05	14	3.49	1.68

Supplementary Table 4 | Transcription factor (TF) binding site enrichment analysis results. Sequences within 50bp of IL-6 associated CpGs were scanned for established motifs using HOMER and compared to a random genomic background matched for GC content. TF name and type are shown alongside the significance of the enrichment (nominal and FDR-adjusted *p*-values). The number of CpGs responsible for the enrichment are also displayed (CpG *n*, CpG %), as well as the percentage of background sequences within 50bp of the respective motif (Background %).

CpG	Gene name	Ensembl ID	beta	p	p_{FDR}	Nearest
cg05575921	*AHRR*	ENSG00000063438	-0.0390	1.11E-108	1.51E-105	TRUE
cg25648203	*AHRR*	ENSG00000063438	-0.0141	1.09E-96	7.47E-94	FALSE
cg08469255	*DDR1*	ENSG00000204580	-0.0113	4.07E-88	1.85E-85	TRUE
cg17901584	*DHCR24*	ENSG00000116133	-0.0207	5.05E-81	1.72E-78	TRUE
cg03984502	*RORC*	ENSG00000143365	-0.0089	7.12E-80	1.94E-77	FALSE
cg06764092	*SLAMF8*	ENSG00000158714	-0.0262	5.72E-78	1.30E-75	TRUE
cg22930808	*PARP9*	ENSG00000138496	-0.0229	2.36E-74	4.59E-72	FALSE
cg26276120	*SPSB2*	ENSG00000111671	-0.0098	4.12E-74	7.03E-72	FALSE
cg10636246	*AIM2*	ENSG00000163568	-0.0162	1.64E-67	2.49E-65	TRUE
cg22930808	*DTX3L*	ENSG00000163840	-0.0213	7.45E-64	1.02E-61	TRUE
cg21161138	*AHRR*	ENSG00000063438	-0.0138	1.33E-58	1.65E-56	FALSE
cg23813257	*IL32*	ENSG00000008517	-0.0123	3.40E-57	3.87E-55	TRUE
cg08216099	*PXDN*	ENSG00000130508	-0.0108	4.20E-51	4.41E-49	TRUE
cg21571160	*C4BPB*	ENSG00000123843	-0.0121	5.12E-50	4.99E-48	TRUE
cg10416593	*SCO2*	ENSG00000130489	-0.0147	3.98E-47	3.62E-45	FALSE
cg10045881	*CHI3L2*	ENSG00000064886	-0.0104	7.29E-43	6.22E-41	TRUE
cg25909396	*PRKCA*	ENSG00000154229	-0.0102	1.60E-39	1.28E-37	TRUE
cg06710464	*BAIAP2*	ENSG00000175866	-0.0096	5.65E-38	4.29E-36	TRUE
cg26916621	*HOXB3*	ENSG00000120093	0.0110	6.49E-34	4.66E-32	FALSE
cg26276120	*LRRC23*	ENSG00000010626	-0.0068	1.33E-33	9.05E-32	FALSE

Supplementary Table 5 | Top 20 expression quantitative trait methylation (eQTM) effects connecting IL-6 associated CpGs and nearby genes (within 100kb), ordered by *p*-value. The CpG and gene identifiers are shown alongside their eQTM effect size and significance.

Database	Gene set	Set size	Overlap	p	p_{FDR}
Reactome (2022)	Immune System	1943	65	1.41E-11	1.04E-07
Reactome (2022)	Disease	1736	58	2.21E-10	8.09E-07
MsigDB (2020)	IL-2/STAT5 Signalling	199	16	1.60E-08	3.92E-05
Elsevier Pathway Collection	Proteins with Altered Expression in Cancer Metabolic Reprogramming	85	11	2.18E-08	4.00E-05
Reactome (2022)	Metabolism	2049	57	2.02E-07	2.82E-04
BioPlanet (2019)	Interleukin-2 signalling pathway	847	32	2.31E-07	2.82E-04
MsigDB (2020)	TNF-alpha Signalling via NF-κB	200	14	6.91E-07	7.23E-04
GeDiPNet (2023)	Lung Adenocarcinoma	336	18	9.62E-07	7.23E-04
Reactome (2022)	Cell Cycle, Mitotic	523	23	9.78E-07	7.23E-04
BioPlanet (2019)	TGF-β regulation of ECM	565	24	1.05E-06	7.23E-04
Reactome (2022)	Infectious Disease	961	33	1.25E-06	7.23E-04
Reactome (2022)	Cell Cycle	654	26	1.28E-06	7.23E-04
GO BP (2023)	Cellular Response To Cytokine Stimulus	308	17	1.28E-06	7.23E-04
GO BP (2023)	DNA Damage Response	384	19	1.57E-06	8.24E-04
Elsevier Pathway Collection	Glutamine in Cancer Metabolism	28	6	1.76E-06	8.62E-04
GO BP (2023)	Regulation Of Cell Population Proliferation	766	28	2.53E-06	1.16E-03
BioPlanet (2019)	Immune system	998	33	2.80E-06	1.21E-03
Reactome (2022)	Adaptive Immune System	733	27	3.33E-06	1.34E-03
BioPlanet (2019)	Adipocytokine signalling pathway	67	8	3.49E-06	1.34E-03
MsigDB (2020)	mTORC1 Signalling	200	13	3.92E-06	1.44E-03

Supplementary Table 6 | Top 20 enriched terms in genes linked to IL-6 associated CpGs, ordered by *p*-value. The set size and overlap is shown alongside their significance (nominal and FDR-adjusted *p*-value).

Confidence	TF	p	p_{FDR}	Regulon size	Overlap	%
A_B_C_D	ETS1	1.30E-11	4.42E-09	9454	264	2.79
A_B_C_D	MAZ	2.80E-10	4.77E-08	488	35	7.12
A_B_C_D	TCF3	2.02E-09	2.30E-07	500	34	6.80
A_B_C_D	TFAP4	1.11E-07	8.64E-06	478	30	6.28
A_B_C_D	NR2F1	1.27E-07	8.64E-06	481	30	6.24
A_B_C_D	ZEB1	2.35E-07	1.34E-05	495	30	6.06
A_B_C_D	STAT3	3.03E-07	1.37E-05	643	35	5.44
A_B_C_D	KLF1	3.23E-07	1.37E-05	475	29	6.11
A_B_C_D	NEUROD1	5.50E-07	2.00E-05	460	28	6.09
A_B_C_D	ELF1	5.87E-07	2.00E-05	489	29	5.93
A_B_C_D	ZNF263	1.10E-06	3.39E-05	3047	101	3.31
A_B_C_D	EGR1	1.24E-06	3.39E-05	10673	270	2.53
A_B_C_D	VDR	1.29E-06	3.39E-05	1476	59	4.00
A_B	EGR1	3.31E-07	3.65E-05	559	32	5.72
A_B_C_D	MXI1	1.56E-06	3.81E-05	485	28	5.77
A_B_C_D	CTCFL	1.98E-06	4.51E-05	491	28	5.70
A_B_C_D	E2F4	3.23E-06	6.89E-05	6448	179	2.78
A_B_C_D	MNT	3.49E-06	6.99E-05	477	27	5.66
A_B_C_D	SNAI2	5.72E-06	1.08E-04	490	27	5.51
A_B_C_D	TFAP2C	9.02E-06	1.62E-04	11222	277	2.47

Supplementary Table 7 | Top 20 enriched TFs targeting genes linked to IL-6 associated CpGs, ordered by *p*-value. The confidence level from DoRothEA, transcription factor name, total regulon size, and size and percentage of overlap with genes identified in our results is shown.

Gene name	Change in RPKM	SE	n	p	p_{FDR}
TPI1	352.8957	43.5953	18	4.01E-05	7.22E-03
CST7	17.9979	2.0314	18	2.08E-05	7.22E-03
S100A11	41.5615	5.3496	18	5.39E-05	7.22E-03
PTTG1	16.9178	2.3099	18	8.20E-05	8.24E-03
TCTA	-2.5093	0.3591	18	1.14E-04	8.65E-03
GPRIN1	1.2432	0.1811	18	1.29E-04	8.65E-03
C2CD4D	0.4046	0.0630	18	2.03E-04	1.04E-02
TRIP11	-0.6523	0.1017	18	2.06E-04	1.04E-02
GAPDH	27.5347	4.4137	18	2.49E-04	1.04E-02
SBNO2	10.3250	1.7073	18	3.07E-04	1.04E-02
BCL3	11.7852	1.9684	18	3.28E-04	1.04E-02
ZNF335	-2.0875	0.3487	18	3.28E-04	1.04E-02
EXOC3	-6.3387	1.0621	18	3.35E-04	1.04E-02
VMP1	18.6956	3.4218	18	5.99E-04	1.70E-02
AIM2	4.6954	0.8782	18	6.88E-04	1.70E-02
C21orf33	4.7578	0.8981	18	7.31E-04	1.70E-02
SLC16A3	4.2623	0.8313	18	8.99E-04	1.70E-02
RHEBL1	1.6392	0.3247	18	9.91E-04	1.70E-02
TAPBPL	-2.0533	0.4075	18	0.001003	1.70E-02
SRCAP	-4.9568	0.9947	18	0.001075	1.70E-02

Supplementary Table 8 | Top 20 genes responding to IL-6 stimulation of isolated CD4+ T-cells, ordered by *p*-value. The change in reads per kilobase million (RPKM), standard error (SE), number of samples analysed (*n*), and significance of the response (nominal and FDR-adjusted *p*-value) is shown.

CpG	Outcome trait	beta	SE	p	p_{FDR}
cg10508317	Ulcerative colitis	0.2283	0.0682	8.23E-04	1.89E-02
cg10508317	Crohn's disease	0.1925	0.0577	8.55E-04	1.97E-02
cg10508317	C-reactive protein	0.0233	0.0069	7.42E-04	3.48E-02
cg18181703	C-reactive protein	0.0167	0.0052	1.39E-03	3.48E-02
cg10508317	Body mass index	0.0186	0.0056	7.97E-04	3.67E-02
cg18181703	Body mass index	0.0147	0.0047	1.69E-03	3.88E-02
cg10508317	LDL cholesterol	0.0097	0.0029	9.03E-04	4.52E-02
cg26775866	Crohn's disease	0.0303	0.0105	4.00E-03	4.59E-02
cg18181703	HDL cholesterol	-0.0218	0.0068	1.35E-03	4.71E-02
cg10508317	HDL cholesterol	-0.0164	0.0053	1.89E-03	4.71E-02
cg10508317	Triglyceride levels	0.0049	0.0015	1.30E-03	6.52E-02
cg27637521	HDL cholesterol	-0.0308	0.0114	6.82E-03	7.56E-02
cg26775866	HDL cholesterol	-0.0019	0.0007	7.42E-03	7.56E-02
cg15901722	HDL cholesterol	0.0057	0.0021	7.56E-03	7.56E-02
cg26775866	Body mass index	0.0029	0.0011	8.04E-03	1.23E-01
cg27305772	Body mass index	0.0072	0.0029	1.18E-02	1.28E-01
cg15901722	Body mass index	-0.0068	0.0028	1.39E-02	1.28E-01
cg26775866	C-reactive protein	0.0018	0.0007	8.47E-03	1.41E-01
cg18181703	LDL cholesterol	0.0034	0.0012	5.67E-03	1.42E-01
cg18181703	Triglyceride levels	0.0032	0.0011	5.82E-03	1.45E-01

Supplementary Table 9 | Top 20 IL-6 effects mediated by DNAm, ordered by *p*-value for the indirect (mediated) effect. The mediating CpG, outcome trait, indirect effect size (beta) are shown, alongside its standard error (SE) and significance (nominal and FDR-adjusted *p*-values; HDL: high density lipoprotein; LDL: low density lipoprotein)

Exposure	Outcome	IEU GWAS ID	beta	SE	p	p_{FDR}
cg26663590	Type 2 diabetes	ebi-a-GCST006867	0.1168	0.0280	3.04E-05	2.74E-04
cg26663590	Ulcerative colitis	ieu-a-970	0.2194	0.0570	1.20E-04	5.40E-04
cg26663590	C-reactive protein	ebi-a-GCST90029070	0.0456	0.0123	2.20E-04	6.60E-04
cg26663590	Rheumatoid arthritis	ukb-d-M13_RHEUMA	0.0012	0.0003	4.45E-04	9.76E-04
cg26663590	Body mass index	ieu-b-40	0.0707	0.0205	5.42E-04	9.76E-04
cg26663590	HDL cholesterol	ieu-b-109	-0.0265	0.0083	1.39E-03	2.09E-03
cg26663590	LDL cholesterol	ieu-b-110	-0.0138	0.0045	1.90E-03	2.44E-03
cg26663590	Crohn's disease	ieu-a-12	0.3363	0.1392	1.57E-02	1.77E-02
cg26663590	Triglyceride levels	ieu-b-111	0.0024	0.0042	5.79E-01	5.79E-01

Supplementary Table 10 | Results from 2SMR investigating effects of DNAm at cg26663590 on nine IL-6 related traits. The outcome trait, IEU GWAS ID, causal effect size (beta), and standard error (SE) are shown, alongside significance (nominal and FDR-adjusted *p*-values; HDL: high density lipoprotein; LDL: low density lipoprotein)

Exposure	IEU GWAS ID	Outcome	beta	SE	p	p_{FDR}
Type 2 diabetes	ebi-a-GCST006867	cg26663590	0.0496	0.0183	6.79E-03	6.11E-02
Ulcerative colitis	ieu-a-970	cg26663590	-0.0182	0.0123	1.39E-01	6.26E-01
HDL cholesterol	ieu-b-109	cg26663590	-0.0381	0.0362	2.92E-01	7.83E-01
Body mass index	ieu-b-40	cg26663590	0.1371	0.1566	3.81E-01	7.83E-01
Triglyceride levels	ieu-b-111	cg26663590	0.0321	0.0411	4.35E-01	7.83E-01
C-reactive protein	ebi-a-GCST90029070	cg26663590	-0.0157	0.0507	7.56E-01	9.17E-01
Crohn's disease	ieu-a-12	cg26663590	0.0017	0.0128	8.92E-01	9.17E-01
LDL cholesterol	ieu-b-110	cg26663590	0.0052	0.0406	8.98E-01	9.17E-01
Rheumatoid arthritis	ukb-d-M13_RHEUMA	cg26663590	-0.9001	8.6455	9.17E-01	9.17E-01

Supplementary Table 11 | Results from 2SMR investigating effects of nine IL-6 related traits on DNAm at cg26663590. The exposure trait, IEU GWAS ID, causal effect size (beta), and standard error (SE) are shown, alongside significance (nominal and FDR-adjusted *p*-values; HDL: high density lipoprotein; LDL: low density lipoprotein)

CHAPTER THREE

Adipokines

DNA methylation of genes involved in lipid metabolism drives adiponectin levels and metabolic disease

Lucy Sinke[1], Thomas Delerue[2], Rory Wilson[2], Xueling Lu[3,4], Yujing Xia[5], Ricardo Costeira[5], M. Kamal Nasr[6,7], Marian Beekman[1], Lude Franke[3], Alexandra Zhernakova[3], Jingyuan Fu[3], Christian Gieger[2,8], Christian Herder[9,10], Wolfgang Koenig[8,11,12], Annette Peters[2,8,13], Jose M. Ordovas[14], Marcus Dörr[6,7], Hans J. Grabe[6], Matthias Nauck[6,7], Jordana T. Bell[5], Alexander Teumer[6,7,15], Harold Snieder[3], Melanie Waldenberger[2,8], P. Eline Slagboom[1], and Bastiaan T. Heijmans[1]

[1] Leiden University Medical Centre, Leiden, The Netherlands
[2] Helmholtz Munich, Neuherberg, Germany
[3] University of Groningen, Groningen, The Netherlands
[4] Shantou University Medical College, Guangdong, China
[5] King's College London, London, United Kingdom
[6] University Medicine Greifswald, Greifswald, Germany
[7] German Centre for Cardiovascular Research (DZHK), Greifswald, Germany
[8] German Centre for Cardiovascular Research (DZHK), Munich, Germany
[9] Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[10] German Centre for Diabetes Research (DZD), Neuherberg, Germany
[11] Technical University of Munich, Munich, Germany
[12] Ulm University, Ulm, Germany
[13] Ludwig Maximilian University of Munich, Munich, Germany
[14] Tufts University, Boston, Massachusetts, United States of America
[15] German Centre for Neurodegenerative Diseases (DZNE), Greifswald, Germany

Published in Diabetologia 69, 127-145 (2026)
DOI: 10.1007/s00125-025-06549-6

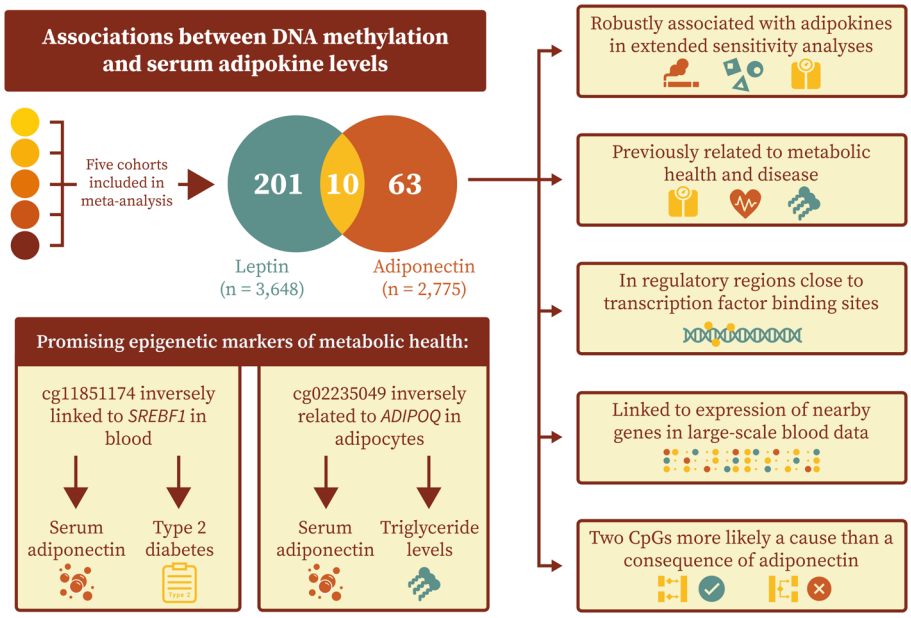
Abstract

Despite playing critical roles in the pathophysiology of metabolic disorders such as type 2 diabetes (T2D), the molecular mechanisms underlying circulating adipokine levels remain poorly understood. By identifying genomic regions involved in the regulation of adipokine levels, we can improve our understanding of T2D pathogenesis and inter-individual differences in metabolic disease risk.

We conducted an epigenome-wide meta-analysis of associations between serum adiponectin (*n* = 2,791) and leptin (*n* = 3,661), and leukocyte DNA methylation at over 400,000 CpG sites across five European cohorts. The resulting methylation signatures were followed up using functional genomics, integrative analyses, and causal inference. Our findings revealed robust associations with adiponectin at 73 CpGs and with leptin at 211 CpGs. Many identified sites were associated with metabolic syndrome risk factors, located in regulatory chromatin regions, and correlated with genes important for lipid transport (e.g. *ABCG1*), metabolism (e.g. *CPT1A*), and biosynthesis (e.g. *DHCR24*).

Bidirectional two-sample Mendelian randomisation further identified two specific sites as plausible drivers of both adiponectin levels and metabolic health: one annotated to *ADIPOQ*, the gene encoding adiponectin, and another linked to *SREBF1* expression, an established modifier of T2D risk known to exert its effects via adiponectin. Taken together, this large-scale and integrative analysis uncovers links between adipokines and widespread yet functionally specific differences in gene regulation with a central role in T2D and its risk factors.

Graphical Abstract



Highlights

- 73 CpGs associate with adiponectin and 211 CpGs link to leptin in blood.
- These CpGs are enriched for residing in enhancers close to relevant transcription factor binding sites.
- Integrative analyses connect adipokine-associated DNA methylation to genes involved in lipid transport, metabolism, and synthesis.
- Triangulation supports DNA methylation at identified CpGs as plausibly driving adiponectin production.
- In line with this, two-sample Mendelian randomisation identifies two putatively causal sites: one at *ADIPOQ*, and another at *SREBF1*.
- Public data uncovers correlation between *ADIPOQ* expression and DNAm at the implicated CpG in adipocytes, the primary producers of adiponectin.
- We propose that shared upstream factors, such as diet, could be inducing tissue-agnostic DNAm effects, enabling large-scale blood-based EWAS to pinpoint regulatory sites even for traits non-hematopoietic in origin.

Keywords: *adiponectin, causal inference, epigenomics, leptin, lipid metabolism, meta-analysis, metabolic health, type 2 diabetes*

Background

Adiponectin and leptin are two adipokines that play central roles in regulating energy homeostasis and metabolism, with influences on both insulin sensitivity and inflammation. Circulating levels of these molecular mediators are directly implicated in the pathogenesis and progression of the metabolic syndrome and type 2 diabetes (T2D). A clearer understanding of their regulation could therefore uncover new avenues for predicting, preventing or treating metabolic disease[1,2].

Epigenetic modifications, such as DNA methylation (DNAm), are established as being both responsive to lifestyle changes and capable of modifying disease risk. Growing evidence also supports epigenetic regulation of adiponectin and leptin specifically as partly driving inter-individual differences in health[3,4]. Blood-based epigenome-wide association studies (EWAS) have previously uncovered robust and biologically coherent correlations between DNAm and metabolic risk factors, even where the investigated traits were non-haematopoietic in origin[5,6]. Supported explanations for detecting such associations in leukocytes include shared upstream drivers such as diet[7], DNAm responses to circulating phenotypes[8], and immune cell mediation of the inflammatory components of metabolic disease[9].

Despite substantial progress, however, research directly examining relationships between adiponectin and DNAm have been limited in sample size[10], and leptin has thus far not been investigated on a genome-wide scale. A comprehensive EWAS of these adipokines in a sufficiently large sample size is warranted to detect subtle but consistent molecular effects. Thorough interpretation of the resulting methylation signatures represents a critical next step in understanding the regulatory architecture of metabolic homeostasis in human health.

Results

Circulating adipokines have largely distinct DNA methylation signatures in blood

We performed EWAS meta-analyses of circulating adiponectin (n = 2,791; 412,224 CpGs) and leptin (n = 3,661; 406,390 CpGs) levels in blood samples from five European cohorts (**Table 1,2**). Mean age was 55.5 years in the leptin analysis and 56.8 years for adiponectin, and the population was predominantly female (adiponectin: 55.2%; leptin: 54.4%). Cohorts represented a combination of fasted (KORA, TwinsUK, SHIP-TREND, and LLD) and non-fasted (LLS) samples. The top 20 CpGs for each adipokine are provided in **Supplementary Tables 1** and **2**.

Characteristic	KORA F4	LLS	LLD	TwinsUK	SHIP-TREND
Sample size	807	718	701	124	441
Age, years	68.8 ± 4.4	58.9 ± 6.7	45.5 ± 13.1	55.1 ± 11.7	50.0 ± 13.4
Sex, female	396 (49.1)	370 (51.5)	411 (58.6)	124 (100.0)	241 (54.6)
BMI, kg/m^2	26.4 (4.2)	25.1 (4.3)	24.7 (4.8)	25.0 (4.7)	27.2 (4.2)
Smoking, current	69 (8.6)	85 (11.8)	130 (18.5)	28 (22.6)	154 (34.9)
Smoking, never	388 (48.1)	199 (27.7)	332 (47.4)	65 (52.4)	173 (39.2)
Adiponectin, µg/ml	9.7 (8.0)	5.3 (3.8)	3.7 (2.7)	7.2 (5.0)	7.0 (5.0)

Table 1 | Characteristics of the five cohorts included in the adiponectin EWAS meta-analysis. Values are shown as mean ± standard deviation for age (in years), as median (IQR) for adiponectin levels (in µg/mL) and body mass index (BMI; in kg/m^2), and as total (%) for sex and smoking status.

Characteristic	KORA F4	LLS	LLD	TwinsUK	SHIP-TREND
Sample size	1702	723	701	94	441
Age, years	60.9 ± 8.9	58.9 ± 6.7	45.5 ± 13.1	55.1 ± 11.9	50.0 ± 13.4
Sex, female	874 (51.4)	372 (51.4)	411 (58.6)	94 (100.0)	241 (54.6)
BMI, kg/m^2	27.4 (4.6)	25.1 (4.3)	24.7 (4.8)	25.0 (4.3)	27.2 (4.2)
Smoking, current	247 (14.5)	85 (11.8)	130 (18.5)	12 (12.8)	154 (34.9)
Smoking, never	711 (41.8)	199 (27.5)	332 (47.4)	50 (53.2)	173 (39.2)
Leptin, ng/ml	13.3 (19.9)	12.1 (20.6)	10.0 (17.0)	15.1 (12.1)	10.1 (14.7)

Table 2 | Characteristics of the five cohorts included in the leptin EWAS meta-analysis. Values are shown as mean ± standard deviation for age (in years), as median (IQR) for leptin levels (in ng/mL) and body mass index (BMI; in kg/m^2), and as total (%) for sex and smoking status.

Circulating levels of adiponectin and leptin were associated with blood-based DNAm at 73 CpG sites and 621 CpG sites, respectively ($p_{FDR} \leq 0.05$, nominal *p*-value thresholds 8.8e−6 for adiponectin, 7.6e−5 for leptin). These results were adjusted for age, sex, technical covariates, and six immune cell types predicted from DNAm data (granulocytes, monocytes, NK cells, CD4+ T cells, CD8+ T, cells and B cells). No CpGs displayed high heterogeneity between cohorts ($I^2 < 80\%$) and test statistics were corrected for bias and inflation.

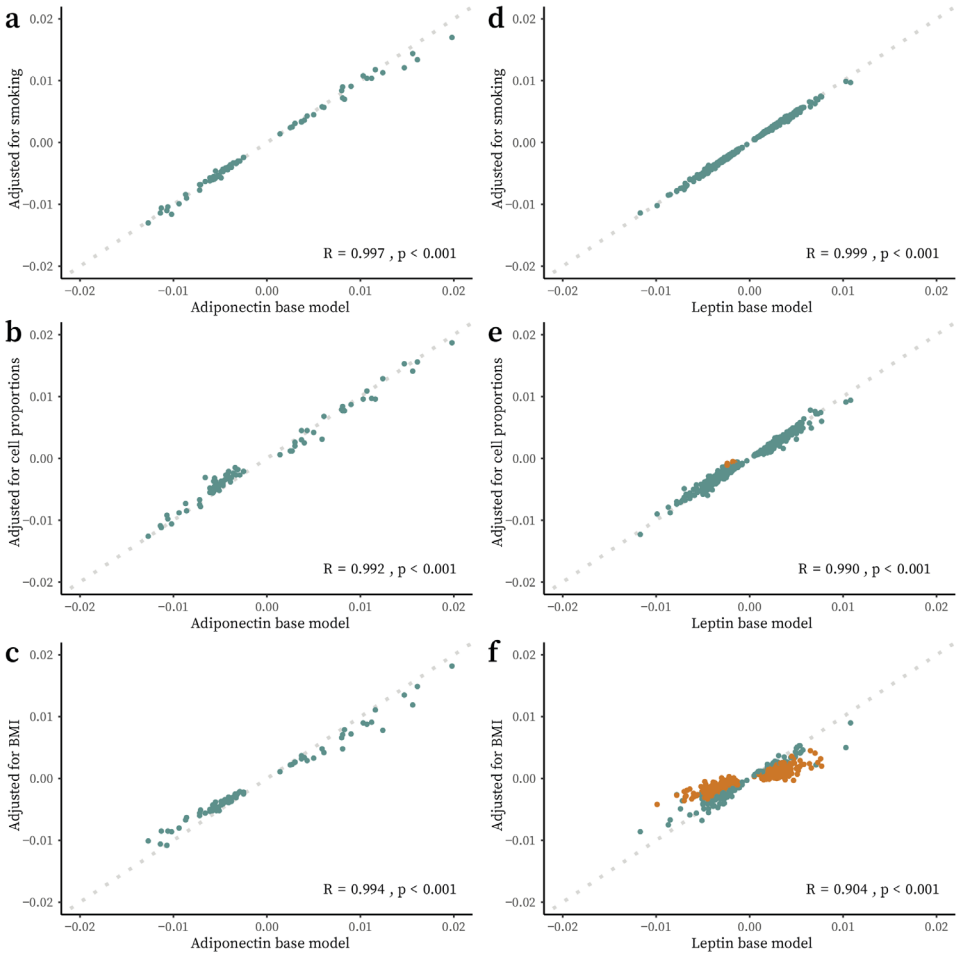


Figure 1 | Relationships between adipokine-associated DNAm effects before and after sensitivity analyses with Pearson correlation coefficients and associated *p*-values. Each point represents a CpG coloured by if they were taken forward in downstream analyses (blue) or removed (orange). A dotted reference line (y = x) indicates no change between models. **a)** Adiponectin-associated effects before and after adjustment for smoking, **b)** for twelve predicted blood-cell-type proportions, **c)** and for body mass index (BMI). **d)** Leptin-associated effects before and after adjustment for smoking, **e)** for twelve predicted blood-cell-type proportions, **f)** and for BMI.

To evaluate the stability of associations between DNAm and adipokines, base models were further adjusted for smoking status, twelve distinct cell types, and body mass index (BMI). For all of the adiponectin CpGs, associations remained statistically significant in these sensitivity analyses ($p_{FDR} \leq 0.05$), with strong correlations between effect sizes before and after adjustment ($R > 0.99$, $p < 0.001$; **Fig. 1a-c**).

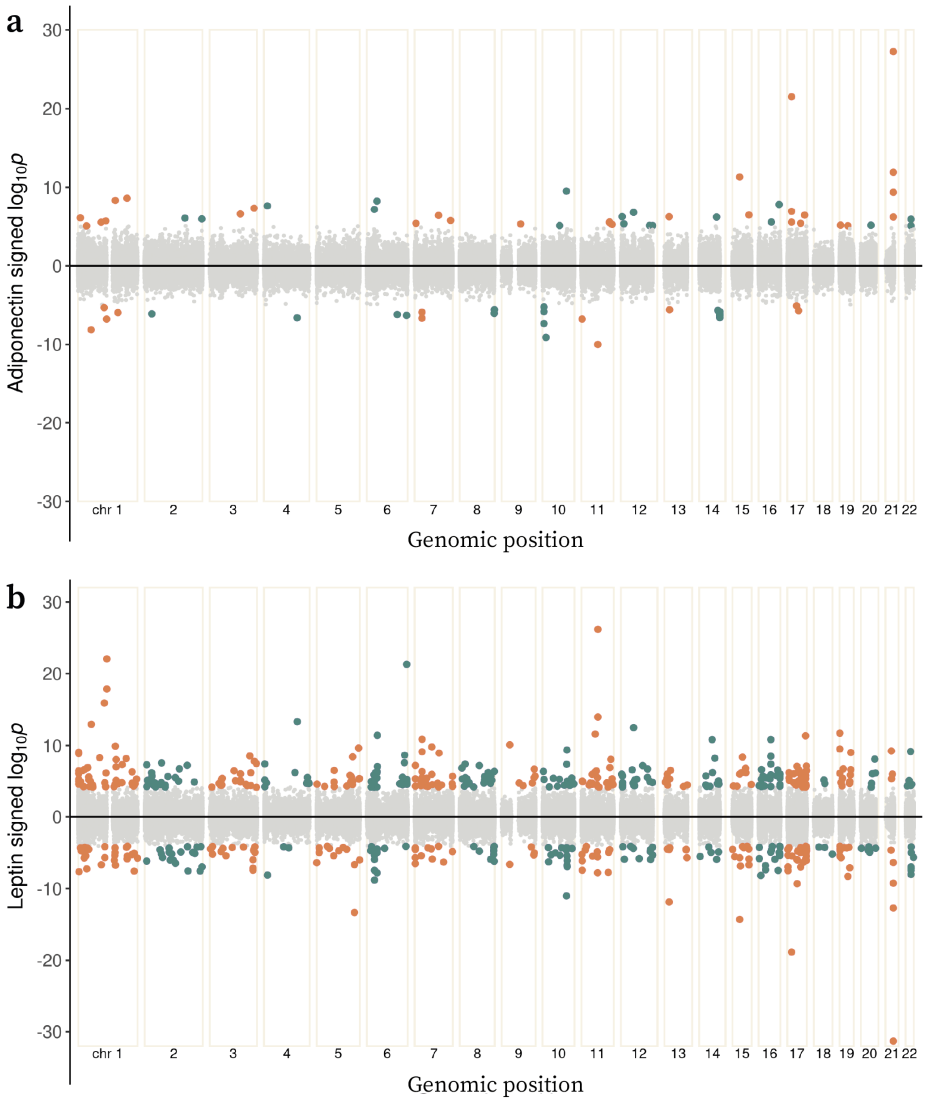


Figure 2 | Bidirectional Manhattan plots of the signed log_{10}p-values for all tested CpGs against their genomic position. CpGs significant at the 5% level after adjusting for multiple testing are shown in orange (odd-numbered chromosomes) or blue (even). Non-significant CpGs are shown in grey. **a)** Results from the adiponectin EWAS meta-analysis, and **b)** from the leptin EWAS meta- analysis.

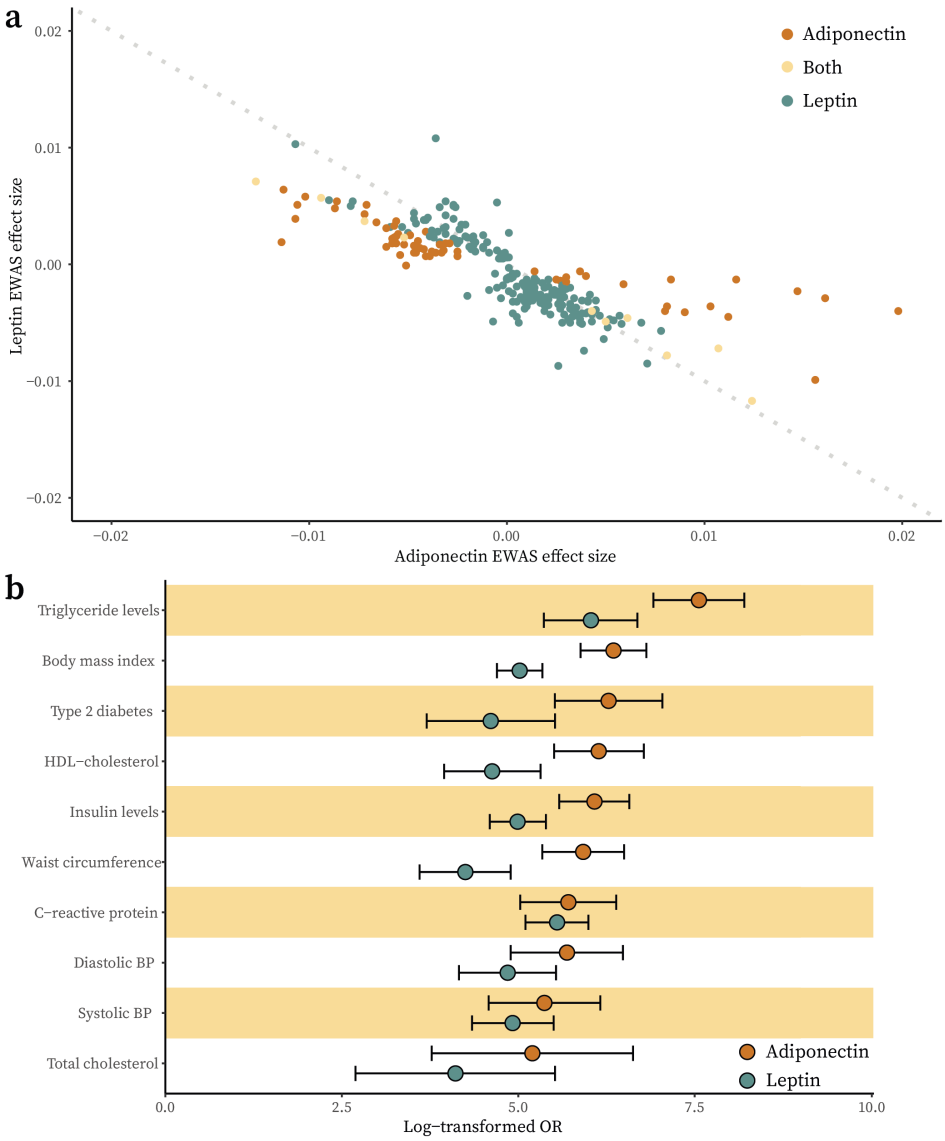


Figure 3 | Relationships between EWAS of different traits. a) Scatter plot of EWAS meta-analysis effect sizes for adiponectin and leptin for all CpGs significant at the 5% level in one or both analyses. A dotted reference line (y = −x) indicates perfectly inverse correlation between the two adipokines. **b)** Forest plot of trait enrichment in the adipokine-associated CpGs showing natural log-transformed ORs and their 95% CIs. Traits that are in the top ten (as determined by OR) for one or both adipokines are included, ordered by maximum OR. All enrichments shown are significant at the 5% level after adjusting for multiple testing.

Effects at leptin CpGs also showed relative independence from smoking and cell-type proportions (**Fig. 1d,e**). Unsurprisingly however, since leptin has stronger and more direct links to obesity[11,12], 401 leptin CpGs were sensitive to BMI adjustment ($p_{FDR} \geq 0.05$; **Fig. 1f**). To ensure focus on adipokine-specific epigenetic links in downstream analyses, these were removed from the results.

The final set of CpGs included 73 adiponectin (**Fig. 2a**) and 211 leptin-associated sites (**Fig. 2b**), representing 65 and 203 distinct loci, respectively. Ten CpGs were associated with both adipokines, and adiponectin and leptin effect sizes were inversely correlated at the 274 uniquely identified CpGs ($R = -0.81$ $p < 0.001$; **Fig. 3a**).

Adipokine-associated methylation is robustly connected to metabolic health

To assess the relevance of the adipokine-associated CpGs to human health and disease, we conducted a search of previous EWAS results (**Supplementary Table 3,4**). Notably, 65 of the 73 adiponectin CpGs (89.0%) and 145 of the 211 leptin CpGs (68.7%) were also associated with at least one other trait. As anticipated *a priori*, adipokines and adiposity were closely related (BMI ORs 571.7 and 151.3 for adiponectin and leptin, respectively; $p_{FDR} < 0.001$). In total, 40 (54.8%) of the adiponectin and 51 (24.2%) of the leptin CpGs were associated with BMI in large-scale blood-based EWAS. This was despite the removal of CpGs no longer significant after adjusting for BMI, supporting shared but independent epigenetic signals underlying BMI and the relevant adipokine at the remaining CpGs. Enrichments also existed for other metabolic risk factors, including HDL-cholesterol (ORs 465.8 and 102.8, $p_{FDR} < 0.001$), triglycerides (ORs 1917.4 and 413.7, $p_{FDR} < 0.001$), systolic BP (ORs 215.2 and 137.4, $p_{FDR} < 0.001$), fasting insulin (ORs 435.5 and 147.5, $p_{FDR} < 0.001$) and glucose levels (ORs 121.4 and 22.0, $p_{FDR} < 0.001$), as well as T2D itself (ORs 533.8 and 100.6, $p_{FDR} < 0.001$), highlighting the relevance of our CpGs to metabolic health as a whole (**Fig. 3b**).

Functional genomics uncovers regulatory potential for adipokine-associated CpGs

We annotated the genomic positions of the 73 adiponectin- and 211 leptin-associated CpGs to 15 chromatin states using Roadmap reference epigenomes[13]. These consist of eight active and seven repressed configurations showing distinct levels of DNAm, accessibility and regulator binding. By testing if adipokine CpGs were enriched for any particular genomic feature in the PBMC reference (E062), we revealed that active chromatin annotations were over-represented and repressive states depleted in our results (**Fig. 4a, Supplementary Table 5,6**).

Both adipokines displayed enhancer enrichment, with 18 adiponectin (24.7%, OR = 8.82) and 33 leptin CpGs (15.6%, OR = 4.87) annotated to enhancers, compared with only 3.6% of total tested CpGs in the adiponectin (n = 14,706) and leptin (n = 14,548)

analyses. This indicated high probability of colocalisation with markers of open chromatin, specifically H2K4me1[14]. To investigate whether this pattern was cell-type specific, enrichment was analysed using reference epigenomes for 22 other immune cell types (Roadmap Epigenomes: E029–48, E050–51) and adipocytes (E063). In all tested epigenomes except one (Treg for adiponectin), CpG genomic locations were enriched for enhancers ($p_{FDR} \leq 0.05$) demonstrating robust regulatory potential across immune and adipose cellular identities.

As the primary mechanism by which DNAm influences nearby expression is by modulating TF binding[15], we tested regions within 50 bp of the adipokine-associated CpGs for TFBS enrichments, revealing links to 14 distinct TFs (**Fig. 4b, Supplementary Table 7**). Several of these were central to immunity and inflammation (e.g. BATF and AP-1)[16-18], while others had specific adipokine relevance, including Fosl2, which promotes leptin expression[19], and MafA, which downregulates adiponectin[20]. Taken together, these functional analyses supported adipokine-related DNAm as occurring at *cis*-regulatory regions with plausible biological relevance.

Integrative analyses relate adipokine-associated CpGs to metabolic gene expression

Correlations between DNAm and expression of nearby genes (±100 kb) was tested using blood-based data from the BIOS consortium (n = 3,152; **Supplementary Table 8,9**). Of the 1,069 tested CpG-gene pairs, 21.2% were connected in this analysis (n = 227, $p_{FDR} \leq 0.05$), with the majority representing inverse relationships (71.5%) in line with previous reports[21]. Thirty-five (47.9%) adiponectin CpGs were associated with expression of 46 genes (**Fig. 4c**), and 100 (47.4%) leptin CpGs were linked to 151 genes (**Fig. 4d**). Of the identified CpG-gene pairs, almost one in six involved the nearest gene in both the adiponectin (15.2%, n = 7) and leptin (15.9%, n = 24) analyses. Additionally, DNAm at two distinct CpGs (cg11851174 and cg20544516) was associated with *SREBF1*, a key regulator of lipid homeostasis[22]. In total, there were eight genes overlapping between leptin and adiponectin analyses, and several of these were central to lipid transport (e.g. *ABCG1*)[23], biosynthesis (e.g. *DHCR24*)[24], and metabolism (e.g. *CPT1A*)[25].

Biological roles for the implicated genes were interrogated using over-representation analysis. Of 16,037 gene sets tested, 79 were enriched in the 46 genes linked to adiponectin DNAm, and 15 were enriched in the 151 leptin genes (**Supplementary Table 10,11**). Findings for both adipokines highlighted links with lipid metabolism (e.g. *cholesterol metabolism* and *metabolism of lipids*). Almost half (n = 7, 46.7%) of the leptin gene sets contained the words terms *metabolic* or *metabolism*, but this pattern was reduced in the adiponectin-related terms (n = 10, 12.7%). Adiponectin gene sets were more closely linked to cellular reprogramming, including via AMPK (p_{FDR} = 3.4e−3) and mTORC1 (p_{FDR} = 9.7e−3) signalling. Overall, these pathway-level results highlighted importance for identified genes in metabolic molecular processes and regulation.

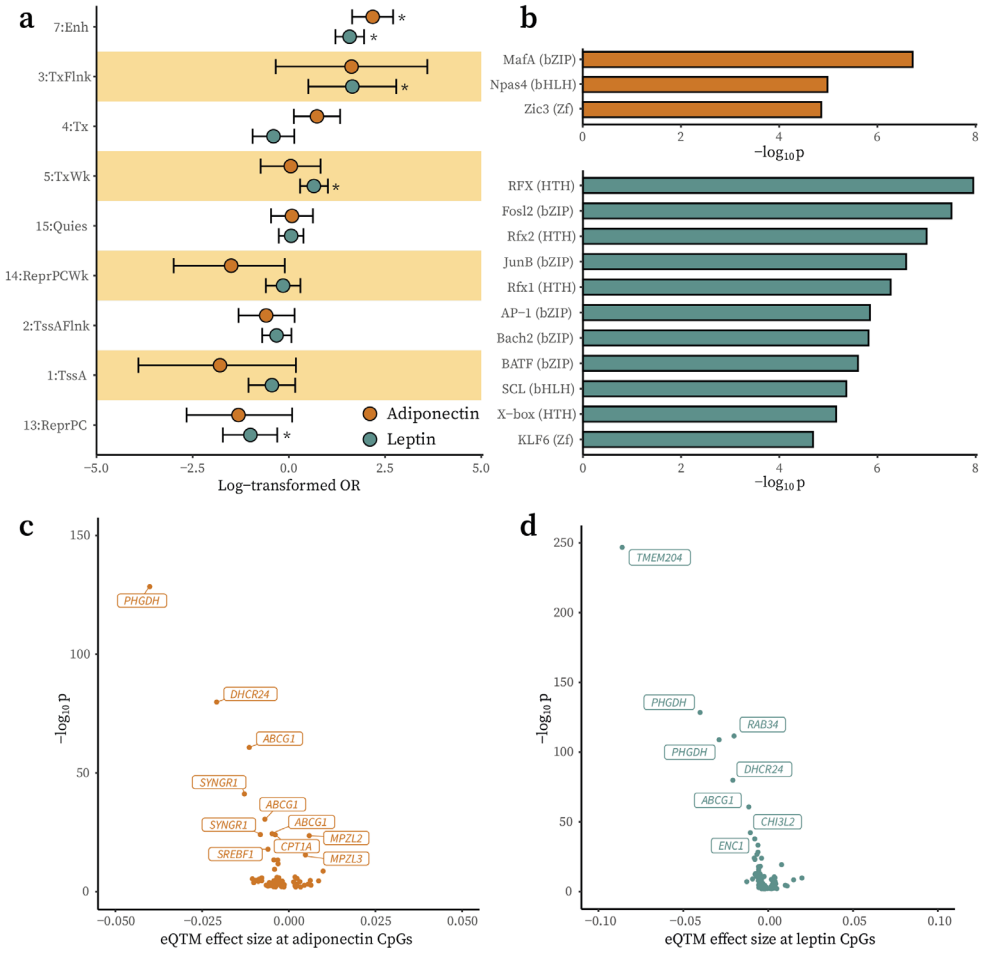


Figure 4 | Regulatory enrichment and integrative analysis results. a) Forest plot of chromatin state enrichments in adiponectin (orange) and leptin (blue) associated CpGs, identified using the PBMC Roadmap reference epigenome (E062). Natural log-transformed ORs and 95% CIs are shown. Six states with very large CIs for one or both adipokines are not shown (12_EnhBiv, 11_BivFlnk, 10_TssBiv, 9_Het, 8_ZNF/Rpts and 6_EnhG). Stars indicate significance at the 5% level after adjusting for multiple testing. **b)** Bar plot showing enriched TFBS motifs and -$\log_{10}$ *p*-values for adiponectin (orange) and leptin (blue) associated CpGs. Regions within 50bp of the CpG sites were scanned using HOMER and compared to a GC-matched random genomic background. All TFBS shown are significant at the 5% level. **c)** Volcano plot of the relationships between expression quantitative trait methylation (eQTM) effect sizes and -$\log_{10}$ *p*-values. DNAm at the adiponectin-associated CpGs and normalised expression levels of genes within 100 kb was investigated in the BIOS consortium. **d)** Volcano plot of the relationships between eQTM effect sizes and -$\log_{10}$ *p*-values for leptin-associated CpGs.

Causal inference supports DNA methylation as driving adiponectin production

Ascertaining the directionality of relationships in EWAS is far from straightforward. To shine light on the most plausible sequence of events, genetic variants were used as proxies for adipokine and DNAm exposures. In line with previous EWAS reporting, we performed bidirectional 2SMR followed by triangulation analysis[5,26]. 2SMR predicts the causal effect of an exposure on an outcome by combining genetically-determined levels of both, using GWAS or quantitative trait loci (QTL) databases. Triangulation expands upon these directional inferences and assumes that, if genetically-determined outcome levels (*observed effects*) are driven by an exposure, then they can be predicted by combining genetically-determined exposure and exposure–outcome associations (the *predicted effects*). The correlation between *observed* and *predicted effects* then quantifies the combined support for a causal direction, even if there is insufficient power at the individual CpG level. By performing both analyses bidirectionally, we comparatively inferred which direction of effect was most strongly supported by the data.

Previous EWAS have frequently concluded that blood-based DNAm is a consequence rather than a cause of investigated traits[5,8,26,27]. Therefore, we explored whether adipokine levels could be driving DNAm at identified CpGs. 2SMR did not suggest that methylation was caused by either adiponectin or leptin, and triangulation consolidated this finding with minimal correlations between *observed* (PGS-DNAm) and *predicted* (PGS-adipokine/adipokine-DNAm) effects ($R < 0.02$; **Fig. 5a,b**). In the reverse direction, 2SMR supported DNAm at two CpGs influencing adiponectin levels. The first, cg11851174 (chr17:17712609), was associated with incident T2D[26] and annotated to active chromatin in both PBMCs (4:Tx) and adipocytes (7:Enh). In the blood-based integrative analyses, this site was linked to *SREBF1* ($\beta = -0.004$, $p_{FDR} = 8.3e{-}5$), which encodes a TF central to lipid homeostasis and biosynthesis and whose expression is decreased in obesity and T2D[28].

The second CpG that putatively drove adiponectin production (cg02235049, chr3:186559186), was a novel site not previously identified in EWAS but strikingly annotated to *ADIPOQ*, the gene encoding adiponectin. Integrative follow-up into its methylation and nearby *ADIPOQ* expression was hindered in the BIOS consortium as adiponectin is not produced by immune cells. However, in publicly available DNAm and expression data from SGBS pre-adipocytes ($n = 38$), this CpG was negatively correlated with *ADIPOQ* expression ($R = -0.36$, $p = 0.029$) expression. This inverse relationship aligned with 2SMR results ($\beta = -0.217$, $p_{FDR} = 2.1e{-}12$) and a functionally repressive effect of DNAm on *ADIPOQ* expression at this adipocyte-specific enhancer.

The 2SMR direction of effect at both CpGs, where DNAm influences adiponectin, was also supported by triangulation analysis (**Fig. 5c,d**), with *observed* and *predicted effects* correlated for adiponectin ($R = 0.335$, $p = 0.030$) but not for leptin ($R = -0.017$, $p = 0.837$).

Taken together, these findings indicated a cell-type specific effect for the two CpGs identified as putative drivers of adiponectin, with evidence of links to expression for cg02235049 and cg11851174 in adipocytes and leukocytes, respectively.

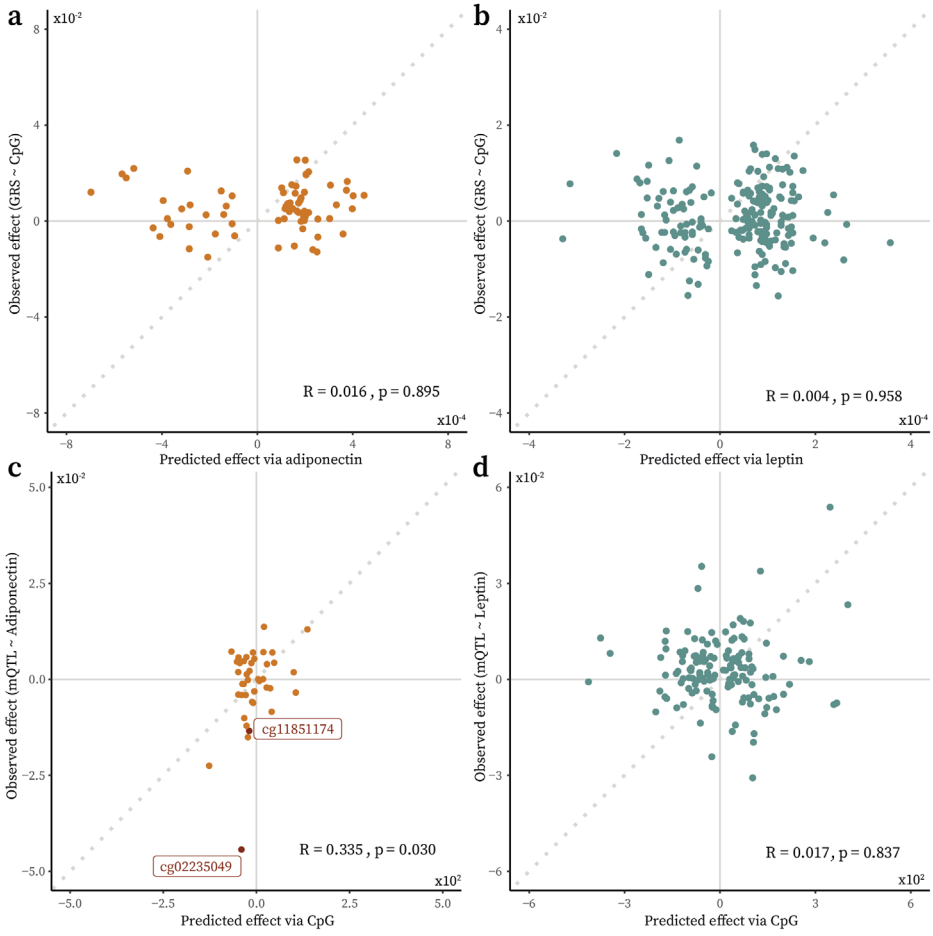


Figure 5 | Triangulation analysis results showing correlations between *predicted* (through an exposure) and *observed* genetic effects on an outcome. Pearson correlation coefficients and their associated *p*-values are shown. **a)** *Predicted* (via adiponectin) and *observed* (PGS-CpG associations) effects for the influence of genetically-determined adiponectin on DNAm at adiponectin-associated CpGs. **b)** *Predicted* (via leptin) and *observed* (PGS-CpG associations) effects for the influence of genetically determined leptin on DNAm at the leptin-associated CpGs. **c)** *Predicted* (via DNAm) and *observed* (mQTL-adiponectin) effects for the influence of DNAm at the adiponectin-associated CpGs on serum adiponectin. Two CpGs significant from the 2SMR analysis are labelled in red. **d)** *Predicted* (via DNAm) and *observed* (mQTL-leptin) effects for the influence of DNAm at the leptin-associated CpGs on leptin.

DNAm influencing adiponectin is also upstream of metabolic risk and disease

Evidence that these two CpGs (cg11851174 and cg02235049) were more likely a cause than a consequence of adiponectin combined with their links to nearby expression in relevant cell -types (*SREBF1* in blood and *ADIPOQ* in adipocytes, respectively; **Fig. 6a**), prompted deeper analysis into their clinical significance. In particular, for the *SREBF1* CpG (cg11851174) there were multiple lines of evidence pointing towards functional regulation, including previous EWAS, genomic annotation, integrative links, and causal inference (**Fig. 6b**).

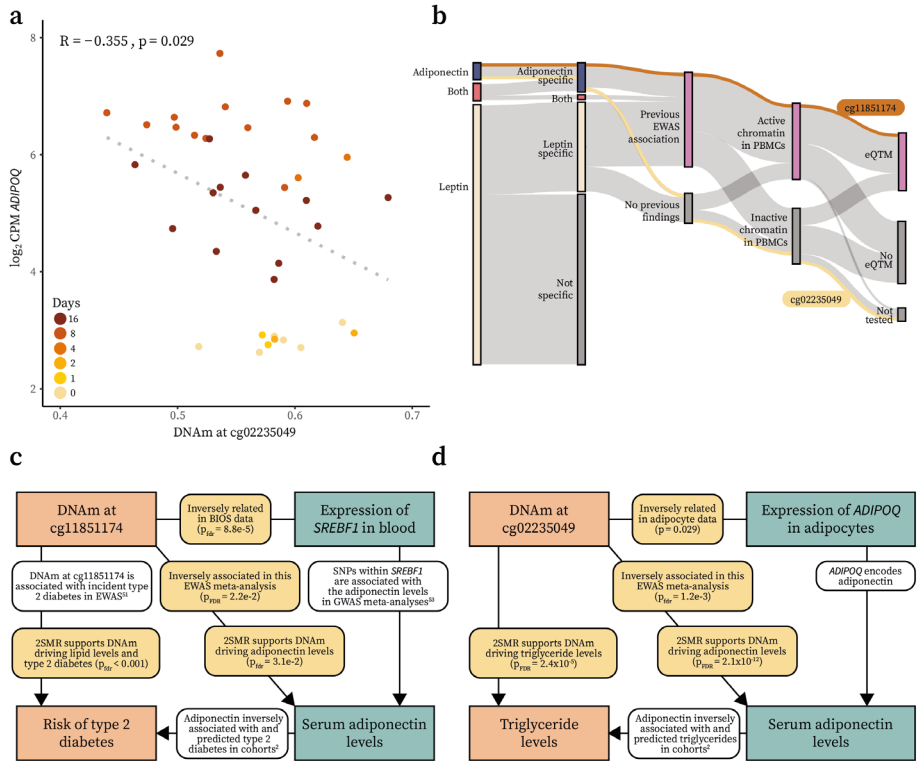


Figure 6 | Combined evidence that DNAm at two loci influences metabolic health. a) Scatter plot of DNAm at cg02235049 and normalised *ADIPOQ* expression levels in SGBS preadipocytes (n = 38). Pearson correlation coefficients and a line of best fit is shown. **b)** Sankey diagram showing multiple downstream investigations into adipokine-associated CpGs. cg11851174 is shown as having evidence supporting its functional relevance in all of these investigations (orange), and cg02235049 is shown as not having such collective evidence (yellow). **c)** Flowchart outlining the evidence linking DNAm at cg11851174 to T2D via *SREBF1* and adiponectin: findings from the current study (yellow); increases in traits (DNAm and T2D risk; orange); decreased traits (*SREBF1* expression and adiponectin; blue); and evidence from previous work (white). **d)** Flowchart showing evidence linking DNAm at cg02235049 to triglyceride levels via *ADIPOQ* and serum adiponectin.

Using 2SMR, we evaluated causal links between these two CpGs, T2D and several metabolic risk factors including fasting insulin and lipid levels. Methylation of the *SREBF1* CpG (cg11851174) plausibly decreased HDL-cholesterol levels (p_{FDR} = 4.11e−3) and increased both fasting insulin (p_{FDR} = 3.84e−3) and T2D risk (p_{FDR} = 2.39e−5). Additionally, both the *SREBF1* and *ADIPOQ* CpGs likely drove triglyceride level increases (p_{FDR} = 2.94e−2 and p_{FDR} = 2.39e−5 for cg11851174 and cg02235049, respectively). In the reverse direction, there was insufficient evidence for DNAm at either CpG responding to investigated metabolic traits or T2D itself. These results showcased these specific loci as upstream epigenetic markers of both adiponectin production and T2D pathogenesis. Coupled with evidence that *SREBF1* expression decreases T2D risk by increasing serum adiponectin[28-30] and the clear relevance of *ADIPOQ* as the gene encoding adiponectin, there is now considerable support for direct regulatory potential at these two loci in human health and metabolic risk (**Fig. 6c,d**).

Discussion

In this study, we performed an EWAS of associations between circulating adipokines and genome-wide DNAm in five blood-based cohorts (adiponectin n = 2,791; leptin n = 3,661). Through sensitivity analyses for cell-type proportions, smoking and BMI, we derived two sets of CpGs robustly associated with adiponectin (n = 73) and leptin (n = 211). Methylation at these CpGs was associated with both T2D and metabolic risk factors, including BMI, fasting insulin and HDL-cholesterol. Additionally, integrative analyses linked adipokine-associated DNAm to expression of genes central to lipid transport (e.g. *ABCG1*)[23], biosynthesis (e.g. *DHCR24*)[24], and metabolism (e.g. *CPT1A*)[25].

Bidirectional 2SMR and triangulation did not indicate a causal relationship between DNAm and leptin in either direction, but did support two CpGs as potentially regulating adiponectin production. The first of these was a novel CpG not previously identified in blood-based EWAS. Annotated to *ADIPOQ*, the gene encoding adiponectin, this CpG lies in a repressed chromatin region in PBMCs, making functional relevance for its methylation in leukocytes unlikely. Indeed, *ADIPOQ* was not expressed in the blood-based BIOS consortium data precluding integrative follow-up of this site in blood. Still, functional genomics data revealed that this CpG resided in an adipocyte-specific enhancer and, since adipocytes are the primary producers of adiponectin, this represented a biologically plausible cell-type specific effect. We therefore investigated the relationship between this CpG and *ADIPOQ* expression in adipocyte data and observed a significant, inverse correlation that aligned with our causal inference results.

This discovery underscores the potential for blood-based epigenetic analyses in large-scale cohorts to identify biologically relevant sites, even where their functional roles may be in less accessible tissues. Such associations could be driven by shared upstream factors, such as diet, causing DNAm in a tissue-agnostic manner. Although these CpGs may only be functional in some tissues (e.g. adipose), their ability to be detected in blood allows well-powered EWAS such as this one to identify biologically meaningful correlations. These sites, and others like them, will ideally serve as focal points for targeted hypothesis-driven investigations into adiponectin production by, for example, experimentally modifying methylation in adipocytes.

The second CpG plausibly driving serum adiponectin levels (cg11851174) resides in active chromatin in both PBMCs and adipocytes close to the *SREBF1* gene, which encodes a TF central to lipid homeostasis that binds to sterol regulatory elements in the promoters of genes including *ADIPOQ*[22,30]. Therefore, similar to the *ADIPOQ* CpG, this *SREBF1* site could represent a functional epigenetic effect in adipocytes being mirrored in blood. However, as we also linked increased DNAm at this site to decreased *SREBF1* expression in the large-scale blood-based BIOS consortium data, there is another plausible explanation for how DNAm could act upstream of adiponectin production. Previous experimental evidence from macrophage-specific sterol regulatory

element-binding protein (SREBP) cleavage-associating protein (SCAP) knockout mice has demonstrated that reduced SREBP-1a activity promotes macrophage polarisation to proinflammatory subtypes[31]. Therefore, circulating monocytes epigenetically primed for lower *SREBF1* expression, could feasibly have proinflammatory cell fates as adipose-tissue macrophages. Since local inflammation is an established inhibitor of adiponectin production from adipocytes, this represents a sequence of events where immune cell DNAm could more directly influence adipokine production[32]. These two hypotheses could be tested in differentiating monocytes or macrophage-adipocyte co-cultures, and our findings offer an indication of plausible mechanisms to follow-up.

Considering the collective evidence at the *ADIPOQ* and *SREBF1* loci, we investigated broader implications of these sites for metabolic disease using bidirectional 2SMR. This indicated that these CpGs may also act upstream of metabolic traits, including triglyceride levels and T2D. These directional associations, coupled with previous work implicating *SREBF1* in T2D risk via adiponectin[28-30] and the plausible relevance of *ADIPOQ* for adiponectin, reinforce these CpGs as promising sites of epigenetic control.

There were limitations to our study. Notably, we explored relationships between leukocyte DNAm and serum adiponectin with only minimal follow-up in adipocytes. Additionally, we did not adjust for smoking in our main analysis due to incomplete data, and instead opted to ensure smoking-independent effects via a two-step sensitivity analysis restricted to the subset with complete data. While sex was included as a covariate to adjust for potential confounding, no sex-stratified analyses were performed. Future research could explore whether these findings apply equally across sex and gender groups. This study was also conducted in European populations, and it remains to be tested whether our findings can be generalised to other ethnicities. Lastly, this study was not immune to the common weaknesses of molecular 2SMR. For the adipokine-DNAm 2SMR, data was not available to test all CpGs, meaning that only 57.5% of the adiponectin-associated CpGs and 61.6% of the leptin-associated CpGs were followed up in this analysis, and not all independent SNPs were available in the mQTL and GWAS datasets. This limited, and could have biased, instruments used for these exposures. In addition, most mQTLs with strong effects lie in close proximity to each other and are highly correlated. Only between one and four independent mQTLs existed to instrument each CpG and this 2SMR approach was expected to have limited success in identifying directional effects with bias towards the null[34].

In summary, this study highlights the potential of integrative, epigenome-wide studies to uncover biologically coherent epigenetic markers of molecular and metabolic traits, and reveals novel insights into the regulatory mechanisms underlying adiponectin production. By highlighting critical loci, we offer focal points for future experimental research aiming to dissect the secretory profiles of adipocytes or to identify therapeutic targets in metabolic disease.

Methods

Cohort descriptions

Leiden Longevity Study (LLS) The LLS is a cohort of long-lived, Dutch, Caucasian siblings (*n* = 944), who were recruited with their offspring (*n* = 1,671) and their offspring's partners (*n* = 744)[35]. Between 2002 and 2006, research nurses collected non-fasted blood samples from living study participants for isolation of DNA, RNA, serum, and plasma, and these were stored at -80˚C until analysis. From these samples, adiponectin levels were determined with the DuoSet ELISA kit (R&D Systems Europe Ltd, Abingdon, United Kingdom) according to the manufacturer's instructions. The lower and upper detection limits of the assay were respectively 62.5 and 4000 pg/ml. Serum leptin was measured using 'two-step' sandwich ELISA kit (Diagnostics Biochem Canada Inc., Dorchester, Ontario, Canada) according to manufacturer's instructions. Assay sensitivity was 0.5 ng/ml.

DNAm data of whole blood samples was generated from 821 unrelated participants from the offspring-partner generation of the LLS by the Human Genotyping facility (HuGe-F, Erasmus MC, Rotterdam, The Netherlands) within the Biobank-Based Integrative Omics Studies (BIOS) consortium, funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007). Genomic DNA (500ng) was bisulfite converted using the Zymo EZ-96 DNA methylation kit (Zymo Research Corp, Irvine, CA, USA). 4 µl was then hybridized on the Infinium HumanMethylation450 BeadChip array (Illumina Inc, San Diego, CA, USA) according to the manufacturer's protocol. Data preprocessing and quality control followed the *DNAmArray* workflow[36].

Kooperative Gesundheitsforschung in der Region Augsburg (KORA) The KORA F4 study (2006-2008) followed up individuals (*n* = 3,080) living in the region of Augsburg, Southern Germany, aged between 32 and 81 years old[37]. Blood samples were drawn after an 8 hour fast and frozen at -80˚C until analysis. Serum leptin concentrations were determined using ELISAs from Mecodia, Stockholm, Sweden. The intra- and inter-assay coefficients of variation for leptin were <10%. Serum adiponectin levels were measured in the subsample aged 61 to 82 years using the Quantikine ELISA kit from R&D Systems (Wiesbaden, Germany)[38]. The intra- and inter- assay coefficients of variation were 3.8% and 8.0%, respectively.

Following isolation according to standard procedures, genomic DNA (750 ng) from 1,707 whole blood samples was bisulfite converted with the Zymo EZ-96 DNA Methylation Kit (Zymo Research Corp, Irvine, CA, USA). 4 µl from each sample underwent amplification, enzymatic fragmentation, and application to the Infinium HumanMethylation450 BeadChip array (Illumina Inc, San Diego, CA, USA) according to the manufacturer's protocol[39]. GenomeStudio with the methylation module was used to extract and process the raw image data and initial quality assessment was conducted using the Control Dashboard. For data pre-processing, a common pipeline was used[40].

TwinsUK Fasting morning serum total adiponectin levels were measured with a two-site DELFIA assay using antibodies and standards from R&D Systems (Minneapolis, MN). The day-to-day coefficient of variation (CV) for adiponectin was 9.9% at a concentration of 3.2 ng/ml, 7.8% at 8.6 ng/ml, and 5.2% at 14.7 ng/ml. Serum leptin concentration was determined after an overnight fast using an RIA (Linco Research, St. Louis, MO).

As described previously, fasting whole blood DNAm of individuals in TwinsUK was profiled using the Infinium HumanMethylation450 BeadChip (Illumina Inc, San Diego, CA)[41]. DNAm was assessed at over 450,000 sites and processing of methylation signals was performed with R Bioconductor software[42]. Briefly, the ENmix package[43] was used for quality control of the data, and the *minfi* package was used to exclude samples with median methylated and unmethylated signal ratio less than 10.5[44]. Background correction, dye bias correction, and quantile normalization were performed with ENmix as described previously[45]. Underperforming probes and outlier samples were identified using standard parameter values and signals with detection *p*-value above 1e-6 and number of beads fewer than three were excluded. Maximum probe and sample missingness were set to 5%.

The Study of Health in Pomerania (SHIP-TREND) The Study of Health in Pomerania – Trend (SHIP-TREND) is a longitudinal population-based cohort study in West Pomerania, a region in the north-east of Germany, which assesses the prevalence and incidence of common population-relevant diseases and their risk factors. Baseline examinations for SHIP-TREND were carried out between 2008 and 2012, comprising 4,420 participants aged 20 to 81 years. Study design and sampling methods were previously described[46].

DNA was extracted from blood samples (n = 508) of SHIP-TREND participants to assess DNA methylation using the Illumina HumanMethylationEPIC BeadChip array. Samples available for this project were randomly selected based on the availability of multiple OMICS data, and excluding T2D. The samples were taken between 7:00 a.m. and 4:00 p.m., and serum aliquots were prepared for immediate analysis and for storage at -80 °C in the Integrated Research Biobank (Liconic, Liechtenstein). Processing of the DNA samples was performed at the Helmholtz Zentrum München. Preparation and normalization of the array data was performed according to the CPACOR workflow[47]. The IDAT files were processed using the *minfi* package[44]. Probes that had a detection p-value above background (sum of per-array methylated and unmethylated intensity valuesbased p-value ≥ 1e-16) were set to missing. Methylation β values were calculated as proportion of methylated intensity value on the sum of methylated + unmethylated + 100 intensities. Arrays with observed technical problems (±4SD outside control probe intensity mean) during steps like bisulfite conversion, hybridization or extension, as well as arrays with mismatch between sex of the proband and sex determined by the chr X and Y probe intensities were removed from subsequent analyses. Additionally, only arrays with a call rate above 95% were processed further, resulting in 495 samples with methylation data on 865,859 sites, and 441 samples with phenotype data available for subsequent analyses. In addition to sex, age, and white blood cell type counts, the first six principal components of the control probe intensities obtained by the CPACOR workflow were included in the model as covariates to account for unmeasured confounders. Details on assessment of the metabolic phenotypes and covariates used in this analysis are provided within the SHIP cohort design publication[46].

LifeLines DEEP (LLD) A total of 701 adults from the Lifelines DEEP cohort were included in this study based on available methylation data and cytokines. Initially, 1539 participants were enrolled in the Lifelines DEEP, which is a subpopulation of the Lifelines cohort in the north of The Netherlands[48,49]. Except for the regular Lifelines procedures, additional deep molecular measurements were performed in Lifelines DEEP participants. The plasma levels of leptin (ng/ml) and adiponectin (µg/ml) were measured by enzyme-linked immunosorbent assay (ELISA) platform.

For genome-wide DNA methylation data, 500 ng of genomic DNA was bisulfite-converted using the EZ DNA Methylation kit (Zymo Research Corp., USA) and hybridized on Illumina HumanMethylation450 BeadChip arrays (Illumina, Inc.) according to the manufacturer's protocols. The original IDAT files were generated by the Illumina iScan BeadChip scanner and performed by the Human Genotyping facility (HugeF) of ErasmusMC, The Netherlands (http://www.glimDNA.org/). Quality control and normalization details are described elsewhere[50]. Methylation levels at each CpG site were expressed as the ratios of the methylated intensity over the total intensity (β-values), which were used for the subsequent statistical analyses.

Cohort-based analyses

Main analysis All contacted cohorts with sufficient data followed a common analysis plan, and all samples analysed were from distinct individuals (i.e. there were no repeat measurements included in the analysis). DNAm was measured either by the Illumina Infinium HumanMethylation450 (in LLS, KORA, TwinsUK, and LLD cohorts) or MethylationEPIC BeadChip array (in SHIP-TREND cohort). Adipokine measurements below the limit of detection and outlying values for DNAm, adiponectin and leptin (more than three IQRs from the nearest quartile) were removed prior to analysis. Cell-type proportions were predicted from DNAm data using the IDOL algorithm[51].

For each of *j* CpGs measured in *i* individuals, a linear regression model (see Equation 1 for general specification) was fitted of DNAm β values on natural log-transformed adiponectin (μg/ml) or leptin (ng/ml). All models were adjusted for age (in years), sex, cell-type proportions predicted from DNAm data (monocytes, CD8+ T cells, CD4+ T cells, natural killer (NK) cells and B cells) and technical covariates (left to the analyst's discretion). Sex was considered in the study design and included as a covariate in statistical models to address potential biological differences. Analyses were not stratified by sex, and gender identity was not recorded.

$$DNAm_{\beta_{ij}} = \beta_0 + \beta_1 \cdot ln(adipokine)_i + \beta_2 \cdot age_i + \beta_3 \cdot sexFemale_i + \beta_4 \cdot CD8T_i + \beta_5 \cdot CD4T_i + \beta_6 \cdot NK_i + \beta_7 \cdot Mono_i + \beta_8 \cdot technical\_factors_i \quad [1]$$

Sensitivity analyses Effects of adjustment for smoking status on the relationship between adipokines and identified CpGs was investigated in a sensitivity analysis, where each cohort added trichotomous smoking to the cohort-specific base models as a categorical fixed effect. In some cases, this resulted in a reduction of the sample size as there was missingness in the smoking data. To distinguish BMI-independent signals, cohorts also ran an additional analysis adjusting for BMI (measured in kg/m^2).

All cohorts also investigated the effect of adjusting for extended cell types, estimated using the epiDISH Bioconductor package release 3.20, which became available in the timeframe of this project[52]. Basophils, memory B cells, naive B cells, CD4+ memory T cells, CD4+ naive T cells, CD8+ memory T cells, CD8+ naive T cells, eosinophils, monocytes, NK cells and regulatory T cells were added to the base model for all cohorts. Neutrophils were excluded to avoid collinearity as proportions for all cells sum to 1.

Meta-analyses

Differentially methylated probes Results from each cohort were inspected and rows were removed if they were estimated from fewer than 50 observations. Probes located on sex chromosomes, in ENCODE Blacklist regions[53], or that contained common genetic variants or that were ambiguously mapped[54] were also removed. To ensure good quality data, we inspected QQ, volcano and Manhattan plots, alongside boxplots of the effect size and SE distributions across cohorts. Following these steps, data were available on 412,224 CpGs from the base adiponectin model and 406,832 CpGs for the base leptin model.

The Bioconductor package *bacon* estimated and adjusted for bias and inflation of the test statistics, using default priors (α = 1.28, β = 0.36)[55]. After running bacon, inflation and bias were estimated at ~1.00 and within ±0.00 for all models, respectively. Bacon-adjusted effect sizes and SEs were used as input in a fixed-effects meta-analysis in *METAL* version 2011-03-25[56]. Separate analyses were performed for each of the base models and each extended model (adjusted for smoking, BMI and extended cell counts). Any CpGs for which there was evidence of high heterogeneity in effect sizes between cohorts ($I^2 \geq 80\%$) would have been removed but there were none. CpGs were regarded as significantly associated with the relevant adipokine if the false discovery rate (FDR)-adjusted *p* value was below 0.05, and only CpGs that still met this criterion in the additional sensitivity analyses for smoking, cell-type proportions and BMI were taken forward into downstream analyses.

Differentially methylated regions To assess distinct genomic loci associated with circulating adipokine levels, differentially methylated regions (DMRs) were identified using the *DMRfinder* algorithm[57], as implemented in the *DNAmArray* workflow version 2.1[36]. DMRs were defined as regions with at least three differentially methylated positions (DMPs) and an inter-CpG distance of less than 1 kb, allowing a maximum of three non-DMPs across a DMR. The number of distinct loci was calculated as the total number of DMPs minus the number of DMPs in DMRs plus the number of DMRs called by *DMRfinder*.

Functional annotation

Phenotype enrichment Using summary data from the EWAS catalogue[58] and EWAS atlas[59], our CpGs were investigated for previous associations with other phenotypes. Any EWAS meeting the following criteria was removed: without an associated PubMed ID; with a sample size under 500; that reported fewer than 100 CpGs in the respective database; missing nominal *p* values; not performed in adults; or not using whole blood or leukocyte samples. Traits were also recoded to ensure consistency between names, for example by combining EWAS of 'BMI' and 'body mass index'. This resulted in a list of 57 traits, which were tested for enrichment of associations with our CpGs using logistic regression.

Chromatin state enrichment Identified CpGs were annotated to chromatin state using immune cell and adipocyte Roadmap reference epigenomes[13]. Logistic regression models were fitted using the *glm* function in R to calculate and test ORs for each of the 15 chromatin states. Nominal p-values were adjusted for multiple testing using FDR and enrichments or depletions were assessed at a 5% significance threshold.

Transcription factor binding site (TFBS) enrichment A 50 bp window around identified CpGs was scanned using *findMotifsGenome.pl* from HOMER version 3.1 for enrichment of known motifs compared with a random genomic background matched for GC content[60]. ENCODE TFBS annotation for TFs and CpGs on the 450k array was used to further investigate the size of binding sites and distance from CpG to summit[54]. TFs associated with enriched TFBS were examined for links with metabolic health and, specifically, adiponectin and leptin pathways and interactions.

eQTM analyses Measurements of blood-based gene expression alongside DNAm from the same samples was available from the BIOS consortium (n = 3152). This dataset comprises six Dutch biobanks: the Cohort on Diabetes and Atherosclerosis Maastricht[61], LifeLines[48], LLS[35], Netherlands Twin Register[62,63], Rotterdam Study[64], and the Prospective ALS Study Netherlands[65]. After filtering out non-autosomal and lowly expressed genes, count data were transformed into $\log_2$ counts per million (CPM) using *edgeR*, and values for each gene were rank inverse normal (RIN)-transformed prior to analysis[66].

Genomic locations of human transcripts, exons, coding sequences and genes were imported from the Ensembl database using *makeTxDbFromEnsembl* from the GenomicFeatures Bioconductor package[67]. These were used to identify the nearest gene to each adipokine-associated CpG and to save a list of all genes within 100 kb of each CpG. To examine links between DNAm and gene expression, linear regression models were fitted with RIN-transformed $\log_2$CPM values as the response variable and methylation β values as the independent variable, adjusting for the effects of age, sex, technical covariates (row, plate, and flowcell) and twelve blood-cell counts predicted from DNAm using EpiDISH release 3.20[52].

For investigations into links between expression and DNAm in Simpson-Golabi-Behmel syndrome (SGBS) pre-adipocytes, publicly available data were downloaded from Gene Expression Omnibus (GEO) using GEOquery in R release 3.20. Data were available for the same samples, with expression profiled using the Illumina HumanHT-12 V4.0 expression BeadChip microarray and DNAm profiled using the Illumina Infinium HumanMethylation450 BeadChip array[68]. Count data were normalised to $\log_2$CPM values and values from probes interrogating *ADIPOQ* (ILMN_1775045) and *SREBF1* (ILMN_1663035, ILMN_1695378 and ILMN_2328986) were extracted. Additionally, β values from cg11851174 and cg02235049 were subset from the DNAm data. Complete information was available for 38 samples across five timepoints (days 0, 1, 2, 4, 8 and 16). Expression and DNAm values were plotted against one another for the relevant comparisons and correlation coefficients were calculated.

Over-representation analysis On the basis of the large-scale blood-based integrative analysis in BIOS, a list of CpGs for which there was evidence for epigenetic regulation of nearby gene expression in leukocytes was saved. The associated gene names were used as input for over-representation analysis using 11 recent (updated in the last 6 years) databases relating to human health and disease downloaded from *Enrichr* (BioPlanet 2019, Elsevier Pathway Collection, GeDiPNet 2023, GO Biological Process 2023, KEGG Human 2021, MSigDB Hallmark 2020, OMIM, PhenGenI Association 2021, PheWeb 2019, Reactome 2022 and WikiPathway Human 2021). These databases were imported into R and analyses were performed using the *Enrichr* function from clusterProfiler release 3.20[69]. *p*-values were FDR-adjusted for multiple testing and significance was assessed at the 5% level.

Causal inference

Two-sample Mendelian Randomization To assess the direction of effects between adipokines and DNAm at identified CpGs, the *TwoSampleMR* package was used to perform bidirectional 2SMR[70]. This instrumental variable (IV)-based method uses genome-wide association study (GWAS) summary statistics to infer whether a risk factor causally influences an outcome. 2SMR relies on several key assumptions, namely that instruments are relevant, independent and that there is no horizontal pleiotropy. To interrogate the effects of DNAm at our CpGs on adiponectin and leptin, we extracted *cis*-methylation quantitative trait locus (mQTL) data from the Genetics of DNA Methylation Consortium (GoDMC)[71] and combined these with summary statistics from recent, large-scale GWAS of both adiponectin[72] and leptin[73]. For some CpGs (42.5% for adiponectin and 38.4% for leptin), there was insufficient data available to interrogate the effects of DNAm at that CpG. For the remaining CpGs, between one and four independent SNPs with data on both their *cis*-association with DNAm and association with the relevant adipokine were used as instruments. These were combined using the Wald ratio (for single mQTL instruments) or inverse variance weighted (IVW) methods (for multiple, independent mQTLs).

To interrogate the influence of adipokine levels on DNAm at identified CpGs, independent GWAS variants from recent, large-scale analyses were used[72,73]. Of the 18 variants that could instrument adiponectin, there were *trans*-mQTL data in GoDMC available for four of them and, of the six variants that could instrument leptin, there was available data for one. Linkage disequilibrium (LD) proxies with $R^2 > 0.8$ for the remaining SNPs were downloaded from the NIH's LDlink tool[74], and GoDMC data[71] were extracted for these where available. This process identified two other instrumental SNPs that could instrument the adipokines, one for each, meaning that leptin was instrumented by two independent SNPs (rs8043757 and rs4665972) and adiponectin was instrumented by five independent SNPs (rs11023332, rs1108842, rs12051272, rs998584 and rs113086489). The GWAS summary statistics and mQTL effects were then combined using the IVW method and the *TwoSampleMR* package in R version 0.6.6. For all analyses, *p*-values were adjusted for multiple testing using the FDR method and potential causal effects were assessed at the 5% significance threshold.

The following cohorts were used to derive both mQTL effects in GoDMC and adiponectin and/or leptin GWAS effects and therefore had overlapping individuals in both the exposure and outcome datasets for the 2SMR analysis: Rotterdam Study (GoDMC 1472 samples, leptin GWAS 3932 samples); and TwinsUK (GoDMC 843 samples, adiponectin GWAS 968 and 1229 samples, leptin GWAS 5654 samples). Overall, the overlap was low considering that all three meta-analyses incorporated data from over 16 cohorts.

The *TwoSampleMR* package version 0.6.6 was also used to interrogate causal links between DNAm at CpGs and metabolic traits. CpGs were instrumented with independent *cis*-mQTLs obtained from GoDMC, and *ieugwasr* was used to extract MR instruments for the metabolic traits. Reference numbers for the investigated traits were as follows: type 2 diabetes (ebi-a-GCST006967); fasting insulin (ebi-a-GCST9002238); triglycerides (ieu-b-111); HDL-cholesterol (ieu-b-109); and BMI (ieu-b-40).

Triangulation analyses To perform triangulation analyses, we interrogated the correlation between the *observed effect* of an IV on an outcome (i.e. mQTL–adipokine or polygenic score (PGS)–DNAm associations) and the predicted effect via the exposure. This analysis assumes that if the effect of an exposure on an outcome is causal, it would be possible to predict the IV's effect on the outcome through a combination of its effect on the exposure and the exposure's association with the outcome (the *predicted effect*).

In detail, when looking at the effect of DNAm on adipokine levels (consequential analysis), the *observed effect* is the association between the top mQTL and ln(adipokine), extracted from the full GWAS summary data. The *predicted effect* combines mQTL and EWAS statistics to estimate the influence of an additional effect allele (EA) on the outcome (i.e. the adipokine). For each additional EA, the expected rise in DNAm at the CpG is equivalent to the mQTL effect size (β_{mQTL}). As the EWAS effect size represents the DNAm effect associated with a one-unit increase in the adipokine level, the expected increase in the adipokine level for a β_{mQTL} increase in DNAm can be calculated as the product of the mQTL and EWAS effects (i.e. $\beta_{mQTL} \times \beta_{EWAS}$). SNP effects on DNAm (mQTL effects) were extracted from GoDMC data[35] and CpG-adipokine effects were extracted from the EWAS meta-analysis presented here.

When looking in the reverse direction (i.e. adipokines as a cause of DNAm), the *observed effect* is a PGS, where the influence of adipokine-associated SNPs on DNAm are weighted by their EA frequency (EAF). The *predicted effect* here uses equivalent EAF weighting and is calculated as PGS~adipokine/adipokine~CpG. The observed and predicted effects in both directions were visualised using scatter plots and correlation was assessed with Pearson correlation coefficients.

Ethics statement

Leiden Longevity Study (LLS) Written informed consent for DNA collection and its use for genetic analyses was obtained from all participants prior to their enrolment into the study. Good clinical practice guidelines were maintained, and the study protocol was approved by the local Medical Ethical Committee of the Leiden University Medical Center.

Kooperative Gesundheitsforschung in der Region Augsburg (KORA) In accordance with the Declaration of Helsinki, written informed consent was obtained from all participants prior to their enrolment into the study. Good clinical practice guidelines were maintained, and the study protocol was approved by the Ethics Committee of the Bavarian Medical Association.

TwinsUK Ethical approval was initially granted by the St. Thomas' Hospital Research Ethics Committee, with subsequent approvals overseen by the NRES Committee London - Westminster following restructuring. In September 2007, ongoing approval was granted, ensuring indefinite continuation.

SHIP-TREND The medical ethics committee of the University of Greifswald approved the study protocol, and oral and written informed consents were obtained from each of the study participants.

LifeLines DEEP (LLD) All participants provided written informed consent. The Lifelines DEEP study was approved by the Medical Ethical Committee of the University Medical Center Groningen (UMCG), Groningen, The Netherlands

Author contributions

L.S. conceived and designed the study, performed data preprocessing and quality control, undertook cohort-specific and main analyses, interpreted results and drafted the manuscript. B.T.H. conceived and designed the study, supervised conduct of the study, iteratively critically revised the manuscript and provided critical intellectual contributions and interpretations of results. T.D. and R.W.

(KORA F4), X.L. (LLD), Y.X. and R.C. (TwinsUK) and M.K.N. (SHIP-TREND) performed cohort specific analyses, provided summary statistics for the meta-analyses, and provided critical intellectual contributions and interpretations of results; J.M.O. provided critical intellectual contributions and interpretations interpretations of results. L.F., A.Z., J.F. and H.S. (LLD), C.G., C.H., W.K., A.P. and M.W. (KORA F4), M.D., H.J.G., M.N. and A.T. (SHIP-TREND), J.T.B. (TwinsUK), and M.B. and P.S. (LLS) provided data, phenotype acquisition and harmonisation for these analyses, and provided critical intellectual contributions and interpretation of results. All authors reviewed the manuscript and approved the final version.

Funding

The work of L.S. was supported by the Joint Programming Initiative 'a Healthy Diet for a Healthy Life' (JPI-HDHL) DIMENSION project [ZonMW project number: 529051021]. Funding for the BIOS consortium was provided by the Netherlands Organization for Scientific Research (NWO 184.021.007 and 184.033.111), made available as a Rainbow Project of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL).

LLD The Lifelines Biobank initiative has been made possible by a subsidy from the Dutch Ministry of Health, Welfare and Sport; the Dutch Ministry of Economic Affairs, the University Medical Centre Groningen (UMCG, the Netherlands); the University of Groningen and the Northern Provinces of the Netherlands. J.F. is supported by the ERC Consolidator grant (grant agreement no. 101001678), NWO-VICI grant VI.C.202.022 and the Ammodo Science Award 2023 for Biomedical Sciences from Stichting Ammodo.

SHIP-TREND SHIP (The Study of Health in Pomerania) 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 (grant 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). DNAm data have been supported by the DZHK (grant 81X3400104). A.T. has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 542489987. The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH.

TwinsUK TwinsUK is funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation (CDRF), Zoe Ltd. and the National Institute for Health Research (NIHR), Clinical Research Network (CRN) and Biomedical Research Centre based on Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London.

KORA F4 The KORA research platform (KORA, Cooperative Health Research in the Region of Augsburg) was initiated and financed by the Helmholtz Munich – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research 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. This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the EU Joint Programming Initiative 'A Healthy Diet for a Healthy Life' (DIMENSION; grant no. 01EA1902A). The German Diabetes Center is supported by the Ministry of Culture and Science of the state of North-Rhine Westphalia (Düsseldorf, Germany) and the German Federal Ministry of Health (Berlin, Germany). This study was supported in part by a grant from the German Federal Ministry of Education and Research to the German Center for Diabetes Research (DZD).

LLS The LLS was supported by a grant from the Innovation-Oriented Research Program on Genomics (SenterNovem IGE01014 and IGE05007), the Centre for Medical Systems Biology, and

the National Institute for Healthy Ageing (grant 05040202 and 05060810), in the framework of the Netherlands Genomics Initiative / Netherlands Organization for Scientific Research, and the VOILA Consortium (ZonMW 457001001).

Funders had no say in the study design, collection, analysis, interpretation, or writing of this work.

Declarations of interest

H.J.G. has received travel grants and speaker's honoraria from Neuraxpharm, Servier, Indorsia and Janssen Cilag. C.H. is Reviews Editor at Diabetologia. The authors declare that there are no other relationships or activities that might bias, or be perceived to bias, their work.

Data & code availability

Individual-level data Regarding individual-level data from the cohorts involved, the informed consents given by KORA study participants does not cover data posting in public databases. However, data are available upon request from KORA Project Application Self-Service Tool (https://helmholtz-muenchen.managed-otrs.com/external/Data). Requests for data can be submitted online and are subject to approval by the KORA Board. The data of the SHIP study cannot be made publicly available due to the informed consent of the study participants but it can be accessed through a data application form available at https://transfer.ship-med.uni-greifswald.de/ for researchers who meet the criteria for access to confidential data. The HumanMethylation450 BeadChip data from the LLD and LLS are available as part of the BIOS Consortium in the European Genome-phenome Archive (EGA), under the accession code EGAD00010000887 (https://ega-archive.org/datasets/EGAD00010000887). Additional genomic and phenotype data are available upon request via the BBMRI-NL BIOS consortium. All data can be requested by bona fide researchers from the respective cohorts.

All other data used in this study is publicly available EWAS summary statistics can be downloaded from the EWAS Catalogue[58] and EWAS atlas[59], PBMC reference epigenome data are available from ROADMAP[13], TFBS data are available within the HOMER software[60], full mQTL summary statistics can be requested from GoDMC[71], adiponectin[72] and leptin[73] GWAS summary statistics are available from the GWAS database, LD proxies and matrices can be accessed using LDlink[74], variances in methylation and expression were calculated from data generated by the BIOS (a full list of investigators is available from https://ega-archive.org/datasets/EGAD00010000887), libraries for GSEA were downloaded directly from Enrichr (https://maayanlab.cloud/Enrichr/), and SGBS adipocyte data are available from GEO (GSE119593 for expression data, GSE119539 for DNAm data)[68].

Custom code availability R scripts for the respective analyses are deposited in a GitHub repository at `nebulyra/adipo_ewas`.

All software used is open source and freely available Unless stated otherwise, all calculations were performed using R version 4.2.2. For all meta-analyses, METAL, version 2011-03-25 was used[56]. TFBS enrichment analyses were performed using HOMER version 3.1[60].

Acknowledgements

The authors thank the staff, participants and related contributing research centres for all cohorts involved in this study. We are additionally grateful to P. S. DeVries (Human Genetics Center, University of Texas Health Science Center at Houston, USA) for his support with the SHIP-TREND EWAS pipeline.

References

1. Fahed, G. *et al.* Metabolic syndrome: updates on pathophysiology and management in 2021. *Int J Mol Sci* **23** (2):786 (2022).
2. Mir, M. M. *et al.* Differential association of selected adipocytokines, adiponectin, leptin, resistin, visfatin and chemerin, with the pathogenesis and progression of type 2 diabetes mellitus (T2DM) in the Asir region of Saudi Arabia: a case control study. *J Pers Med* **12** (5): 735 (2022).
3. Un Nisa, K. *and* Reza, M. I. Key relevance of epigenetic programming of adiponectin gene in pathogenesis of metabolic disorders. *Endocr Metab Immune Disord Drug Targets* **20** (4): 506-517 (2019).
4. Wróblewski, A. *et al.* Molecular insight into the interaction between epigenetics and leptin in metabolic disorders. *Nutrients* **11** (8): 1872 (2019).
5. Wielscher, M., *et al.* DNA methylation signature of chronic low-grade inflammation and its role in cardio-respiratory diseases. *Nat Commun* 13 (1): 2408 (2022).
6. Gomez-Alonso, M. D. C., *et al.* DNA methylation and lipid metabolism: an EWAS of 226 metabolic measures. *Clin Epigenetics* **13** (1): 1–19 (2021).
7. Willmer, T. *et al.* Blood-based DNA methylation biomarkers for type 2 diabetes: potential for clinical applications. *Front Endocrinol (Lausanne)* **9**: 744 (2018).
8. Dekkers, K.F. *et al.* Blood lipids influence DNA methylation in circulating cells. *Genome Biol* **17** (1): 138 (2016)
9. Davis, F. M. *and* Gallagher, K. A. Epigenetic mechanisms in monocytes/macrophages regulate inflammation in cardiometabolic and vascular disease. *Arterioscler Thromb Vasc Biol* **39** (4): 623-634 (2019).
10. Aslibekyan, S. *et al.* CPT1A methylation is associated with plasma adiponectin. *Nutr Metab Cardiovasc Dis* **27** (3) :225 (2016).
11. Suriano, F. *et al.* Novel insights into the genetically obese (ob/ob) and diabetic (db/db) mice: two sides of the same coin. *Microbiome* **9** (1): 147 (2021).
12. Matsubara, M., Maruoka, S., *and* Katayose, S. Inverse relationship between plasma adiponectin and leptin concentrations in normal-weight and obese women. *Eur J Endocrinol* **147** (2): 173-180 (2002).
13. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518** (7539): 317-330 (2015).
14. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39** (3): 311-318 (2007).
15. Kaluscha, S. *et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat Genet* **54** (12): 1895-1906 (2022).
16. Liu, G. *and* Liu, F. Bach2: a key regulator in Th2-related immune cells and Th2 immune response. *J Immunol Res* **2022**: 2814510 (2022).
17. Murphy, T. L., Tussiwand, R., *and* Murphy, K. M. Specificity through cooperation: BATF-IRF interactions control immune-regulatory networks. *Nat Rev Immunol* **13** (7): 499-509 (2013).
18. Liu, Y. *et al.* Clathrin-associated AP-1 controls termination of STING signalling. *Nature* **610** (7933): 761-767 (2022).
19. Wrann, C.D. *et al.* FOSL2 promotes leptin gene expression in human and mouse adipocytes. *J Clin Invest* **122** (3): 1010-1021 (2012).
20. Tsuchiya, M. *et al.* Suppression of MafA mRNA with siRNA prevents adipose cell differentiation in 3T3-L1 cells. *Int J Mol Med* **23** (6): 725-732 (2009).
21. Bonder, M. J. *et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet* **49** (1): 131-138 (2017).
22. Shimano, H. *and* Sato, R. SREBP-regulated lipid metabolism: convergent physiology-divergent pathophysiology. *Nat Rev Endocrinol* **13** (12): 710-730 (2017).
23. Kennedy, M. A. *et al.* ABCG1 has a critical role in mediating cholesterol efflux to HDL and preventing cellular lipid accumulation. *Cell Metab* **1** (2): 121-131 (2005).
24. Zerenturk, E. J. *et al.* Desmosterol and DHCR24: unexpected new directions for a terminal step in cholesterol synthesis. *Prog Lipid Res* **52** (4): 666-680 (2013).
25. Schlaepfer, I. R. *and* Joshi, M. CPT1A-mediated fat oxidation, mechanisms, and therapeutic potential. *Endocrinology* **161** (2): bqz046 (2020).
26. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541** (7635): 81-86 (2017).
27. Hillary, R.F. *et al.* Blood-based epigenome-wide analyses of 19 common disease states: a longitudinal, population-based linked cohort study of 18,413 Scottish individuals. *PLoS Med* **20** (7): e1004247 (2023).
28. Harding, A. H. *et al.* Polymorphisms in the gene encoding sterol regulatory element-binding factor-1c are associated with type 2 diabetes. *Diabetologia* **49** (11): 2642-2648 (2006).
29. Felder, T. K. *et al.* The SREBF-1 locus is associated with type 2 diabetes and plasma adiponectin levels in a middle-aged Austrian population. *Int J Obes (Lond)* **31** (7): 1099-1103 (2007).

30. Seo, J.B. *et al.* Adipocyte determination- and differentiation-dependent factor 1/sterol regulatory element-binding protein 1c regulates mouse adiponectin expression. *J Biol Chem* **279** (21): 22108-22117 (2004).
31. Lee, J. H. *et al.* SCAP deficiency facilitates obesity and insulin resistance through shifting adipose tissue macrophage polarization. *J Adv Res* **45**: 1-13 (2023).
32. Engin, A. B. Message transmission between adipocyte and macrophage in obesity. *Adv Exp Med Biol* **1460**: 273-295 (2024).
33. Lai, L. *et al.* Longitudinal association between DNA methylation and type 2 diabetes: findings from the KORA F4/FF4 study. *Cardiovasc Diabetol* **24** (1): 19 (2025).
34. Burgess, S. *and* Thompson, S. G. Bias in causal estimates from Mendelian randomization studies with weak instruments. *Stat Med* **30** (11): 1312-1323 (2011).
35. Schoenmaker, M. *et al*. Evidence of genetic enrichment for exceptional survival using a family approach: The Leiden Longevity Study. *Eur J Hum Genet* **14** (1): 79-84 (2006).
36. Sinke, L. *et al.* DNAmArray: Streamlined workflow for the quality control, normalization, and analysis of Illumina methylation array data (2.1) *Zenodo* (2019).
37. Meisinger, C. *et al.* Prevalence of undiagnosed diabetes and impaired glucose regulation in 35-59-year-old individuals in Southern Germany: The KORA F4 study. *Diabet Med* **27** (3): 360-362 (2010).
38. Herder, C. *et al.* Association of subclinical inflammation with polyneuropathy in the older population: KORA F4 study. *Diabetes Care* **36** (11): 3663-3670 (2013).
39. Kriebel, J. *et al.* Association between DNA Methylation in whole blood and measures of glucose metabolism: Kora F4 study. *PLoS One* 11 (3): e0152314 (2016).
40. Touleimat, N. *and* Tost, J. Complete pipeline for Infinium® Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics* **4** (3): 325-341 (2012).
41. Costeira, R. *et al.* Metabolomic biomarkers of habitual B vitamin intakes unveil novel differentially methylated positions in the human epigenome. *Clin Epigenetics* **15** (1): 166 (2023).
42. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5** (10): R80 (2004).
43. Xu, Z. *et al.* ENmix: A novel background correction method for Illumina HumanMethylation450 BeadChip. *Nucleic Acids Res* **44** (3): e20 (2016).
44. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30** (10): 1363-1369 (2014).
45. Christiansen, C. *et al.* Adipose methylome integrativeomic analyses reveal genetic and dietary metabolic health drivers and insulin resistance classifiers. *Genome Med* **14** (1): 75 (2022).
46. Völzke, H. *et al.* Cohort Profile Update: The Study of Health in Pomerania (SHIP). *Int J Epidemiol* **51** (6): e372-e383 (2022).
47. Lehne, B. *et al.* A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol* **16** (1): 37 (2015).
48. Scholtens, S. *et al.* Cohort Profile: LifeLines, a threegeneration cohort study and biobank. *Int J Epidemiol* **44** (4): 1172-1180 (2015).
49. Tigchelaar, E. F. *et al.* Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open* **5** (8): e006772 (2015).
50. Lu, X. *et al.* An epigenome-wide association study identifies multiple DNA methylation markers of exposure to endocrine disruptors. *Environ Int* **144**: 106016 (2020).
51. Koestler, D. C. *et al.* Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). *BMC Bioinformatics* **17** (1): 120 (2016).
52. Zheng, S. C. *et al.* Identification of differentially methylated cell types in epigenome-wide association studies. *Nat Methods* **15** (12): 1059-1066 (2018).
53. Amemiya, H. M., Kundaje, A. *and* Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9** (1): 9354 (2019).
54. Zhou, W., Laird, P. W. *and* Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45** (4): e22 (2017).
55. van Iterson, M. *et al.* Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18** (1): 19 (2017).
56. Willer, C. J., Li, Y. *and* Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26** (17): 2190-2191 (2010).

57. Slieker, R. C. *et al.* Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenetics Chromatin* **6** (1): 26 (2013).
58. Battram, T. *et al.* The EWAS Catalog: a database of epigenome-wide association studies. *Wellcome Open Res* **7**: 41 (2022).
59. Li, M. *et al.* EWAS Atlas: A curated knowledgebase of epigenome-wide association studies. *Nucleic Acids Res* **47** (D1): D983-D988 (2019).
60. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38** (4): 576-589 (2010).
61. van Greevenbroek, M. M. J. *et al.* The cross-sectional association between insulin resistance and circulating complement C3 is partly explained by plasma alanine aminotransferase, independent of central obesity and general inflammation (the CODAM study). *Eur J Clin Invest* **41** (4): 372-379 (2011).
62. Willemsen, G. *et al.* The Netherlands twin register biobank: A resource for genetic epidemiological studies. *Twin Res Hum Genet* **13** (3): 231-245 (2010).
63. Van Dongen, J. *et al.* Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun* **7**: 11115 (2016).
64. Hofman, A. *et al.* The Rotterdam Study: 2014 objectives and design update. *Eur J Epidemiol* **28** (11): 889-926 (2013).
65. Huisman, M. H. B. *et al.* Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. *J Neurol Neurosurg Psychiatry* **82** (10): 1165-1170 (2011).
66. Robinson, M. D., McCarthy, D. J., *and* Smyth, G. K. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26** (1): 139-140 (2010).
67. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9** (8): e1003118 (2013).
68. Tini, G. *et al.* DNA methylation during human adipogenesis and the impact of fructose. *Genes Nutr* **15** (1): 21 (2020).
69. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2** (3): 100141 (2021).
70. Pierce, B. L. and Burgess, S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol* **178** (7): 1177-1184 (2013).
71. Min, J. L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat Genet* **53** (9): 1311-1321 (2021).
72. Sarsani, V. *et al.* A cross-ancestry genome-wide meta-analysis, fine-mapping, and gene prioritization approach to characterize the genetic architecture of adiponectin. *HGG Adv* **5** (1): 100252 (2024).
73. Kilpeläinen, T. O. *et al.* Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. *Nat Commun* **7** (1): 1-14 (2016).
74. Machiela, M. J. *and* Chanock, S. J. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* **31** (21): 3555-3557 (2015).

Supplementary Material

CpG	Position	*n*	beta	SE	*p*	p_{FDR}
cg06500161	chr21:43656587	2773	-0.0127	0.0012	5.60E-28	2.31E-22
cg11024682	chr17:17730094	2773	-0.0106	0.0011	3.01E-22	6.20E-17
cg27243685	chr21:43642366	2770	-0.0061	0.0009	1.20E-12	1.65E-07
cg06192883	chr15:52554171	2774	-0.0094	0.0014	4.77E-12	4.92E-07
cg00574958	chr11:68607622	2770	0.0043	0.0007	1.00E-10	8.28E-06
cg07504977	chr10:102131012	2772	-0.0113	0.0018	3.01E-10	2.07E-05
cg01881899	chr21:43652704	2751	-0.0045	0.0007	4.21E-10	2.48E-05
cg00134210	chr10:14644132	2768	0.0040	0.0006	7.90E-10	4.07E-05
cg10717869	chr1:205780912	2768	-0.0057	0.0009	2.44E-09	1.12E-04
cg25217710	chr1:156609523	2775	-0.0052	0.0009	4.64E-09	1.91E-04
cg05945608	chr6:42739639	2773	-0.0072	0.0012	5.76E-09	2.16E-04
cg17901584	chr1:55353706	2775	0.0107	0.0018	7.42E-09	2.55E-04
cg04658841	chr16:85478651	2772	-0.0054	0.0010	1.55E-08	4.92E-04
cg10438589	chr4:14531493	2771	-0.0087	0.0016	2.37E-08	6.97E-04
cg05014727	chr10:6214016	2770	0.0103	0.0019	4.57E-08	1.22E-03
cg02235049	chr3:186559186	2772	-0.0051	0.0009	4.73E-08	1.22E-03
cg13123009	chr6:31681882	2764	-0.0057	0.0011	6.83E-08	1.66E-03
cg20544516	chr17:17717183	2770	-0.0047	0.0009	1.21E-07	2.76E-03
cg06603309	chr11:2724144	2328	0.0059	0.0011	1.72E-07	3.38E-03
cg14476101	chr1:120255992	2759	0.0124	0.0024	1.70E-07	3.38E-03

Supplementary Table 1 | Top 20 CpGs associated with circulating adiponectin levels in blood. CpG identifier, genomic position, sample size (*n*), and association with IL-6 (beta) is shown, alongside standard error (SE) and significance (nominal and FDR-adjusted *p*-value).

CpG	Position	*n*	beta	SE	*p*	p_{FDR}
cg06500161	chr21:43656587	3644	0.0071	0.0006	4.98E-32	2.02E-26
cg00574958	chr11:68607622	3638	-0.0040	0.0004	6.62E-27	1.34E-21
cg14476101	chr1:120255992	3612	-0.0117	0.0012	8.89E-23	1.20E-17
cg17501210	chr6:166970252	3644	-0.0078	0.0008	5.07E-22	5.15E-17
cg11024682	chr17:17730094	3644	0.0051	0.0006	1.41E-19	1.14E-14
cg16246545	chr1:120255941	3638	-0.0085	0.0010	1.41E-18	9.53E-14
cg03725309	chr1:109757585	3638	-0.0049	0.0006	1.29E-16	7.51E-12
cg06192883	chr15:52554171	3646	0.0057	0.0007	5.00E-15	2.54E-10
cg17058475	chr11:68607737	3646	-0.0042	0.0005	1.13E-14	5.09E-10
cg26403843	chr5:158634085	3607	0.0103	0.0014	4.66E-14	1.83E-09
cg06690548	chr4:139162808	3537	-0.0046	0.0006	4.95E-14	1.83E-09
cg17901584	chr1:55353706	3647	-0.0072	0.0010	1.17E-13	3.95E-09
cg27243685	chr21:43642366	3636	0.0031	0.0004	1.94E-13	6.05E-09
cg06559575	chr12:53490352	3643	-0.0043	0.0006	3.34E-13	9.70E-09
cg19750657	chr13:38935967	3643	0.0051	0.0007	1.34E-12	3.62E-08
cg07573872	chr19:1126342	3631	-0.0051	0.0007	1.93E-12	4.91E-08
cg11376147	chr11:57261198	3644	-0.0026	0.0004	2.60E-12	6.21E-08
cg18120259	chr6:43894639	3642	-0.0047	0.0007	3.90E-12	8.81E-08
cg18181703	chr17:76354621	3643	-0.0060	0.0009	4.50E-12	9.62E-08
cg07504977	chr10:102131012	3643	0.0064	0.0009	9.43E-12	1.92E-07

Supplementary Table 2 | Top 20 CpGs associated with circulating leptin levels in blood. CpG identifier, genomic position, sample size (*n*), and association with IL-6 (beta) is shown, alongside standard error (SE) and significance (nominal and FDR-adjusted *p*-value).

Trait	OR	lnOR	SE	p	p_{FDR}	Overlap	%
Body mass index	571.7	6.35	5.15	2.75E-157	2.75E-157	40	54.79
Fasting insulin	435.5	6.08	4.96	5.04E-127	5.04E-127	24	32.88
Triglyceride levels	1917.4	7.56	6.79	4.41E-117	4.41E-117	14	19.18
Waist circumference	372.4	5.92	5.01	3.39E-89	3.39E-89	15	20.55
HDL-cholesterol	465.8	6.14	5.35	8.25E-80	8.25E-80	12	16.44
C-reactive protein	301.3	5.71	5.01	9.11E-61	9.11E-61	10	13.70
Type 2 diabetes	533.8	6.28	5.74	8.76E-59	8.76E-59	8	10.96
Diastolic BP	295.3	5.69	5.21	1.31E-44	1.31E-44	7	9.59
Fasting glucose	121.4	4.80	4.08	1.86E-44	1.86E-44	10	13.70
Systolic BP	215.2	5.37	4.89	2.09E-40	2.09E-40	7	9.59
Age	18.2	2.90	1.73	1.77E-33	1.77E-33	28	38.36
Alcohol	15.9	2.76	1.70	6.57E-26	6.57E-26	20	27.40
Smoking status	11.8	2.46	1.32	1.73E-23	1.73E-23	25	34.25
Sex	8.9	2.19	1.05	1.74E-18	1.74E-18	24	32.88
Dodecanedioate	60.9	4.11	4.00	1.84E-15	1.84E-15	4	5.48
Parkinson's disease	102.1	4.63	4.74	7.20E-15	7.20E-15	3	4.11
Cognition	52.2	3.96	3.84	1.86E-14	1.86E-14	4	5.48
Total cholesterol	181.4	5.20	5.68	8.97E-13	8.97E-13	2	2.74
Dihomo-linolenate	98.4	4.59	5.05	2.19E-10	2.19E-10	2	2.74
PCB exposure	35.5	3.57	3.68	1.57E-09	1.57E-09	3	4.11

Supplementary Table 3 | Top 20 traits from EWAS enrichment analysis. Odds ratio of enrichment in adiponectin-associated CpGs (OR and log-transformed lnOR) is shown, alongside its standard error (SE) and significance (nominal and FDR-adjusted *p*-value). The overlap (total and percentage) with adiponectin-associated CpGs is also presented.

Trait	OR	lnOR	SE	p	p_{FDR}	Overlap	%
Body mass index	151.3	5.02	3.38	1.09E-204	1.09E-204	51	24.17
Fasting insulin	147.5	4.99	3.60	4.05E-134	4.05E-134	30	14.22
C-reactive protein	256.7	5.55	4.30	6.97E-131	6.97E-131	24	11.37
Triglyceride levels	413.7	6.03	5.29	5.75E-71	5.75E-71	11	5.21
Systolic BP	137.4	4.92	4.01	1.90E-62	1.90E-62	13	6.16
Smoking	9.8	2.28	0.53	3.18E-52	3.18E-52	64	30.33
Alcohol	11.7	2.46	0.84	5.60E-49	5.60E-49	46	21.80
Diastolic BP	127.4	4.85	4.16	2.05E-43	2.05E-43	9	4.27
HDL-cholesterol	102.8	4.63	3.94	3.14E-40	3.14E-40	9	4.27
Waist circumference	69.9	4.25	3.48	4.72E-38	4.72E-38	10	4.74
Sex	6.0	1.80	0.12	9.63E-30	9.63E-30	53	25.12
Cognition	40.4	3.70	2.99	5.48E-27	5.48E-27	9	4.27
Type 2 diabetes	100.6	4.61	4.33	2.71E-23	2.71E-23	5	2.37
Mortality	11.4	2.44	1.44	1.54E-18	1.54E-18	14	6.64
Age	5.2	1.65	0.20	8.38E-18	8.38E-18	32	15.17
Fasting glucose	22.0	3.09	2.65	1.16E-13	1.16E-13	6	2.84
Education level	27.3	3.31	3.00	3.98E-13	3.98E-13	5	2.37
Protein biomarker	42.4	3.75	3.85	1.80E-10	1.80E-10	3	1.42
Parkinson's disease	33.9	3.52	3.62	1.91E-09	1.91E-09	3	1.42
Total cholesterol	60.7	4.11	4.57	1.25E-08	1.25E-08	2	0.95

Supplementary Table 4 | Top 20 traits from EWAS enrichment analysis. Odds ratio of enrichment in leptin-associated CpGs (OR and log-transformed lnOR) is shown, alongside its standard error (SE) and significance (nominal and FDR-adjusted *p*-value). The overlap (total and percentage) with leptin-associated CpGs is also presented.

Chromatin State	OR	lnOR	SE	p	p_{FDR}	Full Roadmap name
(1) TssA	0.17	-1.79	-0.64	7.47E-02	2.24E-01	Active TSS
(2) TssAFlnk	0.56	-0.59	-1.18	1.18E-01	2.53E-01	Flanking active TSS
(3) TxFlnk	5.08	1.63	2.78	1.07E-01	2.53E-01	Transcr. at gene 5' and 3'
(4) Tx	2.08	0.73	-0.14	1.67E-02	1.25E-01	Strong transcription
(5) TxWk	1.05	0.05	-0.46	9.09E-01	9.57E-01	Weak transcription
(6) EnhG	0.00	-8.92	113.23	8.87E-01	9.57E-01	Genic enhancers
(7) Enh	8.82	2.18	1.15	1.12E-15	1.69E-14	Enhancers
(8) ZNF/Rpts	0.00	-9.21	323.98	9.57E-01	9.57E-01	ZNF genes + repeats
(9) Het	0.00	-8.63	113.52	8.91E-01	9.57E-01	Heterochromatin
(10) TssBiv	1.99	0.69	0.57	1.82E-01	3.42E-01	Bivalent/poised TSS
(11) BivFlnk	0.00	-9.09	64.73	8.11E-01	9.57E-01	Flanking bivalent TSS/Enh
(12) EnhBiv	0.00	-9.43	112.72	8.80E-01	9.57E-01	Bivalent enhancer
(13) ReprPC	0.27	-1.31	-0.87	6.66E-02	2.24E-01	Repressed polycomb
(14) ReprPCWk	0.22	-1.50	-1.06	3.55E-02	1.78E-01	Weak repressed polycomb
(15) Quies	1.09	0.08	-0.92	7.61E-01	9.57E-01	Quiescent/low

Supplementary Table 5 | Chromatin state enrichment analysis results using the peripheral blood mononuclear cell reference epigenome (PBMC; E062). The odds ratio of enrichment in adiponectin-associated CpGs (OR and log-transformed lnOR) is shown, alongside its standard error (SE) and significance (nominal and FDR-adjusted *p*-value). The abbreviated and full Roadmap names are also displayed (TSS: transcription start site, ZNF: zinc finger).

Chromatin State	OR	lnOR	SE	p	p_{FDR}	Full Roadmap name
(1) TssA	0.64	-0.44	-1.29	1.55E-01	2.91E-01	Active TSS
(2) TssAFlnk	0.73	-0.32	-1.74	1.09E-01	2.72E-01	Flanking active TSS
(3) TxFlnk	5.21	1.65	1.73	4.60E-03	2.03E-02	Transcr. at gene 5' and 3'
(4) Tx	0.67	-0.40	-1.41	1.45E-01	2.91E-01	Strong transcription
(5) TxWk	1.92	0.65	-0.85	4.03E-04	3.02E-03	Weak transcription
(6) EnhG	1.98	0.68	0.53	1.77E-01	2.95E-01	Genic enhancers
(7) Enh	4.87	1.58	0.11	6.23E-17	9.35E-16	Enhancers
(8) ZNF/Rpts	0.00	-10.17	309.90	9.50E-01	9.50E-01	ZNF genes + repeats
(9) Het	1.39	0.33	0.76	6.45E-01	8.73E-01	Heterochromatin
(10) TssBiv	0.00	-10.04	60.86	7.83E-01	8.86E-01	Bivalent/poised TSS
(11) BivFlnk	0.30	-1.20	-0.77	9.03E-02	2.71E-01	Flanking bivalent TSS/Enh
(12) EnhBiv	0.88	-0.13	-0.05	8.27E-01	8.86E-01	Bivalent enhancer
(13) ReprPC	0.37	-1.00	-1.65	5.41E-03	2.03E-02	Repressed polycomb
(14) ReprPCWk	0.86	-0.15	-1.39	5.04E-01	7.55E-01	Weak repressed polycomb
(15) Quies	1.06	0.06	-1.59	6.99E-01	8.73E-01	Quiescent/low

Supplementary Table 6 | Chromatin state enrichment analysis results using the peripheral blood mononuclear cell reference epigenome (PBMC; E062). The odds ratio of enrichment in leptin-associated CpGs (OR and log-transformed lnOR) is shown, alongside its standard error (SE) and significance (nominal and FDR-adjusted *p*-value). The abbreviated and full Roadmap names are also displayed.

Trait	TF	*p*	p_{FDR}	CpG *n*	CpG %	Background %
Leptin	RFX (HTH)	1.12E-08	1.57E-07	4	1.86	0.15
Leptin	Fosl2 (bZIP)	3.11E-08	2.18E-07	6	2.79	0.46
Leptin	Rfx2 (HTH)	9.98E-08	4.66E-07	4	1.86	0.20
Adiponectin	MafA (bZIP)	1.91E-07	6.69E-07	7	9.59	2.22
Leptin	JunB (bZIP)	2.61E-07	7.31E-07	7	3.26	0.76
Leptin	Rfx1 (HTH)	5.41E-07	1.26E-06	5	2.33	0.40
Leptin	AP-1 (bZIP)	1.42E-06	2.66E-06	8	3.72	1.10
Leptin	Bach2 (bZIP)	1.52E-06	2.66E-06	4	1.86	0.27
Leptin	BATF (bZIP)	2.51E-06	3.90E-06	7	3.26	0.90
Leptin	SCL (bHLH)	4.33E-06	6.06E-06	48	22.33	15.41
Leptin	X-box (HTH)	6.95E-06	8.85E-06	3	1.40	0.17
Adiponectin	Npas4 (bHLH)	1.04E-05	1.21E-05	7	9.59	3.05
Adiponectin	Zic3 (Zf)	1.39E-05	1.50E-05	5	6.85	1.68
Leptin	KLF6 (Zf)	2.06E-05	2.06E-05	14	6.51	3.17

Supplementary Table 7 | Transcription factor (TF) binding site enrichment analysis results. Sequences within 50bp of adipokine-associated CpGs were scanned for established motifs using HOMER and compared to a random genomic background matched for GC content. TF name and type are shown alongside the significance of the enrichment (nominal and FDR-adjusted *p*-values). The number of CpGs responsible for the enrichment are also displayed (CpG *n*, CpG %), as well as the percentage of background sequences within 50bp of the respective motif.

CpG	Gene name	Ensembl ID	beta	*p*	p_{FDR}	Nearest
cg14476101	*PHGDH*	ENSG00000092621	-0.0401	3.37E-129	9.90E-127	TRUE
cg17901584	*DHCR24*	ENSG00000116133	-0.0208	1.39E-80	2.05E-78	FALSE
cg06500161	*ABCG1*	ENSG00000160179	-0.0114	1.75E-61	1.72E-59	FALSE
cg17194270	*SYNGR1*	ENSG00000100321	-0.0128	6.85E-42	5.04E-40	FALSE
cg27243685	*ABCG1*	ENSG00000160179	-0.0069	3.02E-31	1.78E-29	FALSE
cg01881899	*ABCG1*	ENSG00000160179	-0.0048	3.59E-25	1.76E-23	FALSE
cg20496314	*SYNGR1*	ENSG00000100321	-0.0082	8.95E-25	3.76E-23	FALSE
cg00574958	*CPT1A*	ENSG00000110090	-0.0039	1.12E-24	4.12E-23	FALSE
cg03234777	*MPZL2*	ENSG00000149573	0.0059	2.64E-24	8.61E-23	FALSE
cg11024682	*SREBF1*	ENSG00000072310	-0.0060	1.31E-18	3.84E-17	FALSE
cg03234777	*MPZL3*	ENSG00000160588	0.0048	3.09E-16	8.27E-15	TRUE
cg14117297	*TMIGD2*	ENSG00000167664	-0.0043	4.05E-14	9.93E-13	FALSE
cg10192877	*ABCG1*	ENSG00000160179	-0.0032	4.73E-14	1.07E-12	FALSE
cg00134210	*FAM107B*	ENSG00000065809	-0.0031	1.69E-12	3.54E-11	FALSE
cg11024682	*TOM1L2*	ENSG00000175662	-0.0041	3.87E-10	7.58E-09	FALSE
cg23594345	*IGHD6-25*	ENSG00000225825	0.0099	2.18E-09	4.01E-08	FALSE
cg26949393	*AGAP3*	ENSG00000133612	0.0018	6.39E-07	1.11E-05	FALSE
cg20544516	*SREBF1*	ENSG00000072310	-0.0034	8.25E-07	1.35E-05	TRUE
cg03234777	*AMICA1*	ENSG00000160593	-0.0030	1.33E-06	2.00E-05	FALSE
cg11147309	*PLEC*	ENSG00000178209	-0.0029	1.36E-06	2.00E-05	TRUE

Supplementary Table 8 | Top 20 expression quantitative trait methylation (eQTM) effects connecting adiponectin-associated CpGs and nearby genes (within 100kb), ordered by *p*-value. The CpG and gene identifiers are shown alongside their eQTM effect size and significance.

CpG	Gene name	Ensembl	beta	*p*	p_{FDR}	Nearest
cg07639376	*TMEM204*	ENSG00000131634	-0.0860	1.45E-247	1.18E-244	TRUE
cg14476101	*PHGDH*	ENSG00000092621	-0.0401	3.37E-129	1.37E-126	TRUE
cg05668853	*RAB34*	ENSG00000109113	-0.0201	2.87E-112	7.79E-110	TRUE
cg16246545	*PHGDH*	ENSG00000092621	-0.0289	1.27E-109	2.58E-107	TRUE
cg17901584	*DHCR24*	ENSG00000116133	-0.0208	1.39E-80	2.27E-78	FALSE
cg06500161	*ABCG1*	ENSG00000160179	-0.0114	1.75E-61	2.38E-59	FALSE
cg10045881	*CHI3L2*	ENSG00000064886	-0.0105	5.35E-43	6.23E-41	TRUE
cg23361127	*ENC1*	ENSG00000171617	-0.0079	1.56E-38	1.59E-36	FALSE
cg17332198	*METTL21B*	ENSG00000123427	-0.0060	5.30E-34	4.80E-32	FALSE
cg18149207	*RORC*	ENSG00000143365	-0.0060	4.95E-29	4.03E-27	FALSE
cg22650271	*SYNGR1*	ENSG00000100321	-0.0066	1.96E-27	1.45E-25	FALSE
cg20496314	*SYNGR1*	ENSG00000100321	-0.0082	8.95E-25	6.08E-23	FALSE
cg00574958	*CPT1A*	ENSG00000110090	-0.0039	1.12E-24	7.03E-23	FALSE
cg18930267	*NOC4L*	ENSG00000184967	-0.0078	7.65E-24	4.45E-22	TRUE
cg13871708	*VARS2*	ENSG00000137411	0.0079	5.00E-20	2.71E-18	FALSE
cg17058475	*CPT1A*	ENSG00000110090	-0.0046	8.18E-19	4.17E-17	FALSE
cg02741985	*CCDC57*	ENSG00000176155	-0.0057	1.50E-18	7.20E-17	TRUE
cg11468085	*ACY3*	ENSG00000132744	-0.0056	1.22E-16	5.51E-15	FALSE
cg23361127	*HEXB*	ENSG00000049860	-0.0056	1.94E-15	8.34E-14	FALSE
cg22534374	*NAV1*	ENSG00000134369	-0.0074	1.16E-13	4.72E-12	FALSE

Supplementary Table 9 | Top 20 expression quantitative trait methylation (eQTM) effects connecting leptin-associated CpGs and nearby genes (within 100kb), ordered by *p*-value. The CpG and gene identifiers are shown alongside their eQTM effect size and significance.

Database	Gene set	Size	Overlap	*p*	p_{FDR}
GeDiPNet (2023)	Micromelia	102	7	6.89E-07	2.42E-03
PheWeb (2019)	Hereditary and idiopathic peripheral neuropathy	18	4	1.55E-06	2.71E-03
WikiPathway (2021)	Cholesterol metabolism (includes both Bloch and Kandutsch-Russell pathways)	46	5	2.94E-06	3.30E-03
Elsevier Pathway Collection	Proteins with Altered Expression in Cancer Metabolic Reprogramming	85	6	3.76E-06	3.30E-03
Elsevier Pathway Collection	Glutamine in Cancer Metabolism	28	4	9.96E-06	6.99E-03
Reactome (2022)	Metabolism Of Lipids R-HSA-556833	732	14	1.40E-05	8.21E-03
Reactome (2022)	Metabolism R-HSA-1430728	2049	25	1.73E-05	8.67E-03
BioPlanet (2019)	Oncostatin M	311	9	2.25E-05	9.86E-03
WikiPathway (2021)	Nuclear Receptors Meta-Pathway	319	9	2.74E-05	1.07E-02
BioPlanet (2019)	Transmembrane transport of small molecules	432	10	5.24E-05	1.75E-02
MsigDB (2020)	Xenobiotic Metabolism	200	7	5.75E-05	1.75E-02
BioPlanet (2019)	Amino acid biosynthesis and interconversion (transamination)	16	3	5.99E-05	1.75E-02
GO BP (2023)	Phosphatidylcholine Metabolic Process (GO:0046470)	55	4	1.50E-04	4.05E-02
Reactome (2022)	Immune System R-HSA-168256	1943	22	1.74E-04	4.37E-02
KEGG (2021)	Insulin resistance	108	5	1.88E-04	4.39E-02

Supplementary Table 10 | Enriched terms in genes linked to leptin-associated CpGs, ordered by *p*-value. The set size and overlap is shown alongside their significance (nominal and FDR-adjusted *p*-value).

Database	Gene set	Size	Overlap	p	p_{FDR}
PheWeb (2019)	Type 1 diabetes	95	7	4.31E-11	5.17E-08
PheWeb_(2019)	Type 1 diabetes with renal manifestations	64	5	2.33E-08	1.40E-05
Elsevier Pathway Collection	Proteins with Altered Expression in Cancer Metabolic Reprogramming	85	5	9.81E-08	3.92E-05
PheWeb (2019)	Celiac disease	90	5	1.31E-07	3.92E-05
WikiPathway (2021)	Cholesterol metabolism (includes both Bloch and Kandutsch-Russell pathways) WP4718	46	4	4.23E-07	1.02E-04
Elsevier Pathway Collection	Metabolic Effects of Oncogenes and Tumor Suppressor in Cancer Cells	68	4	2.07E-06	4.14E-04
KEGG (2021)	AMPK signalling pathway	120	4	1.98E-05	3.40E-03
Reactome (2022)	Signalling By Nuclear Receptors R-HSA-9006931	260	5	2.40E-05	3.61E-03
Reactome (2022)	NR1H2 And NR1H3-mediated Signalling R-HSA-9024446	46	3	3.20E-05	4.27E-03
Elsevier Pathway Collection	Estrogen Deficiency in Female Obesity	9	2	6.20E-05	6.77E-03
Reactome (2022)	NR1H2 And NR1H3 Regulate Gene Expression Linked To Lipogenesis R-HSA-9029558	9	2	6.20E-05	6.77E-03
WikiPathway (2021)	Liver X receptor pathway WP2874	10	2	7.75E-05	7.75E-03
BioPlanet (2019)	AMPK signalling	68	3	1.03E-04	9.07E-03
WikiPathway (2021)	AMP-activated protein kinase (AMPK) signalling WP1403	69	3	1.08E-04	9.07E-03
Elsevier Pathway Collection	Lipodystrophy, Familial Partial	12	2	1.13E-04	9.07E-03
GO BP (2023)	Membrane Depolarization During Action Potential (GO:0086010)	13	2	2.53E-06	1.16E-03
PheWeb (2019)	Open-angle glaucoma	13	2	2.80E-06	1.21E-03
MsigDB (2020)	mTORC1 Signalling	200	4	3.33E-06	1.34E-03
GO BP (2023)	High-Density Lipoprotein Particle Remodelling (GO:0034375)	15	2	3.49E-06	1.34E-03
GO BP (2023)	Regulation Of Cholesterol Biosynthetic Process (GO:0045540)	15	2	3.92E-06	1.44E-03

Supplementary Table 11 | Top 20 enriched terms in genes linked to adiponectin-associated CpGs, ordered by *p*-value. The set size and overlap is shown alongside their significance (nominal and FDR-adjusted *p*-value).

CHAPTER FOUR

Omixer

Omixer: multivariate and reproducible sample randomisation to proactively counter batch effects in genomic studies

Lucy Sinke[1], Davy Cats[1], and Bastiaan T. Heijmans[1]

[1] Leiden University Medical Centre, Leiden, The Netherlands

Published in Bioinformatics 37(18), 3051-2 (2021)
DOI: 10.1093/bioinformatics/btab159

Abstract

Batch effects heavily impact results in genomic research, causing bias and false positive results, yet software to control them pre-emptively is lacking. Sample randomisation prior to measurement is vital for minimising these effects. However, current approaches are often *ad hoc*, poorly documented, and ill-equipped to handle multiple batches and outcomes.

We developed *Omixer*: a Bioconductor package implementing multivariate and reproducible sample randomisation for genomic studies. It proactively counters correlations between technical factors and biological variables of interest by optimising sample distribution across and within batches.

Background

Batch effects can overshadow biological differences and critically influence the results of genomic research[1–3]. Even in benign cases, they decrease power to detect a true biological effect and may contaminate results with false positives[4]. Despite numerous statistical methods developed to adjust for batch effects, these frequently prescribed reactive approaches are often insufficient[5–7]. When technical variables are confounded with experimental factors of interest, batch effect correction will likely mask genuine underlying biological signals[8].

Sample randomisation is a proactive, and arguably more impactful, method for obtaining reproducible results in high-throughput experiments[9]. Despite this, its current implementation suffers from several key issues. Particularly where there are numerous or nested batches composed of a limited number of samples, such as separate microarrays or sequencing lanes, single random draws can inadvertently result in high correlations between technical covariates and biological factors. This is further complicated by an often poorly documented randomisation process that is not by default reproducible. Although stratified randomisation is capable of effectively removing batch effects in microarray experiments, it cannot address all relevant biological variables[10]. Therefore, to adequately combat bias in results, it is imperative that we employ methods capable of handling a wider array of research setups.

We developed *Omixer*: an R package for multivariate and reproducible randomisation in genomic studies. From a large number of randomised sample layouts, it selects the one that optimally balances biological variables across batches. Omixer offers the flexibility required to perform randomisation effectively and reduces the risk of masked or false positive signals across a range of common experimental setups and study designs.

Keywords: *Bioconductor, genomics, randomisation, software*

Results

Multivariate and reproducible sample randomisation

To optimise distribution of samples across batches, randomisation is performed multiple times. The default number of iterations is 1,000 and this can be adjusted if needed (using option `iterNum`; **Fig. 1**). After combining sample lists with the specified plate layout, statistical tests of correlation determine the optimal setup. This is defined as the layout that minimises the absolute sum of correlations between defined biological and technical factors. As a precautionary step, layouts with evidence for any tested batch associations are excluded ($p < 0.05$), although in practice this will not change the result given suitably large iteration numbers (**Fig. 2a**).

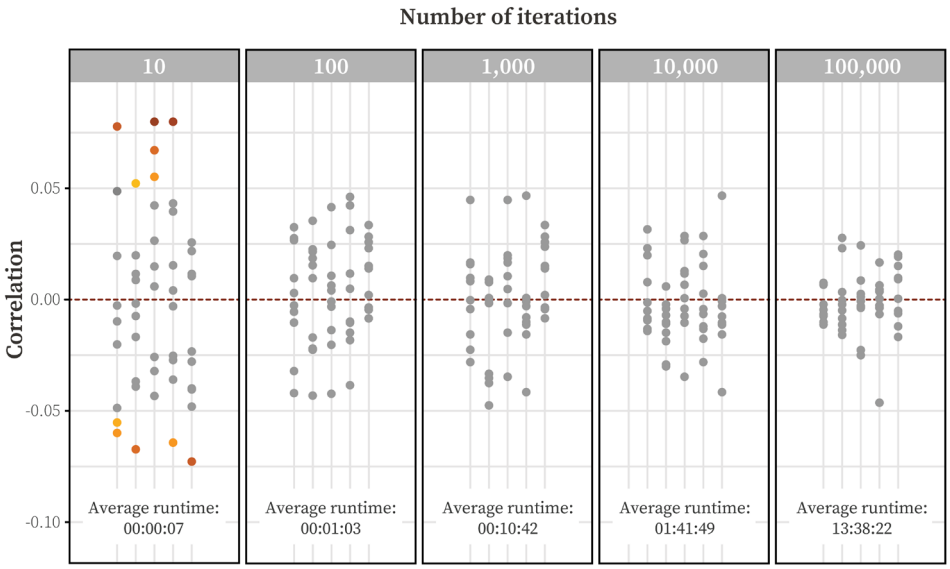


Figure 1 | The default number of iterations Omixer performs is 1,000. Increasing this number may result in smaller overall correlations between technical and biological factors, and complex designs could benefit from such increases above the default. Users should be aware that this choice comes at the cost of increased runtime, exemplified here using 1 CPU with 64 GB RAM to randomise 616 example samples across seven 96-well plates.

To reserve wells for control samples or other studies, a mask can be specified (using option: `mask`), and paired samples such as those from twin studies can be blocked so that they remain together in the same batch (using option: `block`). Non-standard plate layouts can be input, but Omixer also generates commonly used plates (using options: `wells` and `plateNum`). Previously generated layouts can quickly be reproduced using the *omixerSpecific* function and the automatically saved *randomSeed* object.

The main function, *omixerRand*, takes a sample list and plate layout as input and optimises the distribution of specified biological variables (option: `randVars`) across batches (option: `techVars`). The resulting correlations between these two types of variables in the returned layout are then visually displayed as a heatmap for inspection by the user.

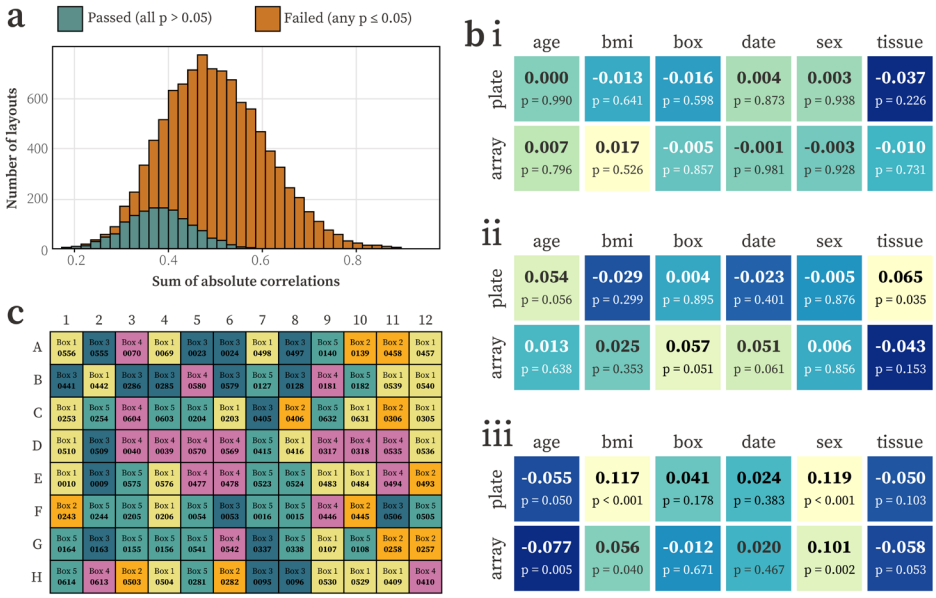


Figure 2 | Overview of Omixer functionality and graphical output. a) Distribution of the sum of absolute correlations from 10,000 randomized layouts, coloured by filtering step outcome. **b)** Resulting correlation matrices from the **i)** optimal Omixer layout, **ii)** median result, and **iii)** worst case scenario after simple randomization. **c)** An example of a lab-friendly sample sheet created by Omixer, showing a 96-well plate coloured by box number.

Omixer outperforms simple randomisation approaches

Particularly when multiple batch types and outcomes are present, a single randomisation is likely to result in significant correlations. To exemplify this, we randomised an example dataset of 616 samples intended to be profiled using the Illumina Infinium MethylationEPIC BeadChip array. Randomisation variables of interest were specified as age, body mass index (BMI), box, isolation date, sex, and tissue. Batches were specified as plate number (plate) and array ID (array). Following 10,000 simple randomizations, 85% of the resulting layouts had at least one significant correlation between a randomisation variable and batch ($p < 0.05$; **Fig. 2a**). The distribution of the sum of absolute correlations in the resulting 10,000 layouts suggested that the expected value for this sum was close to 0.5. Notably, the correlations present in such an average selection

were predominantly small (0.004 to 0.065; **Fig. 2bii**), but significant associations still prevailed ($p < 0.05$).

The worst-case scenario demonstrated that simple randomisation has potential to return layouts with multiple significant associations ($p < 0.05$ for five comparisons in this example; **Fig. 2biii**). This layout could result in substantial batch effects that may bias results. By contrast, *Omixer* would reject all such layouts with significant correlations, and instead return only a layout from the 15% remaining (blue in **Fig. 2a**). In this example, the optimal layout has no significant correlations, and all have absolute values below 0.037 (**Fig. 2bi**).

Omixer produces lab-friendly sample sheets

The *omixerSheet* function converts the output of previous Omixer functions into lab- and colour-blind friendly sample sheets, saving these in the working directory as a printable PDF. Wells can be coloured by specified variables, such as box number or tissue, to further smooth transition into the wet lab (using option: `group`). Such lab-friendly sample sheets improve accessibility of Omixer and reduce the risk of mix-ups when manually pipetting samples (**Fig. 2c**).

Conclusions

In conclusion, Omixer offers an intuitive, reproducible alternative to current randomization practices in genomic research. Its implementation is a key step in combatting batch effects pre-emptively and reducing the risks of sample mix-ups in the wet lab.

Author contributions

L.S. and B.T.H. conceived the project and designed the software. D.C. contributed key expertise in computation and software development.

Funding

This work was supported by the Joint Programming Initiative 'a Healthy Diet for a Healthy Life' (JPI-HDHL) DIMENSION project [ZonMW project number: 529051021].

Declarations of interest

The authors declare no competing interests.

Data & code availability

Scripts and data used to generate the Figures is available upon request. Otherwise, Omixer is a software tool that uses user input data.

References

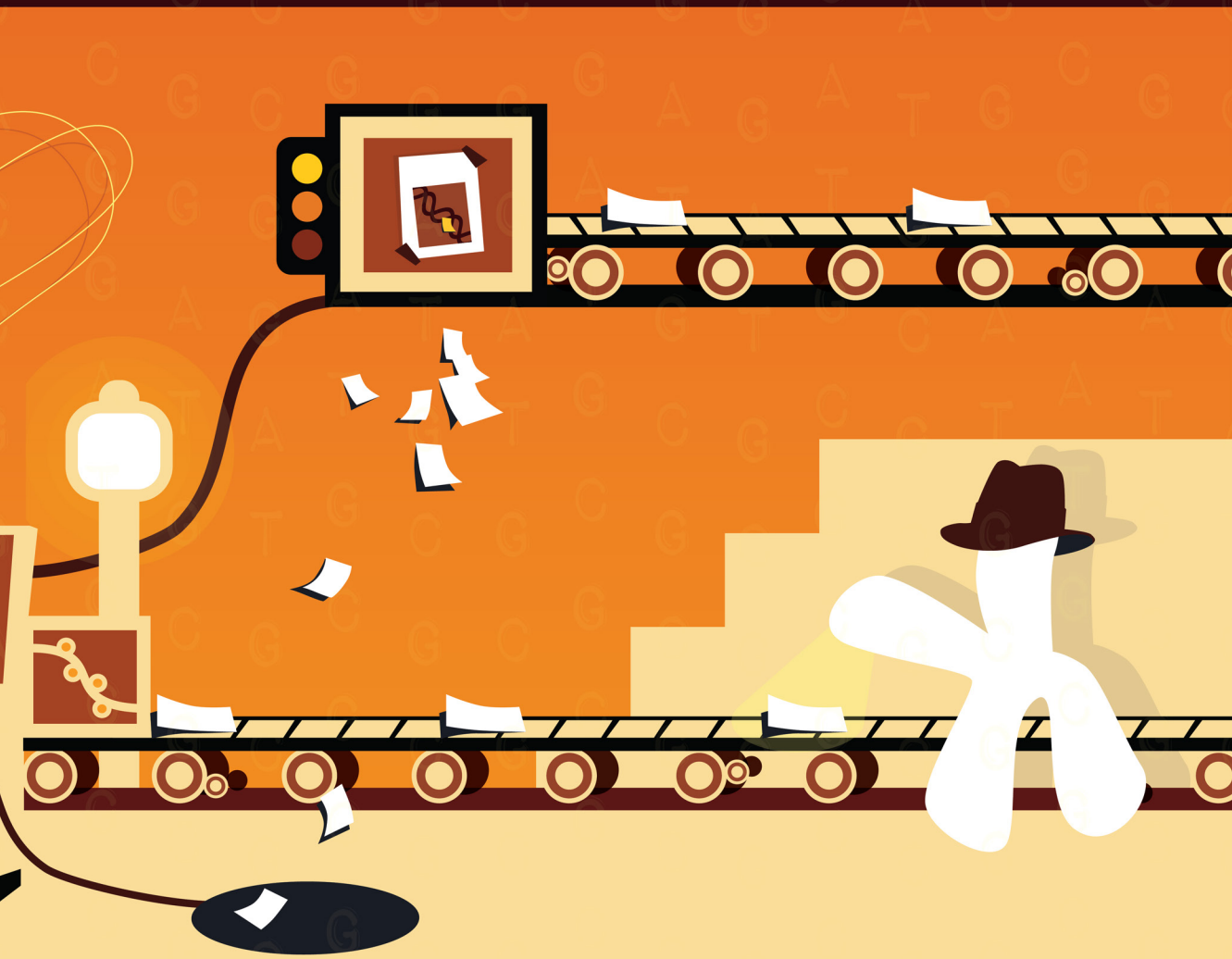
1. Baggerly, K., A., Coombes, K. R., *and* Neeley, E. S. Run batch effects potentially compromise the usefulness of genomic signatures for ovarian cancer. *J Clin Oncol* **26** (7): 1186-1187 (2008).
2. Harper, K. N., Peters, B. A., *and* Gamble, M. V. Batch effects and pathway analysis: Two potential perils in cancer studies involving DNA methylation array analysis. *Cancer Epidemiol Biomarkers Prev* **22** (6): 1052-1060 (2013).
3. Lambert, C. G. *and* Black, L. J. Learning from our GWAS mistakes: From experimental design to scientific method. *Biostatistics* **13** (2): 195–203 (2012).
4. Leek, J. T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* **11** (10): 733–739 (2010).
5. Espín-Pérez, A. *et al.* Comparison of statistical methods and the use of quality control samples for batch effect correction in human transcriptome data. *PLoS One* **13** (8): e0202947 (2018).
6. Johnson, W. E., Li, C., *and* Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8** (1): 118-127 (2007).
7. van Iterson, M. *et al.* Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18** (1): 19 (2017).
8. Goh, W. W. B., Wang, W., *and* Wong, L. Why Batch Effects Matter in Omics Data, and How to Avoid Them. *Trends Biotechnol* **35** (6): 498-507 (2017).
9. Yang, H. *et al.* Randomization in laboratory procedure is key to obtaining reproducible microarray results. *PLoS One* **3** (11): e3724 (2008).
10. Buhule, O. D. *et al.* Stratified randomization controls better for batch effects in 450K methylation analysis: A cautionary tale. *Front Genet* **5:** 354 (2014).

CHAPTER FIVE

DNAmArray

DNAmArray: streamlined workflow for the quality control, normalisation, and analysis of Illumina methylation array data

Lucy Sinke[1], Maarten van Iterson[1], Davy Cats[1], Tom Kuipers[1], and Bastiaan T. Heijmans[1]



[1] Leiden University Medical Centre, Leiden, The Netherlands

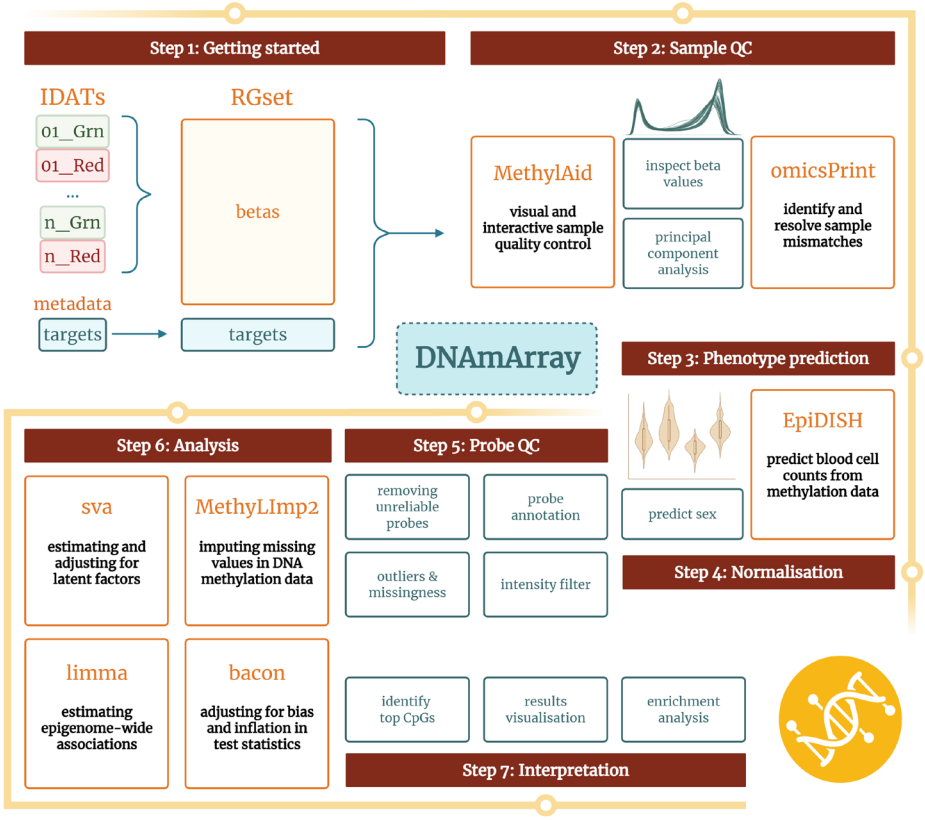
Published on Zenodo (2025)
DOI: 10.5281/zenodo.3332709

Abstract

Available packages for preprocessing, quality control, and analysis of DNA methylation data offer powerful and flexible functionality. However, these tools often exist as isolated elements, requiring users to navigate a complex landscape with limited guidance on how to implement them in a coherent workflow. A streamlined, well-documented pipeline that combines up-to-date processes and packages has potential to improve the accessibility of epigenome-wide analyses, facilitating good quality reporting that can keep up with data generation efforts. With the recent release of the Illumina Infinium Methylation EPIC v2.0 array in mind, we present *DNAmArray*, a well-documented workflow that details necessary fixes and enables the continued use of established methods in epigenomic studies.

DNAmArray is a modular R package and complementary workflow for the preprocessing, quality control, and analysis of DNA methylation array data, tailored for epigenome-wide association studies. Drawing on nearly a decade of experience analysing large-scale genomic data, most notably within the BIOS consortium, DNAmArray integrates widely accepted best practices and practical helper functions to streamline analyses in an up-to-date and user-friendly manner. The workflow takes raw IDAT files as input and transforms them via a series of clearly defined steps. Rather than prescribing a single best approach, DNAmArray transparently presents options, enabling users to make informed choices based on their specific data and research goals. Both the package and workflow are extensively validated on both Illumina 450K and EPIC arrays, with specific notes and adjustments provided for the EPIC v2.0 platform.

Graphical Abstract



Highlights

- *DNAmArray* is a modular workflow for accessible and user-friendly DNA methylation data processing and analysis.
- It is equipped with complementary documentation, which walks through an epigenome-wide analysis on publicly available example data.
- Both the pipeline and package are up-to-date and validated on Illumina 450k, EPIC, and EPIC v2.0 arrays.

Keywords: *DNA methylation; EWAS; data pipeline; reproducibility; Illumina*

Background

DNA methylation (DNAm) is a well-characterized epigenetic modification that plays a central role in regulating gene expression, cellular differentiation, and genomic stability[1–3]. Epigenome-wide association studies (EWAS) leverage DNAm data to identify CpG sites whose methylation is associated with phenotypes of interest, including immunometabolic traits, ageing, and environmental exposures[4–6]. Such studies have potential to offer mechanistic insights into the pathways underlying human health and disease.

To support these efforts, DNAm is often profiled using microarray-based technologies such as the Illumina Infinium HumanMethylation450 BeadChip (450K) and its successors, the Infinium MethylationEPIC (850K) and the recently released MethylationEPIC v2.0 array[7]. These platforms enable cost-effective, high-throughput measurement of methylation at hundreds of thousands of CpGs across the genome. However, despite technological advances, EWAS inherently share many statistical and technical challenges with genome-wide association studies (GWAS), including the need to correct for multiple testing, batch effects, and at times population structure. EWAS also present unique sources of confounding, including cell-type heterogeneity, age, and lifestyle influences[8]. To this end, several R packages have been developed to support various aspects of DNAm analysis, including *methylAid*[9], *omicsPrint*[10], and *bacon*[11], each providing specific tools for quality control (QC), preprocessing, and analysis. While powerful, these packages require thoughtful integration and manual customization to form a complete analysis pipeline compatible with the latest technology[12].

Here, we present *DNAmArray*, an R package and accompanying workflow designed to streamline the analysis of Illumina DNAm array data, including support for the EPIC v2.0 platform. *DNAmArray* integrates widely accepted methods from across the EWAS landscape with practical, in-house functions. It is informed by nearly a decade of work with large-scale datasets, including analysis of almost 4,000 samples from six Dutch biobanks as part of the Biobank-based Integrative Omics Study (BIOS) consortium. Rather than prescribing a rigid pipeline, *DNAmArray* emphasizes transparency, flexibility, and reproducibility by presenting multiple valid options in a modular framework. To aid adoption and implementation, we offer extensive documentation of the analysis workflow for an example EPIC dataset[13].

Results

Step 1 | Getting started

The example dataset used to illustrate the *DNAmArray* workflow is available from the NCBI Gene Expression Omnibus (GSE116339)[13]. It includes DNAm data profiled using the Illumina Infinium MethylationEPIC BeadChip array in 679 whole blood samples. The samples were collected from individuals enrolled in the Michigan PBB Registry, a longitudinal cohort established following an agricultural accident in the 1970s. This accident resulted in widespread exposure to polybrominated biphenyl (PBB) and, because PBB is both lipophilic and biologically stable, those exposed still have detectable levels in their blood many decades later. All samples were collected between 2004 and 2015, approximately 31 to 42 years after the initial exposure event. The resulting DNAm data has been previously analysed in an EWAS of PBB exposure and is well-suited for demonstrating preprocessing, normalisation, and analysis steps within the *DNAmArray* package.

Step 2 | Sample-level quality control

To ensure high data quality for downstream analyses, rigorous sample-level QC is conducted, detecting and excluding samples that exhibit signs of sample mixups or processing errors[14]. The input for this step is an RGset constructed from raw IDAT files of red and green channel intensities combined with sample metadata.

DNAmArray visualises sample quality using plots from the Bioconductor package, *methylAid*[9], utilising the control probes present on Illumina arrays. Each type of control probe is designed to evaluate a distinct stage of the DNAm protocol from bisulphite conversion through extension to hybridization. In addition, signals from negative control probes, which represent randomly permuted sequences and therefore should not hybridize to the DNA template, assess if a user-defined proportion of probes from each sample are visually distinct from background noise[15]. On the basis of these plots, outliers are detected and *DNAmArray* provides updated thresholds for not only the EPIC but also the EPICv2 array. All *MethylAid* plots accept custom thresholds and can be coloured by variables in the sample metadata (**Fig. 1a-d**). Furthermore, *DNAmArray* details how to explore and flag samples by inspecting their β-value distributions and outlines principal component approaches to identify sample clusters or outliers.

As an important part of sample-level QC, *DNAmArray* implements *omicsPrint*[10]. This package detects data linkage errors by comparing recorded sample relationships to those predicted from probes that contain common single nucleotide polymorphisms (SNPs). By examining genetic similarity, samples are inspected for duplicates, underlying family relationships, or sample mix-ups, which can then be resolved or corrected for.

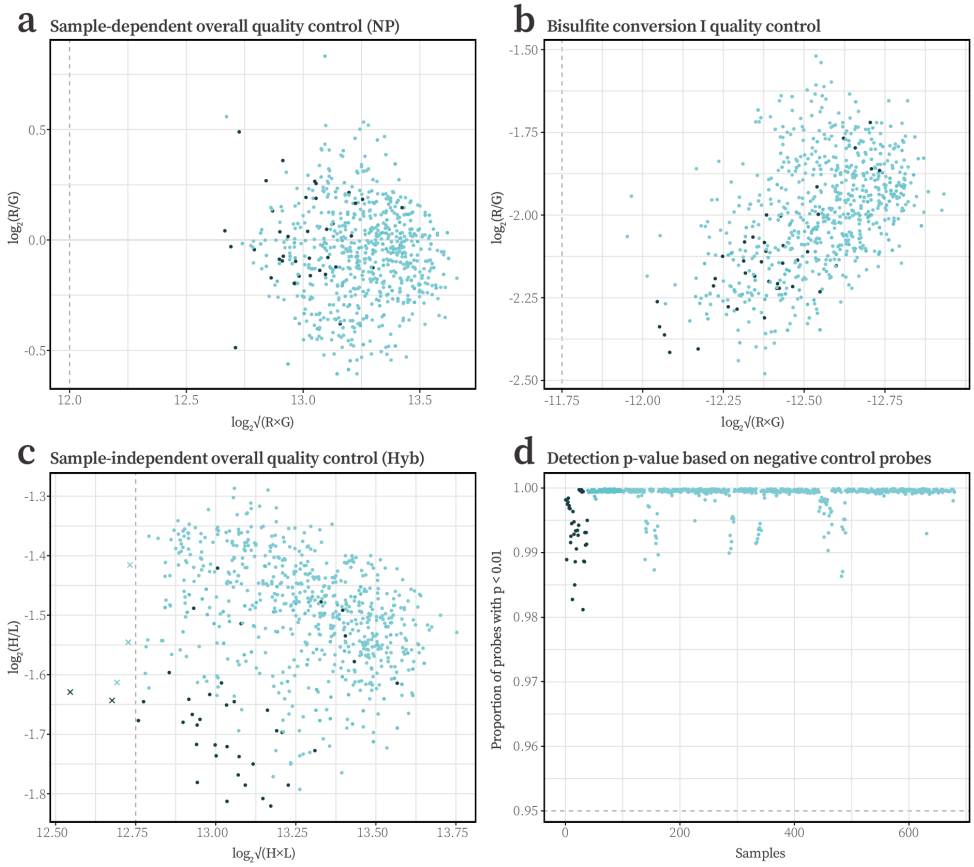


Figure 1 | *MethylAid* plots provided within *DNAmArray*. a) Sample-dependent overall QC based on non-polymorphic (NP) control probes, ensuring that both A and T (red) and C and G (green) nucleotides are detected in sufficient intensities. **b)** Bisulphite-conversion (BC) plot visualises if the conversion reaction was successful for all samples. Converted (C) probes are extended and measured in the defined colour channel (C1 and C2: green; C3 and C4: red). **c)** Sample-independent overall QC measured using hybridization (Hyb) controls. These synthetic targets complement the array perfectly and are present in the hybridization buffer at three concentration levels. This plot ensures distinction between high (H) and low (L) concentrations in the green channel. In the example data, poor hybridization is evident for five samples, which are removed from downstream analysis. **d)** Detection *p*-value plot displaying the proportion of probes (y-axis) that are distinguishable from the background noise (default $p < 0.01$). Background signal is determined using control probes of randomly permuted sequences and thereby not designed to hybridize to the DNA template. The user can specify both the proportion and *p*-value thresholds for this plot.

Step 3 | Predicting phenotypes

The *DNAmArray* workflow implements phenotype prediction, offering tools to validate sample identity, impute missing biological data, and adjust for variable factors. These predictions are particularly important considering the sensitivity of EWAS to biological confounding[22]. DNAm-based predictors of immune cell proportions are increasingly able to capture a larger amount of cellular diversity, and *DNAmArray* outlines how to estimate such subsets and add them to metadata and models (**Fig. 2a**)[23].

Furthermore, *DNAmArray* outlines steps from the *wateRmelon* package for sex prediction, offering a user-friendly way to confirm and resolve sample identity and check for less common karyotypes (**Fig. 2b**)[17]. To ensure optimal predictions, the workflow ensures this step is performed prior to probe masking. Many commonly masked probes, such as those on sex chromosomes, are essential components for prediction algorithms, and implementation on incomplete data is likely to reduce their accuracy. By comparing predicted and measured traits, *DNAmArray* increases confidence in data integrity, enriches metadata, and assists in downstream analyses.

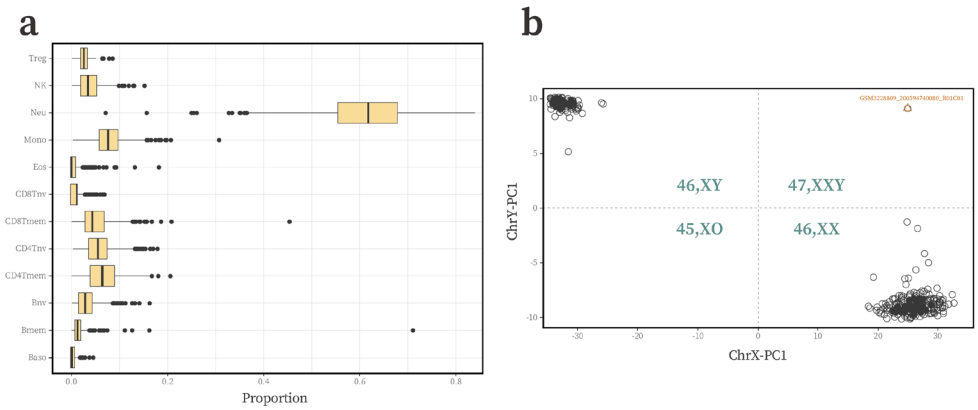


Figure 2 | Phenotype prediction in the *DNAmArray* workflow. a) Twelve immune cell type proportions are predicted from DNAm data using *EpiDISH*. **b)** DNAm at CpGs located on sex chromosomes is used to predict sex, including XXY and XO karyotypes.

Step 4 | Normalisation

Whilst β-values derived from the Illumina arrays are relatively robust due to their ratio-based formulation, this intrinsic stability does not preclude the presence of unwanted technical variation[16]. For studies aiming to detect subtle, biologically meaningful changes, refined correction of such technical variation is essential, as unaddressed technical noise can obscure methylation patterns. This can compromise both statistical power and reproducibility.

DNAmArray outlines implementation of multiple normalisation approaches, including *dasen* and functional normalization (*FunNorm*), using the *wateRmelon* Bioconductor package[17]. *FunNorm* is an approach designed specifically for Illumina methylation arrays that extends quantile normalization through the use of control probes[18]. Unlike traditional quantile normalization, which enforces identical distributions across samples, *FunNorm* retains biologically meaningful differences by estimating and removing variation associated with latent technical factors of the control probe signal. This control-informed approach offers a principled way to improve comparability across samples while minimizing the risk of overcorrection, particularly in biologically informative regions of the methylation spectrum (i.e., intermediate β-values around 0.5). By preserving these subtleties, *FunNorm* enhances the sensitivity of downstream differential methylation analysis[19].

Step 5 | Probe-level quality control

Data-driven probe-level QC is the final data processing step outlined by *DNAmArray*. It provides functions aimed at excluding unreliable data points, including those with zero intensity or based on fewer than three beads. Probes with more than 5% missing data across samples are removed to maximize data completeness and reduce bias. Additionally, probes previously established as unreliable due to cross-hybridization, ambiguous genomic mapping, or the presence of SNPs at critical binding sites are excluded[20]. Users are provided with curated probe masking objects for the 450K, EPIC, and EPIC v2.0 arrays, which flag CpGs located in ENCODE Blacklist regions[21]. To further safeguard against technical artifacts, *DNAmArray* includes options for detecting and handling spurious methylation values using outlier-based filtering approaches, which can be applied to both β and M values.

Following probe QC, the cleaned β-matrix, filtered probe annotations, and corresponding sample metadata are combined into a *SummarizedExperiment* object. This structure maintains internal consistency across assay data and phenotypic annotations and is compatible with existing Bioconductor packages and workflows.

Step 6 | EWAS pipeline

The *DNAmArray* package provides a comprehensive and customisable pipeline for performing EWAS, aimed at improving the accessibility of DNAm data analysis. To facilitate informed model specification, we outline steps to visualise relationships between potential covariates and methylation data, estimate latent factors, and incorporate random effects (**Fig. 3a**). Latent factors address sources of variation not captured by measured variables, and *DNAmArray* implements surrogate variable analyses (SVA) to estimate hidden batch effects in a data-driven approach. Proper specification of full and null models is described, alongside best practices for imputing DNAm using

methyLImp2[25]. By adjusting for both predicted confounding factors, such as estimated cell types, and unmeasured confounders including surrogate variables (SVs), the robustness and interpretation of resulting EWAS findings can be improved.

At its core, the EWAS component relies on the widely used *limma* package to perform linear modeling at all tested CpGs[24]. Steps are outlined to properly specify appropriate models including both fixed and random effects. Yet, even with proper model specification the test statistics from epigenome- and transcriptome-wide studies are susceptible to bias and inflation[11]. Therefore, *DNAmArray* implements *bacon*, a Bayesian method that controls these effects using the empirical null distribution. Correct implementation of this package is outlined, alongside complementary visualisations to assess performance (**Fig. 3b**).

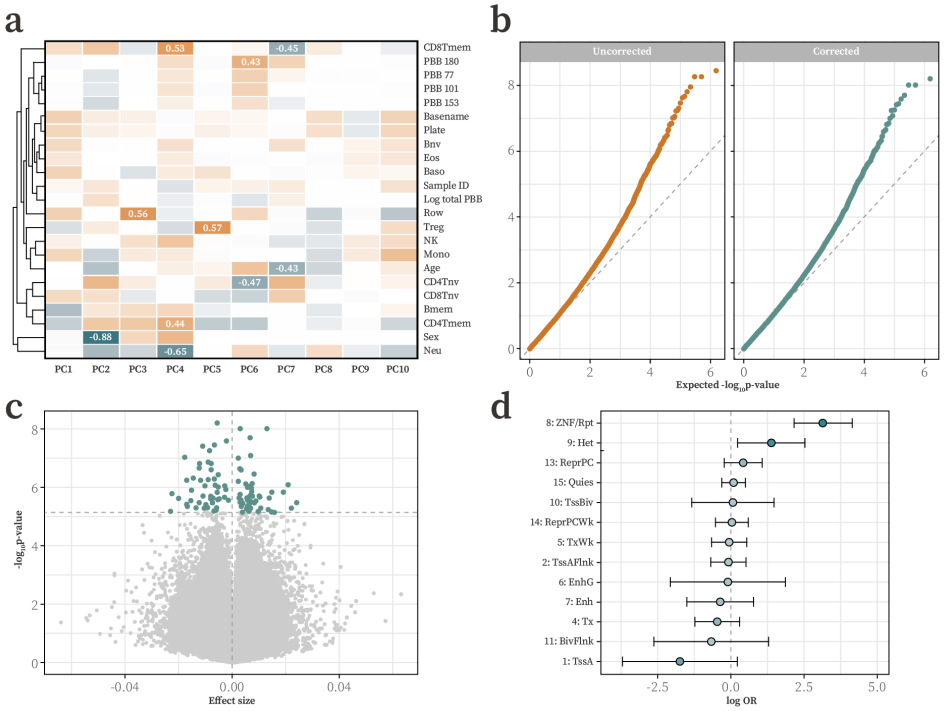


Figure 3 | Analysis steps within the *DNAmArray* workflow. a) Methods to properly specify models including heatmaps and principal component analysis (PCA) are included. **b)** Application of *bacon* allows test statistics to be adjusted for bias and inflation. **c)** EWAS results are visualized using volcano plots. **d)** Enrichment analyses, including for chromatin state, are described and visualised.

Step 7 | Interpretation of results

There is a pressing need to shift from hypothesis generation to biological interpretation in EWAS. Many CpGs have now been robustly associated with human health but remain incompletely characterised. Therefore, *DNAmArray* dedicates its final step to functional interpretation of results. Following standard steps to adjust for multiple testing, identify significant CpGs, and visualise results (**Fig. 3c**), the workflow guides users through probe annotation and chromatin state enrichment analyses (**Fig. 3d**). This serves to exemplify methods and helpful resources that be used to more fully characterise EWAS signals[26]. In this manner, *DNAmArray* paves the way for a user-driven improvement in EWAS interpretation in both novel and existing data.

Discussion

The development of *DNAmArray* addresses a gap in the transparency, reproducibility, and accessibility of EWAS analyses, particularly in the context of rapidly evolving array technologies. As the field transitions from the Illumina 450K to its successors, researchers face ongoing challenges in maintaining robust, reproducible, and up-to-date analysis workflows[27]. *DNAmArray* responds to these challenges by providing a modular, well-documented R package and workflow, integrating best practices from the epigenomics community with practical, experience-driven enhancements.

By validating steps on both 450K and EPIC arrays and providing clear guidance for EPIC v2.0, the workflow ensures continuity for researchers transitioning to new technologies. The emphasis on rigorous QC at both the sample and probe level improves the quality of input data for EWAS, and tools such as *methylAid*[9] and *omicsPrint*[10] alongside curated probe masking resources enable users to identify and address technical artifacts and data linkage errors that could otherwise compromise downstream analyses[20]. The inclusion of flexible, user-adjustable thresholds and visualisation options additionally allows QC procedures to be tailored to a range of datasets. Beyond CpG associations, the workflow highlights the importance of functional interpretation for translating EWAS findings into insights with biological context. By outlining clear steps for such analyses and describing extensions, *DNAmArray* helps bridge the gap between statistical results and biological understanding.

While *DNAmArray* represents a significant advancement, certain limitations remain. The workflow's focus on array-based technologies means that it is not suited to perform analysis on data outside of targeted CpG sites, and the accuracy of phenotype prediction and cell-type deconvolution depends on the quality and relevance of available reference data. As new array platforms and sequencing-based approaches emerge, ongoing maintenance and community engagement will be essential to ensure that *DNAmArray* remains current and widely applicable. Additionally, data pipelines can only process and adjust for measured variables. Other packages, such as *Omixer*[28], allow researchers to proactively reduce the impact of technical variation on the quality of their data and biological signals of interest, and study designs should rely on a combination of proactive measures alongside reproducible preprocessing and analysis to ensure robust results.

In summary, *DNAmArray* provides a timely, flexible, and user-friendly solution for the preprocessing, quality control, and analysis of Illumina DNAm array data. By integrating best practices with practical enhancements, and by supporting the latest array technologies, DNAmArray enables reproducible and interpretable EWAS, facilitating future discoveries in epigenetic epidemiology.

Author contributions

B.T.H. and M.v.I. conceived and initially designed the pipeline. L.S. and B.T.H. updated it to its current iteration, developed the complementary workflow, and tested the pipeline on example datasets. D.C. and T.K. provided bioinformatics support. T.K. additionally contributed scripts to implement *MethylAid* in the workflow.

Funding

Development of the workflow was funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007). The work of L.S. was supported by the Joint Programming Initiative 'a Healthy Diet for a Healthy Life' (JPI-HDHL) DIMENSION project [ZonMW project number: 529051021].

Declarations of interest

The authors declare no competing interests.

Data & code availability

DNAmArray is available from GitHub at `molepi/DNAmArray` and all versions are archived on Zenodo[29].

All software used is open source and freely available Unless stated otherwise, all calculations were performed using R version 4.2.2.

Acknowledgements

We are grateful to Jenny van Dongen for continued testing and improvement of the workflow. In addition, we thank Paul Hop, Jazmin Taubert, Yunfeng Liu, Manhoor Sulaiman, Thomas Jonkman, Helena Rasche, Elmar W. Tobi, Roderick Slieker, Wouter den Hollander, Rene Luijk, and Koen F. Dekkers for contributing to the testing and development of the package and associated workflow.

References

1. Hannon, E. *et al.* Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. *Am J Hum Genet* **103** (5): 654-665 (2018).
2. Suelves, M. *et al.* DNA methylation dynamics in cellular commitment and differentiation. *Brief Funct Genomics* **15** (6): 443-453 (2016).
3. Héberlé, É. *and* Bardet, A. F. Sensitivity of transcription factors to DNA methylation. *Essays Biochem* **63** (6): 727-741 (2019).
4. Wielscher, M. *et al.* DNA methylation signature of chronic low-grade inflammation and its role in cardio-respiratory diseases. *Nat Commun* **13** (1): 2408 (2022).
5. Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541** (7635): 81-86 (2017).
6. Dekkers, K.F. *et al.* Blood lipids influence DNA methylation in circulating cells. *Genome Biol* **17** (1): 138 (2016)
7. Noguera-Castells, A. *et al.* Validation of the new EPIC DNA methylation microarray (900K EPIC v2) for high-throughput profiling of the human DNA methylome. *Epigenetics* **18** (1): 2185742 (2023).
8. Van Rooij, J. *et al.* Evaluation of commonly used analysis strategies for epigenome- A nd transcriptome-wide association studies through replication of large-scale population studies. *Genome Biol* **20** (1): 235 (2019).
9. Van Iterson, M. *et al.* MethylAid: visual and interactive quality control of large Illumina 450k datasets. *Bioinformatics* **30** (23): 3435-3437 (2014).
10. Van Iterson, M. *et al.* omicsPrint: detection of data linkage errors in multiple omics studies. *Bioinformatics* **34** (12): 2142-2143 (2018).
11. van Iterson, M. *et al.* Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol* **18** (1): 19 (2017).
12. Ori, A. P. S. *et al.* Significant variation in the performance of DNA methylation predictors across data preprocessing and normalization strategies. *Genome Biol* **23** (1): 225 (2022).
13. Curtis, S. W. *et al.* Exposure to polybrominated biphenyl (PBB) associates with genome-wide DNA methylation differences in peripheral blood. *Epigenetics* **14** (1): 52-66 (2019).
14. Bhat, B. *and* Jones, G. T. Data Analysis of DNA Methylation Epigenome-Wide Association Studies (EWAS): A Guide to the Principles of Best Practice. *Methods Mol Biol* **2458**: 23–45 (2022).
15. Heiss, J. A. *and* Just, A. C. Improved filtering of DNA methylation microarray data by detection p values and its impact on downstream analyses. *Clin Epigenetics* **11** (1): 15 (2019).
16. Xu, Z. *and* Taylor, J. A. Reliability of DNA methylation measures using Illumina methylation BeadChip. *Epigenetics* **16** (5): 495-502 (2020).
17. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**: 293 (2013).
18. Fortin, J. P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol* **15** (12): 503 (2014).
19. Heiss, J. A. *and* Brenner, H. Between-array normalization for 450K data. *Front Genet* **6:** 92 (2015).
20. Zhou, W., Laird, P. W. *and* Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res* **45** (4): e22 (2017).
21. Amemiya, H. M., Kundaje, A. *and* Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9** (1): 9354 (2019).
22. Teschendorff, A. E. *and* Zheng, S. C. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. *Epigenomics* **9** (5): 757-768 (2017).
23. Zheng, S. C. *et al.* EpiDISH web server: Epigenetic dissection of intra-sample-heterogeneity with online GUI. *Bioinformatics* **36** (6): 1950-1951 (2020).
24. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43** (7): e47 (2015).
25. Plaksienko, A. *et al.* methyLImp2: faster missing value estimation for DNA methylation data. *Bioinformatics* **40** (1): btae001 (2024).
26. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38** (4): 576-589 (2010).
27. Solomon, O. *et al.* Comparison of DNA methylation measured by Illumina 450K and EPIC BeadChips in blood of newborns and 14-year-old children. *Epigenetics* **13** (6): 655-664 (2018).
28. Sinke, L., Cats, D. *and* Heijmans, B. T. Omixer: multivariate and reproducible sample randomization to proactively counter batch effects in omics studies. *Bioinformatics* **37** (18): 3051–3052 (2021).
29. Sinke, L. *et al.* DNAmArray: Streamlined workflow for the quality control, normalization, and analysis of Illumina methylation array data. *Zenodo* (2025).

CHAPTER SIX

Growing Old TOgether

Tissue-specific methylomic responses to a lifestyle intervention in older adults associate with metabolic and physiological health improvements

Lucy Sinke[1], Marian Beekman[1], Yotam Raz[1], Thies Gehrmann[1,2], Ioannis Moustakas[1], Alexis Boulinguiez[3], Nico Lakenberg[1], Eka Suchiman[1], Fatih A. Bogaards[1,4], Daniele Bizzarri[1,5], Erik B. van den Akker[1,5], Melanie Waldenberger[6,7], Gillian Butler-Browne[3], Capucine Trollet[3], C. P. G. M. Lisette de Groot[4], Bastiaan T. Heijmans[1], and P. Eline Slagboom[1]

[1] Leiden University Medical Centre, Leiden, The Netherlands
[2] University of Antwerp, Antwerp, Belgium
[3] Sorbonne University, Paris, France
[4] Wageningen University and Research, Wageningen, The Netherlands
[5] Delft Bioinformatics Lab, Delft, The Netherlands
[6] Helmholtz Munich, Neuherberg, Germany
[7] German Centre for Cardiovascular Research (DZHK), Munich, Germany

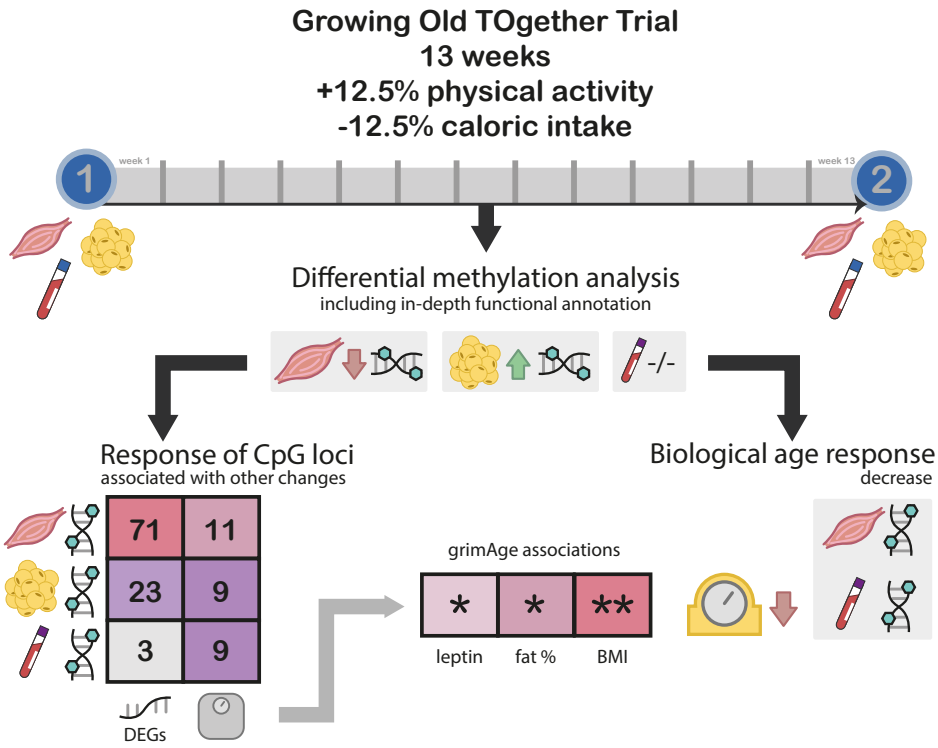
Published in Aging Cell 24(4):e14431 (2025)
DOI: 10.1111/acel.14431

Abstract

Across the lifespan, diet and physical activity profiles substantially influence immunometabolic health. DNA methylation, a tissue-specific epigenetic marker sensitive to behavioural change, may mediate these effects through modulation of transcription factor binding and subsequent gene expression. Despite this, few human studies have profiled DNA methylation and gene expression simultaneously in multiple tissues or examined how molecular levels react and interact in response to lifestyle changes.

The Growing Old Together (GOTO) study was a 13-week lifestyle intervention in older adults, which imparted health benefits to participants. Here, we characterised the DNA methylation response to this intervention at over 750,000 CpGs in muscle, adipose, and blood. Differentially methylated sites were enriched for active chromatin states, located close to relevant transcription factor binding sites, and associated with alterations in insulin sensitivity genes and health parameters. In addition, measures of biological age were consistently reduced by the intervention and decreases in grimAge could be connected to observed health improvements. Taken together, our results reveal and interrogate sets of responsive tissue-specific loci and outline their potential to measure progression and finetune treatment of age-related risks and diseases.

Graphical Abstract



Highlights

- DNA methylation and gene expression are profiled in three tissues before and after a 13-week lifestyle intervention in older adults.
- Differentially methylated sites predominantly reside in active chromatin in close proximity of transcription factor binding sites.
- This methylation response correlates with differential expression of insulin sensitivity genes in muscle and adipose tissues.
- Changes in all three tissues associate with decreases in body fat percentage and leptin levels.
- Tissue-specific epigenetic age algorithms can capture intervention effects and connect them to observed health improvements.

Keywords: *DNA methylation, epigenomics, functional genomics, healthy ageing, lifestyle, metabolism, muscle*

Background

Changes in behaviour across the life course, including adherence to healthy diets and physical activity regimens, have major health impacts. In some cases, they are more effective at improving immunometabolic health than pharmacological interventions[1–3]. Regular aerobic exercise alongside caloric restriction can promote weight loss, insulin sensitivity, and glucose control in both younger and older populations[4–7]. Epigenetic regulation, such as through DNA methylation (DNAm), may mediate a portion of these health benefits by modulating the accessibility of regulatory sites for transcription[8–10]. Physical activity has been found to attenuate the age-dependent decreases in DNAm of the anti-inflammatory *ASC* gene in blood[11,12], and Mendelian randomization has directionally linked epigenetic signatures of a healthy diet with both type 2 diabetes and several of its risk factors[13]. These findings highlight potential for DNAm in measurement and modification of immunometabolic health in individuals of all ages.

Whilst epigenomic responses to intervention studies have revealed promising results in blood, epigenetic reprogramming of other metabolic tissues may have even greater functional consequences on health[14]. Muscle and adipose tissues secrete a plethora of proteins and signalling molecules into the circulation and engage in tissue-to-tissue crosstalk, collectively bringing about clinically meaningful biological changes[15]. Although several experimental studies have investigated the effects of lifestyle interventions on the methylome of muscle[16,17] or adipose[18–20] individually, few researchers have taken a multi-tissue approach despite this showing promise in other-omic fields[21–23]. To advance our understanding of how epigenetics influences immunometabolic health, we must diversify and expand our study designs by incorporating relevant tissues, improving CpG coverage, and including the older adults who represent a growing proportion of our populations.

The Growing Old Together (GOTO) study is a 13-week lifestyle intervention in 164 older adults (mean age 63 years), which expanded on the combined intervention arm of the CALERIE study[24]. Here, we followed up on previous work indicating that the GOTO intervention conferred an immunometabolic health improvement to participants[25,26] and that this benefit associates with changes in the blood[27], adipose, and muscle transcriptomes[28] and the blood metabolome[29]. Using data and biomaterial from before and after the GOTO study, we profiled DNAm at over 750,000 CpG sites across the genome in skeletal muscle (n = 80), subcutaneous adipose (SAT, n = 89), and fasted blood tissues (n = 98). By thoroughly characterising the resulting loci, we examined how methylomic responses to the GOTO intervention relate to genomic regulation and differential gene expression in *cis*, with implications for immunometabolic health and epigenetic measures of chronological and biological age.

Results

The GOTO intervention improves metabolic health comparably across tissue-specific subsets of participants

The GOTO intervention (n = 164) imparted a range of metabolic health benefits, described in detail previously[25,27–29]. Notably, participants saw reductions in their body mass index (BMI, Δ = −1.1 kg/m^2), waist circumference (WC, Δ = −4.3 cm), and total body fat percentage (Δ = −1.8%) alongside improvements in other health measurements (**Table 1**). Baseline characteristics are shown in **Supplementary Table 1**. Individuals were selected for DNAm profiling based on availability of biological material and gene expression data, and for a majority (n = 66, 64.7%), data was collected from all three tissues both before and after the intervention.

	Entire GOTO population (n = 164)		Three tissue overlap (n = 66)		Skeletal muscle (n = 80)		Adipose tissue (n = 89)		Fasted blood (n = 98)	
	Δ(SE)	**Padj**	**Δ(SE)**	**Padj**	**Δ(SE)**	**Padj**	**Δ(SE)**	**Padj**	**Δ(SE)**	**Padj**
Body mass index (kg/m^2)	-1.13 (0.06)	2.0E-39	-1.24 (0.10)	1.2E-17	-1.28 (0.09)	4.2E-23	-1.27 (0.08)	1.5E-26	-1.28 (0.08)	1.9E-29
WC (cm)	-4.32 (0.42)	5.6E-19	-4.83 (0.66)	2.7E-09	-5.10 (0.60)	5.2E-12	-4.53 (0.58)	5.6E-11	-4.65 (0.55)	9.8E-13
Total body fat (%)	-1.76 (0.23)	8.7E-12	-2.13 (0.35)	2.3E-07	-2.03 (0.34)	9.4E-08	-2.18 (0.30)	2.7E-10	-2.05 (0.30)	1.9E-09
Fasting insulin (mU/L)	-0.31 (0.25)	2.7E-01	-0.83 (0.41)	5.8E-02	-0.55 (0.38)	1.7E-01	-0.45 (0.38)	2.7E-01	-0.46 (0.34)	2.0E-01
Systolic BP (mm Hg)	-3.15 (0.94)	1.6E-03	-2.82 (1.68)	1.1E-01	-2.66 (1.46)	1.0E-01	-2.88 (1.34)	4.9E-02	-2.51 (1.25)	6.8E-02
Leptin (µg/L)	-2.32 (0.34)	3.7E-10	-2.75 (0.44)	1.4E-07	-2.72 (0.38)	9.4E-10	-2.95 (0.40)	2.7E-10	-2.86 (0.37)	1.9E-11
Adiponectin (mg/L)	0.26 (0.14)	1.1E-01	0.52 (0.25)	5.8E-02	0.45 (0.21)	6.1E-02	0.27 (0.21)	2.5E-01	0.28 (0.20)	2.0E-01
Interleukin-6 (ng/L)	0.09 (0.11)	4.8E-01	0.30 (0.15)	5.8E-02	0.23 (0.31)	1.1E-01	0.24 (0.11)	4.9E-02	0.26 (0.12)	5.8E-02
HDL cholesterol (mmol/L)	-0.01 (0.02)	6.2E-01	0.00 (0.03)	9.3E-01	0.00 (0.03)	9.4E-01	-0.02 (0.02)	5.3E-01	-0.02 (0.02)	4.8E-01
Fasting HDL size (nm)	0.04 (0.01)	5.7E-09	0.06 (0.01)	1.3E-05	0.07 (0.01)	1.9E-08	0.05 (0.01)	6.6E-07	0.05 (0.01)	3.0E-08

Table 1 | Effects of the 13-week GOTO intervention on ten immunometabolic health measurements in the entire population and each tissue dependent subset. Associations were calculated using linear mixed models with fixed effects for age and sex and a random effect for ID.

Tissue-specific methylation subsets (muscle n = 80, SAT n = 89, and blood n = 98) were representative of the whole study population, with the distribution of changes in ten health parameters from included and excluded individuals being statistically comparable (nonresponse analysis $p_{FDR} \geq 0.05$; **Supplementary Table 2**). The sole exception was a selection bias for individuals with higher high-density lipoprotein (HDL) cholesterol sizes in the muscle subset ($p_{FDR} = 0.001$) urging caution in making inferences

about this trait in muscle analyses. In each tissue, we analysed the genome-wide DNAm consequences of the GOTO intervention, adjusting for age, sex, smoking status, technical covariates, and the first five principal components (PCs). In skeletal muscle and SAT samples, estimated bias and inflation of the test statistics was low ($|\mu| < 0.05$, $\lambda < 1.1$). In the blood samples, there was some deflation in the test statistics ($\lambda = 0.86$) alongside minimal bias ($\mu = 0.03$). For all tissues, bias and inflation of the test statistics was corrected for and residual values were under 0.01 and equal to 1.0 respectively, indicating high quality data.

In skeletal muscle, the GOTO intervention influenced DNAm at 162 predominantly hypomethylated CpGs

To interrogate the muscle methylomic response, we profiled DNAm in skeletal muscle samples biopsied before and after the GOTO intervention (n = 160 samples, 80 individuals). Since cell-type proportions can be an important driver of epigenetic signals, we predicted proportions of seven muscle nuclei types in our samples by applying the *MuSiC* deconvolution algorithm[30] to bulk gene expression data and a publicly available single nuclei reference transcriptome[31]. At baseline, our samples were primarily composed of slow (type I, mean 36.0%) and fast (type II, mean 26.7%) skeletal muscle fibres and endothelial cells (mean 36.6%). Following the intervention, the proportion of predicted endothelial nuclei had increased (Δ = +3.3%, p_{FDR} = 3.5E−03), in line with expected angiogenesis during the intervention[32]. There was insufficient evidence to suggest changes for any other nuclei type ($p_{FDR} \geq 0.05$; **Fig. 1a; Supplementary Table 3**).

In our initial unadjusted model, we identified 354 differentially methylated CpGs following the intervention ($p_{FDR} < 0.05$). However, considering the finding that an increase in endothelial nuclei could have been driving a portion of this methylation signal, we further adjusted our model for predicted endothelial nuclei proportions. This led to the removal of 192 CpGs from our results, leaving 162 predominantly hypomethylated (87.7%) CpGs where DNAm changes were independent of endothelial nuclei proportions ($p_{fdr} < 0.05$; **Fig. 1b; Supplementary Table 4**). Henceforth, we refer to this set of 162 differentially methylated CpGs in skeletal muscle, which represented 160 distinct loci, as the *muscle CpGs*.

CpGs influenced by the GOTO intervention associate with genes important for translocation of GLUT4 to the muscle cell membrane

To investigate potential for the *muscle CpGs* to regulate nearby gene expression, we annotated their genomic positions to 15 chromatin states using reference epigenomes from the Roadmap Epigenomics Consortium[33]. These consisted of eight active and seven repressed configurations that show distinct levels of DNAm, accessibility, and regulator binding. By testing if the *muscle CpGs* were enriched for specific genomic features in the

male (E107) and female (E108) skeletal muscle reference, we revealed that both enhancers (OR_{E107} = 5.83; OR_{E108} = 7.15) and genic enhancers (OR_{E107} = 3.56; OR_{E108} = 4.34) were overrepresented in our results (**Fig. 1c; Supplementary Table 5**). Since the primary mechanism that DNAm influences nearby expression is through transcription factor (TF) binding, we also tested if sequences within 50 bp of the *muscle CpGs* were enriched for known TF binding sites compared to a GC-matched random background (TFBS; **Fig. 1d; Supplementary Table 6**)[34]. The tested regions were enriched for 21 TFBS including ones upregulated by exercise (JunB: 34 CpGs)[35] and critical for muscle regeneration (Fos: 35 CpGs; Fra1: 33 CpGs)[36–38].

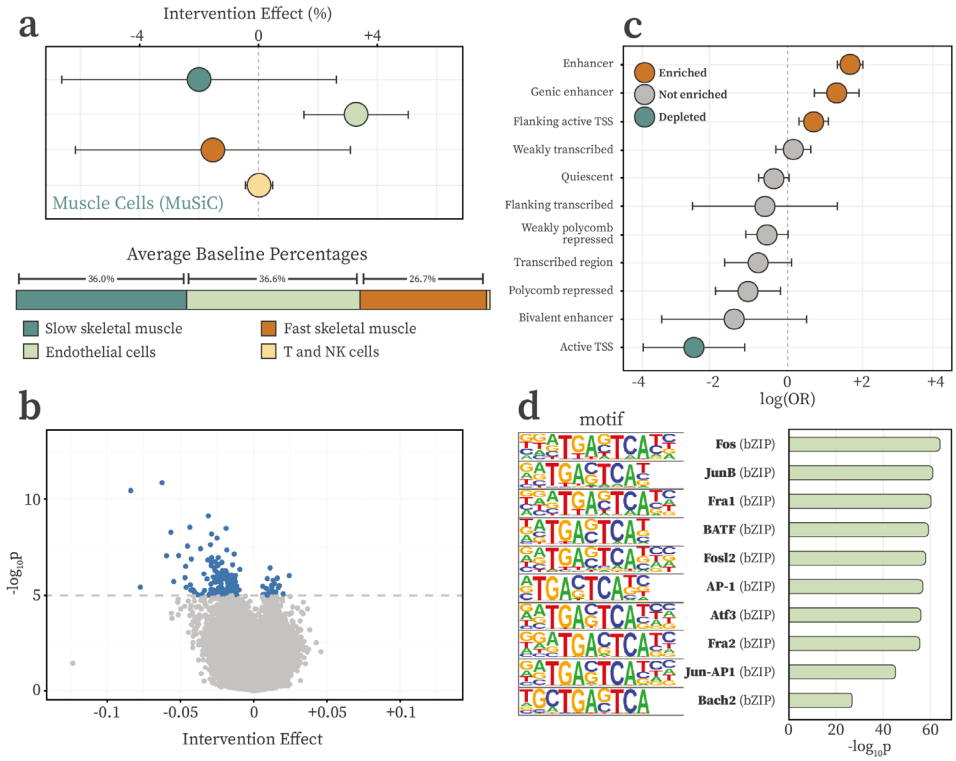


Figure 1 | Characterisation of the muscle cell count and DNAm response to the GOTO intervention. a) Intervention effect on predicted muscle nuclei types alongside baseline proportions (only cells >0.5% at baseline shown). **b)** Volcano plot of the intervention effect on DNAm in muscle at over 750,000 CpGs, showing 162 differentially methylated (blue) and non-significant CpGs (grey). **c)** Forest plot of the odds ratio (OR) and 95% confidence intervals (CIs) for enrichment (orange) or depletion (blue) of 11 chromatin states investigated using the male skeletal muscle reference epigenome from the Roadmap Epigenomics Consortium (E107; four states with extremely wide CIs not shown). **d)** Bar plot of the top ten transcription factor binding site (TFBS) motifs enriched in sequences within 50 bp of identified CpGs when compared to a GC-matched random genomic background.

Having established the *muscle CpGs* as plausibly regulatory, we used a two-step approach to identify their candidate target genes. First, we evaluated if expression of genes in close proximity (±100 kb) to the 162 muscle CpGs was altered by the intervention using available RNA-seq data ($p_{FDR} < 0.05$)[28]. Next, we examined whether these expression changes were associated with differential DNAm at the nearby CpG ($p_{FDR} < 0.05$). In total, there were 454 unique genes within 100kb of a muscle CpG, and 71 of these were both differentially expressed following the intervention and associated with the DNAm response in *cis* (**Supplementary Table 7**). This set of 71 genes included several directly implicated in the translocation of GLUT4 transporters to the muscle cell membrane in response to insulin and contractile activity (*TMOD3, FDFT1,* and *PLEKHG4*)[39–41] alongside an adaptor protein which regulates insulin signalling, specifically in skeletal muscle cells (*GRB10*)[42,43]. Over-representation analysis of these 71 genes revealed enrichment for *Striated Muscle Cell Development* after adjusting for multiple testing ($p_{FDR} = 0.019$), further clarifying the relevance of these genes to muscle-specific functions.

Altered blood-based health markers and grip strength associate with DNAm responses in skeletal muscle

We investigated whether the DNAm at the *muscle CpGs* was associated with changes in the ten immunometabolic health parameters shown in **Table 1** after adjusting for multiple testing (**Fig. 2a; Supplementary Table 8**). Methylation responses at 33 (20.4%) sites were connected to at least one blood-based trait, with eight CpGs linked to improvements in three or more traits ($p_{FDR} < 0.05$).

To explore if methylomic responses in muscle related to physiological adaptations within the same tissue, we expanded our analyses to include three additional muscle-specific phenotypes. This included average dominant hand grip strength as a marker of overall muscle performance, as well as two immunohistochemistry measures ($n = 65$, 81% of the original muscle subset; **Supplementary Table 9**)[44], namely the number of PAX7 positive cells and myonuclei per fibre. PAX7 is a satellite cell marker indicative of muscular regenerative potential[45], and higher numbers of myonuclei per fibre align with larger and stronger muscle fibres[46,47]. Both immunohistochemistry measures were associated with DNAm responses at more than 70 differentially methylated CpGs ($n_{MYO} = 75$, $n_{PAX7} = 84$), and DNAm at 36 CpGs was also correlated with average dominant hand grip strength. In total, over half of the *muscle CpGs* were linked to at least one of the investigated blood- or muscle-based traits, demonstrating relevance for this set of CpGs to observed health improvements ($n = 103$, 61.7%).

More specifically, there were 16 CpGs with multiple lines of evidence supporting their regulatory and clinical potential. These sites were differentially methylated, located in *cis* regulatory regions, and also associated with both differential gene expression and observed health benefits. An example of one such CpG was cg21005024, which was

hypomethylated following the GOTO intervention (β = −0.047) and flanks an active transcription start site (TSS) of *GRB10* within 50 bp of multiple enriched TFBS. DNAm at this CpG is also positively associated with *GRB10* expression (β = 0.083), WC (β = 0.003), and total body fat percentage (β = 0.005), and inversely correlated with both immunohistochemistry measures (β_{PAX7} = −0.220, β_{MYO} = −0.011) and average dominant hand grip strength (β = −0.004; **Fig. 2b**). Taken together, these findings highlighted *GRB10* as a responsive locus in muscle that may be under epigenetic control.

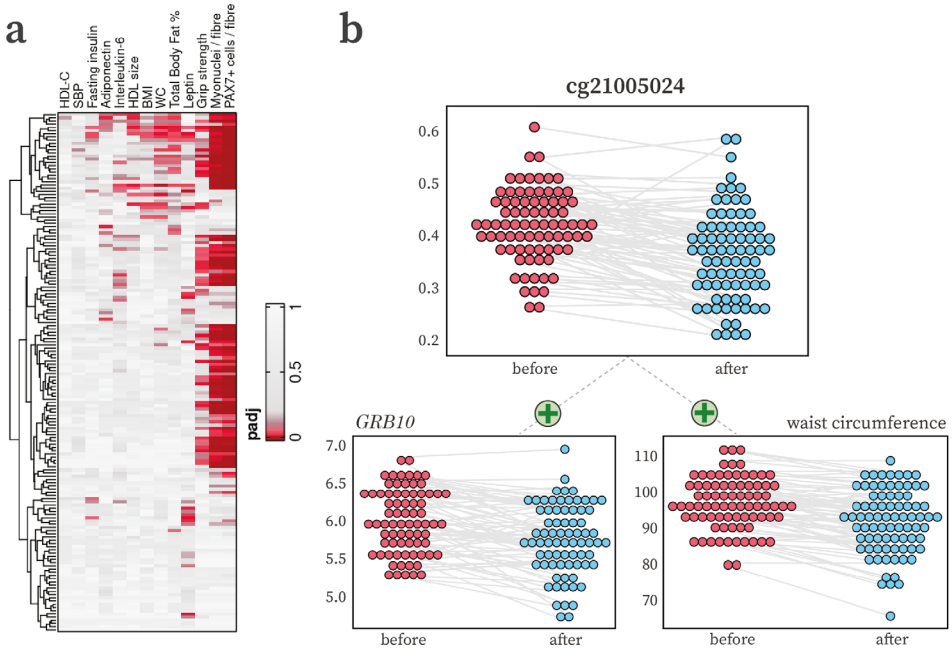


Figure 2 | Integrative analyses in muscle. a) Clustered heatmap of association *p*-values between changes in DNAm at the 162 *muscle CpGs* and thirteen health parameters (ten blood-based and three muscle-specific traits). **b)** Dot plots showing hypomethylation of DNAm in muscle at cg21005024, positively associated with both decreases in muscle *GRB10* gene expression and in waist circumference (HDL: high density lipoprotein; SBP: systolic blood pressure; BMI: body mass index; WC: waist circumference).

In adipose tissue, the GOTO intervention influenced DNAm at 230 predominantly hypermethylated CpGs

Next, we analysed DNAm responses in SAT following the GOTO intervention (n = 89 individuals, 178 samples). We identified 230 differentially methylated CpGs at 201 distinct loci ($p_{FDR} < 0.05$), henceforth referred to as the *adipose CpGs*. To explore whether this methylation signal was partially driven by changes in cell type proportions, we deconvoluted the bulk gene expression data using the *CIBERSORTx* algorithm[48] and a publicly available signature matrix[49] (**Fig. 3a; Supplementary Table 10**). The most prevalent cell types at baseline were adipocytes (mean 72.4%) followed by a large minority of microvascular endothelial cells (MVECs; mean 24.1%), and there was insufficient evidence to support a change in any of the predicted cell types following the intervention ($p_{FDR} \geq 0.05$). Therefore, further adjustment was not appropriate in these paired analyses as cellular composition was contained the individual random effects, and all 230 *adipose CpGs* were taken forward into subsequent analyses.

CpGs influenced by the GOTO intervention correlated with expression of lipid metabolism and insulin resistance genes

The majority (n = 139, 60.4%) of the 230 *adipose CpGs* were hypermethylated following the intervention (**Fig. 3b; Supplementary Table 11**). To determine their functional potential, we annotated their genomic positions to 15 chromatin states using the adipose reference epigenome (E063) from the Roadmap Epigenomics Consortium[33]. Since the *adipose CpGs* were enriched for several repressive marks, such as polycomb repressed regions (OR = 7.76, p_{FDR} = 7.35e−43) and depleted for regulatory states, including enhancers (OR = 0.38, p_{FDR} = 1.10e−02) and active TSS (OR = 0.11, p_{FDR} = 5.21e−04), there was insufficient evidence to suggest that the *adipose CpGs* as a whole were controlling nearby transcription (**Fig. 3c; Supplementary Table 12**). Despite this, sequences within 50 bp of the *adipose CpGs* were enriched for seven known TFBS (**Fig. 3d; Supplementary Table 13**), including for four GATA family TFs involved in the initial stages of adipogenesis and obesity[50].

Although the *adipose CpGs* did not likely represent a regulatory set overall, we evaluated if individual sites correlated with changes in nearby gene expression. Within 100 kb of the 230 *adipose CpGs* there were 412 genes, and 23 of these were both differentially expressed ($p_{FDR} < 0.05$) and associated with the nearby DNAm response ($p_{FDR} < 0.05$; **Supplementary Table 14**). These 23 genes included many relevant for adipogenesis, such as *ZBTB7A51* and *ALX152* and multiple developmental genes including *EN1* and *NR2F153*. Of particular interest were *PITX2* and *DMRT3*, which were associated with responses at 3 and 16 unique *adipose CpGs*, respectively. *PITX2* encodes a TF linked to changes in fasting glucose following weight loss[54] and *DMRT3*, which associates with exercise training and diet[55,56], has been proposed as a marker of insulin resistance specifically in SAT[57].

To further investigate the importance of this set of 23 genes, we performed overrepresentation analysis and revealed enrichment for 27 terms, including many relevant to lipid metabolism and transport (e.g., Phospholipid Efflux p_{FDR} = 0.019; **Supplementary Table 15**)[58].

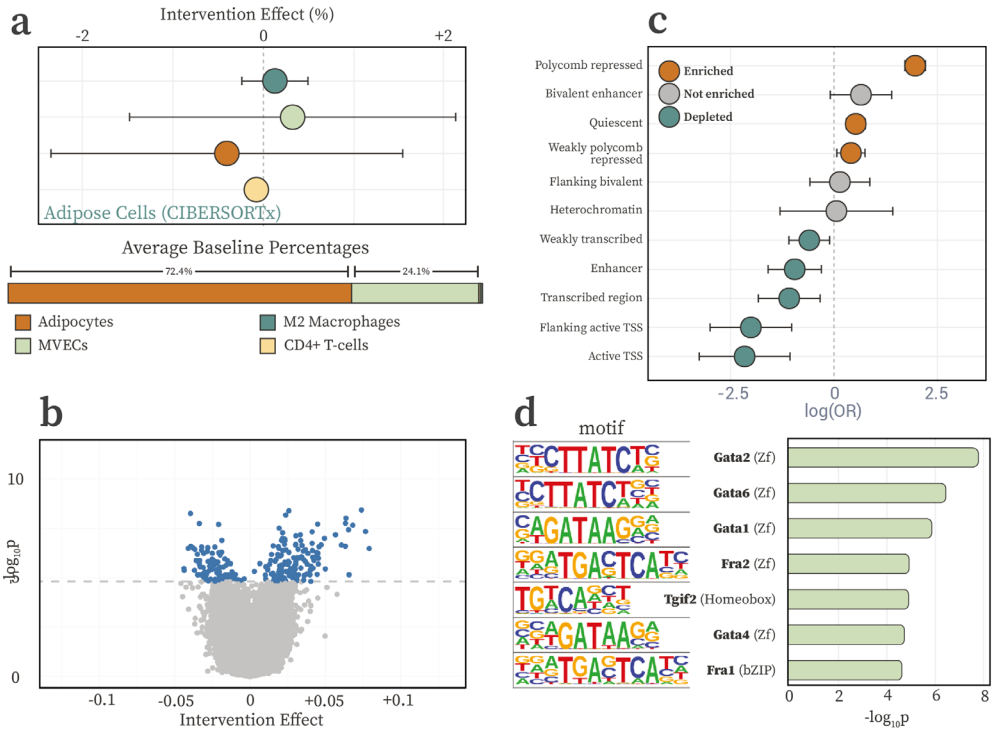


Figure 3 | Characterisation of the adipose cell counts and DNAm response to the GOTO intervention. a) Intervention effects on cell types predicted using *CIBERSORTx* alongside baseline proportions (only cells >0.5% at baseline shown) **b)** Volcano plot of the intervention effect on DNAm in SAT at over 750,000 CpGs, showing significant (blue) and non-significant (grey) CpGs. **c)** Forest plots of the OR and 95% CIs for enrichment or depletion of chromatin states using the E063 adipose reference epigenome from the Roadmap Epigenomics Consortium (four states with extremely wide CIs not shown). **d)** Bar plot of TFBS motifs enriched in sequences within 50 bp of the identified CpGs when compared to a GC-matched random genomic background.

Reductions in total body fat percentage associate with the adipose DNAm response

To link epigenetic findings in SAT to changes in health, we performed paired analyses associating DNAm effects at the *adipose CpGs* and differences in the ten health parameters shown in **Table 1 (Fig. 4a; Supplementary Table 16)**. In total, almost a third of the *adipose CpGs* (n = 75, 32.6%) associated with at least one tested trait, and more than five unique CpGs were correlated with total body fat percentage decreases (39 CpGs), adipocytokine levels (adiponectin: 8 CpGs; interleukin-6 (IL-6): 48 CpGs), BMI (12 CpGs), and WC (12 CpGs). Of these, 12 had been related to nearby gene expression in the previous analyses. Hypomethylation at two of the three *adipose CpGs* at *DMRT3* was also associated with reductions in total body fat percentage (**Fig. 4b**). Notably, these results connected DNAm effects at multiple CpGs to differential expression and improvements in health, indicating the relevance of the identified loci for physiological and molecular responses to lifestyle interventions in older adults.

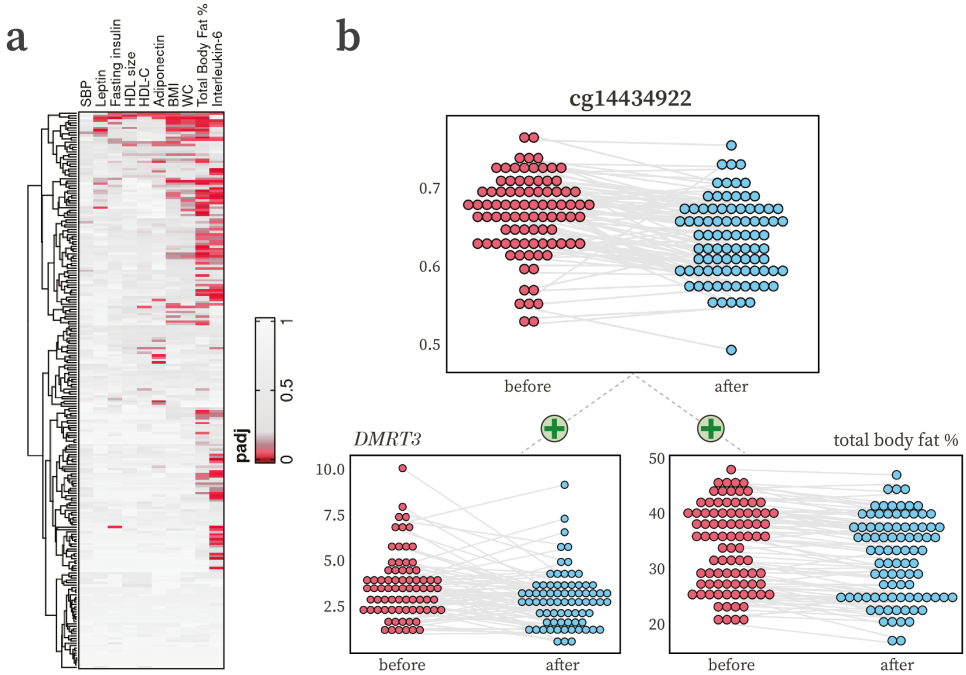


Figure 4 | Integrative analyses in subcutaneous adipose tissue (SAT). a) Clustered heatmap of association *p*-values the DNAm response at the 230 *adipose CpGs* and ten health parameters. **b)** Dot plots showing hypomethylation of DNAm in SAT at cg14434922, positively associated with decreases in adipose *DMRT3* gene expression and total body fat percentage (HDL: high density lipoprotein; SBP: systolic blood pressure; BMI: body mass index; WC: waist circumference).

Altered blood-based health parameters associate with the relatively small DNA methylation response in blood

We analysed DNAm data from paired fasted blood samples (n = 98 individuals, 196 samples), identifying 441 CpGs at distinct loci where methylation was altered following the intervention ($p_{FDR} < 0.05$), henceforth referred to as the *blood CpGs*. In cross-sectional blood-based EWAS, cell-type proportions are a well-known driver of association signals. Therefore, we extensively assessed if changes in cell types that would not be captured by the individual-level random effects existed in our data. To this end, we measured five cell types and predicted a further 36 from DNAm and gene expression data. Neutrophils and their progenitors accounted for the majority of cells (between 47.7% and 51.8% at baseline), followed by lymphocytes (36.9% to 41.1%) and monocytes (7.4% to 10.1%). There was insufficient evidence to suggest that any of the measured or predicted nucleated cell-types changed following the intervention at either the 5% nominal or FDR level. This indicated that cell-type proportions were captured by individual random effects in our models and additional adjustment would be statistically redundant (**Fig. 5a**).

In contrast to the findings in skeletal muscle and SAT, all effect sizes at the 441 b*lood CpGs* were small (β < 4%; **Fig. 5b; Supplementary Table 18**). This alone did not preclude them from functionality. To explore the likelihood that DNAm responses at the *blood CpGs* were regulatory, we performed chromatin state and TFBS enrichment analyses. This set of sites was enriched for regions flanking active TSS (OR = 1.50, p_{FDR} = 1.1e−02; **Fig. 5c; Supplementary Table 19**), although the size of this enrichment was smaller than seen in previous tissues. Sequences within 50 bp of the *blood CpGs* were enriched for three known TFBS (**Fig. 5d; Supplementary Table 20**), including NFE2L2 and MafK, TFs central to the oxidative stress response[59,60]. When we investigated genes within 100 kb of the *blood CpGs*, there were only three genes both differentially expressed and linked to nearby methylation. However, these did include two genes linked to inflammatory and immune responses (*LTBR* and *TNFRSF1A;* **Supplementary Table 21**)[61].

To explore possible distant or pleiotropic effects on health, we evaluated the association between DNAm at the *blood CpGs* and changes in the ten health parameters shown in **Table 1 (Fig. 5e; Supplementary Table 22**). Differential methylation at 66 (15.0%) of the *blood CpGs* associated with changes in nine traits, including total body fat percentage (24 CpGs), BMI (27 CpGs), WC (32 CpGs), HDL cholesterol levels (2 CpGs) and size (18 CpGs), IL-6 (3 CpGs), leptin (19 CpGs), systolic blood pressure (SBP, 2 CpGs), and fasting insulin (5 CpGs). This indicated that, despite the smaller size of the methylomic responses at the *blood CpGs* and the lack of support for *cis* regulatory effects, this set of sites was still able to mark intervention-driven improvements in immunometabolic health.

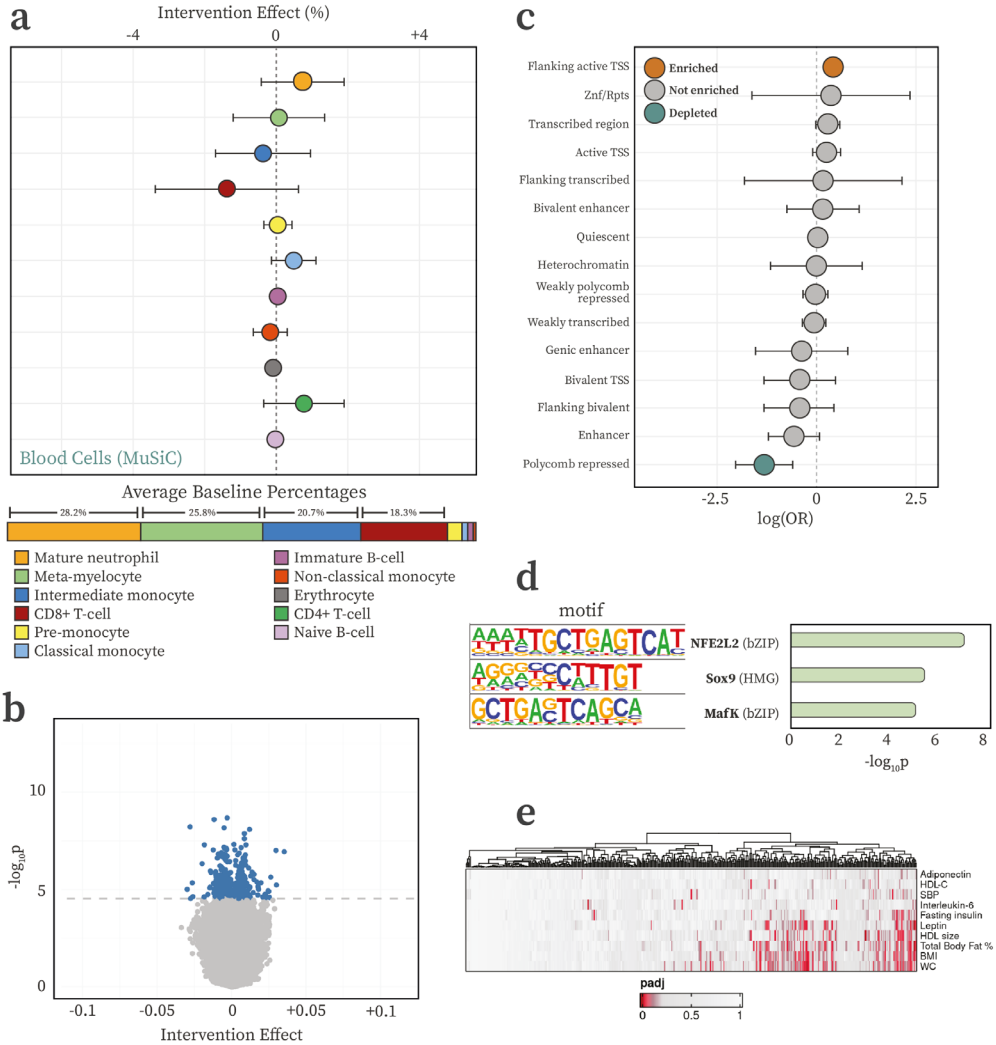


Figure 5 | Characterisation of the blood cell count and DNAm responses to the GOTO intervention. a) Intervention effects on immune cell types predicted using the IDOL algorithm alongside baseline proportions (only cells >0.5% at baseline shown). **b)** Volcano plot of the intervention effect on DNAm at over 750,000 CpGs, showing the 441 significant CpGs (blue) and nonsignificant CpGs (grey). **c)** Forest plots of the the OR and 95% CIs for enrichment (orange) or depletion (blue) of chromatin states using the E062 peripheral blood mononuclear cell (PBMC) reference epigenome. **d)** Bar plot of TFBS motifs enriched in sequences within 50 bp of identified CpGs compared to a GC-matched random genomic background. **e)** Clustered heatmap of association *p*-values between DNAm responses at the *blood CpGs* and ten health parameters.

Chronological age predictors have increased in accuracy but still capture more than just the passage of time

In the field of epigenetics and ageing, several algorithms have been developed to predict chronological age (cAge) from DNAm data. Here, we predicted cAge changes following the GOTO intervention by applying three such clocks to DNAm from pre- and post-intervention blood samples (Horvath[62], Zhang[63], and Bernabeu[64]). Correlations between actual age, which increased by 13 weeks across this longitudinal study, and predicted cAge were calculated. Predictions of cAge using Horvath, one of the original epigenetic clocks, were moderately correlated with actual age ($R = 0.730$, $p_{FDR} = 7.6e{-}34$), showing that from their inception these algorithms have performed well.

Looking at more recent clocks, we observed increases in accuracy over time with the Zhang ($R = 0.888$, $p_{FDR} = 2.7e{-}67$) and Bernabeu ($R = 0.923$, $p_{FDR} = 8.6e{-}82$) cAge predictions both correlating remarkably strongly with actual age at visit date. When looking at the predicted change in chronological age over this 13 week intervention, however, all three clocks did return a reduction in age ranging from a 22.3 week decrease (Horvath clock $p_{FDR} = 0.200$) to a 12.5 week reduction (Bernabeu $p_{FDR} = 0.155$). Overall, these results showed that cAge predictors are well correlated with and increasingly in line with actual age but still have considerably large residuals when compared to actual age. This indicated that such clocks may be swayed by other factors, such as intervention-driven health improvements, and could benefit from further refinement if their intention is to predict calendar age.

GrimAge captured the effect of the GOTO intervention and associated with metabolic and physiological health improvements

In contrast to cAge estimates, recent biological age (bAge) predictors are commonly trained on a combination of age, health parameters, and mortality data. We investigated four bAge clocks (Bernabeu[64], grimAge[65], phenoAge[66], and MEAT[67]). The first three of these were trained using blood samples, and thus we used blood-based DNAm from before and after the GOTO intervention as input. The fourth clock (MEAT) was a muscle-specific algorithm and so was applied to DNAm data from the muscle samples instead. Using a paired analysis, we estimated the effect of the GOTO intervention on biological age, adjusting for age, sex, and technical covariates (**Fig. 6a; Supplementary Table 23**).

All four clocks predicted a decrease in bAge following the GOTO intervention with estimates ranging from a 16.1-week decrease (Bernabeu $p_{FDR} = 2.4e{-}01$) to a 57.9-week decrease (MEAT $p_{FDR} = 2.1e{-}02$). After adjusting for multiple testing, however, only decreases in grimAge ($\beta = -34.6$ weeks, $p_{FDR} = 1.4e{-}04$) and MEAT ($\beta = -57.9$ weeks, $p_{FDR} = 2.1e{-}02$) remained significant at the 5% level. Considering that the strongest methylation

response to GOTO was in muscle, these results supported the development of tissue-specific algorithms for bAge prediction in epigenomic studies. Additionally, since grimAge is established as strongly associated with frailty risk as compared with other epigenetic age measures[68], this finding highlighted differences in available algorithms and supported a specific relevance for grimAge to ageing populations.

To explore whether observed reductions in bAge were connected to immunometabolic health improvements, we associated changes in bAge clocks with significant effects with the ten metabolic health parameters shown in **Table 1 (Supplementary Table 24)**. Seven of the ten tested traits were significantly associated with grimAge at the 5% level after adjusting for multiple testing, including BMI (β = 0.30, p_{FDR} = 5.2e−03), total body fat percentage (β = 0.10, p_{FDR} = 3.9e−02), and leptin levels (β = 0.08, p_{FDR} = 1.3e−02). The directions of effect consistently connected decreasing bAge to improved immunometabolic health (**Fig. 6b**). Together, these results showcased the power of blood-based DNAm markers and bAge algorithms to capture health improvements following lifestyle interventions in older adults.

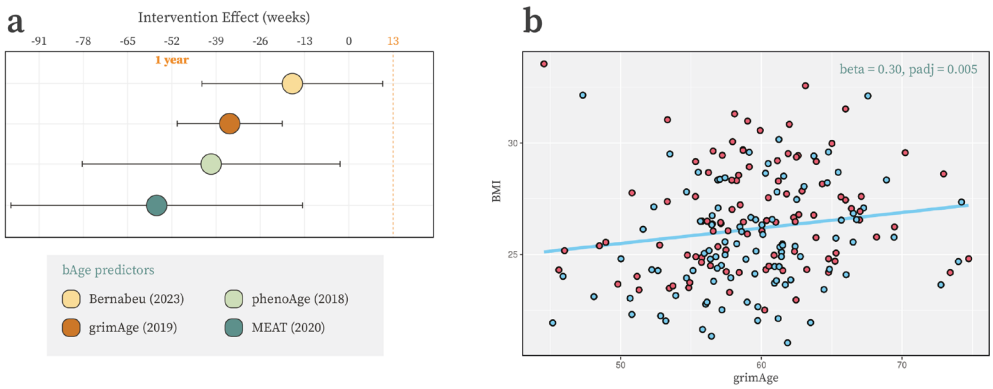


Figure 6 | The GOTO intervention reduces measures of epigenetic age. a) Predicted decreases in biological age following GOTO for four modern bAge algorithms. **b)** Scatter plot showing the association between grimAge and observed BMI both before (red) and after (blue) the GOTO intervention.

Discussion

Tissue-specific DNA methylation responses in muscle, adipose, and blood tissues

Following a 13-week combined lifestyle intervention, we observed DNAm responses at 162 CpGs in skeletal muscle, 230 CpGs in SAT, and 441 CpGs in fasted blood. We characterised the regulatory nature of these responsive CpGs, detecting enrichments for active chromatin states and relevant TFBS. Using matched gene expression and health parameter data, we performed integrative analyses and uncovered relationships between epigenetic changes and metabolic health, with links to insulin sensitivity, muscle regeneration potential, and body composition. On a molecular level, the directions of DNAm responses corresponded with previously observed transcriptomic effects in this study[69]. In skeletal muscle, identified CpGs were predominantly hypomethylated, an established effect of physical activity[70] and aligning with increases in gene expression in muscle. In contrast, methylation increased in SAT, tying in with global decreases in gene expression observed in this tissue. Lastly, in blood, the signal was small in both the methylome and the transcriptome[27,69], possibly due to molecular changes in blood lying further from functionally responding tissues and representing diffuse or systemic alterations.

RNA-Seq data was available to investigate *cis* associations between differentially expressed genes and DNAm responses in the three tissues. In skeletal muscle, expression changes at 71 genes in close proximity to identified CpGs were associated with differential methylation. In contrast, 23 genes were linked to DNAm changes in SAT and only three genes had clear *cis* correlations with CpGs in blood. Looking at the function of the genes identified in muscle and SAT, many were connected to insulin sensitivity and glucose uptake in relevant cell types. This included *GRB10*[42,43], which directly binds to and regulates insulin receptors, *PLEKHG4*[39–41], which is implicated in the translocation of GLUT4 transporters to the membrane in skeletal muscle cells, and *DMRT3*[71], an insulin sensitivity marker specific to SAT. In particular, the lowered expression of *GRB10* observed here enhances insulin-induced PI3K/Akt signalling and glucose uptake in myotubes and increases muscle size[42,70,72,73]. Changes in fasting insulin levels were also associated with DNAm responses in all three tissues (muscle: 2 CpGs, SAT: 3 CpGs; blood: 5 CpGs), although there was no overlap with the identified genes. Caloric restriction and exercise have established effects on insulin resistance with consequences for immunometabolic health, and this study outlined potential molecular mechanisms behind these effects[74–76].

We also observed enrichments for regulatory chromatin states in muscle and blood, including for enhancers and regions flanking active TSS. Relevant TFBS were additionally enriched in sequences within 50 bp of the differentially methylated CpGs in all three tissues. In muscle, we identified Fos and JunB binding sites close to responsive CpGs, TFs with established relevance for muscle health[36]. In SAT, binding sites of four

known GATA family TFs were enriched, pointing to transcriptional regulators connected to adipose tissue remodelling[50]. Since DNAm elicits functional effects by modulating the accessibility of regulatory sites, these findings strengthened the hypothesis that observed DNAm changes were regulatory and may have had functional effects on nearby gene expression[34].

Differential methylation links to metabolic health and decreases in epigenetic age

The GOTO intervention imparted a metabolic health benefit on the study population[25,26], and these improvements also tracked with DNAm adaptations. At over half of the *muscle CpGs* and almost a third of the *adipose CpGs,* DNAm responses were associated with alterations in one or more health parameters. Exemplifying this, DNAm in all three tissues could be connected to eight of the ten tested traits, including total body fat percentage (blood: 24 CpGs, adipose: 39 CpGs, muscle: 13 CpGs), leptin (blood: 19 CpGs, adipose: 4 CpGs, muscle: 14 CpGs), and BMI (blood: 27 CpGs, adipose: 12 CpGs, muscle: 7 CpGs). Favourable changes in body composition, as seen in the GOTO intervention study, are associated with a more balanced secretion of adipokines from adipose tissue, decreasing the risk of insulin resistance and type 2 diabetes. We show here that DNAm responses were associated with key measures of such a reduction in immunometabolic risk.

Previous investigations into the health benefits of the GOTO intervention have demonstrated sex-specific effects, possibly as a result of physiological differences between sexes and the personalised nature of the protocol. Sex-specific differences in muscle performance are partly attributed to larger proportions of type-I fibres in women, characterised by slower oxidative metabolism[77]. However, in this study where we investigated effects at over 750,000 CpGs across the epigenome, we did not consider the two sexes separately. In the future, it will be important to uncover the molecular mechanisms behind observed disparities between men and women using studies more suitably powered for sex stratification.

Lastly, we calculated cAge and bAge in blood and muscle samples using available algorithms[62–67]. These highlighted the impressive precision and accuracy of current epigenetic clocks for calendar age prediction, with estimates from a recent model highly correlated with actual age (R = 0.923)[64]. Yet, all tested cAge algorithms still reported age reductions following the intervention despite participants ageing by 13 weeks. It is plausible, therefore, that these cAge estimates are influenced by other factors such as immunometabolic health improvements. To investigate this further, we evaluated if bAge predictors captured the health benefits of the GOTO intervention. All four algorithms returned a decrease in bAge ranging from −16.1 weeks to −57.9 weeks, larger than the previous cAge estimates. The muscle-specific MEAT algorithm represented the greatest effect and it, alongside reductions in bAge as predicted by grimAge, were

still significant at the 5% level after adjusting for multiple testing. GrimAge, in particular, was associated with observed improvements in seven of the 10 tested health parameters, including BMI, circulating leptin levels, and total body fat percentage. GrimAge and the metabolomics-based score, MetaboHealth, have both previously been reported as good measures of health improvements[69], frailty[68], and mortality[65,78]. The beneficial shifts observed here in these scores indicate potentially global and long-term health improvements from the GOTO intervention protocol, and also highlight the possible value of molecular algorithms for monitoring intervention effects both in general, and specifically in older populations.

It is important to note that this intervention was carried out in healthy, older adults. For some older individuals, for example those with a risk of sarcopenia, this mild intervention may not be the most optimal regime. Other protocols, including ProMuscle and a novel upcoming study VOILA, are better focused on improving muscle mass and strength by including resistance training, increased protein intake, or protein supplementation[79,80]. Clinical study populations of older individuals may also require other response markers due to the higher levels of acute inflammatory proteins in population-based elderly.

Overall, our in-depth study of the methylome, transcriptome, and phenome exemplifies the biological changes that older adults experience following a mild intervention, such as GOTO. The absence of any overlap between the identified sets of CpGs demonstrates strong tissue-specificity in our findings and this, coupled with the distinct directional differences (hypermethylation in adipose tissue and hypomethylation in muscle), highlights the importance of using a multi-tissue approach when investigating the influence of environmental changes on the methylome. As DNAm represents only one form of epigenetic control, more in-depth interpretation of these findings may require examination of other layers of the epigenome, such as chromatin accessibility using ATAC-seq.

This study established the methylomic responses to a 13-week lifestyle intervention in older adults in both circulating cells and relevant metabolic tissues. We identified differential methylation at CpGs located in regulatory regions in close proximity to TFBS. Effects at these CpGs were associated with differential expression of insulin sensitivity genes, including *GRB10* in muscle and *DMRT3* in adipose, and with imparted metabolic health benefits. Identified loci may be investigated to monitor immunometabolic risk, progression of disease, and response to treatment in the future. The GOTO response was also represented by four epigenomic bAge markers and GrimAge, in particular, was able to capture the health improvements imparted to the participants. This study further demonstrated the importance of collecting biologically relevant tissues in intervention studies and highlighted how modifiable molecular markers can capture health improvements following lifestyle changes in older people.

Methods

Recruitment

The GOTO study was nested within the Leiden Longevity Study, a longitudinal cohort of long-lived Caucasian siblings, their offspring, and partners thereof. The GOTO study recruited healthy, older (mean age 63 years) adults (n = 164) between June 2012 and April 2013[25]. Individuals between 46 and 75 years with a BMI between 23 and 35 kg/m^2 from the pool of offspring and partners were eligible for the study.

Exclusion criteria included being on diabetic medication (for type 1 or 2 diabetes), having high fasting blood glucose levels (≥7.0 mmol/L), recent weight change (≥3 kg in the past 6 months), engagement in heavy or intensive physical activity (top sport or physically heavy work), any disease or condition that seriously affected body weight and/or body composition, recent immobilization (for over 1 week in the last 3 months), psychiatric or behavioural problems, use of thyroid medication or immunosuppressive drugs, concurrent participation in any other intervention studies or weight management programs, or not being registered with a general practitioner.

Intervention protocol

Expanding on the combined lifestyle arm of the CALERIE (Comprehensive Assessment of Long-term Effects of Reducing Intake of Energy) study, GOTO participants reduced their energy balance by 25% for 13 weeks, through a combination of caloric restriction and increased physical activity[6]. Informed by baseline questionnaires on energy intake (150-item food frequency questionnaire) and expenditure (IPAQ-SF, International Physical Activity Questionnaire Short Form), dieticians and physiotherapists prescribed individual guidelines to achieve the intervention. Participants were advised to increase physical activity in a way that was compatible with their lifestyle, and dietary guidelines aimed to follow the Dutch Guidelines for a Healthy Diet (2006).

To check and stimulate adherence to the intervention, there was weekly contact with both the dietician and physiotherapist. Participants recorded their adherence to the intervention plan in a diary, and two 24-hour recalls were performed during the first and last month of the intervention. Days of the recall were unannounced to the participants and randomised to obtain a good distribution of weekdays and weekends. During monthly home visits, body weight and composition were measured.

Sample collection

Both prior and post intervention, blood (95 mL) was drawn by venepuncture between 8 and 9 a.m. in the hospital after at least 10 hours of fasting. The participants consumed a SLM Nutridrink TM[72] representative of a typical Northern European meal (300 kcal: 35% energy from fat, 50% from carbohydrate, and 16% protein) between 9 a.m. and noon on the same day. Following this, skeletal muscle biopsies were taken from the musculus vastus lateralis and a subcutaneous adipose biopsy was taken from the abdomen. Biopsies were taken under local anaesthetic and immediately frozen in liquid nitrogen before being stored at −80°C for subsequent analysis.

Of the 164 individuals in the GOTO study, we profiled DNAm from 104 individuals at both timepoints for multiple tissues (sample n = 562). All 562 samples represented distinct samples from a unique timepoint, individual, and tissue combination and not duplicates. DNA from whole blood (n = 206) was isolated using QIAmp DNA Mini kits (QIAgen) and using NucleoMag Tissue kits (Machery Nagel) for adipose (n = 188) and muscle (n = 168) samples. Pairs of samples were shuffled and plated so that they would be adjacent on the same array. These pairs were randomised across eight 96-well plates by tissue, age, and sex using *Omixer*, and sent for profiling using the Infinium MethylationEPIC Kit (Illumina, Helmholtz Institute)[81].

DNA methylation profiling

Following receipt of the methylation data, preprocessing and quality control (QC) followed the *DNAmArray* pipeline[82]. *MethylAid* plots were used to visualize and check sample quality[83]. Due to technical issues with three of the Infinium MethylationEPIC arrays, 25 samples failed quality control checks. These alongside their pairs (n = 30) were removed from the data, and 24 samples with sufficient remaining material were reprofiled and subsequently passed QC checks.

After combining data from both waves, tissue identity was confirmed with PC analysis (PCA) plots, and four outlying samples and their pairs (n = 8) were removed. Sample mismatches were detected and resolved by comparing genotype data with DNAm-derived genotypes using *omicsPrint*[84]. Individuals (n = 6) identified from diary data as non-compliers were also removed. Lastly, methylation-predicted sex was used as a final check of sample identity.

The data underwent functional normalization using four PCs, followed by removal of outlying or unreliable values, such as those based on low bead number (0.20%), intensity (0.08%), that were not distinguishable from background noise (0.37%), or more than 3 interquartile ranges from the nearest quartile per CpG (0.28%). Any probes or samples with over 5% missingness were removed (0.00% of samples, 1.06% of CpGs). Additionally, cross-reactive, polymorphic[85], poorly reproducible[86], blacklisted[87], and sex chromosomal probes were removed. The resulting dataset contained DNAm data at 755,777 CpGs from 534 samples (196 blood, 178 adipose, and 160 muscle) from 102 individuals. For 66 individuals (64.7%), there was complete data from all three tissues at both timepoints.

RNA sequencing

RNA isolation and sequencing has been described previously[69]. In short, libraries were prepared using Illumina TruSeq version 2 library preparation kits. Data processing was performed using the in-house BIOPET Gentrap pipeline[88]. The following steps were part of the data processing: low quality trimming using sickle version 12.00. Cutadapt version 1.1 was used to perform the adapter clipping. The reads were aligned to GRCh37 while masking for single nucleotide polymorphisms common in the Dutch population (Genome of the Netherlands (GoNL) 45 minor allele frequency > 0.01), using STAR version 2.3.0e. Picard version 2.4.1. was used to perform sam to .bam conversion and sorting. Read quantification was performed using htseq-count version 0.6.1.p1 using Ensembl gene annotations version 86 for gene definitions. In blood, the sequencing resulted in an average of 37.2 million reads per sample, of which 97% (±0.4%) were mapped. In SAT, samples had an average of 11.4 million sequenced reads, of which 95% (±1.6%) were mapped. In muscle, an average of 36.9 million sequence reads per sample, of which 98% (±0.4%) were mapped.

Cell-type proportions

In muscle, nuclei types were predicted using the MuSiC algorithm combined with publicly available single nuclei reference transcriptomes and bulk expression data[30,31]. For subcutaneous adipose samples, no suitable single cell reference transcriptome was available. Therefore, adipose cell types were predicted using CIBERSORTx and a publicly available signature matrix provided from studies in TwinsUK and GTEx[48,49]. In fasted blood, the percentage of cell types (neutrophils, lymphocytes, monocytes, eosinophils, and basophils) was measured with a differential test. To investigate the intervention effect on specific immune cell subtypes, we combined a single cell reference atlas with whole blood expression data and the MuSiC algorithm estimating 32 different cell types[30,89]. Furthermore, the Identifying Optimal Libraries (IDOL) and IDOL-extended algorithms deconvoluted six and twelve subtypes from the DNAm data, respectively[90,91].

Statistical analysis

For all statistical investigations, *p*-values were adjusted for multiple testing using the FDR method, and significance was assessed at the 5% level. Intervention effects on DNAm were evaluated using mixed models, with a fixed effect for the intervention (time), an individual random effect (ID), and adjustment for confounders (age, sex, and smoking), technical covariates (plate and array row), and the first five PCs:

$$DNAm_{ijk} \sim time_{ij} + age_i + sex_i + smoke_i + plate1_{ij} + \cdots + plate10_{ij} + arrayRow_{ij} + PC1_{ij} + \cdots + PC5_{ij} + (1|ID_i)$$

Muscle models were additionally adjusted for predicted endothelial nuclei proportions. Bias and inflation in the test statistics was inspected and adjusted for using the Bioconductor package, *bacon*, with default priors ($\alpha = 1.28$, $\beta = 0.36$)[92].

CpG interpretation

Differentially methylated regions To assess the number of distinct genomic loci in our results, differentially methylated regions (DMRs) were calculated using the *DMRfinder* algorithm as implemented in the *DNAmArray* workflow[82,93]. DMRs were defined as regions with at least three differentially methylated probes (DMPs) with an inter-CpG distance less than 1 kb, allowing a maximum of three non-DMPs across a DMR. The number of distinct loci was the total number of DMPs minus the number of DMPs in DMRs plus the number of DMRs called by DMRfinder.

Chromatin state enrichment FDR-significant CpGs were annotated to chromatin states using appropriate reference epigenomes from the Roadmap Epigenomics Consortium (E062 for blood, E063 for adipose, E107 and E108 for muscle)[33]. Logistic regressions models were fitted using the *glm* function in R for each of the 15 chromatin states to assess if the number of significant CpGs annotated to that chromatin state differed significantly from the set of tested CpGs.

TFBS enrichment A 50 bp window around identified CpGs was scanned for enrichment of known motifs compared to background noise matched for GC content using *findMotifsGenome.pl* from HOMER[94]. ENCODE TFBS annotation for TFs and CpGs on the EPIC array was further used to investigate the size of binding sites and distance from CpG to summit[85].

Gene annotation Genomic locations of human transcripts, exons, coding sequences, and genes were imported from the Ensembl database using *makeTxDbFromEnsembl* from the GenomicFeatures Bioconductor package[95]. These both annotated CpGs to their nearest gene and identified all genes within 100 kb of each site.

Differential gene expression Gene expression changes were analysed as described previously but with both sexes combined in a single analysis[69]. Briefly, the differential gene expression analysis used linear mixed models in combination with voom normalization. Models were adjusted for technical factors, age, and sex as fixed effects and included a random effect for ID.

Expression quantitative trait methylation The association between DNAm and differentially expressed genes within 100 kb was investigated using available expression and methylation data. Lowly expressed genes were removed and $\log_2$CPM values were calculated. For each gene, we applied RIN transformations as described previously[96].

$$DNAm_{ij} \sim gene_{ij} + age_i + sex_i + smoke_i + plate1_{ij} + \cdots + plate10_{ij} + arrayRow_{ij} + flowcell_{ij} + (1|ID_i)$$

Over-representation analysis The list of candidate target genes (differentially expressed genes associated with a DNAm response within 100kb) was used as input for over-representation analysis in *clusterProfiler*, combined with recent (updated in the last 6 years) databases relating to human health and disease downloaded from Enrichr (GO Biological Process 2023, KEGG Human 2021, and Reactome 2022)[58].

Health parameter associations Linear mixed models were used to investigate the association between DNAm and ten health parameters, adjusting for age, sex, smoking, technical covariates:

$$DNAm_{ij} \sim trait_{ij} + age_i + sex_i + smoke_i + plate1_{ij} + \cdots + plate10_{ij} + arrayRow_{ij} + (1|ID_i)$$

Epigenetic clock algorithms

cAge prediction Using the *dnaMethyAge* package[97], Horvath[62], Zhang[63], and Bernabeu[64] based cAge was predicted. Correlations between these estimates and actual age was assessed using a Pearson's correlation test with the *cor.test* function in R.

bAge prediction Using the *dnaMethyAge* package[97], LevineM2018 (PhenoAge) was predicted[66]. grimAge was estimated using the coefficients, R, and Python scripts provided by the researchers who developed this measure[65]. Bernabeu's bAge[64] was predicted by combining grimAge, DNAm data, and phenotype data in the *bage_predictor.R* script provided on their GitHub (`elenabernabeu/cage_bage`). Lastly, MEAT bAge[67] was predicted from muscle DNAm data using the MEAT Bioconductor package for R. For all bAge predictions, paired analyses estimated changes following the GOTO intervention, adjusting for age, sex, technical covariates and with a random effect for ID:

$$bAge_{ij} \sim time_i + age_i + sex_i + plate1_{ij} + \cdots + plate10_{ij} + arrayRow_{ij} + (1|ID_i)$$

Associations with health parameters Associations between grimAge and ten health parameters were investigated using a paired analysis, adjusting for age, sex, technical covariates, and with a random effect for ID:

$$grimAge_{ij} \sim trait_{ij} + age_i + sex_i + plate1_{ij} + \cdots + plate10_{ij} + arrayRow_{ij} + (1|ID_i)$$

Ethics statement

The Growing Old TOgether Study protocol was approved by the Medical Ethical Committee of the Leiden University Medical Center before the start of the study (P11.187). In accordance with the Declaration of Helsinki, the Growing Old TOgether Study obtained informed consent from all participants prior to their entering the study. This trial was registered at the Dutch Trial Register (https://onderzoekmetmensen.nl/en) as GOT NL3301.

Author contributions

P.E.S. and C.d.G. designed the study. M.B. collected and curated the health data. N.L. and E.S. generated the DNAm and transcriptome data. L.S., B.T.H., M.B., and P.E.S. designed the data analysis approach, L.S. and F.A.B. performed data analysis, I.M. and L.S. performed the cell deconvolution analyses, D.B. provided expertise on epigenetic clocks, Y.R., A.B., G.B-B., and C.T. provided expertise on muscle physiology, M.W. profiled DNAm in the three tissues and provided vital intellectual contributions, and L.S., Y.R., A.B., G.B-B., C.T., B.T.H., P.E.S., and M.B. performed the research and interpreted the data. All authors were involved in drafting and revising the manuscript.

Funding

The research leading to these results was supported by the Joint Programming Initiative "a Healthy Diet for a Healthy Life" (JPI-HDHL) DIMENSION project (ZonMW project number: 529051021) and ZonMW Project VOILA (ZonMW project number: 457001001). The underlying intervention study was financially supported by the Netherlands Consortium for Healthy Ageing (grant 050-060-810), in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO) and by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007). The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Declarations of interest

The authors declare no competing interests.

Data & code availability

Data supporting the findings of this study are accessible upon reasonable request to the corresponding author. The individual-level data are not publicly available due to privacy or ethical restrictions. All other data used in this study are publicly available: reference epigenomes are available from the Roadmap Epigenomics Consortium[33] and TFBS data is available within the HOMER software[94]. All the software and programmes used to conduct these analyses are freely available. Scripts used during this analysis are provided on GitHub at `nebulyra.github.io/goto_dnam`.

References

1. Fiuza-Luces, C. *et al.* Exercise is the real polypill. *Physiology* **28** (5): 330-358 (2013).
2. Knowler, W. C. *et al.* Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* **346** (6): 393-403 (2002).
3. The Diabetes Prevention Program (DPP): Description of lifestyle intervention. *Diabetes Care* **25** (12): 2165-2171 (2002).
4. Zhang, Y. *et al.* Combined lifestyle factors and risk of incident type 2 diabetes and prognosis among individuals with type 2 diabetes: a systematic review and meta-analysis of prospective cohort studies. *Diabetologia* **63** (1): 21-33 (2020).
5. Leitner, D. R. *et al.* Obesity and type 2 diabetes: Two diseases with a need for combined treatment strategies - EASO can lead the way. *Obes Facts* **10** (5): 483-492 (2017).
6. Kraus, W. E. *et al.* 2 years of calorie restriction and cardiometabolic risk (CALERIE): exploratory outcomes of a multicentre, phase 2, randomised controlled trial. *Lancet Diabetes Endocrinol* **7** (9): 673-683 (2019).
7. Eriksson, J. *et al.* Prevention of Type II diabetes in subjects with impaired glucose tolerance: The Diabetes Prevention Study (DPS) in Finland. Study design and 1-year interim report on the feasibility of the lifestyle intervention programme. *Diabetologia* **42** (7): 793–801 (1999).
8. Voisin, S. *et al.* Exercise training and DNA methylation in humans. *Acta Physiologica* **213** (1): 39-59 (2015).
9. van der Harst, P., de Windt, L. J. *and* Chambers, J. C. Translational Perspective on Epigenetics in Cardiovascular Disease. *J Am Coll Cardiol* **70** (5): 590-606 (2017).
10. Gevaert, A. B. *et al.* Epigenetics in the primary and secondary prevention of cardiovascular disease: influence of exercise and nutrition. *Eur J Prev Cardiol* **29** (17): 2183-2199 (2022).
11. Nakajima, K. *et al.* Exercise effects on methylation of ASC gene. *Int J Sports Med* **31** (9): 671-675 (2010).
12. Butts, B. *et al.* Effects of Exercise on ASC Methylation and IL-1 Cytokines in Heart Failure. *Med Sci Sports Exerc* **50** (9): 1757-1766 (2018).

13. Ma, J. *et al.* Whole Blood DNA Methylation Signatures of Diet Are Associated With Cardiovascular Disease Risk Factors and All-Cause Mortality. *Circ Genom Precis Med* **13** (4): E002766 (2020).
14. Keller, M. et al. DNA methylation signature in blood mirrors successful weight-loss during lifestyle interventions: the CENTRAL trial Genome Med 12: 97 (2020).
15. Stanford, K. I. & Goodyear, L. J. Muscle-Adipose Tissue Cross Talk. Cold Spring Harb Perspect Med 8, a029801 (2018).
16. Jacobsen, S. C. et al. Effects of short-term high-fat overfeeding on genome-wide DNA methylation in the skeletal muscle of healthy young men. Diabetologia 55, 3341–3349 (2012).
17. Barrès, R. et al. Acute exercise remodels promoter methylation in human skeletal muscle. Cell Metab 15, 405–411 (2012).
18. Rönn, T. et al. A Six Months Exercise Intervention Influences the Genome-wide DNA Methylation Pattern in Human Adipose Tissue. PLoS Genet 9, (2013).
19. Gillberg, L., Jacobsen, S. C., Rönn, T., Brøns, C. & Vaag, A. PPARGC1A DNA methylation in subcutaneous adipose tissue in low birth weight subjects - Impact of 5 days of high-fat overfeeding. Metabolism 63, 263–271 (2014).
20. Fabre, O. et al. Exercise training alters the genomic response to acute exercise in human adipose tissue. Epigenomics 10, 1033–1050 (2018).
21. Savikj, M. et al. Exercise timing influences multi-tissue metabolome and skeletal muscle proteome profiles in type 2 diabetic patients – A randomized crossover trial. Metabolism 135, (2022).
22. Moore, T. M. et al. Conserved multi-tissue transcriptomic adaptations to exercise training in humans and mice. Cell Rep 42, (2023).
23. Mill, J. & Heijmans, B. T. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 14, 585–594 (2013).
24. Rickman, A. D. et al. The CALERIE Study: Design and methods of an innovative 25% caloric restriction intervention. Contemp Clin Trials 32, 874–881 (2011).
25. van de Rest, O. et al. Metabolic effects of a 13-weeks lifestyle intervention in older adults: The Growing Old Together Study. Aging 8, 111–126 (2016).
26. Beekman, M. et al. Lifestyle-Intervention-Induced Reduction of Abdominal Fat Is Reflected by a Decreased Circulating Glycerol Level and an Increased HDL Diameter. Mol Nutr Food Res 64, (2020).
27. Gehrmann, T. et al. A Combined Lifestyle Intervention Induces a Sensitization of the Blood Transcriptomic Response to a Nutrient Challenge. Aging Biology 2, 20240036 (2024).
28. Bogaards, F. A. et al. Secondary integrated analysis of multi-tissue transcriptomic responses to a combined lifestyle intervention in older adults from the GOTO nonrandomized trial. Nat Commun 15, (2024).
29. Bogaards, F. A. et al. PLIS: A metabolomic response monitor to a lifestyle intervention study in older adults. FASEB Journal 36, (2022).
30. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun 10, (2019).
31. Perez, K. et al. Single nuclei profiling identifies cell specific markers of skeletal muscle aging, frailty, and senescence. Aging 14, 9393–9422 (2022).
32. Kwak, S.-E., Lee, J.-H., Zhang, D. & Song, W. Angiogenesis: focusing on the effects of exercise in aging and cancer. J Exerc Nutrition Biochem 22, 21–26 (2018).
33. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. Nature 2015 518:7539 518, 317–330 (2015).
34. Kaluscha, S. et al. Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. Nat Genet 54, 1895–1906 (2022).
35. Trenerry, M. K., Carey, K. A., Ward, A. C. & Cameron-Smith, D. STAT3 signaling is activated in human skeletal muscle following acute resistance exercise. J Appl Physiol 102, 1483–1489 (2007).
36. Puntschart, A. et al. Expression of fos and jun genes in human skeletal muscle after exercise. Am J Physiol Cell Physiol 274, (1998).
37. Galvagni, F., Cantini, M. & Oliviero, S. The utrophin gene is transcriptionally up-regulated in regenerating muscle. Journal of Biological Chemistry 277, 19106–19113 (2002).
38. Almada, A. E. et al. FOS licenses early events in stem cell activation driving skeletal muscle regeneration. Cell Rep 34, 108656 (2021).
39. Machin, P. A., Tsonou, E., Hornigold, D. C. & Welch, H. C. E. Rho family gtpases and rho gefs in glucose homeostasis. Cells 10, (2021).
40. Ha, N. T. & Lee, C. H. Roles of farnesyl-diphosphate farnesyltransferase 1 in tumour and tumour microenvironments. Cells 9, 1–33 (2020).
41. Shrestha, M. M., Lim, C. Y., Bi, X., Robinson, R. C. & Han, W. Tmod3 Phosphorylation Mediates AMPK-Dependent GLUT4 Plasma Membrane Insertion in Myoblasts. Front Endocrinol (Lausanne) 12, (2021).
42. Holt, L. J. et al. Ablation of Grb10 Specifically in Muscle Impacts Muscle Size and Glucose Metabolism in Mice. Endocrinology 159, 1339–1351 (2018).

43. Edick, A. M., Auclair, O. & Burgos, S. A. Role of Grb10 in mTORC1-dependent regulation of insulin signaling and action in human skeletal muscle cells. Am J Physiol Endocrinol Metab 318, E173–E183 (2020).
44. Raz, Y. et al. A data-driven methodology reveals novel myofiber clusters in older human muscles. FASEB Journal 34, 5525–5537 (2020).
45. Azhar, M., Wardhani, B. W. K. & Renesteen, E. The regenerative potential of Pax3/Pax7 on skeletal muscle injury. Journal of Genetic Engineering and Biotechnology 20, (2022).
46. Snijders, T., Holwerda, A. M., van Loon, L. J. C. & Verdijk, L. B. Myonuclear content and domain size in small versus larger muscle fibres in response to 12 weeks of resistance exercise training in older adults. Acta Physiologica 231, (2021).
47. Hansson, K. A. et al. Myonuclear content regulates cell size with similar scaling properties in mice and humans. Nat Commun 11, (2020).
48. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 37, 773–782 (2019).
49. Glastonbury, C. A., Couto Alves, A., El-Sayed Moustafa, J. S. & Small, K. S. Cell-Type Heterogeneity in Adipose Tissue Is Associated with Complex Traits and Reveals Disease-Relevant Cell-Specific eQTLs. Am J Hum Genet 104, 1013–1024 (2019).
50. Tong, Q. et al. Function of GATA transcription factors in preadipocyte-adipocyte transition. Science (1979) 290, 134–138 (2000).
51. Laudes, M. et al. Role of the POZ Zinc Finger Transcription Factor FBI-1 in Human and Murine Adipogenesis. Journal of Biological Chemistry 279, 11711–11718 (2004).
52. Breitfeld, J. et al. Developmentally Driven Changes in Adipogenesis in Different Fat Depots Are Related to Obesity. Front Endocrinol (Lausanne) 11, (2020).
53. Singh, S., Rajput, Y. S., Barui, A. K., Sharma, R. & Grover, S. Expression of developmental genes in brown fat cells grown in vitro is linked with lipid accumulation. In Vitro Cell Dev Biol Anim 51, 1003–1011 (2015).
54. Macartney-Coxson, D. et al. Genome-wide DNA methylation analysis reveals loci that distinguish different types of adipose tissue in obese individuals. Clin Epigenetics 9, (2017).
55. Nono Nankam, P. A. et al. Distinct abdominal and gluteal adipose tissue transcriptome signatures are altered by exercise training in African women with obesity. Sci Rep 10, 10240 (2020).
56. Divoux, A. et al. DNA Methylation as a Marker of Body Shape in Premenopausal Women. Front Genet 12, 709342 (2021).
57. Clemente-Olivo, M. P. et al. Four-and-a-half LIM domain protein 2 (FHL2) deficiency protects mice from diet-induced obesity and high FHL2 expression marks human obesity. Metabolism 121, (2021).
58. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation 2, (2021).
59. Ryoo, I. geun & Kwak, M. K. Regulatory crosstalk between the oxidative stress-related transcription factor Nfe2l2/Nrf2 and mitochondria. Toxicol Appl Pharmacol 359, 24–33 (2018).
60. Hwang, Y. J. et al. MafK positively regulates NF-κB activity by enhancing CBP-mediated p65 acetylation. Sci Rep 3, (2013).
61. Piao, W. et al. LTβR Signaling Controls Lymphatic Migration of Immune Cells. Cells 10, 747 (2021).
62. Horvath, S. DNA methylation age of human tissues and cell types. Genome Biol 14, (2013).
63. Zhang, Q. et al. Improved precision of epigenetic clock estimates across tissues and its implication for biological ageing. Genome Med 11, 54 (2019).
64. Bernabeu, E. et al. Refining epigenetic prediction of chronological and biological age. Genome Med 15, 1–15 (2023).
65. Lu, A. T. et al. DNA methylation GrimAge strongly predicts lifespan and healthspan. Aging 11, 303–327 (2019).
66. Levine, M. E. et al. An epigenetic biomarker of aging for lifespan and healthspan. Aging 10, 573–591 (2018).
67. Voisin, S. et al. An epigenetic clock for human skeletal muscle. J Cachexia Sarcopenia Muscle 11, 887–898 (2020).
68. Kuiper, L. M. et al. Epigenetic and Metabolomic Biomarkers for Biological Age: A Comparative Analysis of Mortality and Frailty Risk. J Gerontol A Biol Sci Med Sci 78, 1753 (2023).
69. Bogaards, F. A. et al. Secondary integrated analysis of multi-tissue transcriptomic responses to a combined lifestyle intervention in older adults from the GOTO nonrandomized trial. Nature Communications 2024 15:1 15, 1–18 (2024).
70. Plaza-Diaz, J. et al. Impact of Physical Activity and Exercise on the Epigenome in Skeletal Muscle and Effects on Systemic Metabolism. Biomedicines 10, (2022).
71. Pujar, M. K., Vastrad, B. & Vastrad, C. Integrative analyses of genes associated with subcutaneous insulin resistance. Biomolecules 9, (2019).
72. Mokbel, N. et al. Grb10 Deletion Enhances Muscle Cell Proliferation, Differentiation and GLUT4 Plasma Membrane Translocation. J Cell Physiol 229, 1753–1764 (2014).

73. Holt, L. J. et al. Grb10 regulates the development of fiber number in skeletal muscle. The FASEB Journal 26, 3658–3669 (2012).
74. Johnson, M. L. et al. Mechanism by which caloric restriction improves insulin sensitivity in sedentary obese adults. Diabetes 65, 74–84 (2016).
5. Dubé, J. J. et al. Effects of weight loss and exercise on insulin resistance, and intramyocellular triacylglycerol, diacylglycerol and ceramide. Diabetologia 54, 1147–1156 (2011).
76. Roberts, C. K., Hevener, A. L. & Barnard, R. J. Metabolic syndrome and insulin resistance: Underlying causes and modification by exercise training. Compr Physiol 3, 1–58 (2013).
77. Haizlip, K. M., Harrison, B. C. & Leinwand, L. A. Sex-based differences in skeletal muscle kinetics and fiber-type composition. Physiology 30, 30–39 (2015).
78. Deelen, J. et al. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. Nat Commun 10, (2019).
79. Leenders, M. et al. Protein supplementation during resistance-type exercise training in the elderly. Med Sci Sports Exerc 45, 542–552 (2013).
80. Van Dongen, E. J. et al. Translation of a tailored nutrition and resistance exercise intervention for elderly people to a real-life setting: adaptation process and pilot study. BMC Geriatr 17, 1–15 (2017).
81. Sinke, L., Cats, D. & Heijmans, B. T. Omixer: multivariate and reproducible sample randomization to proactively counter batch effects in omics studies. Bioinformatics 37, 3051–3052 (2021).
82. Sinke, L., van Iterson, M., Cats, D., Kuipers, T. & Heijmans, B. DNAmArray: Streamlined workflow for the quality control, normalization, and analysis of Illumina methylation array data. doi:10.5281/ZENODO.16962439.
83. Van Iterson, M. et al. MethylAid: visual and interactive quality control of large Illumina 450k datasets. Bioinformatics 30, 3435–3437 (2014).
84. Van Iterson, M., Cats, D., Hop, P. & Heijmans, B. T. omicsPrint: detection of data linkage errors in multiple omics studies. Bioinformatics 34, 2142–2143 (2018).
85. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res 45, e22 (2017).
86. Sugden, K. et al. Patterns of Reliability: Assessing the Reproducibility and Integrity of DNA Methylation Measurement. Patterns 1, (2020).
87. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Scientific Reports 2019 9:1 9, 1–5 (2019).
88. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. Nat Genet 49, 139–145 (2017).
89. Xie, X. et al. Single-cell transcriptomic landscape of human blood cells. Natl Sci Rev 8, (2021).
90. Koestler, D. C. et al. Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL). BMC Bioinformatics 17, (2016).
91. Salas, L. A. et al. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. Nature Communications 2022 13:1 13, 1–13 (2022).
92. van Iterson, M. et al. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol 18, 1–13 (2017).
93. Slieker, R. C. et al. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. Epigenetics Chromatin 6, (2013).
94. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589 (2010).
95. Lawrence, M. et al. Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol 9, (2013).
96. Bonder, M. J. et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet 49, 131–138 (2017).
97. Wang, Y., Grant, O. A., Zhai, X., Mcdonald-Maier, K. D. & Schalkwyk, L. C. Insights into ageing rates comparison across tissues from recalibrating cerebellum DNA methylation clock. Geroscience 46, 39–56 (2024).

Supplementary Material

Supplementary Figure 1 | Cross-tissue effect sizes. a. CpGs with significant responses in muscle, ordered by effect size in muscle and shown for all tissues (row 1: muscle, row 2: adipose, row 3: blood). **b.** CpGs with significant responses in adipose, ordered by effect size in adipose and shown for all tissues. **c.** CpGs with significant responses in blood, ordered by effect size in blood and shown for all tissues. Bars are coloured by FDR-adjusted *p*-value.

	Entire GOTO population (n = 164)		Three tissue overlap (n = 66)		Skeletal muscle (n = 80)		Adipose tissue (n = 89)		Fasted blood (n = 98)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Body mass index (kg/m^2)	26.9	2.5	26.9	2.4	26.8	2.3	26.9	2.4	26.8	2.4
WC (cm)	96.1	8.0	97.0	7.3	96.9	7.4	96.1	7.4	96.1	7.5
Total body fat (%)	32.6	8.2	32.2	7.7	31.1	7.8	33.4	7.7	32.5	8.0
Fasting insulin (mU/L)	9.2	4.6	9.4	4.4	9.0	4.2	9.1	4.4	9.0	4.4
Systolic BP (mm Hg)	139.9	17.1	140.4	15.7	139.7	15.7	138.3	15.7	138.1	15.6
Leptin (µg/L)	14.0	10.6	13.0	9.0	11.9	8.6	14.9	10.6	14.1	10.4
Adiponectin (mg/L)	10.4	5.3	9.8	4.5	9.8	4.7	10.6	4.8	10.7	4.9
Interleukin-6 (ng/L)	1.4	1.8	1.1	0.7	1.2	1.3	1.0	0.7	1.2	1.2
HDL cholesterol (mmol/L)	9.6	0.2	9.6	0.2	9.5	0.2	9.6	0.2	9.6	0.2
Fasting HDL size (nm)	1.5	0.4	1.5	0.4	1.4	0.4	1.5	0.4	1.5	0.4

Supplementary Table 1 | Baseline characteristics in the GOTO population and tissue subsets.

	Three tissue overlap (n = 66)		Skeletal muscle (n = 80)		Adipose tissue (n = 89)		Fasted blood (n = 98)	
	Δ (SE)	p_{FDR}	Δ (SE)	p_{FDR}	Δ (SE)	p_{FDR}	Δ (SE)	p_{FDR}
Body mass index (kg/m^2)	-0.18 (0.13)	3.1E-01	-0.28 (0.12)	1.3E-01	-0.30 (0.13)	1.7E-01	-0.36 (0.13)	6.1E-02
WC (cm)	-0.83 (0.85)	4.1E-01	-1.46 (0.83)	2.6E-01	-0.39 (0.83)	8.0E-01	-0.74 (0.84)	5.4E-01
Total body fat (%)	-0.62 (0.47)	3.1E-01	-0.50 (0.46)	4.3E-01	-0.90 (0.47)	2.3E-01	-0.70 (0.47)	2.8E-01
Fasting insulin (mU/L)	-0.83 (0.52)	3.1E-01	-0.43 (0.50)	4.9E-01	-0.27 (0.49)	8.0E-01	-0.35 (0.50)	5.4E-01
Systolic BP (mm Hg)	0.48 (2.00)	8.1E-01	0.88 (1.89)	6.4E-01	0.56 (1.87)	8.5E-01	1.57 (1.88)	5.4E-01
Leptin (µg/L)	-0.65 (0.65)	4.1E-01	-0.70 (0.67)	4.3E-01	-1.27 (0.69)	2.3E-01	-1.23 (0.74)	2.4E-01
Adiponectin (mg/L)	0.45 (0.30)	3.1E-01	0.38 (0.29)	4.3E-01	0.05 (0.28)	8.6E-01	0.07 (0.29)	8.1E-01
Interleukin-6 (ng/L)	0.35 (0.22)	3.1E-01	0.27 (0.22)	4.3E-01	0.33 (0.23)	3.3E-01	0.42 (0.24)	2.4E-01
HDL cholesterol (mmol/L)	0.02 (0.03)	7.3E-01	0.02 (0.03)	6.4E-01	-0.02 (0.03)	8.0E-01	-0.02 (0.03)	5.4E-01
Fasting HDL size (nm)	0.02 (0.01)	3.1E-01	0.05 (0.01)	1.4E-03	0.02 (0.01)	3.3E-01	0.03 (0.01)	7.2E-02

Supplementary Table 2 | Non-response analysis comparing changes in health parameters in included and excluded individuals.

Nuclei type	Baseline	SD	beta	SE	p	p_{FDR}
Endothelial	36.57 %	7.06	3.26	0.90	5.03E-04	3.52E-03
Fibro-adipogenic progenitor	0.02 %	0.11	0.13	0.05	1.65E-02	5.76E-02
Satellite	<0.001 %	<0.001	0.08	0.04	3.74E-02	8.74E-02
Slow (type I) fibre	35.96 %	17.32	-2.01	2.35	3.95E-01	6.91E-01
Fast (type II) fibre	26.75 %	19.90	-1.51	2.36	5.24E-01	7.33E-01
Immune	0.69 %	1.56	0.05	0.22	8.34E-01	8.34E-01
Smooth muscle	0.01 %	0.11	0.004	0.02	8.10E-01	8.34E-01

Supplementary Table 3 | Baseline percentages and changes in predicted muscle nuclei types.

CpG	Position	beta	SE	p	p_{FDR}	n
cg12394201	chr11: 43942417	-0.0624	0.0085	1.40E-11	1.06E-05	147
cg24161080	chr18: 33889786	-0.0839	0.0117	3.56E-11	1.35E-05	148
cg05008948	chr7: 99160713	-0.0308	0.0046	7.40E-10	1.86E-04	148
cg01668986	chr8: 21541541	-0.0435	0.0068	2.86E-09	4.92E-04	148
cg21342383	chr10: 127661991	-0.0187	0.0029	3.26E-09	4.92E-04	148
cg13585930	chr10: 72027356	-0.0564	0.0089	5.23E-09	6.59E-04	148
cg07626206	chr7: 55200656	-0.0287	0.0046	6.37E-09	6.88E-04	147
cg02632441	chr2: 218711231	-0.0249	0.0040	1.06E-08	9.97E-04	148
cg25834201	chr1: 59949992	-0.0294	0.0049	2.30E-08	1.93E-03	148
cg12772738	chr20: 36616990	-0.0451	0.0075	2.76E-08	2.08E-03	148
cg25981106	chr1: 209824796	-0.0362	0.0061	3.75E-08	2.58E-03	148
cg07327489	chr9: 134554052	-0.0183	0.0031	4.33E-08	2.73E-03	148
cg02331902	chr5: 90610302	-0.0132	0.0023	7.14E-08	4.15E-03	148
cg22898055	chr7: 75581160	-0.0511	0.0088	8.59E-08	4.37E-03	148
cg04451259	chr11: 35358923	-0.0594	0.0102	8.68E-08	4.37E-03	148
cg25225070	chr11: 9587742	-0.0286	0.0050	1.08E-07	4.84E-03	148
cg18648613	chr4: 102153422	-0.0241	0.0042	1.09E-07	4.84E-03	148
cg01266377	chr8: 128920442	-0.0291	0.0051	1.20E-07	5.06E-03	148
cg16477554	chr3: 24198188	-0.0425	0.0074	1.29E-07	5.12E-03	148
cg03128029	chr2: 203143287	-0.0312	0.0055	1.45E-07	5.49E-03	148

Supplementary Table 4 | Top 20 CpGs differentially methylated in muscle after the intervention accounting for changes in predicted endothelial nuclei proportions.

	Male skeletal muscle reference epigenome (E107)			Female skeletal muscle reference epigenome (E108)		
Chromatin state	**OR**	p	p_{FDR}	**OR**	p	p_{FDR}
Active TSS	0.0893	6.75E-04	3.37E-03	0.1171	2.24E-03	1.12E-02
Flanking active TSS	1.3482	2.02E-01	3.78E-01	1.6381	4.78E-02	1.02E-01
Flanking transcribed	1.7736	5.68E-01	8.52E-01	2.3216	4.02E-01	6.29E-01
Transcribed region	0.4532	8.11E-02	1.74E-01	0.2334	3.81E-02	1.02E-01
Weakly transcribed	1.1973	4.24E-01	7.06E-01	1.2288	4.19E-01	6.29E-01
Genic enhancer	4.3410	2.61E-06	1.96E-05	3.5642	1.09E-03	8.20E-03
Enhancer	7.1539	4.53E-35	6.79E-34	5.8304	2.50E-21	3.75E-20
Znf/Rpts	0.0001	9.38E-01	9.38E-01	0.0001	9.46E-01	9.46E-01
Heterochromatin	0.0001	8.95E-01	9.38E-01	0.0001	9.08E-01	9.46E-01
Bivalent TSS	0.0001	8.22E-01	9.38E-01	0.0001	8.44E-01	9.46E-01
Flanking bivalent	0.0001	8.14E-01	9.38E-01	0.0001	7.61E-01	9.46E-01
Bivalent enhancer	0.0001	8.20E-01	9.38E-01	0.0001	8.42E-01	9.46E-01
Polycomb repressed	0.1976	2.25E-02	5.64E-02	0.2593	5.44E-02	1.02E-01
Weak polycomb repressed	0.3959	1.00E-02	3.01E-02	0.4559	4.27E-02	1.02E-01
Quiescent	0.4737	1.40E-03	5.26E-03	0.6117	4.41E-02	1.02E-01

Supplementary Table 5 | Chromatin state enrichment analysis for muscle CpGs using the male (E107) and female (E108) skeletal muscle reference epigenomes from the Roadmap Epigenomics Consortium.

Motif Name	Consensus	p	Target n (%)	Background (%)
AP-1	VTGACTCATC	1.44E-25	36 (22.22)	2.18
Atf3	DATGASTCATHN	3.34E-25	34 (20.99)	1.91
Bach1	AWWNTGCTGAGTCAT	1.13E-04	4 (2.47)	0.15
Bach2	TGCTGAGTCA	1.27E-12	14 (8.64)	0.58
BATF	DATGASTCAT	1.33E-26	35 (21.60)	1.87
Fos	NDATGASTCAYN	9.07E-29	35 (21.60)	1.61
Fosl2	NATGASTCABNN	4.08E-26	28 (17.28)	0.99
Fra1	NNATGASTCATH	4.22E-27	33 (20.37)	1.52
Fra2	GGATGACTCATC	5.28E-25	30 (18.52)	1.35
HIF-1a	TACGTGCV	4.26E-03	7 (4.32)	1.24
Jun-AP1	GATGASTCATCN	6.20E-21	22 (13.58)	0.76
JunB	RATGASTCAT	1.97E-27	34 (20.99)	1.62
MafA	TGCTGACTCA	6.75E-05	13 (8.02)	2.19
MafB	WNTGCTGASTCAGCANWTTY	7.25E-04	7 (4.32)	0.90
MafK	GCTGASTCAGCA	4.36E-04	6 (3.70)	0.59
NF-E2	GATGACTCAGCA	2.43E-07	6 (3.70)	0.16
NFkB-p65-Rel	GGAAATTCCC	2.51E-04	3 (1.85)	0.07
Nrf2	HTGCTGAGTCAT	9.78E-08	6 (3.70)	0.13
RUNX1	AAACCACARM	1.74E-05	14 (8.64)	2.22
RUNX-AML	GCTGTGGTTW	6.67E-04	9 (5.56)	1.45

Supplementary Table 6 | TFBS enrichment analysis in sequences within 50bp of muscle CpGs. The percentage of background sequences within 50bp of the respective motif is also shown.

CpG	Gene	Ensembl ID	log_2FC	p_{FDR}	eQTM	SE	p_{FDR}
cg17411016	*MCFD2*	ENSG00000180398	-0.1464	7.59E-06	0.1574	0.0172	6.50E-13
cg09312464	*FSCN1*	ENSG00000075618	0.3990	1.41E-05	-0.0325	0.0043	3.54E-10
cg13585930	*LRRC20*	ENSG00000172731	-0.2115	2.16E-06	0.1950	0.0257	3.54E-10
cg12595459	*EXTL3*	ENSG00000012232	-0.1339	8.65E-04	0.0883	0.0121	1.22E-09
cg21730012	*INPP5A*	ENSG00000068383	-0.2053	3.58E-06	0.0813	0.0128	1.01E-07
cg02849956	*DPP9*	ENSG00000142002	-0.1718	7.13E-06	0.1341	0.0208	1.01E-07
cg20748397	*FLII*	ENSG00000177731	-0.2214	2.86E-06	0.0837	0.0135	1.63E-07
cg03045139	*LHFPL2*	ENSG00000145685	0.3691	4.54E-06	-0.0294	0.0048	1.72E-07
cg02233071	*RUNX1*	ENSG00000159216	0.7576	4.54E-06	-0.0196	0.0035	2.46E-06
cg20748397	*ALKBH5*	ENSG00000091542	-0.1247	9.39E-04	0.0934	0.0169	2.86E-06
cg25981106	*HSD11B1*	ENSG00000117594	-0.4646	3.79E-07	0.0364	0.0066	3.51E-06
cg17411016	*TTC7A*	ENSG00000068724	0.2008	2.61E-03	-0.0476	0.0086	3.51E-06
cg21005024	*GRB10*	ENSG00000106070	-0.2544	2.13E-04	0.0831	0.0157	7.08E-06
cg14426392	*ATP1A1*	ENSG00000163399	0.2351	5.87E-05	-0.0735	0.0150	3.44E-05
cg17357895	*EXOC3L1*	ENSG00000179044	0.2516	2.58E-04	-0.0554	0.0114	3.44E-05
cg12402318	*TMOD3*	ENSG00000138594	0.2408	7.59E-06	-0.0346	0.0071	3.52E-05
cg20748397	*GID4*	ENSG00000141034	-0.1100	6.51E-03	0.0759	0.0159	4.59E-05
cg07827395	*FAM220A*	ENSG00000178397	-0.2584	5.87E-05	0.0502	0.0106	5.09E-05
cg07827395	*RAC1*	ENSG00000136238	-0.1316	1.83E-04	0.0875	0.0190	8.74E-05
cg20617626	*ARHGEF17*	ENSG00000110237	0.1426	5.99E-04	-0.0406	0.0089	9.56E-05

Supplementary Table 7 | Top 20 differentially expressed genes in muscle that were also associated with DNAm changes at nearby CpGs. The CpG and genes are shown alongside the log_2FC and its significance (FDR-adjusted *p*-value), as well as the eQTM correlation, SE, and adjusted p-value.

CpG	Trait	Effect size	SE	*p*	p_{FDR}
cg27187848	Waist circumference	0.0023	0.0005	1.15E0-5	1.15E-04
cg16350675	Leptin levels	-0.0025	0.0006	4.12E-05	4.12E-04
cg17411016	Waist circumference	0.0022	0.0006	1.25E-04	1.25E-03
cg25114611	Leptin levels	-0.0021	0.0005	1.98E-04	1.98E-03
cg13733654	Waist circumference	-0.0007	0.0002	2.90E-04	2.90E-03
cg11835462	Leptin levels	-0.0023	0.0006	3.30E-04	3.30E-03
cg25114611	Waist circumference	-0.0015	0.0004	8.41E-04	4.21E-03
cg27187848	HDL cholesterol size	-0.0987	0.0290	8.99E-04	4.49E-03
cg13306815	Leptin levels	-0.0028	0.0008	5.67E-04	5.67E-03
cg12386285	Leptin levels	-0.0009	0.0002	5.72E-04	5.72E-03
cg12772738	Leptin levels	-0.0037	0.0011	7.35E-04	7.35E-03
cg17411016	Body fat percentage	0.0035	0.0011	2.00E-03	8.01E-03
cg17411016	HDL cholesterol size	-0.1015	0.0327	2.40E-03	8.01E-03
cg14435903	Adiponectin levels	-0.0028	0.0008	1.01E-03	1.01E-02
cg18710458	Body fat percentage	0.0046	0.0014	1.03E-03	1.03E-02
cg02501746	Leptin levels	-0.0023	0.0007	1.04E-03	1.04E-02
cg25628315	Leptin levels	-0.0026	0.0008	1.10E-03	1.10E-02
cg25225070	HDL cholesterol size	-0.1073	0.0326	1.34E-03	1.21E-02
cg25225070	Adiponectin levels	-0.0035	0.0011	2.41E-03	1.21E-02
cg19768360	HDL cholesterol size	0.0860	0.0263	1.41E-03	1.41E-02

Supplementary Table 8 | Top 20 associations between muscle CpGs and immunometabolic health markers.

CpG	Trait	beta	SE	p	p_{FDR}
cg09312464	PAX7+ cells per fibre	-0.1575	0.0239	1.62E-09	4.85E-09
cg15982707	PAX7+ cells per fibre	-0.2593	0.0397	2.44E-09	7.33E-09
cg24161080	PAX7+ cells per fibre	-0.4751	0.0743	4.34E-09	1.30E-08
cg17633300	PAX7+ cells per fibre	-0.1146	0.0182	7.11E-09	2.13E-08
cg11012616	PAX7+ cells per fibre	-0.2104	0.0333	9.42E-09	2.83E-08
cg17411016	PAX7+ cells per fibre	-0.1907	0.0304	1.00E-08	3.01E-08
cg25981106	PAX7+ cells per fibre	-0.2723	0.0438	1.19E-08	3.56E-08
cg01668986	PAX7+ cells per fibre	-0.3013	0.0493	1.82E-08	5.45E-08
cg17370665	PAX7+ cells per fibre	-0.2303	0.0379	2.38E-08	7.15E-08
cg17357895	PAX7+ cells per fibre	-0.3072	0.0513	2.89E-08	8.68E-08
cg21730012	PAX7+ cells per fibre	-0.1683	0.0283	3.74E-08	1.12E-07
cg00545756	PAX7+ cells per fibre	-0.2086	0.0354	4.51E-08	1.35E-07
cg27461254	PAX7+ cells per fibre	-0.1996	0.0340	5.53E-08	1.66E-07
cg14343713	PAX7+ cells per fibre	-0.1446	0.0246	6.67E-08	2.00E-07
cg25132536	PAX7+ cells per fibre	-0.3268	0.0562	7.41E-08	2.22E-07
cg12166519	PAX7+ cells per fibre	-0.1852	0.0321	8.66E-08	2.60E-07
cg27187848	PAX7+ cells per fibre	-0.1576	0.0276	1.22E-07	3.66E-07
cg20617626	PAX7+ cells per fibre	-0.1231	0.0217	1.30E-07	3.90E-07
cg04814966	myonuclei per fibre	-0.0075	0.0014	3.29E-07	5.15E-07
cg04814966	PAX7+ cells per fibre	-0.1258	0.0231	3.43E-07	5.15E-07

Supplementary Table 9 | Top 20 associations between DNAm at muscle CpGs and immunohistochemistry markers.

Cell type	Baseline	SD	Delta	SE	p	p_{FDR}
Adipocytes	72.45 %	7.83	-0.40	0.99	6.87E-01	7.28E-01
CD4+ T-cells	0.15 %	0.50	-0.08	0.06	2.00E-01	7.28E-01
MVECs	24.06 %	6.53	0.32	0.92	7.28E-01	7.28E-01
M1 Macrophages	0.03 %	0.16	0.03	0.04	3.73E-01	7.28E-01
M2 Macrophages	3.31 %	2.26	0.12	0.19	5.20E-01	7.28E-01

Supplementary Table 10 | Baseline percentages and changes in predicted adipose tissue cell types.

CpG	Position	beta	SE	p	p_{FDR}	n
cg12544951	chr20: 21695342	0.0738	0.0119	3.66E-09	1.24E-03	178
cg06524692	chr1: 28864470	0.0257	0.0041	4.03E-09	1.24E-03	178
cg01733176	chr4: 111561069	-0.0396	0.0064	5.50E-09	1.24E-03	178
cg03475429	chr10: 94982392	0.0236	0.0039	6.58E-09	1.24E-03	178
cg08156066	chr20: 21695333	0.0636	0.0105	9.56E-09	1.44E-03	178
cg05844247	chr20: 21694427	0.0632	0.0106	1.72E-08	1.64E-03	178
cg07931189	chr2: 53463157	-0.0338	0.0057	1.77E-08	1.64E-03	178
cg10629004	chr20: 21696466	0.0447	0.0075	1.90E-08	1.64E-03	178
cg13655674	chr1: 119522385	-0.0210	0.0035	1.97E-08	1.64E-03	178
cg07213060	chr20: 21694826	0.0681	0.0115	2.17E-08	1.64E-03	178
cg18574731	chr10: 116415067	0.0192	0.0033	3.04E-08	2.09E-03	178
cg17596409	chr4: 153186304	0.0652	0.0112	3.67E-08	2.31E-03	178
cg15556943	chr20: 21694616	0.0765	0.0133	4.50E-08	2.51E-03	178
cg06032603	chr9: 38127746	0.0411	0.0071	4.65E-08	2.51E-03	178
cg22888671	chr17: 26799382	0.0215	0.0037	5.13E-08	2.58E-03	178
cg17908503	chr20: 21690788	0.0566	0.0099	6.55E-08	2.91E-03	178
cg19404433	chr20: 21695772	0.0333	0.0058	6.55E-08	2.91E-03	177
cg18059621	chr2: 105486198	0.0303	0.0053	7.46E-08	3.13E-03	178
cg00497086	chr16: 23869599	0.0215	0.0038	9.94E-08	3.76E-03	178
cg20432507	chr20: 21689860	0.0493	0.0088	9.96E-08	3.76E-03	178

Supplementary Table 11 | Top 20 CpGs differentially methylated after the intervention in adipose tissue.

Chromatin state	OR	p	p_{FDR}
Active TSS	0.1051	1.04E-04	5.21E-04
Flanking active TSS	0.1271	3.45E-05	2.59E-04
Flanking transcribed	0.0001	9.22E-01	9.87E-01
Transcribed region	0.3265	3.51E-03	1.05E-02
Weakly transcribed	0.5178	1.35E-02	2.90E-02
Genic enhancer	0.0001	7.91E-01	9.87E-01
Enhancer	0.3798	4.41E-03	1.10E-02
Znf/Rpts	0.0001	9.18E-01	9.87E-01
Heterochromatin	0.9895	9.88E-01	9.88E-01
Bivalent TSS	0.0001	8.71E-01	9.87E-01
Flanking bivalent	1.1540	7.09E-01	9.87E-01
Bivalent enhancer	1.8753	8.07E-02	1.34E-01
Polycomb repressed	7.7623	4.90E-44	7.35E-43
Weak polycomb repr.	1.5008	2.09E-02	3.92E-02
Quiescent	1.6535	4.06E-04	1.52E-03

Supplementary Table 12 | Chromatin state enrichment analysis for identified CpGs using the adipocyte reference epigenome from the Roadmap Epigenomics Consortium (E063). The chromatin state, odds ratio (OR), and significance of the enrichment is shown (nominal and FDR-adjusted *p*-values; TSS: transcription start site).

Motif Name	Consensus	*p*	Target *n* (%)	Background (%)
Gata2(Zf)	BBCTTATCTS	3.73E-04	11 (4.78)	1.36
Gata6(Zf)	YCTTATCTBN	1.44E-03	12 (5.22)	1.86
Gata1(Zf)	SAGATAAGRV	2.58E-03	9 (3.91)	1.24
Fra2(bZIP)	GGATGACTCATC	6.62E-03	6 (2.61)	0.72
Tgif2(Homeobox)	TGTCANYT	6.75E-03	30 (13.04)	8.10
Gata4(Zf)	NBWGATAAGR	8.03E-03	11 (4.78)	2.03
Fra1(bZIP)	NNATGASTCATH	8.92E-03	6 (2.61)	0.76

Supplementary Table 13 | TFBS enrichment analysis in sequences within 50bp of identified CpGs in adipose tissue. The TF name and type is shown alongside its consensus motif, the significance of the enrichment, and the number and percentage of CpGs responsible for the enrichment. The percentage of background sequences within 50bp of the respective motif is also shown.

CpG	Gene	Ensembl ID	$\log_2$FC	p_{FDR}	eQTM	SE	p_{FDR}
cg02649849	*DMRT3*	ENSG00000064218	-0.3375	1.63E-03	0.0199	0.0018	8.13E-18
cg03148184	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0592	0.0058	1.21E-16
cg26708319	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0564	0.0057	5.18E-16
cg19370653	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0621	0.0065	1.68E-15
cg07790170	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0696	0.0074	8.49E-15
cg03943773	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0507	0.0056	1.89E-14
cg23646776	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0365	0.0040	3.67E-14
cg24005685	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0582	0.0065	4.44E-14
cg10895452	*EN1*	ENSG00000163064	-0.1853	4.64E-03	-0.0146	0.0017	1.64E-13
cg24925400	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0502	0.0059	6.84E-13
cg01951086	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0654	0.0079	2.24E-12
cg23806894	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0485	0.0063	2.44E-11
cg17242937	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0510	0.0067	4.32E-11
cg21299542	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0309	0.0041	4.71E-11
cg01733176	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0428	0.0060	4.01E-10
cg26023087	*DMRT3*	ENSG00000064218	-0.3375	1.63E-03	0.0157	0.0022	5.90E-10
cg19849728	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0380	0.0059	1.23E-08
cg14434922	*DMRT3*	ENSG00000064218	-0.3375	1.63E-03	0.0134	0.0022	8.69E-08
cg05581451	*PITX2*	ENSG00000164093	-0.2134	1.23E-02	0.0341	0.0058	2.06E-07
cg18792984	*TREM1*	ENSG00000124731	-0.3144	1.27E-02	0.0170	0.0029	2.33E-07

Supplementary Table 14 | Top 20 differentially expressed genes in adipose tissue that were also associated with DNAm changes at nearby CpGs. The CpG and associated gene identifiers are shown alongside the $\log_2$FC and its significance (FDR-adjusted *p*-value), and the eQTM effect size (eQTM), standard error (SE), and significance (FDR-adjusted p-value).

Database	Term	p_{FDR}	Overlapping genes
GO BP (2023)	Embryonic Organ Morphogenesis (GO:0048562)	1.27E-02	*ALX1, HOXB3, HOXB4*
GO BP (2023)	Phospholipid Efflux (GO:0033700)	1.86E-02	*APOC1, APOE*
GO BP (2023)	Regulation Of Transcription By RNA Polymerase II (GO:0006357)	1.86E-02	*ALX1, DMRT3, EN1, HOXB2, HOXB3, HOXB4, HOXB8, NR2F1, PITX2, SAP30L*
GO BP (2023)	High-Density Lipoprotein Particle Remodelling (GO:0034375)	1.86E-02	*APOC1, APOE*
Reactome (2022)	Plasma Lipoprotein Assembly R-HSA-8963898	1.86E-02	*APOC1, APOE*
Reactome (2022)	Activation Of HOX Genes During Differentiation R-HSA-5619507	1.86E-02	*HOXB2, HOXB3, HOXB4*
GO BP (2023)	Cholesterol Efflux (GO:0033344)	2.68E-02	*APOC1, APOE*
GO BP (2023)	Regulation Of DNA-templated Transcription (GO:0006355)	2.68E-02	*ALX1, APOE, DMRT3, EN1, HOXB3, HOXB4, HOXB8, PITX2, SAP30L*
GO BP (2023)	Hematopoietic Progenitor Cell Differentiation (GO:0002244)	3.32E-02	*ARMC6, HOXB4*
GO BP (2023)	Embryonic Skeletal System Morphogenesis (GO:0048704)	3.32E-02	*HOXB3, HOXB4*
GO BP (2023)	Positive Regulation Of DNA-templated Transcription (GO:0045893)	3.32E-02	*ALX1, APOE, HOXB2, HOXB3, HOXB4, NR2F1, PITX2*
GO BP (2023)	Embryonic Skeletal System Development (GO:0048706)	3.32E-02	*HOXB3, HOXB4*
Reactome (2022)	Plasma Lipoprotein Clearance R-HSA-8964043	3.32E-02	*APOC1, APOE*
Reactome (2022)	NR1H3 And NR1H2 Regulate Gene Expression Linked To Cholesterol Transport And Efflux	3.32E-02	*APOC1, APOE*
Reactome (2022)	Transcriptional Regulation By AP-2 (TFAP2) Family Of Transcription Factors	3.32E-02	*APOE, PITX2*
GO BP (2023)	Positive Regulation Of Transcription By RNA Polymerase II (GO:0045944)	3.32E-02	*ALX1, HOXB2, HOXB3, HOXB4, NR2F1, PITX2*
GO BP (2023)	Acylglycerol Metabolic Process (GO:0006639)	3.85E-02	*APOC1, APOE*
GO BP (2023)	Skeletal System Morphogenesis (GO:0048705)	4.00E-02	*HOXB3, HOXB4*
GO BP (2023)	Triglyceride Metabolic Process (GO:0006641)	4.12E-02	*APOC1, APOE*
Reactome (2022)	NR1H2 And NR1H3-mediated Signalling R-HSA-9024446	4.12×1-02	*APOC1, APOE*

Supplementary Table 15 | Top 20 enriched genes sets in differentially expressed genes in adipose tissue also associated with DNAm changes at nearby CpGs.

CpG	Trait	beta	SE	p	p_{FDR}
cg07219853	Interleukin-6 levels	-0.0211	0.0051	5.06E-05	5.06E-04
cg04715165	Interleukin-6 levels	-0.0211	0.0052	7.34E-05	7.34E-04
cg04543233	Body fat percentage	-0.0045	0.0011	8.96E-05	8.96E-04
cg06161697	Body fat percentage	0.0032	0.0008	9.22E-05	9.22E-04
cg03160217	Interleukin-6 levels	-0.0120	0.0030	1.02E-04	1.02E-03
cg20907471	Body fat percentage	-0.0062	0.0016	2.02E-04	2.02E-03
cg02355868	Fasting insulin levels	0.0026	0.0007	2.22E-04	2.22E-03
cg23994043	Interleukin-6 levels	0.0195	0.0052	2.29E-04	2.29E-03
cg02468230	Body fat percentage	0.0030	0.0008	2.81E-04	2.81E-03
cg04290158	Body fat percentage	0.0031	0.0009	3.15E-04	3.15E-03
cg23247845	Interleukin-6 levels	-0.0172	0.0047	3.33E-04	3.33E-03
cg02355868	Body fat percentage	0.0027	0.0008	7.26E-04	3.63E-03
cg12022722	Leptin levels	-0.0013	0.0003	3.69E-04	3.69E-03
cg17193941	Interleukin-6 levels	-0.0145	0.0040	4.21E-04	4.21E-03
cg10629004	Body fat percentage	-0.0055	0.0015	4.24E-04	4.24E-03
cg07213060	Interleukin-6 levels	0.0307	0.0086	4.50E-04	4.50E-03
cg23660235	Interleukin-6 levels	-0.0191	0.0054	4.76E-04	4.76E-03
cg07452809	Interleukin-6 levels	-0.0188	0.0053	5.04E-04	5.04E-03
cg02355868	Waist circumference	0.0012	0.0004	1.66E-03	5.55E-03
cg02621151	Interleukin-6 levels	0.0158	0.0045	5.60E-04	5.60E-03

Supplementary Table 16 | Top 20 associations between identified CpGs in adipose tissue and immunometabolic health markers. The CpG identifier and associated trait is shown alongside its effect size (beta), standard error (SE), and significances (nominal and FDR-adjusted *p*-value)

Measured cell type	Baseline	SD	Beta	SE	p	p_{FDR}
Neutrophils	50.37 %	8.50	0.215	0.666	7.48E-01	9.46E-01
Lymphocytes	36.89 %	8.36	-0.067	0.563	9.06E-01	9.46E-01
Monocytes	8.62 %	2.15	-0.034	0.209	8.73E-01	9.46E-01
Eosinophils	3.57 %	2.50	-0.128	0.145	3.78E-01	9.46E-01
Basophils	0.53 %	0.30	0.017	0.030	5.76E-01	9.46E-01
MuSiC cell types	**Baseline**	**SD**	**Beta**	**SE**	p	p_{FDR}
Mature neutrophils	28.89 %	5.35	0.812	0.607	1.84E-01	9.46E-01
Meta-myelocyte	24.48 %	4.92	0.065	0.637	9.18E-01	9.46E-01
Intermediate monocytes	20.87 %	5.37	-0.535	0.676	4.31E-01	9.46E-01
CD8+ T-cells	18.80 %	8.75	-1.604	0.948	9.45E-02	9.46E-01
Pre-monocyte	3.13 %	1.60	0.044	0.208	8.32E-01	9.46E-01
Classical monocytes	1.17 %	2.04	0.505	0.329	1.29E-01	9.46E-01
Immature B-cells	1.12 %	2.10	0.031	0.115	7.92E-01	9.46E-01
Non-classical monocytes	0.70 %	1.42	-0.201	0.230	3.84E-01	9.46E-01
CD4+ T-cells	0.29 %	2.00	1.012	0.679	1.38E-01	9.46E-01
Erythrocytes	0.20 %	0.27	-0.094	0.027	7.51E-04	2.07E-02
Naïve B-cells	0.11 %	1.02	-0.015	0.066	8.16E-01	9.46E-01
Common monocyte progenitors	0.09 %	0.46	-0.003	0.043	9.46E-01	9.46E-01
Myelocyte	0.08 %	0.71	-0.050	0.085	5.63E-01	9.46E-01
Plasma cells	0.03 %	0.06	-0.001	0.008	9.20E-01	9.46E-01
Monocyte-dendritic progenitors	0.02 %	0.10	0.012	0.023	6.09E-01	9.46E-01
Cytotoxic natural killer cells	0.01 %	0.03	0.001	0.003	6.46E-01	9.46E-01
Regulatory B-cells	<0.001 %	0.01	-0.001	0.001	4.22E-01	9.46E-01
IDOL cell types	**Baseline**	**SD**	**Beta**	**SE**	p	p_{FDR}
Neutrophils	51.84 %	7.66	-0.117	0.673	8.62E-01	9.46E-01
CD4+ T-cells	15.74 %	5.72	0.073	0.338	8.31E-01	9.46E-01
CD8+ T-cells	12.98 %	3.94	0.037	0.210	8.59E-01	9.46E-01
Monocytes	10.13 %	2.37	0.057	0.193	7.69E-01	9.46E-01
B-cells	6.52 %	4.57	0.130	0.128	3.12E-01	9.46E-01
Natural killer cells	5.88 %	3.09	-0.202	0.175	2.52E-01	9.46E-01
IDOLext cell types	**Baseline**	**SD**	**Beta**	**SE**	p	p_{FDR}
Neutrophils	47.72 %	8.78	-0.190	0.748	8.00E-01	9.46E-01
CD4+ memory T-cells	10.42 %	3.70	0.313	0.269	2.48E-01	9.46E-01
Monocytes	7.35 %	2.01	0.023	0.165	8.92E-01	9.46E-01
CD8+ memory T-cells	7.27 %	4.78	-0.102	0.215	6.35E-01	9.46E-01
CD4+ naïve T-cells	6.81 %	4.43	-0.122	0.185	5.12E-01	9.46E-01
Natural killer cells	5.59 %	2.43	-0.142	0.155	3.60E-01	9.46E-01
Naïve B-cells	3.60 %	1.73	0.096	0.081	2.41E-01	9.46E-01
Memory B-cells	2.38 %	4.91	-0.011	0.085	8.94E-01	9.46E-01
Eosinophils	2.37 %	2.45	-0.052	0.198	7.93E-01	9.46E-01
Regulatory T-cells	1.27 %	0.98	0.027	0.074	7.17E-01	9.46E-01
CD8+ naïve T-cells	0.76 %	1.27	0.083	0.071	2.43E-01	9.46E-01
Basophils	0.68 %	0.63	0.099	0.055	7.42E-02	9.46E-01

Supplementary Table 17 | Baseline percentages and changes in blood cell types using measured cell types and three deconvolution algorithms.

CpG	Position	beta	SE	p	p_{FDR}	n
cg23961638	chr4:153700210	-0.0029	0.0006	2.06E-09	9.50E-04	194
cg16209776	chr12:63177893	-0.0117	0.0023	2.51E-09	9.50E-04	196
cg26201957	chr11:82601012	-0.0278	0.0056	5.98E-09	1.23E-03	195
cg13479371	chr18:72530930	-0.0050	0.0010	6.72E-09	1.23E-03	196
cg25040733	chr5:141704733	0.0121	0.0024	8.12E-09	1.23E-03	194
cg08769073	chr11:12214803	0.0086	0.0018	1.32E-08	1.67E-03	196
cg17279445	chr4:188428636	0.0086	0.0018	2.42E-08	2.61E-03	193
cg09925408	chr19:1578728	0.0016	0.0003	4.72E-08	3.78E-03	196
cg06860177	chr10:122938108	-0.0182	0.0039	4.98E-08	3.78E-03	191
cg15280030	chr3:127637280	0.0097	0.0021	5.00E-08	3.78E-03	196
cg21196038	chr6:169560205	-0.0051	0.0011	6.60E-08	4.08E-03	194
cg07318969	chr9:101366405	0.0064	0.0014	6.96E-08	4.08E-03	196
cg09404334	chr6:86304179	-0.0049	0.0011	7.31E-08	4.08E-03	196
cg12484370	chr9:14314969	-0.0034	0.0007	7.62E-08	4.08E-03	192
cg17707984	chr6:31830405	-0.0033	0.0007	8.91E-08	4.08E-03	194
cg12559900	chr15:85268268	0.0023	0.0005	9.12E-08	4.08E-03	196
cg09910998	chr7:94285941	-0.0123	0.0027	9.17E-08	4.08E-03	196
cg13679891	chr1:177669461	0.0298	0.0065	9.72E-08	4.08E-03	194
cg16868350	chr19:55064372	0.0354	0.0077	1.13E-07	4.49E-03	190
cg00784718	chr1:221916685	-0.0069	0.0015	1.25E-07	4.74E-03	194

Supplementary Table 18 | Top 20 CpGs differentially methylated after the intervention in blood. The CpG identified and its position are shown alongside the intervention effect size (beta), its standard error (SE), significance (nominal and FDR-adjusted *p*-value), and sample size used to derive the effect (*n*).

Chromatin state	OR	p	p_{FDR}
Active TSS	1.2061	3.38E-01	7.23E-01
Flanking active TSS	1.4998	1.47E-03	1.11E-02
Flanking transcribed	1.1703	8.75E-01	9.67E-01
Transcribed region	1.2934	7.44E-02	3.72E-01
Weakly transcribed	0.9176	5.74E-01	9.43E-01
Genic enhancer	0.6663	4.84E-01	9.07E-01
Enhancer	0.6273	1.27E-01	4.75E-01
Znf/Rpts	1.3934	7.40E-01	9.43E-01
Heterochromatin	0.9704	9.59E-01	9.67E-01
Bivalent TSS	0.6468	3.33E-01	7.23E-01
Flanking bivalent	0.6379	3.17E-01	7.23E-01
Bivalent enhancer	1.1514	7.54E-01	9.43E-01
Polycomb repressed	0.3420	7.50E-04	1.11E-02
Weak polycomb repression	0.9326	6.37E-01	9.43E-01
Quiescent	0.9958	9.67E-01	9.67E-01

Supplementary Table 19 | Chromatin state enrichment analysis for identified CpGs using the PBMC reference epigenome from the Roadmap Epigenomics Consortium (E062). The chromatin state, odds ratio (OR), and significance of the enrichment is shown (nominal and FDR-adjusted *p*-values; TSS: transcription start site).

Motif Name	Consensus	*p*	Target *n* (%)	Background (%)
NFE2L2(bZIP)	AWWWTGCTGAGTCAT	1.00E-03	3 (0.68)	0.04%
Sox9(HMG)	AGGVNCCTTTGT	1.00E-02	13 (2.95)	1.23%
MafK(bZIP)	GCTGASTCAGCA	1.00E-02	6 (1.36)	0.36%

Supplementary Table 20 | TFBS enrichment analysis in sequences within 50bp of identified CpGs in blood. The TF name and type is shown alongside its consensus motif, the significance of the enrichment, and the number and percentage of CpGs responsible for the enrichment. The percentage of background sequences within 50bp of the respective motif is also shown.

CpG	Gene	Ensembl ID	$\log_2$FC	p_{FDR}	eQTM	SE	p_{FDR}
cg18032191	*LTBR*	ENSG00000111321	0.155	3.97E-03	-0.019	0.004	3.26E-05
cg18032191	*TNFRSF1A*	ENSG00000067182	0.123	1.65E-02	-0.018	0.004	7.55E-05
cg21095811	*ATG2B*	ENSG00000066739	-0.086	3.81E-02	-0.004	0.001	8.85E-03

Supplementary Table 21 | Differentially expressed genes in blood that were also associated with DNAm changes at nearby CpGs. The CpG and associated gene identifiers are shown alongside the $\log_2$FC and its significance (FDR-adjusted p-value), and the eQTM effect size (eQTM), SE and significance (FDR-adjusted *p*-value).

CpG	Trait	Beta	SE	*p*	p_{FDR}
cg16209776	Waist circumference	0.0011	0.0002	4.80E-06	4.80E-05
cg09222461	HDL cholesterol size	-0.0277	0.0061	1.62E-05	1.62E-04
cg16657397	Leptin levels	-0.0008	0.0002	2.73E-05	1.69E-04
cg16657397	Body fat percentage	-0.0014	0.0003	3.38E-05	1.69E-04
cg06612452	Waist circumference	0.0003	0.0001	4.58E-05	4.58E-04
cg07735194	Interleukin-6 levels	0.0086	0.0021	5.82E-05	5.82E-04
cg16610939	HDL cholesterol size	-0.0540	0.0131	6.46E-05	6.46E-04
cg17061897	Body mass index	0.0032	0.0008	7.04E-05	7.04E-04
cg08943940	HDL cholesterol size	-0.0423	0.0110	1.79E-04	1.79E-03
cg18945877	HDL cholesterol size	-0.0753	0.0197	1.91E-04	1.91E-03
cg17061897	Waist circumference	0.0007	0.0002	4.90E-04	2.45E-03
cg04103122	Leptin levels	0.0011	0.0003	3.52E-04	3.52E-03
cg27120531	Waist circumference	0.0011	0.0003	7.17E-04	3.79E-03
cg27120531	Body mass index	0.0043	0.0012	7.57E-04	3.79E-03
cg15815515	Body fat percentage	0.0060	0.0017	5.22E-04	5.22E-03
cg27422762	Body fat percentage	0.0036	0.0010	5.35E-04	5.35E-03
cg25440376	Systolic blood pressure	0.0002	0.0001	5.53E-04	5.53E-03
cg16209776	Leptin levels	0.0009	0.0003	1.42E-03	5.91E-03
cg16209776	Body fat percentage	0.0016	0.0005	1.77E-03	5.91E-03
cg24722351	Body fat percentage	0.0014	0.0004	6.31E-04	6.31E-03

Supplementary Table 22 | Top 20 associations between identified CpGs in blood and immunometabolic health markers. The CpG identifier and associated trait is shown alongside its effect size (beta), standard error (SE), and significances (nominal and FDR-adjusted *p*-value; HDL: high density lipoprotein)

cAge clock	Delta	Change in weeks	SE	p	p_{FDR}
Horvath (2013)	-0.430	-22.349	0.333	2.004E-01	2.004E-01
Zhang (2019)	-0.337	-17.503	0.116	4.542E-03	1.363E-02
Bernabeu cAge (2023)	-0.241	-12.545	0.147	1.031E-01	1.547E-01
Bernabeu bAge (2023)	-0.3092	-16.0803	0.2639	2.44E-01	2.44E-01
grimAge	-0.6658	-34.6191	0.1542	3.61E-05	1.44E-04
phenoAge	-0.7788	-40.4974	0.3751	4.04E-02	5.39E-02
MEAT	-1.1125	-57.8510	0.4230	1.03E-02	2.05E-02

Supplementary Table 23 | Predictions of change in biological (bAge) and chronological (cAge) age following the GOTO intervention using various algorithms. The estimated change (shown in years and weeks) is shown alongside the standard error (SE) and significance (nominal and FDR-adjusted *p*-values).

bAge clock	Trait	beta	SE	p	p_{FDR}
grimAge	Body mass index	0.2983	0.0844	5.21E-04	5.21E-03
grimAge	Waist circumference	0.0733	0.0199	3.25E-04	5.21E-03
grimAge	Leptin levels	0.0847	0.0268	1.89E-03	1.26E-02
grimAge	Body fat percentage	0.1013	0.0406	1.37E-02	3.91E-02
grimAge	Fasting insulin	0.1064	0.0411	1.07E-02	3.91E-02
grimAge	Interleukin-6 levels	0.3027	0.1179	1.13E-02	3.91E-02
grimAge	HDL cholesterol size	-3.2202	1.2836	1.30E-02	3.91E-02
grimAge	Adiponectin levels	-0.1249	0.0572	3.02E-02	7.56E-02
grimAge	HDL cholesterol levels	-0.9049	0.5631	1.10E-01	2.44E-01
MEAT	Leptin levels	0.0619	0.0535	2.50E-01	5.00E-01
MEAT	HDL cholesterol levels	1.1047	1.0718	3.05E-01	5.54E-01
MEAT	Interleukin-6 levels	0.1971	0.2185	3.69E-01	6.15E-01
grimAge	Systolic blood pressure	0.0063	0.0114	5.84E-01	7.67E-01
MEAT	Waist circumference	0.0181	0.0400	6.52E-01	7.67E-01
MEAT	Body fat percentage	0.0465	0.0798	5.61E-01	7.67E-01
MEAT	Fasting insulin	-0.0392	0.0786	6.19E-01	7.67E-01
MEAT	Systolic blood pressure	-0.0099	0.0217	6.47E-01	7.67E-01
MEAT	Body mass index	0.0439	0.1430	7.60E-01	8.44E-01
MEAT	Adiponectin levels	-0.0148	0.0897	8.69E-01	8.90E-01
MEAT	HDL cholesterol size	0.3155	2.2756	8.90E-01	8.90E-01

Supplementary Table 24 | Associations between bAge as measured by MEAT and grimAge, and ten health parameters. The algorithm used and associated trait is shown alongside the effect size (beta), its standard error (SE), and significance (nominal and FDR-adjusted *p*-value; HDL: high density lipoprotein)

CHAPTER SEVEN

Discussion

Aims of this thesis

A substantial body of epidemiological work positioned DNA methylation (DNAm) as a central component of ageing and age-related disease[1,2]. Large-scale epigenome-wide association studies (EWAS) robustly connected CpGs at metabolically relevant loci to a range of related factors, from dietary patterns to diabetes risk[2–4]. These findings cemented EWAS as a cornerstone of population epigenetics, and a compelling framework through which to dissect the molecular underpinnings of human health[5].

In parallel, circulating signalling molecules emerged as sensitive readouts of immunometabolic health across the lifespan[6]. These metabolic and inflammatory mediators, including interleukin-6 (IL-6), adiponectin, and leptin, were implicated in disease pathogenesis, predictive of diabetes onset and mortality, and could quantify intervention-driven health improvements[7–9]. Their responsiveness, prognostic value, and strong links to longevity rendered them prime candidates for capturing heterogeneity in healthy ageing[10–12]. Yet, the epigenetic architecture of such measures remained largely unexplored.

This thesis aimed to bridge the domains of epigenomics and immunometabolic health, uncovering how their intersection limits or leads to age-related disease. Specifically, it sought to map blood-based DNAm signatures of three related markers: IL-6, adiponectin, and leptin, and characterise multi-tissue epigenetic responses to a lifestyle intervention, in which older adults improved their immunometabolic health through a combination of increased physical activity and calorie restriction[13]. By assimilating multiple genomic layers, this work strived to trace causal pathways and reveal their regulatory points of control. Furthermore, by systematically embedding biological and clinical context into this integrative approach, it hoped to move beyond the nucleotide resolution of CpGs back up to higher-order insights with direct translational relevance.

To support these goals, this thesis endeavoured to develop and disseminate open-source computational tools. Equipped with clear and accessible documentation, this software set out to support epidemiological researchers throughout the analytical workflow, from study design to statistical inference. In sum, this work aspired to provide both empirical evidence and practical resources, taking meaningful strides towards realising the full potential of EWAS to understand, and ultimately improve, human health in ageing.

Summary of chapters

This thesis began by presenting EWAS meta-analyses of three blood-based markers of immunometabolic health. In a combined sample of over 4,000 individuals, **Chapter 2** identified 401 CpGs associated with IL-6 levels and mapped these sites to inflammatory and metabolic genes, such as *AIM2*, *MTOR*, and *IL6R*. IL-6 is secreted by circulating cells, meaning this methylation could have reflected direct epigenetic reprogramming of their transcriptional profiles and IL-6 production capabilities. In line with this, two-sample Mendelian randomisation (2SMR) implicated one locus (*NFATC2IP*) as plausibly influencing not only IL-6 production but also multiple immunometabolic traits, including body mass index (BMI) and type 2 diabetes (T2D).

Nevertheless, this CpG represented an exception to the trend supported by triangulation, in which DNAm was typically a consequence rather than a cause of IL-6. Follow-up mediation analysis uncovered two sites at *SOCS3* that not only responded to IL-6, but also partially transmitted its effects on C-reactive protein (CRP), BMI, and inflammatory bowel disease. This Chapter therefore broadly concluded that blood-based DNAm reflects immune cell exposure to IL-6 in the circulation, and that these responses can in turn propagate effects on inflammatory risk factors and disease.

Chapter 3 extended our approach to adiponectin and leptin, two adipokines secreted into the circulation by adipocytes. In this context, three causal scenarios were evaluated: (i) reverse causation, where immune cell DNAm responds to adipokine levels, (ii) leukocyte-mediated regulation of adipokine production, and (iii) shared upstream factors driving correlated epigenetic patterns in both blood and adipose tissues. Meta-analysis identified 73 CpGs associated with adiponectin and 211 sites linked to leptin, with triangulation revealing that, in contrast to IL-6, DNAm may play a role in regulating adiponectin production.

Two loci stood out: a CpG near *ADIPOQ*, the gene encoding adiponectin itself, and another connected to *SREBF1* expression, a transcription factor (TF) central to lipid metabolism that promotes pro-inflammatory polarisation of macrophages[14,15]. While the *ADIPOQ* CpG was discovered in blood, its links to *ADIPOQ* expression could not be examined here as immune cells do not produce adiponectin. Instead, we connected this CpG to *ADIPOQ* expression in adipocytes, and hypothesized that a shared upstream factor, such as diet, may be driving co-ordinated DNAm in blood and adipose tissues.

While such a causal structure was also plausible at the *SREBF1* site, an alternative model supported DNAm as epigenetically priming immune cells to amplify adipose tissue inflammation upon immune infiltration. This Chapter therefore showcased how large-scale blood-based studies can uncover biologically meaningful loci for traits non-hematopoietic in origin, and exemplified their validation in a targeted follow-up analysis.

The next section of this thesis shifted focus to software development, setting out to facilitate evaluation of DNAm in a novel experimental setting. **Chapter 4** introduced *Omixer*, an open-source Bioconductor package that proactively mitigated technical confounding in genomic studies. Designed to accommodate paired and complex study designs, *Omixer* outperformed standard randomization approaches, reducing associations between technical and biological factors, and thereby enhancing confidence in downstream results.

In **Chapter 5**, we presented *DNAmArray*, a flexible and scalable computational workflow for processing array-based DNAm data. Building on tools developed by our research group, including *methylAid*[16], *omicsPrint*[17], and *bacon*[18], this Chapter showcased streamlined and modular steps for quality control, normalisation, and analysis, providing high quality data and reproducible results in subsequent studies.

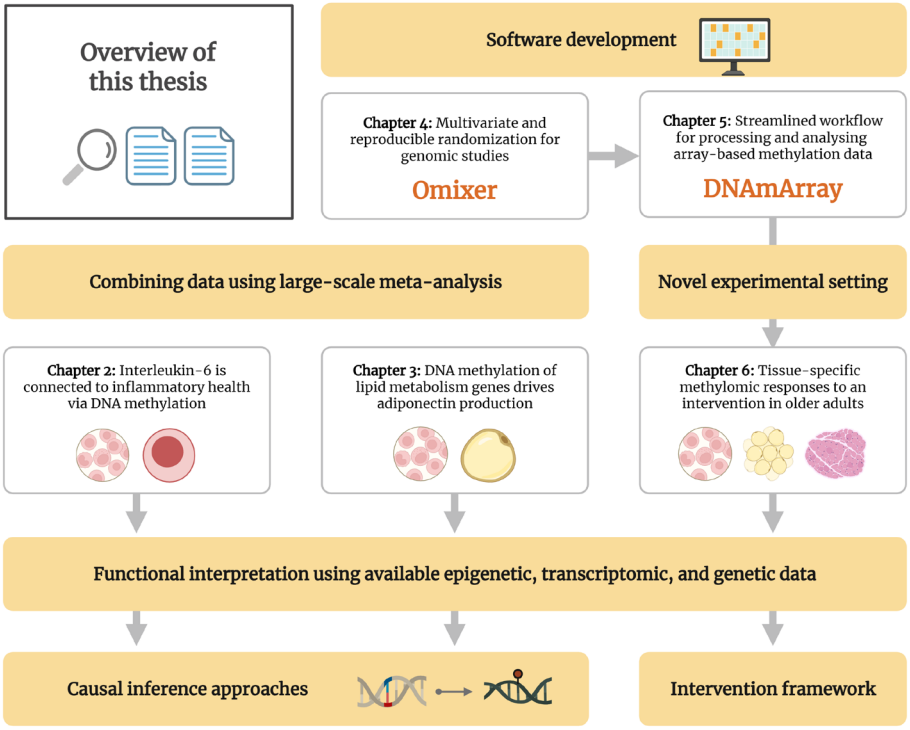


Figure 1 | Overview of this thesis. *Chapters 2 and 3* combined data using large-scale meta-analysis and evaluated DNAm signatures of IL-6, adiponectin, and leptin. *Chapters 4 and 5* developed software for use in *Chapter 6*, a novel experimental setting that interrogated tissue-specific methylomic responses to a lifestyle intervention in older adults. Tissues and approaches used are illustrated. *Figure created in BioRender.*

We made effective use of our integrative framework and software in **Chapter 6**, profiling DNAm before and after the Growing Old TOgether (GOTO) lifestyle intervention, where older adults improved their immunometabolic health through a combination of increased physical activity and calorie restriction. Three metabolically relevant tissues were investigated, revealing 162 differentially methylated CpGs in muscle, 230 sensitive sites in adipose, and 441 loci responsive to the intervention in blood. Yet, despite high tissue-specificity in these epigenetic signatures, all three could be connected to immunometabolic and physiological health changes, including reductions in BMI and leptin levels.

This last Chapter also applied epigenetic clock algorithms, uncovering consistent decreases in biological age (bAge) following the GOTO intervention. GrimAge, in particular, was associated with alterations in health, including circulating adipokine levels. Collectively, this Chapter concluded the work of this thesis, showcasing the value of our computational toolkit and integrative approach for dissecting the epigenetics of immunometabolic health (**Fig. 1**).

From association to interpretation

Common CpGs

In this thesis, we conducted EWAS meta-analyses of IL-6, adiponectin, and leptin, and evaluated DNAm responses to a lifestyle intervention in three tissues. Across these investigations, 44 CpGs were associated with more than one outcome, thereby representing loci whose biological relevance was supported by converging lines of evidence (**Fig. 2a**). The majority of these (n = 26) were shared between the IL-6 and leptin EWAS, with a further five overlapping between each of (i) IL-6 and adiponectin, (ii) the adipokines, and (iii) all three blood-based EWAS of *Chapters 2 and 3*.

Notably, all shared CpGs exhibited concordant effects for IL-6 and leptin, whereas their adiponectin associations were in the opposite direction. This pattern was biologically coherent: IL-6 and leptin are both pro-inflammatory and metabolically unfavourable signals, often increasing with adiposity, while adiponectin is anti-inflammatory and typically diminished in immunometabolic disease. The persistence of this directional signature across multiple loci pointed to common mechanisms connecting these biochemical markers and highlighted these sites as potential nodes of regulatory control.

Three loci emerged as particularly compelling exemplars within the context of established literature: cg06500161, cg17901584, and cg00574958. Each of these CpGs had been robustly reported in over 25 previous EWAS, spanning a spectrum of immunometabolic phenotypes including BMI[19–22], lipid levels, and T2D[23]. This thesis linked all three sites to adipokine levels, while also associating cg17901584 with IL-6 (*Chapters 2 and 3*).

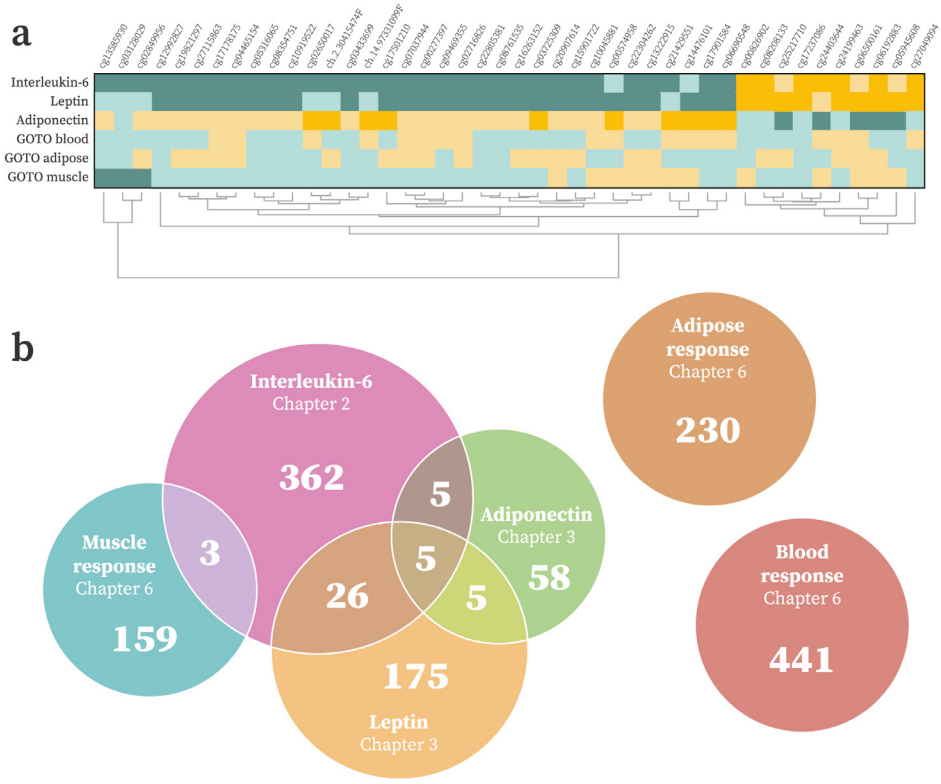


Figure 2 | CpG-level findings across this thesis. a) Heatmap of 44 CpGs identified in multiple investigations, clustered and coloured by effect size (positive: yellow, negative: blue). CpGs not reaching significance are shown in muted tones. **b)** Venn diagram illustrating number and overlap of CpGs found in each Chapter.

Functional annotation connected DNAm at these CpGs to genes critical for lipid transport (*ABCG1*)[24], biosynthesis (*DHCR24*)[25], and metabolism (*CPT1A*)[26], reinforcing methylation at lipid-related loci as a robust molecular readout of immunometabolic health.

The overlap between investigations in this thesis varied markedly (**Fig. 2b**). The pronounced tissue-specificity of DNAm responses to GOTO underscored how critical cellular context is to epigenetic patterns indicating that, in this setting, changes in blood do not mirror adaptations in the other tissues. Nevertheless, DNAm patterns did align with expression findings from this study, with the adipose and muscle methylome responding substantially, alongside smaller changes in blood[27]. Furthermore, the limited concordance between cross-sectional and experimental Chapters suggested that certain CpGs may capture long-term physiology, potentially shaped by cumulative lifetime exposure, whereas others reflect dynamically modifiable loci. This thesis, therefore, contributed to disentangling stable from dynamic sites, a necessary distinction for understanding heterogeneity in human health and intervention responsiveness.

Shared regulatory architecture

While DNAm was traditionally regarded as a repressive mark, particularly when located in promoters or enhancers, its regulatory consequences are now recognized as strongly context dependent. To characterise the chromatin landscape surrounding identified CpGs, we leveraged cell-type and tissue-specific reference epigenomes from the Roadmap Epigenomics Consortium[28]. These integrated epigenetic features to define eight active and seven repressive states, thereby positioning CpGs within their broader genomic environment and evaluating their functional potential (**Fig. 3a**).

In the adipocytokine EWAS (*Chapters 2 and 3*), as well as the muscle response to the GOTO intervention (*Chapter 6*), identified CpGs resided disproportionately within enhancers, consistent with a potential to disrupt TF binding and modulate nearby gene expression in relevant cell types. By contrast, the other tissues examined in the GOTO intervention exhibited alternate patterns: blood-derived CpGs showed no substantial deviation from the background distribution, suggesting that DNAm responses in this tissue reflect diffuse or systemic effects rather than focal transcriptional regulation. CpGs underlying the adipose tissue response were enriched in polycomb-repressed regions, a chromatin state linked to developmental gene silencing that accumulates DNAm as polycomb repression wanes with age[29]. This pattern raised intriguing questions about whether intervention-induced weight loss may have reinforced mature adipocyte identity through silencing of developmental transcriptional programs.

To gain complementary insights into the regulatory context of our DNAm associations, we also performed TF binding site (TFBS) enrichment analyses. Across Chapters, we identified twelve TFs whose binding sites were enriched within 50bp of CpG sets from multiple investigations (**Fig. 3b**). Notably, eleven of these belonged to the basic leucine

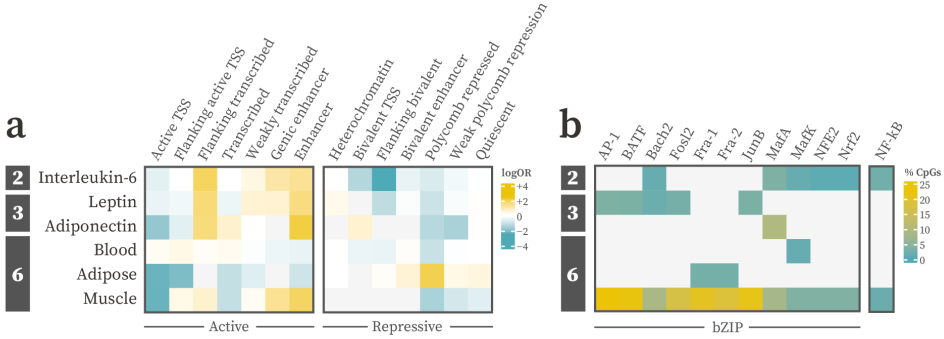


Figure 3 | Regulatory context of results across this thesis. a) Heatmap of chromatin state enrichment, separated into active and repressive states. Colours represent the log-transformed odds ratios (ORs). **b)** TFs whose binding sites were enriched within 50bp of identified CpGs in each Chapter. Colours indicate the percentage of CpGs underlying the enrichment, with all shown significant at the 5% level after adjusting for multiple testing.

zipper (bZIP) family, a group of dimeric TFs that recognise short palindromic motifs through a basic DNA-contact region (*b*) coupled to a leucine zipper dimerization domain (*ZIP*). Members of the bZIP family are established as regulators of cellular stress responses and metabolic flexibility[30], and their recurrent enrichment across this thesis pointed to a potentially conserved regulatory architecture underpinning investigated immunometabolic traits.

Previous work has demonstrated that most bZIP TFs exhibit reduced or abolished binding at methylated motifs, including several identified in our analysis including MafA, JunB, and Bach2[31]. Such sensitivity may have particular relevance at enhancers, where decreased DNAm could thereby facilitate TF binding and modulate target gene expression[32]. Across this thesis, traits associated with CpGs enriched for enhancers and bZIP TFBS were also predominantly negatively correlated with DNAm (70.3% of IL-6 CpGs in *Chapter 2*; 64.4% of adiponectin CpGs and 67.8% of leptin CpGs in *Chapter 3*; and 87.7% of CpGs that responded in muscle in *Chapter 6*). Our results therefore supported a model where elevated adipocytokine levels and intervention effects in muscle were accompanied by reduced DNAm at enhancers, increasing bZIP TF binding propensity and expression of stress-related and metabolic target genes.

Convergence on immunometabolism

Annotating CpGs to their target genes remains a major challenge in epigenetic research, reflecting both the context-dependent nature of DNAm and the crude nature of nearest gene pairings[33,34]. A more evidence-based strategy is expression quantitative trait methylation (eQTM) analysis, which directly correlates expression and methylation in matched samples. In the adipocytokine EWAS, we calculated eQTMs using the BIOS consortium, a large-scale resource comprising DNAm and RNA-seq profiles from over 3,000 individuals. This approach revealed several biologically plausible examples where CpGs were linked to their non-nearest genes, including the IL-6 receptor (*IL6R; Chapter 2*), and key lipid homeostatic genes such as *ABCG1* and *CPT1A* (*Chapter 3*). We adapted this framework to the tissue-specific experimental setting in *Chapter 6*, conducting paired analyses to assess whether genes responsive to the GOTO intervention in each tissue were associated with local DNAm adaptations, directly linking epigenetic changes to observed transcriptional responses.

To further refine CpG–gene links, we incorporated genetic association data in colocalization analyses, combining methylation quantitative trait loci (QTLs) from the Genetics of DNA Methylation Consortium (GoDMC)[35] with expression QTLs from eQTLGen (*Chapter 2*)[36]. This specific approach evaluated whether two traits shared causal variants, revealing common underlying regulatory architecture. This was particularly valuable given that DNAm can reflect transcriptional potential even for genes with low or absent expression at the time of sampling. Collectively, integration of large-scale population

data, dynamic tissue-specific eQTMs, and colocalisation analyses provided a more comprehensive and functionally grounded map of CpG-gene relationships across this thesis, and these strategies could serve as a starting point for data-driven annotation initiatives moving forward.

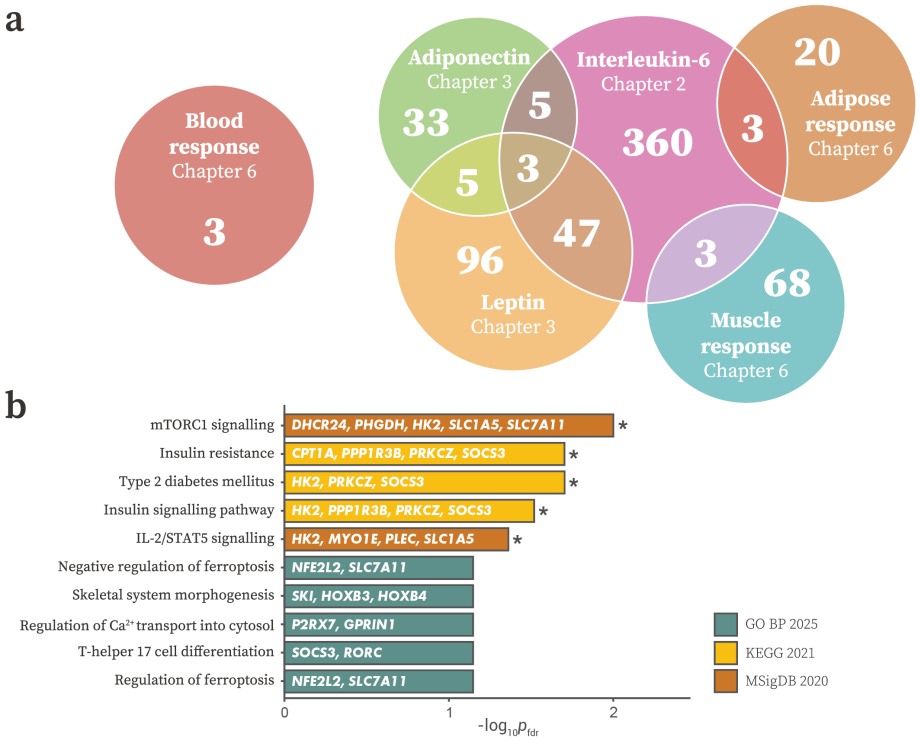


Figure 4 | Gene-based results across this thesis. a) Venn diagram showing number and overlap of genes identified in each Chapter. **b)** Top ten terms from an over-representation analysis of the 66 genes implicated in more than one investigation. Stars indicate significance at the 5% level after adjusting for multiple testing.

To explore common biological pathways implicated across Chapters, we performed an over-representation analysis of the 66 genes detected in multiple investigations (**Fig. 4a,b**). Enriched terms included insulin resistance (p_{FDR} = 0.021) and signalling (p_{FDR} = 0.032), mTORC1 signalling (p_{FDR} = 0.011), and T2D (p_{FDR} = 0.021). Insulin resistance represents a critical driver of immunometabolic dysfunction, contributing to the development of T2D, while mTORC1 functions as a nutrient-sensing hub with strong links to longevity and whose inhibition is emerging as a multimodal therapeutic[37–39]. Notably, many of the overlapping genes mapped to canonical mTORC1-driven effects, including translational upregulation, lipid accumulation, oxidative stress, and inflammation (**Supplementary Figures 1-4**). The convergence of multiple analyses on these pathways therefore highlighted immunometabolic regulation as a biological theme unifying this thesis.

Chapter-specific findings corroborated this pattern, with systematic interrogation of gene set databases and previous EWAS revealing enrichments for immunometabolic pathways and risk factors, including BMI, high-density lipoprotein (HDL) cholesterol, and fasting insulin, as well as T2D itself (**Fig. 5a,b**). Adiponectin-associated CpGs in particular displayed consistent and pronounced enrichment patterns, reinforcing findings that these sites may hold translational relevance for diabetes pathogenesis (*Chapter 3*). These coordinated relationships thereby consolidated the broader conclusion that epigenetic signatures across this thesis converged onto core immunometabolic phenotypes.

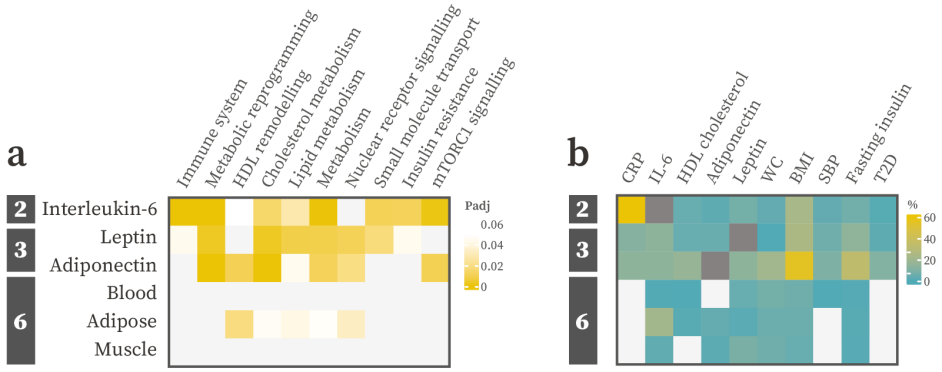


Figure 5 | Higher-order enrichments across this thesis. a) Heatmap showing pathway-level enrichments, coloured by FDR-adjusted *p*-value. **b)** Heatmap of trait-level enrichments (*Chapter 2 and 3*) and longitudinal associations (*Chapter 6*) across immunometabolic phenotypes. Colour represents the percentage of identified CpGs additionally associated with the labelled trait. Dark grey indicates index traits. All enrichments are significant at the 5% level after adjusting for multiple testing.

Considering causality

By altering chromatin accessibility and DNA interactions with regulatory proteins, CpG methylation can modulate local gene expression[40]. Yet, this regulatory *potential* by no means guarantees that DNAm-trait associations identified in EWAS arise from direct, functional effects within studied cells. A central challenge in epigenetic epidemiology is distinguishing DNAm that actively influences the investigated trait (**Fig. 6a,b**) from changes driven by reverse causation (**Fig. 6c**)[5,41,42]. While study design can assist by delineating parts of the causal pathway, as showcased by the GOTO intervention which preceded DNAm by design (*Chapter 6*), proposing plausible chains of events in EWAS requires integration of causal inference approaches with functional genomics and targeted follow-up.

A central component of this strategy was 2SMR, an instrumental variable framework that exploits the fact that genotypes are fixed at conception and almost entirely unaffected by exposures or phenotypic variation[43]. For each identified CpG, we leveraged

independent single nucleotide polymorphisms (SNPs) strongly associated with DNAm as genetic proxies. By combining these with large-scale genome-wide association study (GWAS) data, we interrogated not only if DNAm was likely driving the relevant adipocytokine, but also whether it might influence additional health-related phenotypes. These analyses identified one CpG that plausibly modulated IL-6 levels, and two loci that may alter adiponectin production, with evidence supporting all three sites as consequential for immunometabolic disease (*Chapters 2 and 3*). This thesis therefore highlighted both the value of embedding causal inference within EWAS, and the benefits of extending such analyses to encompass a range of clinically relevant outcomes.

This work was further strengthened by the emergence of QTL databases cataloguing associations between SNPs across the genome and molecular traits, including DNAm (GoDMC) and expression (eQTLGen)[35,36]. These resources provided unprecedented access to genetic association information, and we exploited both *cis*- and *trans*- methylation-QTL effect estimates to apply 2SMR bidirectionally. By comparatively inferring which direction of effect was most strongly supported by the data, we improved confidence in our 2SMR-based conclusions (*Chapters 2 and 3*).

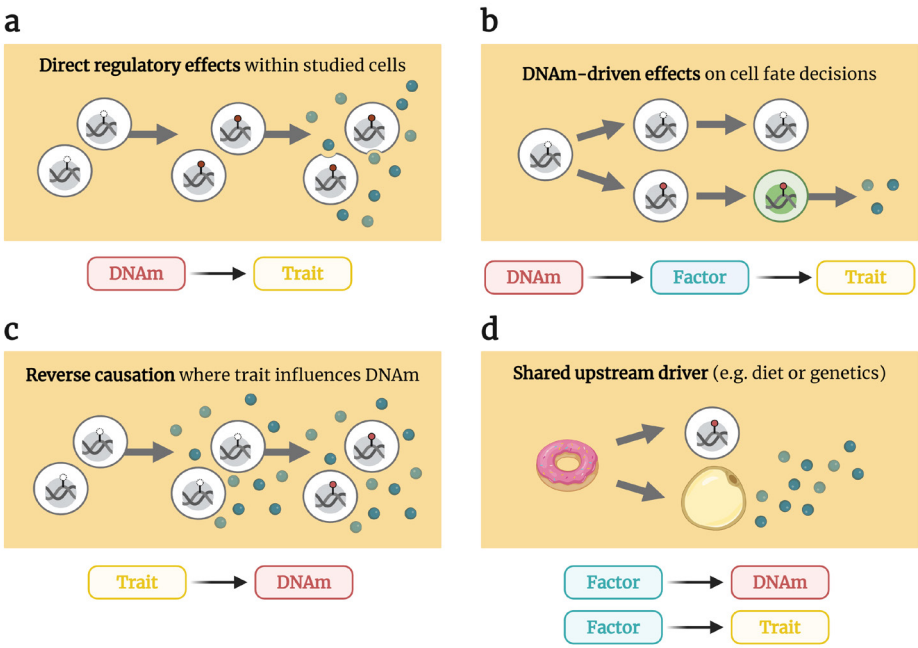


Figure 6 | Four potential causal structures underlying DNAm-trait associations in EWAS. a) Direct regulatory effects within studied cells alters expression and thereby influences the investigated trait. **b)** DNAm directed cell fate decisions lead to changes in sample composition that subsequently alter the trait. **c)** The trait influences DNAm in studied cells. **d)** Shared upstream factors like diet cause independent effects on DNAm in studied cells and the trait, including through DNAm in a non-studied cell type. *Figure created in BioRender.*

To complement 2SMR, this thesis also employed triangulation; a framework that adopts a more holistic view of causality. This approach posits that if an exposure is generally causal for an outcome, then genetically determined outcome levels should correlate with those predicted from genetically driven exposure levels and outcome-exposure associations. In this manner, a direction of effect can be inferred even if statistical power is limited for individual CpGs[21,44]. Bidirectional triangulation served to cross-validate our 2SMR findings. In the IL-6 EWAS, the methods diverged with triangulation concluding that DNAm most often reflected circulating IL-6 levels rather than driving them. As such, the CpG identified as influencing IL-6 production by 2SMR represented an exception to this general rule (*Chapter 2*). By contrast, triangulation corroborated 2SMR findings for adiponectin, indicating that DNAm typically drives adiponectin production and thus strengthening evidence for the two plausibly causal loci implicated by 2SMR (*Chapter 3*).

Since triangulation analyses indicated that DNAm was more often a consequence rather than a cause of circulating IL-6 levels, we employed mediation analysis to position DNAm between IL-6 and its downstream health effects, assessing whether methylation might act as a mediating factor. For two CpGs, both linked to *SOCS3* expression, the evidence supported this sequence of events. This suggested that these sites are more than markers of health status, and play a functional role in mediating IL-6–driven disease risk (*Chapter 2*). As GWAS and QTL resources continue to expand, the power of approaches such as bidirectional 2SMR, triangulation, and mediation analysis will increase, enabling more precise disentanglement of the causal relationships underlying complex molecular pathways.

To further strengthen our directional conclusions, we incorporated functional genomics throughout this thesis. This overlaid identified CpGs onto chromatin state and TFBS annotations and assessed enrichment for regulatory marks. In this manner, evidence for a genuine functional role was strengthened when CpGs implicated as causal fell within active chromatin close to relevant TFBS in studied cells (**Fig. 6a,b**). Sites in repressive regions were deemed more likely to reflect reverse causation or confounding (**Fig. 6c,d**). For example, 2SMR identified two CpGs plausibly influencing adiponectin: one close to *ADIPOQ* and the other associated with *SREBF1*. Functional genomic annotation revealed that the *SREBF1* site resided in a regulatory region within circulating cells, supporting a direct functional role in blood. In contrast, the *ADIPOQ* CpG lay in a repressive chromatin state in leukocytes, making such an effect less likely. In both cases, functional genomics provided essential information for interpreting causality and guiding follow-up analyses, highlighting the importance of considering the context-specific nature of DNAm in causal inference (*Chapter 3*).

Leveraging the reference epigenomes generated by the Roadmap Epigenomics Consortium[28], we examined the *ADIPOQ* site in greater detail. Since adiponectin is almost

exclusively produced by adipocytes, a direct effect of DNAm on its production in immune cells was biologically implausible; a view corroborated by this CpG's location in a repressive chromatin region in circulating cells. In adipocytes, however, this same locus was situated within an enhancer. This observation led us to hypothesize that a shared upstream driver, such as diet, could induce correlated epigenetic responses in a tissue-agnostic manner, generating DNAm-trait associations in blood that mirror functional effects in adipose tissues (**Fig. 6d**). To evaluate this hypothesis, we performed a targeted follow-up in adipocyte data, uncovering connections between *ADIPOQ* expression and methylation in these cells[45]. This finding not only strengthened support for our hypothesis at this CpG but also illustrated the utility of large-scale blood-based data to detect meaningful loci even for traits non-hematopoietic in origin (*Chapter 3*).

Countering confounding

As epidemiological investigations, EWAS are bound by the traditional rules of observational study design. As such, they are vulnerable to multiple forms of confounding (**Fig. 7a,b**)[46]. Variations in sample handling, array position, or environmental conditions during processing can be associated with both the trait of interest and DNAm, even when standard randomisation protocols are implemented[47]. Such technical artifacts do not reflect underlying biology in the organism or test system sampled and, if not properly addressed, can obscure signals or generate misleading results[48,49].

Although addressing technical confounding proactively is an established improvement over *post hoc* approaches, accessible tools for sample layout optimisation were limited[48,50]. To fill this gap, this thesis introduced *Omixer*, a Bioconductor package developed to minimise associations between biological variables and technical factors at the experimental design stage (*Chapter 4*). In the GOTO intervention study, *Omixer* ensured that all investigated correlations were both non-significant and had absolute values below 0.1 (**Fig. 7c**). *Omixer* also handled the paired design of this study, in which the intervention response was our primary focus, by positioning pre- and post-intervention samples from the same tissue and individual adjacently on the EPIC array (**Fig. 7d**). This arrangement largely eliminated concerns about technical confounding for intervention effects (*Chapter 6*). By developing user-friendly software to proactively control confounding and by demonstrating its practical application, this thesis showcased a strategy for routinely implementing best-practice recommendations to manage technical effects in high-throughput genomic studies moving forward.

Beyond technical factors, EWAS are also sensitive to biological confounding, where differences in cellular composition across samples are associated with both methylation and the trait of interest[46,51]. Blood, the most accessible and widely studied tissue in epigenetic epidemiology, offers exceptional opportunities for powerful large-scale analyses, but its inherent heterogeneity complicates interpretation of results.

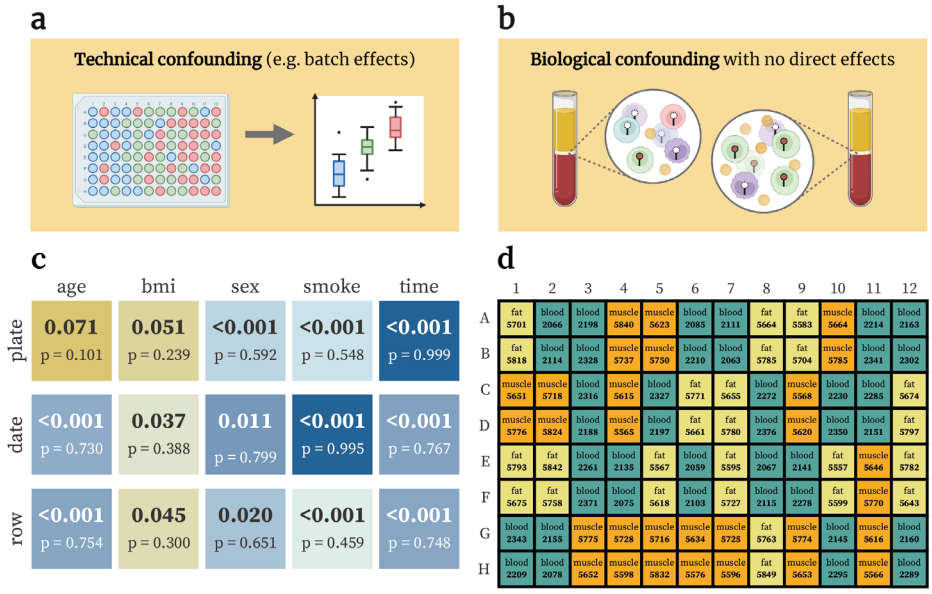


Figure 7 | Sources of and solutions to confounding in EWAS. a) Associations between the trait and technical factors induce systematic DNAm differences. **b)** Associations between sample composition and trait levels (represented by yellow circles) leads to apparent DNAm effects at cell-type specific CpGs. **c)** Omixer-output associations between biological variables and technical factors for DNAm samples profiled before and after the GOTO intervention in *Chapter 6*. **d)** Example of a lab-friendly plate layout generated by *Omixer* showing adjacent sample pairs along columns which were transferred to the eight vertical wells of the EPIC array. *Figure partly made in BioRender.*

Recent advances in algorithmic deconvolution, including within the IDOL-ext and UniLIFE frameworks, allow estimation of a wide range of immune cell subtypes from DNAm data[52,53]. This thesis leveraged these modern prediction tools to perform rigorous sensitivity analyses excluding CpGs whose effects were no longer significant after adjustment for estimated cell-type proportions, thereby ensuring focus remained on sites independently associated with the relevant adipocytokine (*Chapters 2 and 3*). While deconvolution methods continue to improve, they do not yet reach the resolution of purified cell assays. Future advances in single-cell DNAm profiling may provide the granularity needed to fully resolve the cellular sources of identified signals in EWAS.

In the GOTO intervention study, we had access to both these DNAm-derived cell-type proportion estimates, and cell counts measured by routine clinical testing. Using these, we were able to conclude, at multiple resolutions, that the immune cell proportions of participants was stable. Furthermore, in this multi-tissue setting, we extended deconvolution by integrating bulk RNA-sequencing profiles with single-cell and single-nuclei reference transcriptomes[54,55]. This enabled prediction of a distinct set of immune cell subtypes and allowed estimation of cell-type proportions in adipose and muscle tissues. Here, we detected an increase in endothelial cells in muscle post-intervention, consistent

with angiogenic adaptations to physical activity (*Chapter 6*)[56]. Taken together, these examples demonstrated the value of deploying deconvolution strategies in EWAS and illustrated how genomic data can derive meaningful readouts in complex and heterogenous tissues.

Leveraging genomic resources

The impact of integrative DNAm analyses is greatly amplified when paired with large-scale resources, collaborative initiatives, and rich genomic information. In this thesis, we combined in-depth functional annotations (**Fig. 8**) with a broad portfolio of available and novel data (**Table 2**) to strengthen our results. This included leveraging large-scale population cohorts and consortia for eQTM and meta-analyses (BBMRI-BIOS, TwinsUK, SHIP-TREND, and KORA F4; *Chapters 2 and 3*). By aggregating data across populations in this manner, we pooled statistical power and revealed modest but consistent DNAm associations that may have remained undetected in individual cohorts.

The Gene Expression Omnibus (GEO) proved particularly valuable as an accessible data repository. From GEO, we sourced not only the adipocyte and immune cell follow-up data (*Chapters 2 and 3*) but also single-cell and single-nuclei reference transcriptomes (*Chapter 6*)[57,58], and EPIC array data to test the external validity and applicability of *DNAmArray* (*Chapter 5*)[59]. This open-source workflow also incorporated publicly available probe information, excluding probes that were cross-reactive or contained common SNPs (from the Zhou-lab GitHub[60]), and flagging unreliable genomic regions (ENCODE blacklist[61]). This ensured that generated data was of high-quality with reduced risk of unreliable signals. Collectively, the combined utility of these resources showcased the value of accessible repositories such as GEO and Bioconductor and highlighted the importance of data and software democratisation in science.

A detailed understanding of positional context is essential for disentangling molecular architecture. Across this work, genomic locations and ranges were combined with cell-type reference epigenomes (from the Roadmap Epigenomics Project[28]) and TFBS motifs (accessed via HOMER[62]) to situate DNAm findings within regulatory landscapes. To explore causal directions, we extended this further, integrating SNP associations with DNAm (from GoDMC[35]), expression (from eQTLgen[36]), phenotypes (from GWAS catalog and IEU GWAS databases), and each other (via LDlink[62]), in instrumental variable based causal inference approaches. Overall, by systematically layering genetic, epigenetic, transcriptomic, and phenotypic resources in complementary ways, we strengthened the biological interpretation of our findings and ensured their relevance beyond the immediate scope of this thesis.

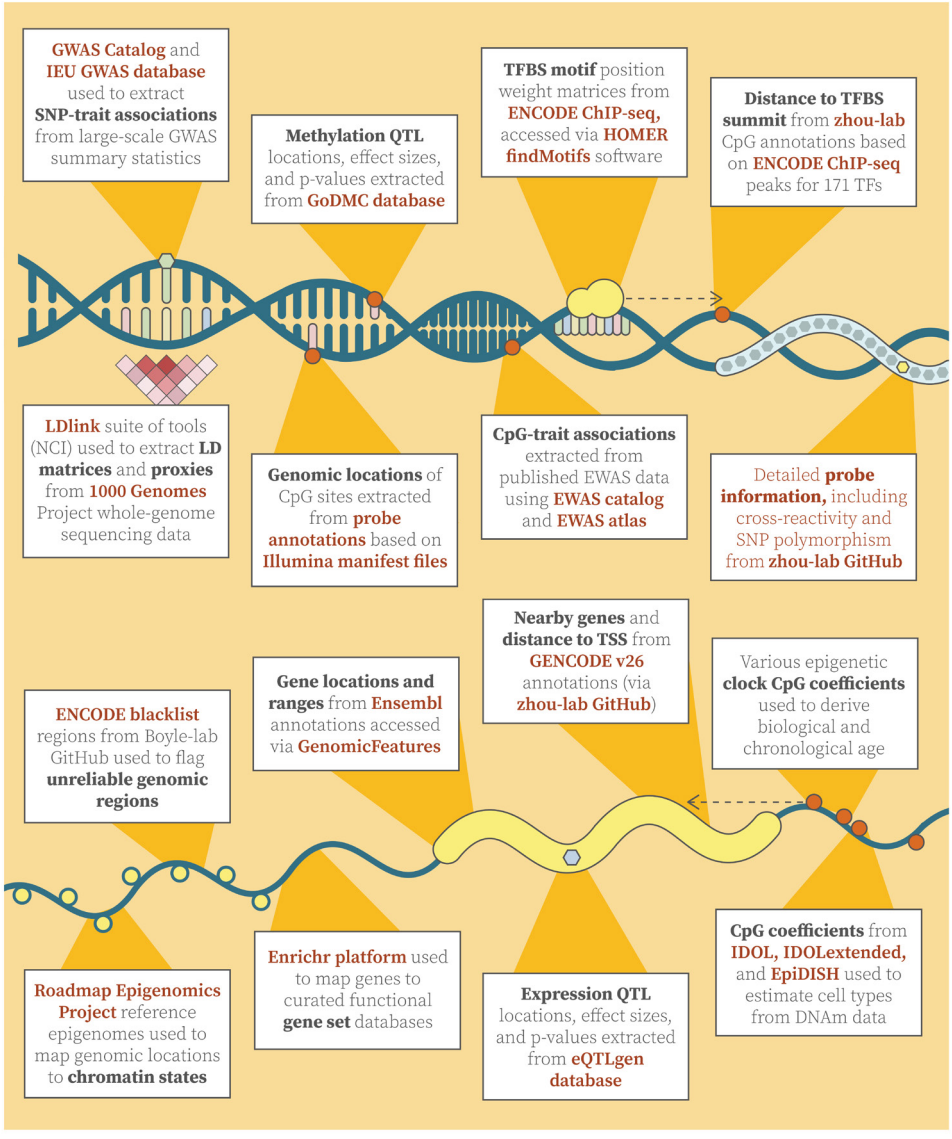


Figure 8 | Annotation information used throughout this thesis describing SNP-associations (LDlink, GWAS Catalogues, GoDMC, and eQTLgen), CpG associations (EWAS catalogues, CpG coefficients for clocks and cell counts), genomic co-ordinates (TFBS motifs, CpGs, genes, ENCODE blacklist, and chromatin states) and distances (CpG to TFBS summit, CpG to gene TSS), as well as detailed probe and gene set information.

Data	DNAm	Expression	Phenotypes	Chapters
BBMRI-BIOS	450k array blood	RNA-seq blood	age, sex, smoking, hsCRP, BMI, IL-6, adipokines	2, 3
KORA F4	450k array blood	-	age, sex, smoking, hsCRP, BMI, IL-6, adipokines	2, 3
SHIP-TREND	EPIC array blood	-	age, sex, smoking, BMI, adipokines	3
TwinsUK	450k array blood	-	age, sex, smoking, BMI, adipokines	3
GSE119593 GSE119539	450k array adipocytes	microarray adipocytes	-	3
GSE116339	EPIC array blood	-	age, sex, exposure status	5
Growing Old TOgether data	EPIC array 3 tissues	RNA-seq 3 tissues	age, sex, smoking, multiple immunometabolic traits	6
GSE149938	-	scRNA-seq blood	-	6
GSE167186	-	scRNA-seq muscle	-	6

Table 2 | DNAm and expression datasets used across this thesis. (WC: waist circumference, SBP: systolic blood pressure)

Drawing on our experience with novel EPIC data and large-scale population cohort analyses, we developed a modular and transparent pipeline providing current guidelines for processing and analysing DNAm data from 450k, EPIC, and the latest EPICv2 arrays. *DNAmArray* integrated several Bioconductor packages from our research group, including *MethylAid* for interactive quality control[16], *omicsPrint* for sample identity verification[17], and *bacon* for bias and inflation correction[18]. This workflow was designed to be flexible and scalable, enabling researchers to tailor analyses to match their specific study designs and aims (*Chapter 5*). By combining best-practice computational tools with clear, accessible documentation, *DNAmArray* facilitated data processing and analysis for the GOTO intervention study (*Chapter 6*), and now serves to exemplify of how well-curated, open-source pipelines can keep pace with evolving platforms and analytical challenges.

7

Healthy ageing

Investigating the epigenetic component of age-related traits was a central aim of this thesis. While chronological age remains the single strongest risk factor for most common chronic diseases, individuals exhibit marked heterogeneity in ageing trajectories[64,65]. DNAm has emerged as a critical molecular layer that both reflects and potentially influences this inter-individual variability. Certain age-associated CpGs undergo predictable remodelling across the lifespan, with a subset of these forming the basis of epigenetic clocks[66,67]. These DNAm-based models have been refined not only to capture calendar

age but also biological age (bAge), a measure more intricately linked to health and physiological function. In the GOTO intervention, we applied four such bAge predictors with all registering intervention-driven reductions. This demonstrated that even modest, short-term lifestyle changes can recalibrate molecular aging processes across tissues. Crucially, these bAge shifts correlated with improvements in clinical health parameters, including BMI, total body fat percentage, and leptin levels. Yet, despite compelling findings, challenges remain to uncover the specific biological processes captured by these algorithms, with emerging evidence suggesting that they may be driven at least in part by leukocyte population remodelling[68,69].

The application of muscle and blood-based epigenetic clocks exemplified another core tenet of this work: that both the progression and modulation of health can be tracked across multiple tissues. In this thesis, we characterised blood-based signals (*Chapters 2, 3, and 6*); explored targeted adipocyte-focused follow-up and adipose tissue signatures (*Chapters 3 and 6*); and analysed skeletal muscle, including by evaluating of the muscle-specific MEAT deconvolution algorithm (*Chapter 6*). Such a multi-tissue perspective is critical to build a comprehensive understanding of ageing, as physiological decline manifests not only in the circulation but also within less accessible tissues, where it can impair metabolic function and homeostasis. Furthermore, both adipocytes and myocytes are far from passive targets of signalling molecules; they actively secrete pro- and anti-inflammatory mediators shaping systemic immunity and organismal energy balance[70,71]. By presenting findings across distinct tissues sitting at the crossroads of immune surveillance and metabolic regulation, we emphasised the physiological heterogeneity of age-related phenotypes, advocating for continued, tissue-targeted strategies to uncover the molecular mechanisms that limit or lead to age-related disease.

By interrogating both inter-individual differences via large-scale EWAS meta-analyses (*Chapters 2 and 3*), and intra-individual dynamics through a multi-tissue intervention study (*Chapter 6*), this thesis identified molecular features that reflected cross-sectional physiological states distinct from those responsive to short-term lifestyle changes. Taken together, these findings established both a conceptual framework and a practical toolkit for distinguishing and dissecting an individual's intrinsic predisposition to a healthier later life from the epigenetic plasticity required to improve it through behavioural change.

Future perspectives

As the great regulatory filter that governs when, where, and how our DNA is expressed, epigenetic mechanisms promise to function as a molecular babel fish, translating the dialogue between our environment, our genes, and ultimately our physiology[38]. This strategic position at the genomic interface endows epigenetic research with immense potential to illuminate, and perhaps reshape, human health trajectories. Yet, the epigenome is inherently complex and heterogenous, making clear signals difficult. Much progress to date, including in this thesis, has relied on large-scale consortia pooling data, resources, expertise, and crucially, statistical power. Only within such collaborative frameworks can we detect the modest but consistent signals required to chart the DNAm landscape across the lifecourse and in disease. The future of epigenetic epidemiology therefore hinges on preserving and nurturing these partnerships, recognising that the task of understanding human health is a collective scientific pursuit.

Numbers alone will be insufficient. The field must also prioritise longitudinal initiatives. Despite originally being hailed as a stable archive of lifelong exposure and long-term physiological condition, this thesis and other recent research has uncovered dynamic subsections of the methylome responsive to even mild and short-term perturbations. This duality between stability and plasticity raises fundamental questions about the biological character of DNAm: is it a sturdy barometer of past exposures, a sensitive readout of environmental influence, or both? Integration of observations from longitudinal, experimental, and cross-sectional findings will be needed to disentangle the temporal variability of different sets of CpGs.

Transitioning from statistical associations to mechanistic insight will require more precise mapping of CpGs to plausible target genes. Simple proximity-based mapping is inadequate for a regulatory system as context-dependent and tissue-specific as DNAm. The field requires a rigorous, evidence-based framework that assigns graded confidence scores to CpG-gene links proportional to the weight of supporting data. Core evidence pillars could encompass longitudinal, tissue-specific, large-scale, and colocalisation analyses. Taken together, the result would provide a regulatory atlas, capable of connecting DNAm signals to their most likely functional targets and enhancing the biological interpretation of EWAS results past, present, and future.

By annotating CpGs more precisely, we will also be better equipped to tackle one of the field's most enduring challenges: causality. Instrumental variable-based methods, such as 2SMR and triangulation, can suggest directionality, but their genetic basis makes them vulnerable to tissue pleiotropy. Embedding these approaches within functional genomics, mediation frameworks, and tissue-specific analyses offers more refined dissection of underlying causal structures. Such integrative strategies can help clarify whether loci identified in blood represent direct epigenetic regulation inside immune cells, or instead reflect pleiotropic effects in other, less accessible tissues.

Concrete causal inference will ultimately require functional experiments and technological innovation, however. *In vitro* affinity-based assays have been invaluable for uncovering how TFs differentially recognise methylated motifs, but they have clear limitations: by operating outside the context of the chromatin landscape, they fail to capture the dependencies that govern TF binding *in vivo*[72]. In contrast, emerging CRISPR-based epigenetic editing platforms achieve targeted, single-CpG modifications within the native genomic environment[73]. These versatile systems support combinatorial perturbation of multiple loci, and their effects can be profiled by other molecular readouts[74]. Although the application of such cutting-edge tools remains restricted to cellular models, they provide a cornerstone for causality. Together with epigenetic epidemiology of human cohorts, they can reveal not just that DNAm matters, but precisely how and why.

With momentum behind epigenetic epidemiology, the field must also orient itself towards clinical translation. This thesis indicated the capacity of epigenetic clocks to track bAge reductions following lifestyle modifications, underscoring their potential to serve as dynamic biomarkers for monitoring intervention efficacy. Expanding these algorithms across multiple tissues could reveal organ-specific ageing processes that contribute to systemic decline, offering a more granular view of physiology. Moreover, integrating DNAm with genetic data to derive polygenic epigenetic risk scores could enhance patient stratification by capturing both inherited predisposition and environmental responsiveness. Such composite metrics may detect the subtle subclinical shifts that underlie heterogeneity in immunometabolic health and healthy ageing.

The future of epigenetic research, then, is not a list of incremental improvements but an integrated vision combining complementary strengths: large-scale consortia providing power and diversity, longitudinal studies capturing temporal dynamics, functional genomics revealing mechanistic insight, and experimental validation of causality. When woven together and translated to the clinic, these threads promise to create a comprehensive framework for understanding how environment and genome co-author human health. Within this vision, epigenetic signals can serve as critical focal points for mechanistic inquiry and generate testable hypotheses about how molecular phenotypes interact. The challenge, then, lies in bridging the gap between discovery and delivery, not only refining our science but translating it into better health for our ageing populations.

Conclusions

This General Discussion has synthesised the main findings of this thesis within the broader framework of immunometabolic health and ageing, emphasizing translational relevance, causal inference, and functional interpretation. By tracing results from the nucleotide resolution of CpGs back up to higher-order biology, it underscored the value of integrative approaches in EWAS, shifting focus from static associations to a multidimensional understanding of epigenetics in human health and disease. Cross-Chapter comparisons revealed convergent methylation signatures linked to immunometabolic risk and highlighted clear focal points for targeted experimental validation.

This thesis set out to interrogate relationships between DNAm and three adipocytokines, whilst also identifying CpG sites sensitive to lifestyle changes in older adults. By expanding EWAS to evaluate both intra- and inter-individual variation, and by employing longitudinal, multi-tissue, and large-scale analyses, it illuminated molecular markers of the subtle shifts in health that accompany ageing and disease pathogenesis. Recognising the context-dependent nature of DNAm, results were situated within a broader regulatory landscape encompassing TF binding, chromatin accessibility, and underlying genetic architecture, enabling biologically grounded interpretations.

Beyond foundational insights, this thesis also developed and disseminated computational tools, supporting researchers from design to data analysis. Leveraging this software in conjunction with the wealth of available genomic resources, it demonstrated that when DNAm is processed with care and rigorously contextualised, it can reveal critical biological pathways and prioritise promising loci. In sum, this thesis made meaningful strides towards realising the potential of EWAS to deepen our understanding of immunometabolic health in ageing.

References

1. Jones, M. J., Goodman, S. J. & Kobor, M. S. DNA methylation and healthy human aging. Aging Cell 14, 924 (2015).
2. Walaszczyk, E. et al. DNA methylation markers associated with type 2 diabetes, fasting glucose and HbA1c levels: a systematic review and replication in a case–control sample of the Lifelines study. Diabetologia 61, 354 (2017).
3. Ling, C. & Rönn, T. Epigenetics in Human Obesity and Type 2 Diabetes. Cell Metab 29, 1028 (2019).
4. Szarc Vel Szic, K., Declerck, K., Vidaković, M. & Vanden Berghe, W. From inflammaging to healthy aging by dietary lifestyle choices: Is epigenetics the key to personalized nutrition? Clin Epigenetics 7, (2015).
5. Bock, C. Analysing and interpreting DNA methylation data. Nature Reviews Genetics 2012 13:10 13, 705–719 (2012).
6. Wagner, K. H., Cameron-Smith, D., Wessner, B. & Franzke, B. Biomarkers of Aging: From Function to Molecular Biology. Nutrients 8, 338 (2016).
7. Hirata, T. et al. Associations of cardiovascular biomarkers and plasma albumin with exceptional survival to the highest ages. Nat Commun 11, 1–17 (2020).
8. Mossmann, M. et al. Increased serum IL-6 is predictive of long-term cardiovascular events in high-risk patients submitted to coronary angiography: an observational study. Diabetol Metab Syndr 14, (2022).
9. Yadav, R., Yadav, R. K., Khadgawat, R. & Pandey, R. M. Comparative efficacy of a 12 week yoga-based lifestyle intervention and dietary intervention on adipokines, inflammation, and oxidative stress in adults with metabolic syndrome: A randomized controlled trial. Transl Behav Med 9, 594–604 (2019).
10. Zeng, Y. et al. Novel loci and pathways significantly associated with longevity. Sci Rep 6, (2016).
11. Zhou, L. et al. Centenarians Alleviate Inflammaging by Changing the Ratio and Secretory Phenotypes of T Helper 17 and Regulatory T Cells. Front Pharmacol 13, (2022).
12. Minciullo, P. L. et al. Inflammaging and Anti-Inflammaging: The Role of Cytokines in Extreme Longevity. Arch Immunol Ther Exp (Warsz) 64, 111–126 (2016).
13. van de Rest, O. et al. Metabolic effects of a 13-weeks lifestyle intervention in older adults: The Growing Old Together Study. Aging 8, 111–126 (2016).
14. Shimano, H. & Sato, R. SREBP-regulated lipid metabolism: Convergent physiology-divergent pathophysiology. Nat Rev Endocrinol 13, 710–730 (2017).
15. Seo, J. B. et al. Adipocyte determination- and differentiation-dependent factor 1/sterol regulatory element-binding protein 1c regulates mouse adiponectin expression. Journal of Biological Chemistry 279, 22108–22117 (2004).
16. Van Iterson, M. et al. MethylAid: visual and interactive quality control of large Illumina 450k datasets. Bioinformatics 30, 3435–3437 (2014).
17. Van Iterson, M., Cats, D., Hop, P. & Heijmans, B. T. omicsPrint: detection of data linkage errors in multiple omics studies. Bioinformatics 34, 2142–2143 (2018).
18. van Iterson, M. et al. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol 18, 1–13 (2017).
19. Sun, D. et al. Body Mass Index Drives Changes in DNA Methylation: A Longitudinal Study. Circ Res 125, 824–833 (2019).
20. Mendelson, M. M. et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. PLoS Med 14, (2017).
21. Wahl, S. et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 2016 541:7635 541, 81–86 (2016).
22. Demerath, E. W. et al. Epigenome-wide association study (EWAS) of BMI, BMI change and waist circumference in African American adults identifies multiple replicated loci. Hum Mol Genet 24, 4464–4479 (2015).
23. Hillary, R. F. et al. Blood-based epigenome-wide analyses of 19 common disease states: A longitudinal, population-based linked cohort study of 18,413 Scottish individuals. PLoS Med 20, (2023).
24. Kennedy, M. A. et al. ABCG1 has a critical role in mediating cholesterol efflux to HDL and preventing cellular lipid accumulation. Cell Metab 1, 121–131 (2005).
25. Zerenturk, E. J., Sharpe, L. J., Ikonen, E. & Brown, A. J. Desmosterol and DHCR24: unexpected new directions for a terminal step in cholesterol synthesis. Prog Lipid Res 52, 666–680 (2013).
26. Schlaepfer, I. R. & Joshi, M. CPT1A-mediated Fat Oxidation, Mechanisms, and Therapeutic Potential. Endocrinology 161, (2020).
27. Bogaards, F. A. et al. Secondary integrated analysis of multi-tissue transcriptomic responses to a combined lifestyle intervention in older adults from the GOTO nonrandomized trial. Nature Communications 2024 15:1 15, 1–18 (2024).
28. Roadmap Epigenomics Consortium et al. Integrative analysis of 111 reference human epigenomes. Nature 2015 518:7539 518, 317–330 (2015).

29. Dozmorov, M. G. Polycomb repressive complex 2 epigenomic signature defines age-associated hypermethylation and gene expression changes. Epigenetics 10, 484–495 (2015).
30. Hai, T. et al. ATF3, a Hub of the Cellular Adaptive-Response Network, in the Pathogenesis of Diseases: Is Modulation of Inflammation a Unifying Component?
31. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science (1979) 356, (2017).
32. Llorca, C. M., Potschin, M. & Zentgraf, U. bZIPs and WRKYs: Two large transcription factor families executing two different functional strategies. Front Plant Sci 5, (2014).
33. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how and why? Mol Cell 49, 10.1016/j.molcel.2013.01.038 (2013).
34. Hannon, E. et al. Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. Am J Hum Genet 103, 654–665 (2018).
35. Min, J. L. et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nature Genetics 2021 53:9 53, 1311–1321 (2021).
36. Võsa, U. et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nature Genetics 2021 53:9 53, 1300–1310 (2021).
37. Mannick, J. B. & Lamming, D. W. Targeting the biology of aging with mTOR inhibitors. Nat Aging 3, 642–660 (2023).
38. Johnson, S. C., Rabinovitch, P. S. & Kaeberlein, M. MTOR is a key modulator of ageing and age-related disease. Nature 493, 338–345 (2013).
39. Passtoors, W. M. et al. Gene expression analysis of mTOR pathway: Association with human longevity. Aging Cell 12, 24–31 (2013).
40. Moore, L. D., Le, T. & Fan, G. DNA Methylation and Its Basic Function. Neuropsychopharmacology 38, 23 (2012).
41. Birney, E., Smith, G. D. & Greally, J. M. Epigenome-wide Association Studies and the Interpretation of Disease -Omics. PLoS Genet 12, e1006105 (2016).
42. Mill, J. & Heijmans, B. T. From promises to practical strategies in epigenetic epidemiology. Nat Rev Genet 14, 585–594 (2013).
43. Teschendorff, A. E. & Relton, C. L. Statistical and integrative system-level analysis of DNA methylation data. Nature Reviews Genetics 2017 19:3 19, 129–147 (2017).
44. Wielscher, M. et al. DNA methylation signature of chronic low-grade inflammation and its role in cardio-respiratory diseases. Nature Communications 2022 13:1 13, 1–14 (2022).
45. Tini, G. et al. DNA methylation during human adipogenesis and the impact of fructose. Genes Nutr 15, 21 (2020).
46. Teschendorff, A. E. & Zheng, S. C. Cell-type deconvolution in epigenome-wide association studies: a review and recommendations. Epigenomics 9, 757–768 (2017).
47. Leek, J. T. et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 11, 733–739 (2010).
48. Harper, K. N., Peters, B. A. & Gamble, M. V. Batch effects and pathway analysis: Two potential perils in cancer studies involving DNA methylation array analysis. Cancer Epidemiology Biomarkers and Prevention 22, 1052–1060 (2013).
49. Zindler, T., Frieling, H., Neyazi, A., Bleich, S. & Friedel, E. Simulating ComBat: how batch correction can lead to the systematic introduction of false positive results in DNA methylation microarray studies. doi:10.1186/s12859-020-03559-6.
50. Price, E. M. & Robinson, W. P. Adjusting for batch effects in DNA methylation microarray data, a lesson learned. Front Genet 9, (2018).
51. Loyfer, N. et al. A DNA methylation atlas of normal human cell types. Nature 613, 355–364 (2023).
52. Salas, L. A. et al. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. Nature Communications 2022 13:1 13, 1–13 (2022).
53. Guo, X. et al. Unified high-resolution immune cell fraction estimation in blood tissue from birth to old age. Genome Med 17, 1–26 (2025).
54. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun 10, 1–9 (2019).
55. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 37, 773–782 (2019).
56. Kwak, S.-E., Lee, J.-H., Zhang, D. & Song, W. Angiogenesis: focusing on the effects of exercise in aging and cancer. J Exerc Nutrition Biochem 22, 21–26 (2018).
57. Xie, X. et al. Single-cell transcriptomic landscape of human blood cells. Natl Sci Rev 8, (2021).
58. Perez, K. et al. Single nuclei profiling identifies cell specific markers of skeletal muscle aging, frailty, and senescence. Aging 14, 9393–9422 (2022).
59. Curtis, S. W. et al. Exposure to polybrominated biphenyl (PBB) associates with genome-wide DNA methylation differences in peripheral blood. Epigenetics 14, 52–66 (2019).

60. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res 45, e22 (2017).

61. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Scientific Reports 2019 9:1 9, 1–5 (2019).

62. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589 (2010).

63. Machiela, M. J. & Chanock, S. J. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 31, 3555–3557 (2015).

64. Niccoli, T. & Partridge, L. Ageing as a risk factor for disease. Current Biology 22, (2012).

65. Partridge, L., Deelen, J. & Slagboom, P. E. Facing up to the global challenges of ageing. Nature 2018 561:7721 561, 45–56 (2018).

66. Bernabeu, E. et al. Refining epigenetic prediction of chronological and biological age. Genome Med 15, 1–15 (2023).

67. Horvath, S. DNA methylation age of human tissues and cell types. Genome Biol 14, (2013).

68. Jonkman, T. H., Consortium, B., Zwet, E. W. van & Heijmans, B. T. Probing epigenetic clocks as a rational markers of biological age using blood cell counts. medRxiv 2025.05.12.25327213 (2025).

69. Guo, X. et al. Variations in Innate Immune Cell Subtypes Correlate with Epigenetic Clocks, Inflammaging and Health Outcomes. Advanced Science e05922 (2025).

70. Maury, E. & Brichard, S. M. Adipokine dysregulation, adipose tissue inflammation and metabolic syndrome. Mol Cell Endocrinol 314, 1–16 (2010).

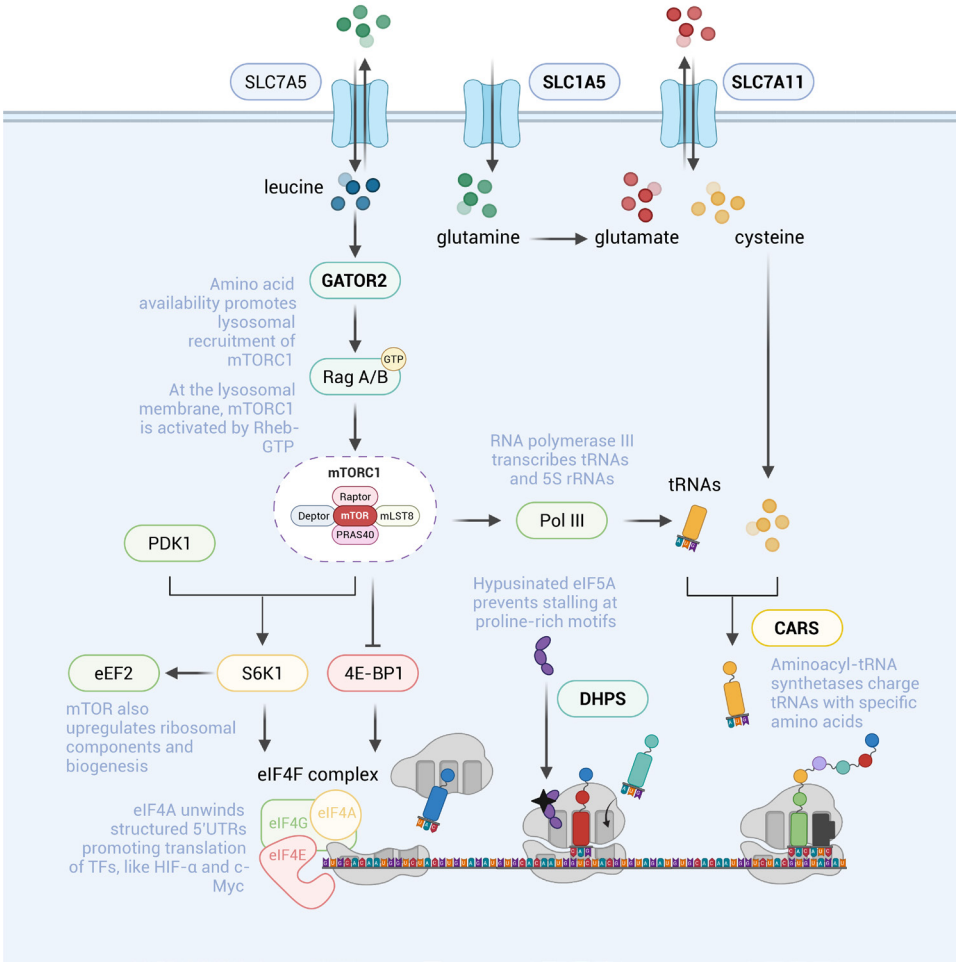
71. Leal, L. G., Lopes, M. A. & Batista, M. L. Physical exercise-induced myokines and muscle-adipose tissue crosstalk: A review of current knowledge and the implications for health and metabolic diseases. Front Physiol 9, (2018).

72. Liesenfelder, S. et al. Epigenetic editing at individual age-associated CpGs affects the genome-wide epigenetic aging landscape. Nature Aging 2025 5:6 5, 997–1009 (2025).

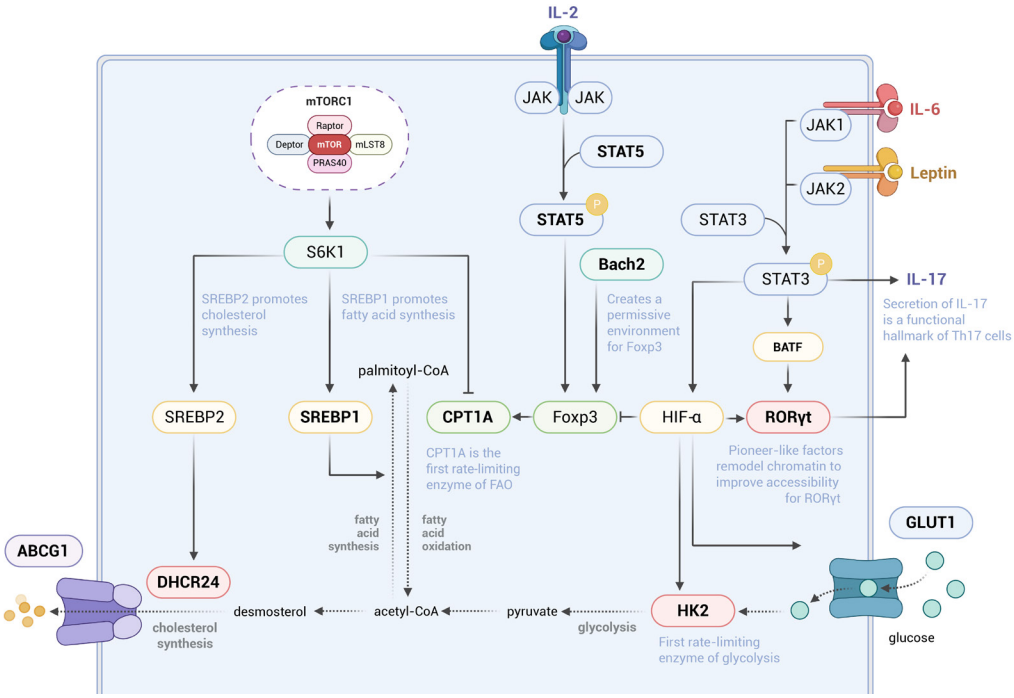
73. Zsido, B. and Hetenyi, C. Molecular structure, binding affinity, and biological activity in the epigenome. Int J Mol Sci. 21(11):4134 (2020).

74. de Groote, M., Vershure P., and Rots, M. Epigenetic editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. Nucleic Acids Res. 40(21):10596-613 (2012).

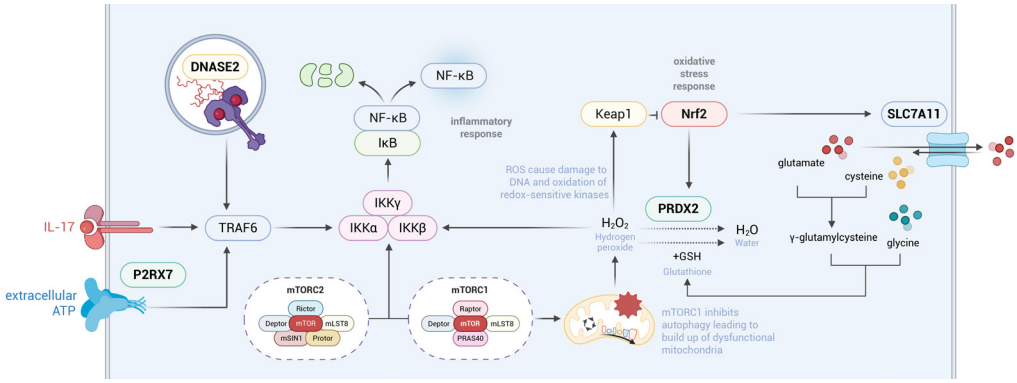
Supplementary Material



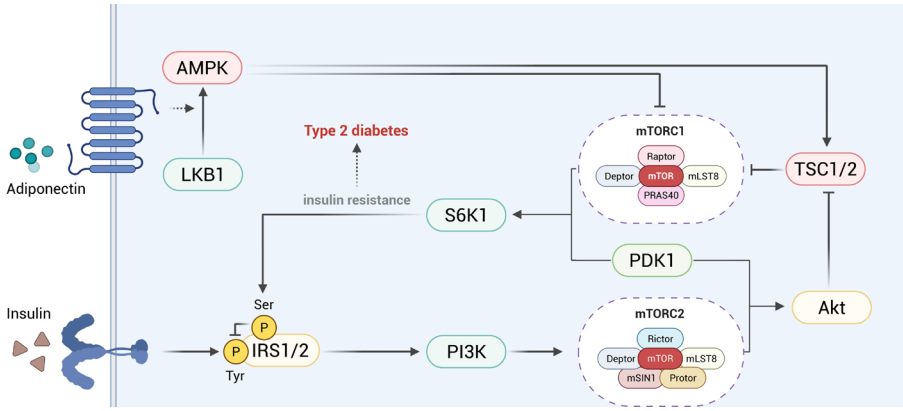
Supplementary Figure 1 | Amino acid supply promotes mTORC1 activation and downstream protein synthesis. Amino acids enter cells via transporters, such as SLC7A5, SLC1A5, and SLC67A11, where they trigger GATOR2 and RagA/B•GTP to position the mTORC1 complex at the lysosomal membrane ready for activation. Activated mTORC1 upregulates translation, fuelled by amino acids ligated to tRNAs by aminoacyl-tRNA synthetases, such as CARS, GARS, and VARS2, ribosomal biogenesis, and factors that prevent elongation stalling, such as hypusinated eIF5A generated via DHPS. By stimulating the eIF4F complex, mTORC1 preferentially upregulates the translation of mRNAs with structured 5'UTRs, including HIF-a and c-Myc. *Genes or proteins implicated in this thesis shown in the Figure in bold. Figure created in BioRender.*



Supplementary Figure 2 | Crosstalk between mTORC1, IL-6, and leptin in T-cell fate regulation. Th17 differentiation is promoted by IL-6 and leptin signalling, and TFs BATF, RORγt, and HIF-a. These upregulate glycolytic machinery, including GLUT1 and HK2 and suppress Treg pathways via inhibition of Foxp3. IL-2/STAT5 signalling and Bach2 promote Treg phenotypes by upregulating Foxp3 and fatty acid oxidation, including via CPT1A. mTORC1 actively undermines this by directly inhibiting CPT1A and further altering lipid handling through fatty acid synthesis (SREBP1) and cholesterol synthesis (SREBP2 and DHCR24). *Genes and proteins implicated in this thesis shown in the Figure in bold. Figure created in BioRender.*



Supplementary Figure 3 | mTOR drives both oxidative stress and inflammation. By inhibiting autophagy, mTORC1 promotes a build-up of dysfunctional mitochondria fuelling reactive oxygen species formation. This is characterised by Nrf2 activation of target genes, including the peroxiredoxin PRDX2 and cysteine/glutamate anti-porter SLC7A11. Oxidative stress drives DNA damage and inflammatory responses, the latter through oxidation of IKKB. NF-kB and inflammatory signalling is also triggered by misplaced self-DNA, whose clearance is impaired by reduced DNASE2, and extracellular ATP, detected by P2RX7. These highlight a role for the "garbaging" model of inflammaging in our results. *Genes and proteins identified in this thesis shown in the Figure in bold. Figure created in BioRender.*



Supplementary Figure 4 | Adiponectin dampens mTOR activity through AMPK activation. AMPK inhibits mTORC1 both through TSC1/2 activation and direct inhibitory phosphorylation of mTORC1 itself. This also has potential to counteract insulin resistance resulting from hyperactive mTORC1, mediated by S6K1 phosphorylation at inhibitory serine residues that blocks tyrosine phosphorylation essential for insulin signal transduction. *Figure created in BioRender.*

CHAPTER EIGHT

Appendix

English summary

Introduction

As we get older, many of us face similar challenges. We struggle to stay energetic and active, take longer to recover from illness, and find it harder to maintain a healthy weight. Beneath these everyday experiences lies a biological transition that alters how our bodies respond to the world around us. Two tightly connected aspects of this process are changes to the *immune* system, which detects and eliminates threats, and to our *metabolism*, which controls how we use and store energy. Together, these age-related shifts are called ***immunometabolic dysfunction***, and they can eventually lead to severe age-related diseases, such as cardiovascular disease or type 2 diabetes.

One way to monitor *immunometabolic* health is by measuring levels of ***messenger molecules*** in the blood. These tiny signals allow different tissues to communicate with each other, sharing updates about hunger, stress, and inflammation. *Insulin*, for instance, is released by the pancreas after a meal to indicate energy availability. Other examples include *leptin*, which helps to control appetite by carrying messages from our fat and stomach to our brain, and *interleukin-6*, a danger signal released during infection or injury (**Box 1**). Breakdown in how these *messenger molecules* are produced, transmitted, or received plays a central role in ageing and disease.

Box 1. Key messenger molecules that help measure and maintain immunometabolic health

Insulin is released by the pancreas after a meal in response to rising sugar levels, signalling that your body should use the available sugar as energy or store it for later. In this way, insulin helps keep blood sugar levels within a healthy range.

Interleukin-6 (IL-6) is involved in immune responses and inflammation. Short-term IL-6 signalling helps co-ordinate defence and repair during infections or injury, but prolonged or dysregulated activity is linked to chronic low level inflammation in ageing and disease.

Leptin is produced primarily by fat tissue and helps regulate long-term energy balance. A small portion is also released by the stomach after a meal. It communicates with the brain to control appetite in line with the body's energy stores.

Adiponectin is also released by fat tissue and is associated with beneficial metabolic effects. It enables tissues to respond more effectively to insulin and limits inflammation, thereby supporting immunometabolic health.

Epigenetic control

To produce these *messenger molecules*, our bodies rely on instructional blueprints called ***genes*** encoded in our DNA. Although all tissues carry the same DNA inherited from our parents, they do not transmit the same signals. Instead, they each perform specific roles in the body. Almost all *leptin* comes from fat, while insulin is created exclusively in the pancreas. This is because *containing* a gene is not enough; a tissue must also *activate* it. This is where ***epigenetic control*** comes in.

Epigenetic control determines which *genes* are switched on or off in each tissue, thereby shaping their identity and function. Importantly, it can also change over time in response to diet, exercise, and age. This is why genetics alone does not determine our *immunometabolic* health and ageing trajectory, and why lifestyle can still have a powerful impact. By altering how tissues read their DNA, *epigenetic control* provides a critical connection between lifestyle, biology, and *immunometabolic* health.

Aims of this thesis

This thesis set out to discover whether ***DNA methylation***, a specific form of *epigenetic control*, helps to shape *immunometabolic* health. It examined how *DNA methylation* relates to *messenger molecules* in blood (**Chapters 2 and 3**), whether it can respond to lifestyle changes in older adults (**Chapter 6**), and introduced new tools to support similar work in the future, both in the lab (**Chapter 4**) and using a computer (**Chapter 5**).

Studies in blood

We explored whether *immunometabolic* health was linked to altered *DNA methylation* in blood and, if it was, whether this reflected disruption in the transmission or reception of *messenger molecules*. Our first focus was *interleukin-6 (IL-6)*, a danger signal strongly linked to ageing and inflammatory disease. This revealed that *DNA methylation* was primarily responding to *IL-6* in the blood. Circulating cells spend their lives exposed to *IL-6* and this alters how they read their DNA. We traced a clear path from elevated *IL-6* through *DNA methylation* to obesity and disease, revealing that *DNA methylation* mediates the harmful effects of inflammation.

When we shifted our focus to *adiponectin* and *leptin*, our findings took an unexpected turn. Although these *messenger molecules* are not produced in blood, the signals identified were tied to adiponectin production rather than reception. Crucially, these patterns occurred right next to the adiponectin *gene* and could also be detected in fat. This revealed two important insights. Firstly, *DNA methylation* can influence how adiponectin is produced and this directly shapes immunometabolic health. Secondly, representative signals of molecular changes in fat can also be detected in blood. This finding opens the

door to using existing large-scale blood biobanks to study age- and disease-related molecular regulation and *immunometabolic* health throughout the body.

Lifestyle changes

Beyond identifying *DNA methylation* connected to messenger molecules, we also asked whether it could respond when people alter their lifestyles. To answer this, older adults joined the GOTO study where they improved their *immunometabolic* health through a combination of increased physical activity and reduced caloric intake. Samples of their blood, fat, and muscle were collected both before and after the study and showed *DNA methylation* changes in all three tissues. These alterations could also all be linked to observed reductions in body mass index (BMI) and leptin levels.

Combining our findings across studies, we found a similar pattern: *genes* identified in multiple Chapters were involved in insulin signalling and type 2 diabetes, strengthening confidence that *DNA methylation* is biologically relevant for both responding to and driving immunometabolic health. Taken together, this thesis uncovered ***where*** DNA methylation matters, that it ***can*** respond to lifestyle changes, and ***how*** it connects *messenger molecules* to disease. By revealing how lifestyle, biological messages, and epigenetic control intersect, this work helped to lay a foundation for understanding and improving *immunometabolic* health in our ageing populations.

Nederlandse samenvatting

Inleiding

Naarmate we ouder worden, komen velen van ons voor vergelijkbare uitdagingen te staan. We hebben moeite om energiek en actief te blijven, het duurt langer om te herstellen van ziekte en het wordt steeds lastiger om een gezond gewicht te behouden. Achter deze veranderingen in het dagelijks leven schuilt een biologische transitie die vormgeeft aan hoe ons lichaam reageert op de wereld om ons heen. Dit proces wordt beïnvloed door twee nauw verbonden aspecten: veranderingen in het *immuunsysteem*, dat ziekteverwekkers detecteert en elimineert, en in onze *stofwisseling*, die regelt hoe we energie gebruiken en opslaan. Samen worden deze leeftijdsgebonden veranderingen ***immunometabole disfunctie*** genoemd en kunnen ze uiteindelijk leiden tot ouderdomsziekten zoals hart- en vaatziekten of diabetes type 2.

Een manier om inzicht te krijgen in de *immunometabole* gezondheid is door het gehalte aan ***boodschappermoleculen*** in het bloed te meten. Deze kleine signalen zorgen ervoor dat verschillende weefsels met elkaar kunnen communiceren en informatie kunnen uitwisselen over honger, stress en ontstekingen. Een bekend voorbeeld is *insuline*, dat na een maaltijd door de alvleesklier wordt afgegeven als signaal dat er energie beschikbaar is. Andere voorbeelden zijn *leptine*, dat helpt de eetlust te beheersen door berichten van ons vet en onze maag naar onze hersenen te sturen, en *interleukine-6*, een gevaarsignaal dat vrijkomt bij infecties of verwondingen (**Tekst box 1**). Een verstoring in de manier waarop deze *boodschappermoleculen* worden geproduceerd, verzonden of ontvangen, speelt een centrale rol bij veroudering en ziekte.

Epigenetica

Om deze *boodschappermoleculen* te produceren, vertrouwt ons lichaam op instructieve blauwdrukken genaamd ***genen*** die in ons DNA zijn gecodeerd. Hoewel alle weefsels hetzelfde DNA bevatten dat we van onze ouders hebben geërfd, geven ze niet dezelfde signalen door. In plaats daarvan vervullen ze elk een specifieke rol in het lichaam. Bijna alle leptine komt uit vet, terwijl insuline uitsluitend in de alvleesklier wordt aangemaakt. Dit komt omdat het niet voldoende is om een gen te bevatten; een weefsel moet het ook *activeren*. Dit is waar ***epigenetische controle*** een belangrijke rol speelt.

Epigenetische controle bepaalt welke *genen* in elk weefsel worden in- of uitgeschakeld, wat bepalend is voor de identiteit en functie ervan. Belangrijk is dat dit ook in de loop van de tijd kan veranderen als reactie op voeding, lichaamsbeweging en leeftijd. Daarom hebben niet alleen genetische, maar ook overige factoren zoals leefstijl een krachtige invloed op onze immunometabole gezondheid en verouderingstraject. Door te veranderen hoe DNA wordt gelezen in verschillende weefsels, zorgt *epigenetische controle* voor een cruciale verbinding tussen levensstijl en *immunometabole* gezondheid.

Doel van dit proefschrift

Dit proefschrift heeft als doel te onderzoeken of ***DNA-methylatie***, een specifieke vorm van *epigenetische controle*, bijdraagt aan de vorming van *immunometabole* gezondheid. Er werd onderzocht hoe *DNA-methylatie* verband houdt met *boodschappermoleculen* in het bloed (**hoofdstukken 2 en 3**) en of het kan reageren op veranderingen in levensstijl bij oudere volwassenen (**hoofdstuk 6**). Daarnaast werden er nieuwe hulpmiddelen geïntroduceerd om soortgelijk werk in de toekomst te ondersteunen, zowel in het laboratorium (**hoofdstuk 4**) als met behulp van een computer (**hoofdstuk 5**).

Onderzoek in bloed

We hebben onderzocht of *immunometabole* gezondheid verband houdt met gewijzigde *DNA-methylatie* in het bloed en of deze veranderingen wijzen op een verstoring in de overdracht of ontvangst van *boodschappermoleculen*. Onze eerste focus lag op *interleukine-6 (IL-6)*, een gevaarsignaal dat is gerelateerd aan veroudering en ontstekingsziekten. Hieruit bleek dat *DNA-methylatie* voornamelijk reageerde op *IL-6* in het bloed. Circulerende cellen worden hun hele leven blootgesteld aan *IL-6* en dit verandert de manier waarop ze hun DNA lezen. We hebben een duidelijk verband gelegd tussen verhoogde *IL-6*-waarden, *DNA-methylatie*, obesitas en ziekte, waaruit bleek dat DNA-methylatie de schadelijke effecten van ontstekingen medieert.

Toen we ons richtten op *adiponectine* en *leptine*, namen onze bevindingen een onverwachte wending. Hoewel deze *boodschappermoleculen* niet in het bloed worden geproduceerd, hielden de geïdentificeerde *DNA-methylatiesignalen* verband met de productie van adiponectine in plaats van met de ontvangst ervan. Cruciaal hierbij is dat deze patronen zich vlak naast het adiponectine-gen voordeden en ook in vetweefsel konden worden gedetecteerd. Dit leverde twee belangrijke inzichten op. Ten eerste kan *DNA-methylatie* invloed hebben op de manier waarop adiponectine wordt geproduceerd, waardoor het de immunometabole gezondheid direct beïnvloedt. Ten tweede kunnen moleculaire signalen die hun oorsprong vinden in het vetweefsel, in het bloed worden gedetecteerd. Deze bevinding opent de deur naar het gebruik van bestaande grote bloedbiobanken om leeftijds- en ziektegerelateerde moleculaire regulatie en *immunometabole* gezondheid in het hele lichaam te bestuderen.

Veranderingen in levensstijl

Naast het identificeren van *DNA-methylatie* die verband houdt met boodschappermoleculen, vroegen we ons ook af of deze zou veranderen wanneer mensen hun levensstijl aanpassen. Om dit te beantwoorden, namen oudere volwassenen deel aan het GOTO-onderzoek, waarbij ze hun *immunometabole* gezondheid verbeterden door een combinatie van meer fysieke activiteit en minder calorieën. Voor en na deze leefstijl

aanpassingen werden monsters van hun bloed, vet en spieren verzameld, waaruit bleek dat in alle drie de weefsels DNA-methylatieveranderingen optraden. Deze veranderingen konden ook allemaal in verband worden gebracht met een lagere body mass index (BMI) en leptinespiegel.

Door onze bevindingen uit verschillende hoofdstukken te combineren, vonden we een consistent patroon: genen die in meerdere hoofdstukken terugkwamen, waren betrokken bij de insulinesignalering en diabetes type 2. Dit ondersteunt de biologische relevantie van *DNA-methylatie* voor *immunometabole* gezondheid. Al met al heeft dit proefschrift aan het licht gebracht ***waar*** DNA-methylatie van belang is, dat het ***kan*** reageren op veranderingen in levensstijl en ***hoe*** het *boodschappermoleculen* in verband brengt met ziekte. Door te onthullen hoe levensstijl, biologische boodschappen en epigenetische controle elkaar kruisen, heeft dit werk een basis gelegd voor het begrijpen en verbeteren van de *immunometabole* gezondheid van onze vergrijzende bevolking.

Tekst box 1. Belangrijke boodschappermoleculen die helpen bij het meten en behouden van de immunometabole gezondheid.

Insuline wordt na een maaltijd door de alvleesklier afgegeven als reactie op een stijgende suikerniveaus, waarmee het lichaam wordt gesignaleerd dat het de beschikbare suiker als energie moet gebruiken of moet opslaan voor later. Op deze manier helpt insuline de bloedsuikerspiegel binnen een gezond bereik te houden.

Interleukine-6 (IL-6) is betrokken bij immuunreacties en ontstekingen. Kortdurende IL-6-signalering helpt bij het coördineren van de afweer en het herstel tijdens infecties of verwondingen, terwijl langdurige of ontregelde activiteit in verband wordt gebracht met chronische milde ontstekingen bij veroudering en ziekte.

Leptine wordt voornamelijk geproduceerd door vetweefsel en helpt bij het reguleren van de energiebalans op lange termijn. Een klein deel wordt ook door de maag afgegeven na een maaltijd. Het communiceert met de hersenen om de eetlust te reguleren in overeenstemming met de energievoorraden van het lichaam.

Adiponectine wordt ook afgegeven door vetweefsel en wordt in verband gebracht met gunstige metabolische effecten. Het zorgt ervoor dat weefsels effectiever reageren op insuline en beperkt ontstekingen, waardoor het de immunometabole gezondheid ondersteunt.

List of publications

1. **Sinke, L.**, van Dongen, J., Delerue, T., et al. "Epigenome-wide association study of circulating interleukin-6 connects DNA methylation to immunometabolic and inflammatory disease." ***Commun Bio.*** 9(1):242 (2026)

2. **Sinke, L.,** Delerue, T., Wilson, R., et al. "DNA methylation of genes involved in lipid metabolism as a driver of adiponectin levels and metabolic disease." ***Diabetologia.*** 69(1):127-145 (2026)

3. Liu, Y., Vukic, M., Hannon, E., Hailing, M., Walker, E., **Sinke, L.,** et al. "Identification of SENP7 and UTF1/VENTX as new loci influencing clustered protocadherin methylation across blood and brain using a genome-wide association study." ***Mol Psych.*** 31(3):1556-1568 (2026)

4. **Sinke, L.,** van Iterson, M., Cats, D., et al. "DNAmArray: streamlined workflow for the quality control, normalization, and analysis of Illumina methylation array data (v3.0)." ***Zenodo.*** (2025)

5. **Sinke, L.,** Beekman, M., Raz, Y., et al. "Tissue-specific methylomic responses to a lifestyle intervention in older adults associate with metabolic and physiological health improvements." ***Aging Cell.*** 24(4): e14431 (2025)

6. Reilly, N. A., Sonnet, F., Dekkers, K. F., Kwekkeboom, J. C., **Sinke, L.,** et al. "Oleic acid triggers metabolic rewiring of T cells poising them for T helper 9 differentiation." ***iScience.*** 27(4): 109496 (2024)

7. Costeira, R., Daimiel Ruiz, L., Gehrmann, T., Bogaards, F., Villicaña S., **Sinke, L.,** et al. "DNA methylation and gene expression trajectories of human postprandial metabolism." ***bioRxiv***. (2024)

8. Lange de Luna, J., Nounu, A., Neumeyer, S., **Sinke, L.,** et al. "Epigenome-wide association study of dietary fatty acid intake." ***Clin Epigenetics.*** 16(1): 29 (2024)

9. Costeira, R., Evangelista, L., Wilson, R., Yan, X., Hellbach, F., **Sinke, L.,** et al. "Metabolomic biomarkers of habitual B vitamin intakes unveil novel differentially methylated positions in the human epigenome." ***Clin Epigenetics.*** 15(1): 166 (2023)

10. Liu, Y., **Sinke, L.,** Jonkman, T. H., et al. "The inactive X chromosome accumulates widespread epigenetic variability with age." ***Clin Epigenetics.*** 15(1): 135 (2023)

11. Hellbach, F., **Sinke, L.,** Costeira, R., et al. "Pooled analysis of epigenome-wide association studies of food consumption in KORA, TwinsUK, and LLS." ***Eur J Nutr.*** 62(3): 1357-1375 (2023)

12. **Sinke, L.,** Cats, D., and Heijmans, B. T. "Omixer: multivariate and reproducible sample randomization to proactively counter batch effects in omics studies." ***Bioinformatics.*** 37(18): 3051-3052 (2021)

13. Mndeme, F. G., Mmbaga, B. T., Kim, M. J., **Sinke, L.,** et al. "Red reflex examination in reproductive and child health clinics for early detection of paediatric cataract and ocular media disorders: cross-sectional diagnostic accuracy and feasibility studies from Kilimanjaro, Tanzania." ***Eye (Lond).*** 35(5): 1347-1353 (2021)

14. Mangtani, P., Nguipdop-Djomo, P., Keogh, R., **Trinder, L.,** et al. "Observational study to estimate the changes in the effectiveness of bacillus Calmette-Guérin (BCG) vaccination with time since vaccination for preventing tuberculosis in the UK." ***Health Technol Assess.*** 21(39): 1-54 (2017)

Curriculum Vitae

Lucy Sinke was born in London on the 17th of March, 1990. She completed her secondary school education at James Allen's Girls' School in London, England. In 2008, she began her academic journey, enrolling in a Neuroscience B.Sc. at King's College London. It was here, during an internship simulating visual cortical neurons in MATLAB, that she discovered her passion for computational biology. To build on this interest, she pursued additional statistical qualifications and went on to complete an M.Sc. in Medical Statistics at the London School of Hygiene and Tropical Medicine (LSHTM) in 2014. During this period, she became proficient in statistical programming, including with R, STATA, and SAS, applying these tools to uncover how Huntington's disease progression affects cognition and mood. Lucy then continued her work as a Research Fellow at the International Centre for Eye Health (ICEH), where she supported clinical Ph.D. candidates by designing and statistically analysing studies.

In 2019, following emigration to The Netherlands, Lucy embarked on her Ph.D. journey in the Molecular Epidemiology group at Leiden University Medical Centre (LUMC). Here, under the guidance of Prof. dr. Bas Heijmans and Prof. dr. P. Eline Slagboom, she focused on developing and refining genomic study designs, data processing pipelines, and integrative analytical methodology. Lucy used these approaches to examine how DNA methylation impacts immunometabolic health across the life course, with the results of this research outlined in this thesis. Her work was part of the DIMENSION consortium and supported by the Joint Programming Initiative on a Healthy Diet for a Healthy Life (JPI-HDHL). During her Ph.D., Lucy collaborated with research groups both within the Netherlands and abroad, and presented her findings at national and international scientific conferences.

In 2024, Lucy began a new chapter as a postdoctoral researcher at the Leiden Academic Centre for Drug Research (LACDR). Here, in the group of Prof. dr. Bob van de Water and Dr. Giulia Callegaro, she is advancing the field of toxicogenomics and risk assessment by exploring how co-regulatory gene networks respond to chemical stressors across a diverse spectrum of *in vitro* test systems.

Acknowledgements

We actually did it! After half a decade of curiosity, spirit, and perserverance, the credits of my Ph.D. journey are officially rolling. Looking back, at all the triumphant victories and trying setbacks, I'm immensely grateful to have been granted the opportunity to develop into the scientist I am today. I would love nothing more than to continue immersing myself into the fascinating world of molecular biology for all of the years to come.

Naturally, I did not reach this milestone alone. It would have been singularly impossible without the support of my family, friends, and colleagues. Your belief in me made all the difference.

First and foremost, to my supervisors. I will always remember how you entrusted me with this project and gave me space to thrive. Bas, your endless enthusiasm for science has been invaluable since our very first meeting. Throughout all of our wild, wonderful, and wacky ideas, you have always made time for me and I will truly miss our discussions. Eline, your incredible work ethic coupled with such a genuine sense of fun is admirable and I am deeply proud to have been a part of your research group.

To my fellow epigenomics enthusiasts: Jazmin, Yunfeng, Manhoor, Laura, Thomas, Tom, and Koen. Thank you for making our work life warm and continuing to make me feel welcome whenever our paths cross. Pia and Nathalie, from the very beginning you both shared this journey with me and reminded me that I was not alone.

To my Molepi colleagues: Fatih, Stelios, Rick, Daniele, Thies, Onur, Giorgia, Ilja, Niels, Marcella, Margo, Eka, and Nico. Marian, you were my unofficial guide and I am so grateful for your help. Inge, you were there for me when I had no clue what I was doing (which was most of the time). And of course, my guidance committee, Jelle and Diana: you both provided support when I needed it most, both scientific and non-scientific.

To all my collaborators during this project: Fabian, Rory, Thomas, Jenny, Xueling, Davy, Maggie, Gibran, Josine, Erik, and Melanie. Jordana, you brought us together and made DIMENSION everything that it was. Gillian, Capucine, Alexis, and Yotam; I learned so much about myology and medicine in our evening muscle calls. And of course, to all of the participants and clinicians from the studies involved. Without your trust, this research would not have been possible.

Giulia and Bob, thank you for so warmly welcoming me into your research group. It has been a genuine pleasure to share in your passion for toxicogenomics, and to see science and research from a fresh and new perspective. And to my (not so) new colleagues: Imke, Steven, Sibel, Mick, Tamara, Kirsten, Xuesong, Martijn, Tessa, Sylvia, Mazène, Natasha, Vera, Lukas, Hugo, Hanneke, Gerhard, and Dario. I'm excited to find out what we can accomplish together!

Mummy and daddy, I miss you both more than words can say. Your instilled a love in me for science, teaching, and mathematics from such a young age. From practicing times tables on the bus to helping you grade homework on photosynthesis (although I don't know how much help I was!). Without you, I quite literally would not be here. Some days are easier than others, but not being able to share this book with you breaks my heart. I take comfort in the knowledge that you would both be so proud of me, and I will cherish the time we had and the love you gave me for the rest of my life.

Cherry, you have pushed me forwards and comforted me even and especially when I didn't know I needed it. Your continuous compassion and determination lifts me up, and I can't wait to have tea in our doctor mugs together! To all of my beautiful family scattered around the U.K., but most of all to my Uncle Malc. You were my shelter in the darkest times and I will never forget it. Milou, bedankt voor al je geduld en hulp met mijn abominabele Nederlands. Jaap en Jeanette, mijn schoonfamilie, dat jullie mij als jullie eigen dochter omarmen, is een gevoel dat ik niet in woorden kan uitdrukken. Het betekent de wereld voor mij. And of course, to my ever-growing fluffy family: Nola, Kwakje, Boefje, Sam, and Max. Life would be far less cuddly and far more cold without you to warm my heart and my keyboard.

And above all, Mark. How you put up with me I will never know. I count myself endlessly lucky to have such a spectacular partner in love and life. You have stood by my side even when the stress of this Ph.D. made me unrecognisable. You make everything worth it and I feel your love, curiosity, and kindness every single day. We've been down the rabbit hole together and back up again so many times. You're my absolute hero.

"Everything's got to end sometime. Otherwise nothing would ever get started." -- The Doctor.


SOCS3
ADIPOQ
SREBF1
G
C
C