doi:10.2533/chimia.2022.1052 Chimia 76 (2022) 1052–1062

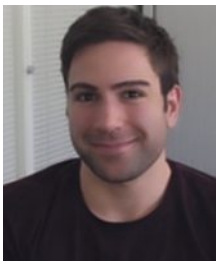
PhenoExplorer: An Interactive Web-based Platform for Exploring (Epi)Genome-Wide Associations Using a Swiss Population-based Study

Jean-Pierre Ghobril[a], Dusan Petrovic[b], Georg Ehret[c], Belén Ponte[d], Menno Pruijm[e], Daniel Ackermann[f], Bruno Vogt[f], Silvia Stringhini[g], Aurélien Thomas[h], Jonviea Chamberlain[a], Semira Gonseth-Nusslé[a], and Murielle Bochud*[a]

Abstract: The recent advent of high-throughput sequencing technologies has allowed the exploration of the contribution of thousands of genomic, epigenomic, transcriptomic, or proteomic variants to complex phenotypic traits. Here, we sought to conduct large-scale (Epi)Genome-Wide Association Studies (GWAS/EWAS) to investigate the associations between genomic (Single Nucleotide Polymorphism; SNP) and epigenomic (Cytosine-Phospho-Guanine; CpG) markers, with multiple phenotypic traits in a population-based context. We used data from SKIPOGH, a family- and population-based cohort conducted in the cities of Lausanne, Geneva, and Bern (N = 1100). We used 7,577,572 SNPs, 420,444 CpGs, and 825 phenotypes, including anthropometric, clinical, blood, urine, metabolite, and metal measures. GWAS analyses assessed the associations between SNPs and metabolites and metals (N = 279), using regression models adjusted for age, sex, recruitment center, and familial structure, whereas EWAS analyses explored the relations between CpGs and 825 phenotypes, additionally adjusting for the seasonality of blood sampling and technical nuisance. Following the implementation of GWAS and EWAS analyses, we developed a web-based platform, PhenoExplorer, aimed at providing an open access to the obtained results. Of the 279 phenotypes included in GWAS, 103 displayed significant associations with 2804 SNPs (2091 unique SNPs) at Bonferroni threshold, whereas 109 of the 825 phenotypes included in EWAS analyses were associated with 4893 CpGs (2578 unique CpGs). All of the obtained GWAS and EWAS results were eventually made available using the in-house built web-based PhenoExplorer platform, with the purpose of providing an open-access to the tested associations. In conclusion, we provide a comprehensive outline of GWAS and EWAS associations performed in a Swiss population-based study. Further, we set up a web-based PhenoExplorer platform with the purpose of contributing to the overall understanding of the role of molecular variants in regulating complex phenotypes.

Keywords: EWAS · GWAS · PhenoExplorer · Population-based · SKIPOGH

Dr. Jean-Pierre Ghobril (PhD) is currently working as a post-doctoral fellow in the group of Murielle Bochud at Unisanté in Lausanne. In the framework of NCCR TransCure, he performs research on associations between transporter genes and biological phenotypes using data from the Swiss Kidney Project on Genes in Hypertension population-based study (SKIPOGH). As part of TransCure, Dr. Ghobril's interests are related to the relation between genome-wide molecular variants (SNPs) and plasma and urinary phenotypes, with a particular interest for small elements (*i.e.* zinc, iodine, copper, manganese). Currently, Dr. Ghobril is working on opening extensive data for hypothesis-driven interrogation by other TransCure researchers.



Dr. Dusan Petrovic (PhD) is a post-doctoral researcher working at the Department of Epidemiology and Health Systems at the Center for Primary Care and Public Health – Unisanté. Trained as a biologist, with a PhD in social epidemiology, he is currently investigating the underlying mechanisms of chronic diseases using population-based data. His interests within the TransCure project include the relation between epigenetic alterations (DNA methylation markers) and transmembrane transport of small elements and solutes (urea, zinc, copper, selenium).

**Correspondence:* Prof. M. Bochud[a], E-mail: murielle.bochud@unisante.ch
[a]Department of Epidemiology and Health Systems (DESS), University Center for General Medicine and Public Health (UNISANTE), Lausanne, Switzerland; [b]University Center for General Medicine and Public Health (UNISANTE), Lausanne, Switzerland; [c]Department of Cardiology, Geneva University Hospitals (HUG), Geneva, Switzerland; [d]Department of Nephrology and Hypertension, Geneva University Hospitals, Geneva, Switzerland; [e]Department of Nephrology and Hypertension, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland; [f]University Clinical for Nephrology, Hypertension, and Clinical Pharmacology, Inselspital, Bern, Switzerland; [g]Unit of Population Epidemiology, Division of Primary Care Medicine, Geneva University Hospitals, Geneva, Switzerland; [h]Unit of Forensic Toxicology and Chemistry, CURML, Lausanne University Hospital-Geneva University Hospitals, Lausanne-Geneva, Switzerland

Prof. Georg Ehret is a cardiologist and Assistant Professor associated with the Department of Cardiology at the Geneva University Hospitals and the University of Geneva. His clinical and research interests include cardiac imagery, cardiovascular risk factors, prevention, and disease mechanisms, as well as hypertension.

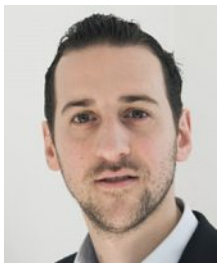
Prof. Belén Ponte works as a doctor at the University Hospitals of Geneva in the Nephrology and Hypertension Service. She is specialized in internal medicine and nephrology, with a title of European Hypertension Specialist and a certification in abdominal ultrasound. She also obtained a MSc and PhD in Epidemiology. She became assistant professor at the Faculty of Medicine of Geneva's University in 2021. She is both a clinician and researcher. Her research is mainly focused on biomarkers and complications of chronic kidney disease and hypertension.

PD Dr. Menno Pruijm is full-time staff member (médecin adjoint) and Head of the chronic dialysis unit of the Service of Nephrology and Hypertension of the University Hospital of Lausanne (CHUV). His research focuses on the application of imaging techniques to unravel the pathophysiology and hemodynamic alterations in CKD. He is one of the main drivers of several imaging (ultrasound) cohort studies including SKIPOGH and he is an international expert in functional renal MRI and part of the European Union COST Action ('PARENCHIMA'). He has published >150 peer-reviewed papers, was cited >3000 times and has an H-index of 32.

PD Dr. Daniel Ackermann is a nephrologist and full-time staff member at the Nephrology and Hypertension Clinic of the Bern University Hospital (Inselspital). His clinical and research interests include hypertension mechanisms, with a particular focus on the role of endogenous steroid compounds and metabolites.

Prof. Bruno Vogt is Head of the Nephrology and Hypertension Clinic of the Bern University Hospital (Inselspital). Throughout his career, Prof. Vogt has worked on the clinically relevant problem of renal sodium retention and edema formation in nephrotic syndrome and cirrhosis of the liver, developing new mice models for this purpose. In clinical research, he focuses on the problem of renal oxygenation and perfusion using magnetic resonance techniques and classic renal hemodynamic human studies. He has authored >270 publications in peer-reviewed journals.

Prof. Silvia Stringhini is Head of the Unit of Population Epidemiology of Geneva University Hospitals. As an experienced epidemiologist with a keen interest for socioeconomic inequalities in health, Prof. Stringhini has been one of the PI's of the Pan-European Lifepath project, and a senior member of the COVID-19 task force in Switzerland during the 2020–2021 SARS-COV-2 pandemic. She has authored >200 peer-reviewed publications in prestigious journals, including the very first Seroprevalence study of anti-SARS-CoV-2 antibodies published in the Lancet.



Prof. Aurélien Thomas is Associate Professor, head of the Faculty Unit of Toxicology and deputy head of the Unit of Forensic Toxicology of the CURML Institute, Lausanne-Geneva. His main research interest lies in developing omic strategies in order to investigate complex molecular perturbation occurring in pathologic phenotypes, including toxicological and metabolic disorders. He has published over 100 peer-reviewed articles in peer-reviewed journals.

Dr. Jonviea D. Chamberlain (PhD) is an epidemiologist at Unisanté studying epigenetic epidemiology and works closely with the Swiss start-up Genknowme as part of an Innosuisse grant. With an interest in chronic disease epidemiology and prevention, Dr. Chamberlain's research seeks to identify modifiable lifestyle exposures to support healthy aging. Prior to her position at Unisanté, Dr. Chamberlain worked at the University of Bordeaux (Inserm center U1219) investigating alternative hypotheses of the causal chain leading to dementia. During her PhD and first postdoc, she studied differences in all-cause and cause-specific mortality outcomes in the Swiss traumatic spinal cord injured (TSCI) population.

Dr. Semira Gonseth Nusslé, MD, MSc, board-certified in prevention and public health has been working as a senior resident in Unisanté during 2017–2022. She was a co-principal investigator of epidemiological population-based studies in Switzerland such as the Swiss health study pilot and the Serocovid study on Covid-19. Previously, she studied cancer genetics and epigenetics at the University of California, San Francisco and Berkeley. She led a technology-transfer project to develop blood-based epigenetic signatures of ageing and lifestyle based on analyses of the SKIPOGH cohort, which are now commercialized in the startup 'Genknowme'.

Prof. Murielle Bochud is professor of public health at the Faculty of Biology and Medicine of the University of Lausanne. She is a public health physician specialized in genetic and nutrition epidemiology. She is Head of the Department of Epidemiology and Health Systems at the Center for Primary Care and Public Health - Unisanté, Lausanne. Her research focuses primarily

on the epidemiology of cardiometabolic risk factors, nutrition epidemiology from population-based data.

1. Introduction

The recent advent of high-throughput 'OMICS' technologies has provided a new conceptual framework for characterizing genetic, biological, and pathophysiological processes occurring in human populations.[1,2] Unlike the classical 'gene candidate approach', which investigates the relation between a candidate gene and a phenotype of interest in an experimental setting, large-scale OMICS technologies allow implementing a 'hypothesis-free' or 'agnostic' approach, whereby the associations between a *large number* of pre-identified molecular variants (*i.e.* >30'000 variants) and a phenotype of interest are explored in a population-based setting.[2,3]

Typical examples of OMICS analyses include Genome-Wide Association Studies (GWAS), which investigate the associations between Single Nucleotide Polymorphisms (SNP) on one hand, and phenotypes of interest on the other hand (*i.e.* presence of disease, life-span, anthropometric traits).[1,4] To date, the overall impact of OMICS-based GWAS studies has been substantial in biomedical research, highlighting *common* genomic (or genetic) variants related to heart disease, age-related macular degeneration, chronic kidney disease, neurodegenerative disorders, as well as many other diseases and phenotypic traits.[1,4–6] Genome-wide polygenic risk scores, which combine the effects of millions of markers across the human genome, have now reached potential for clinical utility in people of European and Asian ancestries to predict risk of future common chronic diseases, yet they currently have limited transferability and clinical utility in people of African ancestry.[7,8]

In addition to *genomic* molecular variants, OMICS-Wide Association Studies further include *transcriptomic*, *proteomic*, *metabolomic*, as well as *epigenomic* molecular markers, which have been the object of keen interest in recent years, although their tissue-specific and time-dependent characteristics make them much more complex to explore. Epigenomic (or epigenetic) changes include histone modifications, chromatin remodeling, microRNAs, as well as DNA methylation, which constitutes the most well-described epigenetic process.[9] DNA methylation changes refer to the addition or removal of methyl groups to CpG dinucleotides across the genome, occurring either naturally (*i.e.* development, senescence), or as a result of environmental exposures, such as lifestyle factors, nutrition, adversity, or pollution.[9–12] In particular, previous studies have suggested that DNA methylation changes may constitute an intermediate process through which the external environment gets biologically embedded, whereby various environmental exposures lead to DNA.[13] Epigenome-wide association studies (EWAS), based on easily accessible white blood cells DNA methylation data can provide insight into the molecular and functional understanding of GWAS loci and further our understanding of complex genotype-phenotype relationships.[14,15]

In this research, we sought to conduct a large-scale multi-OMICS wide analysis using data from SKIPOGH, a multicentric population-based cohort conducted in Switzerland, mainly focusing on kidney and blood pressure related phenotypes.[16–25] The overarching objectives of this work are to document the associations between genomic/epigenomic variants available in SKIPOGH, including 7M SNPs, 420K CpGs, and over 800 different phenotypes (*i.e.* plasma and urinary solutes, lipids, steroids, metabolites, metals, clinical outcomes), as well as to develop a web-based platform, PhenoExplorer, providing open access to the obtained GWAS and EWAS associations results, thereby promoting the findability, accessibility and use of this publicly funded resource. Specifically focusing on the TransCure project, the goal of this work is to provide candidate (epi)genetic markers, genes, and phenotypes to TransCure groups, with the purpose of establishing a multi-disciplinary approach in studying biological mechanisms related to membrane transporters.

2. Methods

2.1 *Study Population*

We used data from the Swiss Kidney Project on Genes in Hypertension (SKIPOGH), a Swiss family- and population-based multicentric cohort investigating the genetic and environmental determinants of health-related outcomes in the Swiss population (Fig. 1).[23,26] Study participants were recruited in the city of Lausanne and the cantons of Geneva and Bern between 2009 and 2013 (SKIPOGH 1, baseline visit), and came for a follow-up visit three years later (SKIPOGH 2: 2013–2016).[23] Inclusion criteria were: (1) written informed consent; (2) ≥18 years of age; (3) Caucasian origin; (4) at least one first-degree family member willing to participate to the study. We excluded women who reported being pregnant. Upon recruitment, SKIPOGH 1 baseline visit eventually included 1129 participants grouped within 275 families, SKIPOGH 2 follow-up visit included 1034 participants coming from 270 families, whereas 983 individuals participated to both study waves. At both visits, participants attended in-depth medical and anthropometric examinations after an overnight fast, provided blood and urine samples, and completed a self-administered questionnaire inquiring about their living standards, socioeconomic and financial circumstances across the life-course, lifestyle factors, and medical history. All participants provided written informed consent.

2.2 *SNP Data Collection and Pre-processing*

Genome-wide Single Nucleotide Polymorphism (SNP) data was obtained from white blood cells' DNA (SKIPOGH 1), and generated using the Illumina Human Omni 2.5 platform.[27] We subsequently performed SNP data pre-processing and quality control checks by applying an in-house built algorithm.[28–31] Briefly, after an automatic clustering in GenomeStudio (Illumina Inc. San Diego), we selected samples (participants) with a call rate (proportion of non-missing SNPs) >0.99 to update the SNP statistics. Once re-clustered, we retained markers with a call rate >0.95. The quality control and pre-processing procedures yielded 979 samples with 1,637,659 SNPs, whose call rate was >95%, whose minor allele frequency (MAF) was >2%, and whose Hardy-Weinberg Equilibrium P value was >0.001. We then performed multiple imputation for SNP *missing values* using the Minimac3 imputation algorithm, yielding a final set of 7,577,572 SNPs available for analyses.[30,32,33]

2.3 *CpG Data Collection and Pre-processing*

Epigenome-wide Cyto-Phospho-Guanine (CpG) DNA methylation from white blood cells was measured in 242

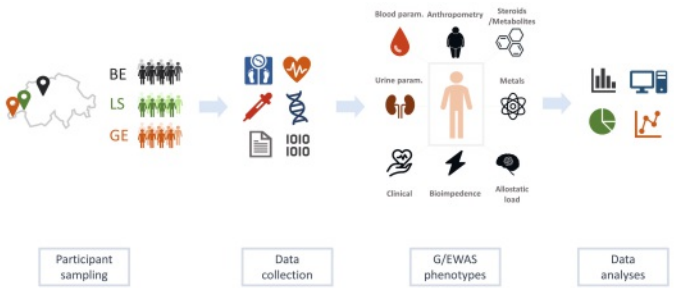


Fig. 1. Infographics summarizing the SKIPOGH study design and groups of phenotypes included in GWAS and EWAS analyses.

SKIPOGH 2 participants (follow-up visit) using the Infinium HumanMethylation450 BeadChip microarray of Illumina (HM450), assessing the methylation status at 485,512 CpG sites. For a different set of 442 SKIPOGH 2 participants, epigenome-wide DNA methylation was measured using a more recent Infinium MethylationEPIC v1.0 microarray (EPIC), including >90% of the CpG sites from the HM450 and an additional 413,743 CpGs (865,859 CpGs in total).[34] For both arrays, CpG methylation data were summarized as β coefficients representing a ratio of the average signal for methylated CpG sites to the sum of methylated *and* unmethylated sites. CpG data pre-processing included multiple imputation of missing data using the nearest averaging multiple imputation method,[35] logit transformation of imputed values, and a 'denoising' procedure aimed at accounting for the variance introduced by participant's familial structure (random-effect confounder), whereby the resulting residuals were directly added to the transformed CpG methylation data, enabling the implementation of fixed-effect regression models. To control for the nuisance introduced by technical factors (methylation array type, array position, and plate level), the CPACOR procedure was applied, yielding 30 principal components to be used as fixed-effect covariates in the regression models.[36] The data pre-processing procedures eventually yielded 420,444 CpG sites available for Epigenome-Wide Association Study analyses (EWAS) in the SKIPOGH 2 sample.

2.4 *Phenotypes*

2.4.1 *SKIPOGH 1 GWAS Phenotypes*

In SKIPOGH 1, 981 participants had available samples for plasma metabolites measurement. Samples were analyzed using Liquid Chromatography-Multiple Reaction Monitoring/Mass Spectrometry (LC-MRM/MS - 100µL samples), with measures performed in positive and negative electrospray ionization, and using the Mass Spectrometry Metabolite Library as reference material for standard metabolites.[37] Of the 606 targeted raw metabolites, 232 were eventually measured and used for GWAS analyses following quality control and correction procedures to account for the correlated nature of the data.[37,38]

In addition to plasma metabolites, GWAS analyses included associations for metal elements measured in 24 h urine collections ('metallomics' phenotypes). Briefly, 24 elements (Ag, Al, As, Be, Bi, Cd, Co, Cr, Cu, Hg, I, Li, Mn, Mo, Ni, Pb, Pd, Pt, Sb, Se, Sn, Ti, V, Zn) were measured in day *and* night urine samples using the imaging mass spectrometry method (ICP-MS), enabling the detection and quantification of metals in biological samples (N = 47 additional phenotypes).[39]

To account for the non-normal distribution of plasma metabolites and urinary metal measures, phenotypes were transformed using the log10 transformation.

2.4.2 *SKIPOGH 2 EWAS Phenotypes*

We included a total of 825 phenotype variables in EWAS analyses (Fig. 1 and Supplementary Information SI_I). We subdivided phenotype variables into nine categories: (1) anthropometric measures (N = 19 variables), (2) bioelectrical impedance measures (N = 8), (3) blood parameters (N = 74), (4) clinical outcomes and health behaviors (N = 7), (5) general urinary parameters (N = 196 – including day, night and 24 h measures), (6) urinary steroids and enzymatic activities (N = 184 – including day, night and 24 h measures), (7) plasma metabolites (N = 239), (8) plasma and urinary metals (N = 96), and (9) allostatic load scores (N = 2). Considering that physiological phenotypes may display a non-normal distribution and potentially yield spurious associations with CpG markers, we assessed the distribution of each phenotype and performed a non-linear transformation where necessary, in order to obtain normal, or close-to-normal distribution. Applied non-linear transformations included square-root, $\log_{10}$, inverse, inverse square-root, square, inverse square, cubic, and inverse cubic transformations.[40,41]

2.5 *Statistical Analyses*

2.5.1 *Genome-Wide Association Study*

We tested the associations between genome-wide SNP markers (predictor variables; 0: homozygous minor allele genotype, 1: heterozygous genotype, 2: homozygous major allele genotype) and 279 phenotypes from SKIPOGH 1 using linear regression models, adjusting for age, sex, and recruitment center. We accounted for random-effect familial relations by using a kinship matrix generated based on imputation genotypes using PLINK.[42] After discarding participants with missing data for the main outcome and covariate variables (N = 241), the final analytical set included 738 individuals. We accounted for multiple testing by applying the Bonferroni correction method ensuring a control of the family wise error rate below 0.05 (P<5E-08).[43] The GWAS statistical analyses were carried out using EPACTS, Efficient and Parallelizable Association Container Toolbox (University of Michigan, Michigan, United States), while data preparation and pre-processing was conducted using the R statistical software.[44] For each GWAS, Manhattan plot and corresponding regional plots were generated, the latter including every putative local association peak under association strength of P≤1.0E-5 (chromosome-wide significance). Lookup of genetic functional consequence and possible known related clinical phenotypes was performed querying ENSEMBL.[45]

2.5.2 *Epigenome-Wide Association Study (EWAS)*

We applied fixed-effect linear regression models for the EWAS analysis.[46] We ran 825 *multiple* linear regression models, using SKIPOGH 2 phenotypes as successive, independent variables, and pre-processed CpG DNA methylation data as a combined set of dependent variables (N = 420,444 CpG markers). Regression models were adjusted for age, sex, recruitment center (Lausanne, Geneva, Bern), seasonality of blood sampling (spring, summer, fall, winter), chip type (HM450, EPIC), 30 CPACOR principal components, and Houseman-estimated white blood cell composition (CD8, CD4, NK, B cells, Monocytes, Granulocytes).[47] The EWAS statistical analyses were carried out using the R statistical software and relevant CRAN and Bioconductor packages (R Foundation for Statistical Computing, Vienna, Austria). For each phenotype, EWAS results were represented using a Volcano plot (x-axis: beta regression coefficient, y-axis: P value), a Manhattan plot (x-axis: CpG chromosomal position, y-axis: P value), and a summary table including beta, standard-error (SE) and P value coefficients between each of the 420,444 CpG marker and the phenotype of interest (Supplementary Information SI_IV). Statistical significance threshold were set at P<0.05/420,444 following Bonferroni correction for multiple testing, as well as the Benjamini-Hochberg (BH) correction.

2.6 *PhenoExplorer*

We developed PhenoExplorer, a web-based, interactive platform leveraging EPACTS, and 'LocusZoom Standalone' pipelines, and RStudio's 'Shiny' web app, with the purpose of visualizing GWAS/EWAS association results introduced previously.[44,48] Briefly, the first development step comprised GWAS and EWAS analyses ran on local machines aimed at generating summary statistics along with Manhattan and Volcano plots (7,577,572 SNPs, EWAS: 420,444 CpGs). The second development step included summary statistics filtering in order to obtain local association minima and generating regional plots for every filtered local minimum SNP (GWAS only). Third, annotated lookup tables with genetic and clinical consequences were eventually generated from the raw

data querying ENSEMBL database,[45] whilst the fourth and final development step included results grouping, structuring, and web publishing *via* 'Shiny' RStudio app.

3. Results

3.1 *General Characteristics*

We present the GWAS and EWAS samples characteristics in Table 1. Overall, 979 individuals (87%) were included in GWAS analyses (SKIPOGH 1: 2009–2012), of which 52% were women. The mean age was 48.3 years, with 41% of the study sample being recruited in Lausanne, 43% in Geneva, and 16% in Bern. 65% of included participants were actively working at the time when the interview was conducted, while 5% were students, 20% were retired or disabled, and 10% were either unemployed or stay-at-home. The proportion of current smokers was 24%, the average BMI 25.2 kg/m^2, whereas the proportion of individuals self-reporting chronic diseases was 12% for heart disease, 13% for kidney disease, 5% for diabetes, 32% for hypertension, and 1% for cancer (any type of malignancy). A subset of 684 participants were included in EWAS analyses using SKIPOGH 2 data (2013–2016), with a mean age of 52.6 years. Whilst we found similar distributions for sex, recruitment center, professional activity, and tobacco smoking when compared to GWAS participants, the EWAS sample included more individuals reporting heart disease (19%), and fewer individuals reporting kidney disease (4%).

3.2 *GWAS*

Of the 279 phenotypes included in the GWAS analyses, 103 were significantly associated with 2804 SNPs (2091 unique SNPs) at the genome-wide Bonferroni threshold (P<5.0E-08), including 90 plasma metabolites and 13 urine metal phenotypes (Supplementary Information SI_II and SI_III). In Table 2, we show a list of *non-exhaustive* GWAS associations, displaying 26 phenotypes along with their top-associated SNPs (threshold: P<9.4E-10). Further, among the SNPs displayed in Supplementary Information SI_III, we found that none of the highlighted markers were located within the genes of interest for the TransCure project (SLC17, SLC9, DMT, FPN, SLC11, SLC40, ABCG, FABP, TRPM4, SLC7). For each GWAS analysis, the obtained results are represented using a Manhattan plot and a summary table reporting the regression coefficients (beta, standard error, P value) as displayed in Fig. 2 (Phenotype: Bilirubin (plasma metabolite)).

3.3 *EWAS*

Out of the 825 EWAS-related phenotypes, 109 were significantly associated with 4893 CpG markers at the Bonferroni threshold (Supplementary Information SI_IV: 2578 unique CpGs), including 34 blood parameters, 23 urinary parameters, 17 urinary steroids/enzymatic activity measures, 15 plasma metabolites, 9 metallomics markers, 5 bioimpedance parameters, 3 anthropometry measures, 2 clinical outcomes/health behaviors, and 1 biological health score (allostatic load). The phenotypes displaying the highest number of associated CpG markers were related to white blood cell composition, including the number of eosinophils (N = 1733/N = 1204 for the absolute number of eosinophils), neutrophils (N = 600/N = 5 for the absolute number of neutrophils), basophils (N = 22/N = 139 for the absolute number of basophils), and leukocytes (N = 79), followed by body water amount (N = 199), smoking status (N = 120), and urinary methylxanthine excretion (urinary paraxanthine: N = 146; urinary theophylline: N = 83). Specifically examining CpGs located within TransCure genes of interest, we found 30 significant associations, involving 12 phenotypes (Table 3). Overall, 20 associations were found for the number/percentage of eosinophils or neutrophils, two CpGs were associated with blood triglyceride levels (ABCG1-cg06500161, SLC7A11-cg06690548), and respectively one CpG for blood insulin (ABCG1-cg06500161), HDL cholesterol ABCG1-cg06500161), Gamma-Glutamyl Transferase (SLC7A11-cg06690548), basal metabolism (SLC17A6-cg04954559), body water (SLC9A6-cg04675306), mean corpuscular haemoglobin (SLC7A11-cg06690548), and urinary magnesium excretion (SLC7A6-cg09194755). The results for each EWAS analysis are presented using a Volcano plot and a Manhattan plot, as illustrated in Fig. 3 (Phenotype: current smoking), as well as a table reporting regression coefficients (beta, standard error, P value) between each phenotype and CpG marker (see PhenoExplorer). In the example of current smoking (binary phenotype: Yes/No), 122 CpGs were found to be associated at Bonferroni threshold (Top five associated CpGs: AHRR-cg05575921, β=-1.17, P=4.52E-65; F2RL3-cg03636183 β=-0.95, P=5.09E-45; cg21566642, β=-1.17, P=1.22E-44; AHRR-cg21161138 β=-1.00, P=2.16E-44; cg01940273, β=-0.81, P=3.79E-42).

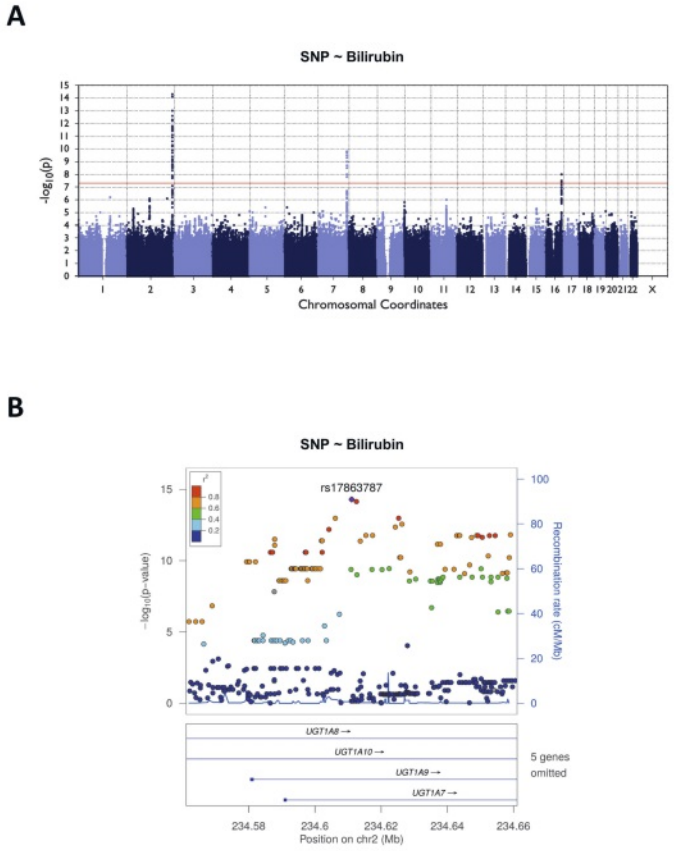


Fig. 2. Genome-Wide Association Study results for plasma bilirubin concentration represented using a Manhattan plot (A) and a regional plot (B). Linear regression model for the association between GWAS markers and phenotypes of interest, adjusting for age, sex, recruitment center, and familial structure (random-effect covariable).

3.4 *PhenoExplorer*

We present the inner structure of PhenoExplorer as well as the user's layout in Fig. 4. As described in the Methods section, the development process included four main steps, including running GWAS/EWAS regressions on a local server (R, EPACTS), Manhattan/Volcano/regional plot generation using summary statistics, and web-based platform organization and structuring using 'Shiny'. From the user's perspective, PhenoExplorer allows performing queries using four different criteria: (1) Phenotype of interest (*i.e.* urinary zinc levels, blood glucose, CRP, smoking

Table 1. General characteristics of GWAS/EWAS included participants.

		SKIPOGH 1 (GWAS)	SKIPOGH 2 (EWAS)
		N=979	N=684
Women (N, %)		589 (52%)	359 (52%)
Age (μ±SD)		47.4 (±17.5)	52.6 (±15.5)
Center (N, %)			
	Lausanne	416 (37%)	307 (45%)
	Geneva	426 (38%)	279 (41%)
	Bern	287 (25%)	98 (14%)
Professional activity			
	Working (employed, self-employed)	732 (66%)	449 (66%)
	Student	58 (5%)	7 (1%)
	Retired/Disabled	207 (19%)	148 (22%)
	Unemployed/Stay-at-home	119 (11%)	75 (11%)
	Military service	1 (0%)	0 (0%)
Tobacco smoking (N, %)			
	Never	498 (45%)	287 (42%)
	Former	346 (31%)	222 (33%)
	Current	272 (24%)	171 (25%)
BMI (μ±SD)		25 (±4.5)	25.6 (±4.6)
Chronic disease (N, %)			
	Heart disease	128 (12%)	130 (19%)
	Kidney disease	147 (13%)	26 (4%)
	Diabetes	51 (5%)	39 (6%)
	Hypertension	346 (31%)	204 (30%)
	Cancer	13 (1%)	8 (1%)
Assessed phenotypes (N)			
	Plasma metabolites	232	239
	Metallomics	47	96
	Anthropometry		19
	Bioelectrical impedence		8
	Blood parameters		74
	Clinical outcomes/Health behaviors		7
	Urinary parameters		196
	Urinary steroids		184
	Allostatic load		2

status, history of heart disease); (2) Genetic markers (CpG, SNP, gene name); (3) Statistical significances; and (4) genomic position. The resulting overview includes a Manhattan plot, a summary table (beta, standard error, *P* value, chromosome, chromosomal position, and location description (*i.e.* intragenic, gene body, transcription start site, exon, intron)), a Volcano plot (EWAS only), as well as regional plots displaying local minima of interest on a 1000bp-wide sliding window (GWAS only). Whilst PhenoExplorer includes >800 SKIPOGH phenotypes, additional GWAS and EWAS associations results (summary tables, Manhattan, Volcano, regional plots) can be directly added as a result of the modular conception of PhenoExplorer (use case scenario given in the Supplementary Information SI-Video). The PhenoExplorer users' URL shall be communicated upon motivated request addressed to the authors (jean-pierre.ghobril@unisante.ch).

4. Discussion

In this research, we conducted two large-scale multi-OMICS analyses using data from the SKIPOGH population-based cohort, including over 7M genetic variants, 420K epigenetic markers, and 825 phenotypes. We found that 2804 SNPS (2091 unique

Table 2. Non-exhaustive selection of SNPs associated with plasma metabolite phenotypes (SKIPOGH 1).

Phenotype	*Type*	*top SNP*	*Chromosome*	*Gene & consequence*	*P* [a,b]
BILIVERDIN	Plasma metabolite	rs887829	2	UGT1A8 intron drug response	4.20E-30
N.ACETYL.L.PHENYLALANINE	Plasma metabolite	rs13538	2	NAT8 / ALMS1P1 missense variant	7.71E-25
N.ACETYL.L.PHENYLALANINE	Plasma metabolite	rs6546838	2	ALMS1P1 missense variant	1.87E-23
URIDINE.5.MONOPHOSPHATE	Plasma metabolite	rs35724886	14	ACOT4 missense variant	2.50E-21
THYMIDINE.5.MONOPHOSPHATE	Plasma metabolite	rs35724886	14	ACOT4 missense variant	4.53E-21
5.CMP	Plasma metabolite	rs35724886	14	ACOT4 missense variant	5.40E-17
hexanoylcarnitine	Plasma metabolite	rs1061337	1	ACADM synonymous variant	7.70E-17
L.CARNITINE	Plasma metabolite	rs1171614	10	SLC16A9 5 prime UTR variant	8.94E-16
BILIRUBIN	Plasma metabolite	rs887829	2	UGT1A8 intron drug response	4.50E-15
BILIRUBIN	Plasma metabolite	rs1137617	7	KCNH2 stop gained	4.50E-15
DEOXYCARNITINE	Plasma metabolite	rs7969761	12	SLC6A13 intron	2.01E-13
TRANS.4.HYDROXYPROLINE	Plasma metabolite	rs78703287	10	APB1IP intron variant	1.70E-12
DEOXYCARNITINE	Plasma metabolite	rs10848766	12	KDM5A intron	3.72E-12
isovalerylcarnitine	Plasma metabolite	rs142651837	8	RAB2A intron variant	1.83E-11
INOSINE.5.PHOSPHATE	Plasma metabolite	rs61292904	3	ACAD11	2.72E-11
GALACTARATE	Plasma metabolite	rs2229540	1	AKR1A1	5.20E-11
isovalerylcarnitine	Plasma metabolite	rs113899565	8	LINC01301 non coding transcript exon variant	8.46E-11
octanoylcarnitine	Plasma metabolite	rs748826784	1	ACADM frameshift variant	1.32E-10
GUANINE	Plasma metabolite	rs184011454	2	intron RHOQ in LD with ATP6V1E2	1.63E-10
LEUCINE	Plasma metabolite	rs78703287	10	APBB1IP intron variant	2.23E-10
CYTIDINE	Plasma metabolite	rs62363602	5	SLC38A9 intron	4.57E-10
THEOPHYLLINE	Plasma metabolite	rs2787566	6	GRIK2 non coding exon transcript	5.30E-10
L.ISOLEUCINE	Plasma metabolite	rs151287024	13	FAM12A intron variant	7.62E-10
3.HYDROXY.3.METHYLGLUTARATE	Plasma metabolite	rs1171616	10	SLC16A9 intron variant	8.07E-10
L.PHENYLALANINE	Plasma metabolite	rs75338359	14	KCNH5 intron variant	8.86E-10
LEUCINE	Plasma metabolite	rs151287024	13	FAM12A intron variant	9.40E-10

a Linear regression model for the association between SNP marker and phenotypes of interest, adjusting for age, sex, recruitment center, and familial structure (random-effect covariable)
b Significance threshold was set using the Bonferroni correction method (P <5.0E-08).

markers) were associated with 103 phenotypes in the GWAS analyses at the genome-wide Bonferroni threshold, whereas 4893 CpGs (2578 unique CpGs) were associated with 109/825 phenotypes in the EWAS analyses. Finally, using 'Shiny' R-based app, we developed a user-friendly, web-based platform named 'PhenoExplorer', providing an open access to GWAS and EWAS results obtained in this research.

Whilst the purpose of this research was to conduct large-scale GWAS and EWAS analyses without specifically interpreting the obtained results, we used some of the previously established associations to verify whether the models implemented here could partially replicate former findings. The large-scale exploratory GWAS yielded associations in line with previous investigations, but also novel findings. In particular, we reproduced known associations involving plasma biliverdin and bilirubin with rs887829,[49] rs13538 with N-acetyl-phenylalanine,[50] rs1061337 with hexanoylcarnitine,[51] as well SLC2A9 variants and uric acid.[52,53] Using tobacco smoking in EWAS analyses, we found that 109 of the 120 (91%) identified CpGs were found to be associated with smoking status in former smoking-related EWAS studies, with top hits consistently including AHRR-cg05575921, F2RL3-cg03636183, or GFI1-cg09935388 in SKIPOGH.[12,54,55] Finally, *all* of the GWAS and EWAS analyses presented in this research are comprehensively reported and organized within PhenoExplorer, a web-based platform providing a fast turnover and visualization of millions of associations, and offering a modular structure for the incorporation of additional results.

4.1 Strengths and Limitations

This study has several strengths, the first being the richness of the physiological, clinical, lifestyle, and (epi)genetic data available in the SKIPGOH cohort, allowing to investigate multiple research questions related to the contribution of molecular variants to complex phenotypes. Second, we developed a user-friendly, web-based platform with the purpose of openly sharing GWAS and EWAS results with researchers interested in either specific genetic or epigenetic markers, or phenotypes available in SKIPOGH. The added values of PhenoExplorer is the real time interaction with the data as well as a rapid visualization of GWAS/EWAS associations results, whereby user queries may include phenotypes, (epi)genetic marker identifiers (SNPs, CpGs), and gene names. Finally, the modular structure of PhenoExplorer allows a rapid integration of additional GWAS/EWAS/XWAS results, further allowing the platform to grow.

This study also has important limitations to acknowledge. First, the small sample size (N_{GWAS} = 979, N_{EWAS} = 684) restricts the ability to detect smaller effect-size associations and limits the overall statistical power. Second, GWAS and EWAS models were minimally adjusted for potential confounders (age, sex, recruitment center, familial structure, technical nuisance), but did not include covariates that may specifically confound relations

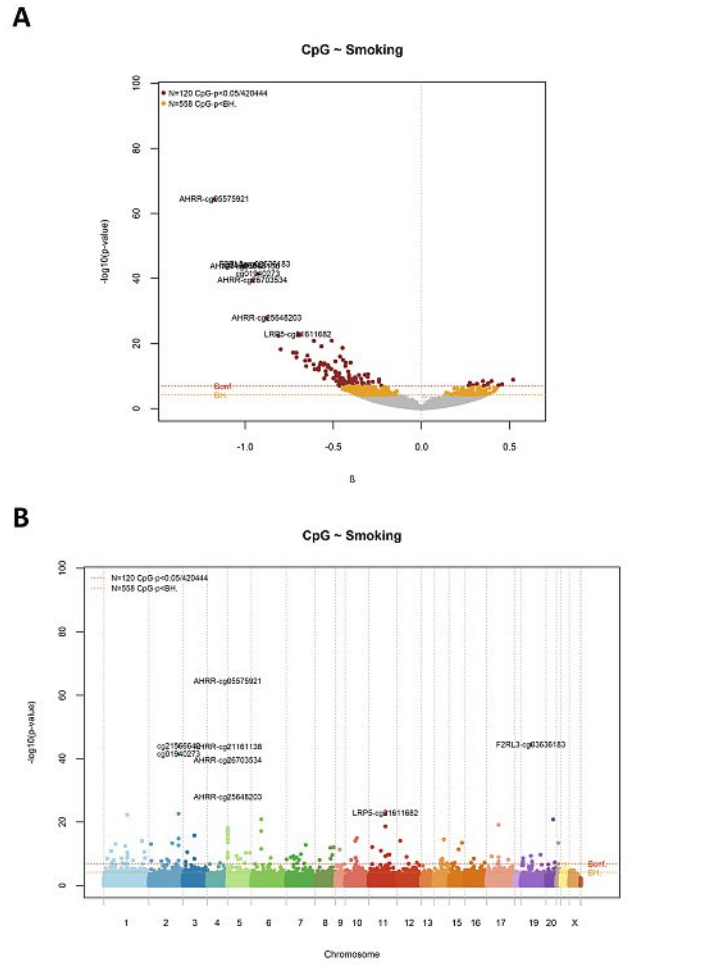


Fig. 3. Epigenome-Wide Association Study results for current smoking represented using a Volcano plot (A) and Manhattan plot (B) (SKIPOGH 2 – EWAS). Linear regression model for the association between CpG markers and phenotypes of interest, adjusting for age, sex, recruitment center, seasonality of blood sampling, and familial structure (random-effect covariable).

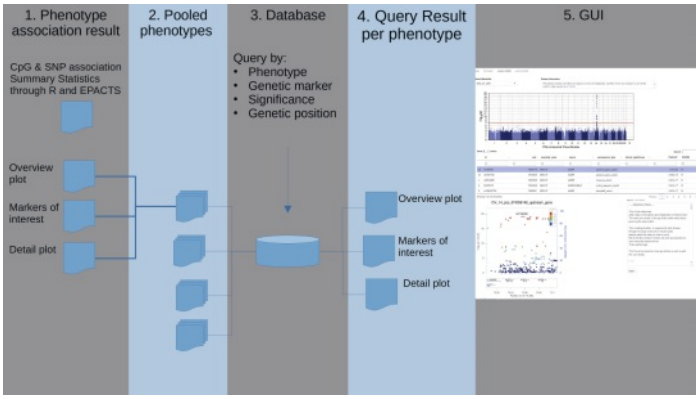


Fig. 4. PhenoExplorer development steps and data organization for GWAS/EWAS results. (1) GWAS/EWAS analyses running on local machines (R, EPACTS); (2) Database organization, structuring and pooling; (3) Query description; (4) Description of Query results; (5) Graphical User Interface presentation.

between (epi)genetic variants and given phenotypes, eventually leading to spurious associations. Disentangling these effects would thus require repeating GWAS/EWAS analyses by including additional covariates, depending on the choice of phenotypes and the research question. Third, unlike genetic variants, epigenetic modifications are tissue-specific and influenced by multiple factors, such as genetic makeup, aging, and environmental exposures, and constitute a dynamic process over the life-course.[55,56] Consequently, the present EWAS analyses preclude directly establishing a cause-to-effect relation between CpG markers and included phenotypes, and may be affected by time discrepancies between the collection periods of the phenotype on one hand, and the epigenetic profile on the other hand (*i.e.* metabolites/metals were measured at SKIPOGH 1 while CpG data was measured at SKIPOGH 2). Fourth, although multi GWAS and EWAS analyses tend to be appointed as *agnostic* or *hypotheses-free*, they actually imply hidden hypotheses that need to be accounted for while conducting OMICS studies.[2,3] In a previous research by Barker,[3] the authors have identified three hidden hypotheses underlying EWAS studies: (1) insufficient EWAS coverage: implying that not *all* epigenetic markers are included in available EWAS arrays; (2) biological relevance of the tissue in which epigenetic markers are measured: whereby the tissues in which epigenetic measures are performed may not be relevant for the assessed phenotypes (*i.e.* DNA methylation measured in blood and tested in relation to complex neurological disease); (3) biological relevance for the phenotype of interest: postulating that the biological nature of diseases and most phenotypes is often complex, with *individual* epigenetic modifications generally playing a modest role in the global understanding of the trait or the pathophysiological process of interest.

4.2 *Conclusion*

In conclusion, this work provides a comprehensive outline of GWAS and EWAS-based associations performed in the context of a population-based study conducted in adults of European ancestry living in Switzerland, including a large number of physiological, clinical, and lifestyle-related phenotypes. Further, by providing the obtained results through the development of the PhenoExplorer web-based platform, we aim to contribute to the overall understanding of the role of genetic and epigenetic variants in the regulation of multiple phenotypes, assessed in a population-based observational setting. We believe that the use of PhenoExplorer as part of the TransCure project shall allow highlighting potential associations of interest between markers located within or in the vicinity of TransCure genes/transporters of interest and phenotypes sampled in the SKIPOGH population-based cohort. We show here a practical example of how publicly funded results can be made easily findable, accessible and usable for future projects by the researchers with interest in the phenotypes available, in particular kidney-related phenotypes.

Conflicts of interest

None to declare.

Acknowledgements

This work was supported as a part of NCCR TransCure, a National Centre of Competence in Research, funded by the Swiss National Science Foundation (grant number 185544).

Supplementary Information

Supplementary Information for this publication is available on https://chimia.ch/chimia/article/view/2022_1052.

Received: August 2, 2022

[1] J. Hardy, A. Singleton, *N. Engl. J. Med.* **2009**, *360*, 1759, https://doi.org/10.1056/NEJMra0808700.
[2] G. D. Kitsios, E. Zintzaras, *Transl. Res.* **2009**, *154*, 161, https://doi.org/10.1016/j.trsl.2009.07.001.
[3] E. D. Barker, S. Roberts, E. Walton, *Curr. Opin. Psychol.* **2019**, *27*, 13, https://doi.org/10.1016/j.copsyc.2018.07.009.

Table 3. Summary of intragenic CpGs (NCCR TransCure genes) associated with EWAS-included phenotypes (SKIPOGH 2)

CpG	β [a]	SE [a]	*P* [a,b]	Gene	Intragenic region	Chromosome	Position	Strand	Phenotype	Transcure project
cg04954559	9.49E-04	1.69E-04	3.23E-08	SLC17A6	Body	chr11	22365428	-	Basal metabolism (kcal/day)	1 Astrocytes AVolterra
cg21627181	-9.42E-01	6.46E-02	1.49E-37	SLC17A4	TSS1500	chr6	25754190	+	Eosinophils	1 Astrocytes AVolterra
cg21627181	6.86E-02	8.20E-03	1.83E-15	SLC17A4	TSS1500	chr6	25754190	+	Neutrophils	1 Astrocytes AVolterra
cg21627181	-1.50E+00	1.22E-01	1.30E-28	SLC17A4	TSS1500	chr6	25754190	+	Abs. n. eosinophils	1 Astrocytes AVolterra
cg04675306	4.46E-01	7.29E-02	1.61E-09	SLC9A6	TSS200-TSS200	chrX	135067450	-	Bioimpedance - Body water (L)	2 NHE DFuster
cg01520853	-5.66E-01	6.79E-02	2.03E-15	SLC11A1	Body	chr2	219259478	+	Eosinophils	3 Iron DMT FPN RDutzler
cg18854666	2.90E-01	5.33E-02	1.08E-07	SLC11A1	1stExon-5'UTR	chr2	219247055	+	Eosinophils	3 Iron DMT FPN RDutzler
cg01520853	4.67E-02	7.78E-03	5.24E-09	SLC11A1	Body	chr2	219259478	+	Neutrophils	3 Iron DMT FPN RDutzler
cg01520853	-8.44E-01	1.25E-01	6.62E-11	SLC11A1	Body	chr2	219259478	+	Abs. n. eosinophils	3 Iron DMT FPN RDutzler
cg06500161	-1.48E+00	2.71E-01	6.16E-08	ABCG1	Body	chr21	43656587	+	Blood HDL chol. (mmol/L)	4 Multidrug ABCG2 KLocher
cg06500161	8.98E-01	1.41E-01	3.78E-10	ABCG1	Body	chr21	43656587	+	Blood triglyceride (mmol/L)	4 Multidrug ABCG2 KLocher
cg06500161	5.72E-01	1.02E-01	3.39E-08	ABCG1	Body	chr21	43656587	+	Blood insulin (mU/L)	4 Multidrug ABCG2 KLocher
cg21827955	-5.82E-01	8.72E-02	1.07E-10	ABCG5	Body	chr2	44062010	-	Eosinophils	4 Multidrug ABCG2 KLocher
cg01176028	3.90E-01	5.35E-02	2.27E-12	ABCG1	Body	chr21	43653234	+	Eosinophils	4 Multidrug ABCG2 KLocher
cg01881899	6.31E-01	6.72E-02	1.07E-18	ABCG1	Body	chr21	43652704	-	Eosinophils	4 Multidrug ABCG2 KLocher
cg01176028	-3.42E-02	6.04E-03	3.25E-08	ABCG1	Body	chr21	43653234	+	Neutrophils	4 Multidrug ABCG2 KLocher
cg21827955	-9.42E-01	1.57E-01	5.60E-09	ABCG5	Body	chr2	44062010	-	Abs. n. eosinophils	4 Multidrug ABCG2 KLocher
cg01176028	6.30E-01	9.68E-02	2.81E-10	ABCG1	Body	chr21	43653234	+	Abs. n. eosinophils	4 Multidrug ABCG2 KLocher
cg01881899	1.11E+00	1.20E-01	3.50E-18	ABCG1	Body	chr21	43652704	-	Abs. n. eosinophils	4 Multidrug ABCG2 KLocher
cg04748546	-5.28E-01	6.83E-02	1.28E-13	TRPM4	Body	chr19	49696260	-	Eosinophils	6 Cation TRPM4 HAbriel
cg04748546	-8.08E-01	1.24E-01	2.96E-10	TRPM4	Body	chr19	49696260	-	Abs. n. eosinophils	6 Cation TRPM4 HAbriel
cg06690548	-1.03E+00	1.67E-01	1.44E-09	SLC7A11	Body	chr4	139162808	-	Blood triglyceride (mmol/L)	7 AA SLC7 DFotiadis
cg06690548	-7.32E-01	1.10E-01	5.51E-11	SLC7A11	Body	chr4	139162808	-	Blood GGT (U/L)	7 AA SLC7 DFotiadis
cg19928703	-9.20E-01	6.27E-02	7.14E-38	SLC7A1	5'UTR	chr13	30143971	-	Eosinophils	7 AA SLC7 DFotiadis
cg02999224	2.55E-01	3.96E-02	4.33E-10	SLC7A7	1stExon-5'UTR	chr14	23284559	-	Eosinophils	7 AA SLC7 DFotiadis
cg19928703	6.56E-02	8.06E-03	8.65E-15	SLC7A1	5'UTR	chr13	30143971	-	Neutrophils	7 AA SLC7 DFotiadis
cg06690548	-4.87E-05	8.68E-06	3.07E-08	SLC7A11	Body	chr4	139162808	-	Blood MCH (pg)	7 AA SLC7 DFotiadis
cg19928703	-1.40E+00	1.21E-01	4.98E-26	SLC7A1	5'UTR	chr13	30143971	-	Abs. n. eosinophils	7 AA SLC7 DFotiadis
cg09194755	-5.44E+00	9.13E-01	4.41E-09	SLC7A6	5'UTR-5'UTR	chr16	68298776	-	Urinary magnesium day excretion (mmol/h)	7 AA SLC7 DFotiadis
cg06690548	-2.91E-01	5.21E-02	3.48E-08	SLC7A11	Body	chr4	139162808	-	Allostatic load score	7 AA SLC7 DFotiadis

a Linear regression model for the association between CpG markers and phenotypes of interest, adjusting for age, sex, recruitment center, seasonality of blood sampling, and familial structure (random-effect covariable)
b Significance threshold was set using the Bonferroni correction method ($P < 0.05/420'444$)

[4] NIH, Genome-Wide Association Studies Fact Sheet, https://www.genome.gov/about-genomics/fact-sheets/Genome-Wide-Association-Studies-Fact-Sheet, **2020**.

[5] C. A. Böger, M. Gorski, M. Li, M. M. Hoffmann, C. Huang, Q. Yang, A. Teumer, V. Krane, C. M. O'Seaghdha, Z. Kutalik, H.-E. Wichmann, T. Haak, E. Boes, S. Coassin, J. Coresh, B. Kollerits, M. Haun, B. Paulweber, A. Köttigen, G. Li, M. G. Shlipak, N. Powe, S.-J. Hwang, A. Dehghan, F. Rivandeneira, A. Uitterlinden, A. Hofman, J. S. Beckmann, B. K. Krämer, J. Witteman, M. Bochud, D. Siscovick, R. Rettig, F. Kronenberg, C. Wanner, R. I. Thadhani, I. M. Heid, C. S. Fox, W. H. Kao, The CKDGen Consortium, *PLoS Genetics* **2011**, *7,* e1002292 , https://doi.org/10.1371/journal.pgen.1002292.

[6] Z. Chen, H. Schunkert, *J. Intern. Med.* **2021**, *290*, 980, https://doi.org/10.1111/joim.13362.

[7] N. Mars, S. Kerminen, Y.-C. A. Feng, M. Kanai, K. Läll, L. F. Thomas, A. H. Skogholt, P. della Briotta Parolo, The Biobank, Japan Project, FinnGen, B. M. Neale, J. W: Smoller, M. E: Gabrielsen, K. Hveem, R. Mägi, K. Matsuda, Y. Okada, M. Pirinen, A. Palotie, A. Ganna, A. R. Martin, S. Ripatti, *Cell Genomics* **2022**, *2*, 100118, https://doi.org/10.1016/j.xgen.2022.100118.

[8] O. Weissbrod, M. Kanai, H. Shi, S. Gazal, W. J. Peyrot, A. V. Khera, Y. Okada, The Biobank Japan Project, A. R: Martin, H. K. Finucane, A. L. Price, *Nat. Gen.* **2022**, *54*, 450, https://doi.org/10.1038/s41588-022-01036-9.

[9] S.-W. Choi, S. Friso, *Adv. Nutrition* **2010**, *1*, 8, https://doi.org/10.3945/an.110.1004.

[10] A. Bird, *Genes Devel.* **2002**, *16*, 6, https://doi.org/10.1101/gad.947102.

[11] N. Mostafavi, R. Vermeulen, A. Ghantous, G. Hoek, N. Probst-Hensch, Z. Herceg, S. Tarallo, A. Naccarati, J. C. S. Kleinjans, M. Imboden, A. Jeong, D. Morley, A. F. S. Amaral, E. van Nunen, J. Gulliver, M. Chadeau-Hyam, P. Vineis, J. Vlaanderen, *Environ. Int.* **2018**, *120*, 11, https://doi.org/10.1016/j.envint.2018.07.026.

[12] D. Petrovic, B. Bodinier, S. Dagnino, M. Whitaker, M. Karimi, G. Camanella, T. Haugdahl Nost, S. Polidoro, D. Palli, V. Krogh, R. Tumino, C. Sacerdote, S. Panico, E. Lund, P.-A. Dugué, G. G. Giles, G. Severi, M. Southey, P. Vineis, S. Stringhini, M. Bochud, T. M. Sandanger, R. C. H. Vermeulen, F. Guida, M. Chadeau-Hyam, *Eur. J. Epidemiol.* **2022**, *37*, 629, https://doi.org/10.1007/s10654-022-00877-2.

[13] J. A. McKay, J. C. Mathers, *Acta Physiologica* **2011**, *202*, 103, https://doi.org/10.1111/j.1748-1716.2011.02278.x.

[14] P. Schlosser, A. Tin, P. R. Matias-Garcia, C. H. L. Thio, R. Joehanes, H. Liu, A. Weihs, Z. Yu, A. Hoppmann, F. Grundner-Culemann, J. L. Min, A. A. Adeyemo, C. Agyemang, J. Ärnlöv, N. A. Aziz, A. Baccarelli, M. Bochud, H. Brenner, M. M. B. Breteler, C. Carmeli, L. Chaker, J. C. Chambers, S. A. Cole, J. Coresh, T. Corre, A. Correa, S. R. Cox, N. de Klein, G. E. Delgado, A. Domingo-Relloso, K.-U. Eckardt, A. B. Ekici, K. Endlich, K. L. Evans, J. S. Floyd, M. Fornage, L. Franke, E. Fraszczyk, X. Gao, X. Gào, M. Ghanbari, S. Ghasemi, C. Gieger, P. Greenland, M. L. Grove, S. E. Harris, G. Hemani, P. Henneman, C. Herder, S. Horvath, L. Hou, M. A. Hurme, S.-J. Hwang, M.-R. Jarvelin, S. L. R. Kardia, S. Kasela, M. E. Kleber, W. Koenig, J. S. Kooner, H. Kramer, F. Kronenberg, B. Kühnel, T. Lehtimäki, L. Lind, D. Liu, Y. Liu, D. M. Lloyd-Jones, K. Lohman, S. Lorkowski, A. T. Lu, R. E. Marioni, W. März, D. L. McCartney, K. A. C. Meeks, L. Milani, P. P. Mishra, M. Nauck, A. Navas-Acien, C. Nowak, A. Peters, H. Prokisch, B. M. Psaty, O. T. Raitakari, S. M. Ratliff, A. P. Reiner, S. E. Rosas, B. Schöttker, J. Schwartz, S. Sedaghat, J. A. Smith, N. Sotoodehnia, H. R. Stocker, S. Stringhini, J. Sundström, B. R. Swenson, M. Tellez-Plaza, J. B. J. van Meurs, J. V. van Vliet-Ostaptchouk, A. Venema, N. Verweij, R. M. Walker, M. Wielscher, J. Winkelmann, B. H. R. Wolffenbuttel, W. Zhao, Y. Zheng, Estonian Biobank Research Team, Genetics of DNA Methylation Consortium, M. Loh, H. Snieder, D. Levy, M. Waldenberger, K. Susztak, A. Köttgen, A. Teumer, *Nat. Commun.* **2021**, *12*, 1, https://doi.org/10.1038/s41467-021-27234-3.

[15] J. L. Min, G. Hemani, E. Hannon, K. F. Dekkers, J. Castillo-Fernandez, R. Luijk, E. Carnero-Montoro, D. J. Lawson, K. Burrows, M. Suderman, A. D. Bretherick, T. G. Richardson, J. Klughammer, V. Iotchkova, G. Sharp, A. Al Khleifat, A. Shatunov, A. Iacoangeli, W. L. McArdle, K. M. Ho, A. Kumar, C. Söderhäll, C. Soriano-Tárraga, E. Giralt-Steinhauer, N. Kazmi, D. Mason, A. F. McRae, D. L. Corcoran, K. Sugden, S. Kasela, A. Cardona, F. R. Day, G. Cugliari, C. Viberti, S. Guarrera, M. Lerro, R. Gupta, S. Bollepalli, P. Mandaviya, Y. Zeng, T.-Kim Clarke, R. M. Walker, V. Schmoll, D. Czamara, C. Ruiz-Arenas, F. I. Rezwan, R. E. Marioni, T. Lin, Y. Awaloff, M. Germain, D. Aïssi, R. Zwamborn, K. van Eijk, A. Dekker, J. van Dongen, J.-J. Hottenga, G. Willemsen, C.-J. Xu, G. Barturen, F. Català-Moll, M. Kerick, C. Wang, P. Melton, H. R. Elliott, J. Shin, M. Bernard, I. Yet, M. Smart, T. Gorrie-Stone, BIOS Consortium, C. Shaw, A. Al Chalabi, S. M. Ring, G. Pershagen, E. Melén, J. Jiménez-Conde, J. Roquer, D. A. Lawlor, J. Wright, N. G. Martin, G. W. Montgomery, T. E. Moffitt, R. Poulton, T. Esko, L. Milani, A. Metspalu, J. R. B. Perry, K. K. Ong, N. J. Wareham, G. Matullo, C. Sacerdote, S. Panico, A. Caspi, L. Arseneault, F. Gagnon, M. Ollikainen, J. Kaprio, J. F. Felix, F. Rivadeneira, H. Tiemeier, M. H. van IJzendoorn, A. G. Uitterlinden, V. W. V. Jaddoe, C. Haley, A. M. McIntosh, K. L. Evans, A. Murray, K. Räikkönen, J. Lahti, E. A. Nohr, T. I. A. Sørensen, T. Hansen, C. S. Morgen, E. B. Binder, S. Lucae, J. R. Gonzalez, M. Bustamante, J. Sunyer, J. W. Holloway, W. Karmaus, H. Zhang, I. J. Deary, N. R. Wray, J. M. Starr, M. Beekman, D. van Heemst, P. E. Slagboom, P.-E. Morange, D.-A. Trégouët, J. H. Veldink, G. E. Davies, E. J. C. de Geus, D. I. Boomsma, J. M. Vonk, B. Brunekreef, G. H. Koppelman, M. E. Alarcón-Riquelme, R.-C. Huang, C. E. Pennell, J. van Meurs, M. A. Ikram, A. D. Hughes, T. Tillin, N. Chaturvedi, Z. Pausova, T. Paus, T. D. Spector, M. Kumari, L. C. Schalkwyk, P. M. Visscher, G. D. Smith, C. Bock, T. R. Gaunt, J. T. Bell, B. T. Heijmans, J. Mill, C. L. Relton, *Nat. Gen.* **2021**, *53*, 1311, https://doi.org/10.1038/s41588-021-00923-x.

[16] V Rousson, D. Ackermann, B. Ponte, M. Pruijm, I. Guessous, C. H. d'Uscio, G. Ehret, G. Escher, A. Pechère-Bertschi, M. Groessl, P.-Y. Martin, M. Burnier, B. Dick, M. Bochud, B. Vogt, N. A. Dhayat, *PloS one* **2021**, *16*, e0253975, https://doi.org/10.1371/journal.pone.0253975.

[17] M. Pruijm, B. Ponte, D. Ackermann, F. Paccaud, I. Guessous, G. Ehret, A. Pechère-Bertschi, B. Vogt, M. G. Mohaupt, P.-Y. Martin, S. C. Youhanna, N. Nägele, P. Vollenweider, G. Waeber, M. Burnier, O. Devuyst, M. Bochud, *Clin. J. Am. Soc. Nephrol.* **2016**, *11*, 70, https://doi.org/10.2215/CJN.04230415.

[18] H. Alwan, M. Pruijm, B. Ponte, D. Ackermann, I. Guessous, G. Ehret, J. A. Staesson, K. Asayama, P. Vuistiner, S. Estoppey Younes, F. accaud, G. Wuerzner, A. Pechère-Bertschi, M. Mohaupt, B. Vogt, P.-Y. Martin, M. Burnier, M. Bochud, *PloS one* **2014**, *9*, e92522, https://doi.org/10.1371/journal.pone.0092522.

[19] Z.-Y. Zhang, C. Carmeli, B. Ponte, M. Purijm, D. Ackermann, G. Ehret, I. Guessous, D. Petrovic, A. Pechère-Bertschi, B. Vogt, P.-Y. Martin, M. Burnier, S. Lenglet, M. Augsberger, A. Thomas, M. Bochud, *Hypertension* **2020**, *75*, 1133, https://doi.org/10.1161/HYPERTENSIONAHA.119.13649.

[20] N. A. Dhayat, M. Purijm, B. Ponte, D. Ackermann, A. B. Leichle, O. Devuyst, G. Ehret, I. Guessous, A. Pechère-Bertschi, J. Pastor, P.-Y. Martin, M. Burnier, G.-M. Fiedler, B. Vogt, O. W. Moe, M. Bochud, D. G. Fuster, *J. Clin. Endocrinol. Metabol.* **2020**, *105*, e1135, https://doi.org/10.1210/clinem/dgz232.

[21] B. Ponte, M. Pruijm, D. Ackermann, G. Ehret, N. Ansermot, J. A. Staessen, B. Vogt, A. Pechère-Bertschi, M. Burnier, P.-Y. Martin, C. B. Eap, M. Bochud, I. Guessous, *Mayo Clinic Proc.* **2018,** *93*, 586, https://doi.org/10.1016/j.mayocp.2017.12.010.

[22] M. Bochud, J. Jenny-Burri, M. Purijm, B. Ponte, I. Guessous, G. Ehret, D. Petrovic, V. Dudler, M. Haldimann, G. Escher, B. Dick, M. Mohaupt, F. Paccaud, M. Burnier, A. Péchère-Bertschi, P.-Y. Martin, B. Vogt, D. Ackermann, *J. Clin. Endocrinol. Metabol.* **2018**, *103*, 748, https://doi.org/10.1210/jc.2017-01540.

[23] D. Petrovic, E. Pivin, B. Ponte, N. Dhayar, M. Purijm, G. Ehret, D. Ackermann, I. Guessous, S. Estoppey Younes, A. Pechère-Bertschi, B. Vogt, M. Mohaupt, P.-Y. Martin, F. Paccaud, M. Burnier, M. Bochud, S. Stringhini, *Psychoneuroendocrinology* **2016**, *67*, 76, https://doi.org/10.1016/j.psyneuen.2016.02.003.

[24] D. Petrovic, M. Pruijm, B. Ponte, N. A. Dhayat, D. Ackermann, G. Ehret, N. Ansermot, B. Vogt, P.-Y. Martin, S. Stringhini, S. Estoppey-Younès, L. Thijs, Z. Zhang, J. D. Melgarejo, C. B. Eap, J. A. Staesson, M. Bochud, I. Guessous, C. B. Eap, M. Bochud, I. Guessous, *Mayo Clinic Proc.* **2021,** *96*, 3071 https://doi.org/10.1016/j.mayocp.2021.05.030.

[25] D. Petrovic, S. Estoppey Younes, M. Purijm, B. Ponte, D. Ackermann, G. Ehret, N. Nasermot, M. Mohaupt, F. Paccaud, B. Vogt, A. Pechère-Bertschi, P.-Y. Martin, M. Burnier, C. B. Eap, M. Bochud, I. Guessous, *Nutrition & Metabolism* **2016**, *13*, 1, https://doi.org/10.1186/s12986-016-0144-4.

[26] Maelstrom-Research, Swiss Kidney Project on Genes in Hypertension, https://www.maelstrom-research.org/study/skipogh, **2022**.

[27] N.-T. Ha, S. Freytag, H. Bickeboeller, *Eur. J. Human Gen.* **2014**, *2*, 1124, https://doi.org/10.1038/ejhg.2013.304.

[28] H. Hor, Z. Kutalik, Y. Dauvilliers, A. Valsesia, G. J. Lammers, C. E. H. M. Donjacour, A. Iranzo, J. Santamaria, R. Peraita Adrados, J. L. Vicario, S. Overeem, I. Arnulf, I. Theodorou, P. Jennum, S. Knudsen, C. Bassetti, J. Mathis, M. Lecendreux, G. Mayer, P. Geisler, A. Beneto, B. Petit, C. Pfister, J. Vienne Bürki, G. Didelot, M. Billiard, G. Ercilla, W. Verduijn, F. H. J. Claas, P. Vollenweider, G. Waeber, D. M. Waterworth, V. Mooser, R. Heinzer, J. S. Beckmann, S. Bergmann, M. Tafti, *Nat. Gen.* **2010**, *42*, 786, https://doi.org/10.1038/ng.647.

[29] E. Reed, S. Nunez, D. Kulp, J. Qian, M. P. Reilly, A. S. Foulkes, *Statistics in Medicine* **2015**, *34*, 3769, https://doi.org/10.1002/sim.6605.

[30] The Haplotype Reference Consortium, *Nat. Gen.* **2016**, *48*, 1279, https://doi.org/10.1038/ng.3643.

[31] C. Blauwendraat, F. Faghri, L. Pihlstrom, J. T. Geiger, A. Elbaz, S. Lesage, J.-C. Corvol, P. May, A. Nicolas, Y. Abramzon, N. A. Murphy, R. Gibbs, M. Ryten, R. Ferrari, J. Bras, R. Guerreiro, J. Williams, R. Sims, S. Lubbe, D. G. Hernandez, K. Y. Mok, L. Robak, R. H. Campbell, E. Rogaeva, B. J. Traynor, R. Chia, S. Ju Chung, International Parkinson's Disease Genomics Consortium (IPDGC), COURAGE-PD Consortium, J. A. Hardy, A. Brice, N. W. Wood, H. Houlden, J. M. Shulman, H. R. Morris, T. Gasser, R. Krüger, P. Heutink, M. Sharma, J. Simón-Sánchez,

M. A.Nalls, A. B. Singleton, S. W. Scholz, *Neurobiol. Aging* **2017**, *57*, 247, https://doi.org/10.1016/j.neurobiolaging.2017.05.009.
[32] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, *Am. J. Human Gen*. **2007**, *81*, 559, https://doi.org/10.1086/519795.
[33] S: Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, C. Fichsberger, *Nat. Gen*. **2016**, *48*, 1284, https://doi.org/10.1038/ng.3656.
[34] R. Pidsley, E. Zotenko, T. J. Peters, M. G. Lawrence, G. R. Risbridger, P. Molloy, S. Van Djik, B. Muhlhausler, C. Stirzaker, S. J. Clark, *Genome Biol*. **2016**, *17*, 1, https://doi.org/10.1186/s13059-016-1066-1.
[35] T. Hastie, R. Tibshirani, G. Sherlock, M. Eisen, P. Brown, D. Botstein, 'Imputing missing data for gene expression arrays', Technical Report, **1999**, https://hastie.su.domains/Papers/missing.pdf.
[36] B. Lehne, A. W. Drong, M. Loh, W. Zhang, W. R. Scott, S.-T. Tan, U. Afzal, J. Scott, M.-R. Jarvelin, P. Elliot, M. I. McCarthy, J. S. Kooner, J. C. Chambers, *Genome Biol*. **2015**, *16*, 1, https://doi.org/10.1186/s13059-015-0600-x.
[37] G. Magliocco, J. Desmeules, A. Matthey, L. M. Quiros-Guerrero, N. Barapour, T. Joye, L. Marcourt, E. F. Queiroz, J.-L. Wolfender, Y. Gloor, A. Thomas, Y. Daali, *Brit. J. Pharmacol*. **2021**, *178*, 4708, https://doi.org/10.1111/bph.15651.
[38] N. Bararpour, F. Gilardi, C. Carmeli, J. Sidibe, J. Ivanisevic, T. Caputo, M. Augsburger, S. Grabherr, B. Desvergne, N. Guex, M. Bochud, A. Thomas, *bioRxiv* **2020**, https://doi.org/10.1101/2020.01.22.914051.
[39] E: Lauer, M. Villa, M. Jotterand, R. Vilarino, M. Bollmann, K. Michaud, S. Grabherr, M. Augsberger, A. Thomas, *Int. J. Legal Med*. **2017**, *131*, 497, https://doi.org/10.1007/s00414-016-1414-4.
[40] N.J.Cox,*StataJ*.**2005**,*5*,259,https://doi.org/10.1177/1536867X0500500210.
[41] A. Kassambara, https://www.datanovia.com/en/lessons/transform-data-to-normal-distribution-in-r, **2021**.
[42] E. Tessitore, K. Dobretz, N. A. Dhayat, I. Kern, B. Ponte, M. Purijm, D. Ackermann, S. Estoppey, M. Burnier, P.-Y. Martin, B. Vogt, N. Vuilleumier, M. Bochud, F. Mach, G. Ehret, *Eur. J. Clin. Invest*. **2022**, *52*, e13699, https://doi.org/10.1111/eci.13699.
[43] N. Risch, K. Merikangas, *Science* **1996**, *273*, 1516, https://doi.org/10.1126/science.273.5281.1516.
[44] H. M. Kang, EPACTS (efficient and parallelizable association container toolbox). EPACTS: efficient and parallelizable association container toolbox, **2014**.
[45] F. Cunningham, J. E Allen, J. Allen, J. Alvarez-Jarreta, M. R. Amode, I. M. Armean, O. Austine-Orimoloye, A. G. Azov, I. Barnes, R. Bennett, A. Berry, J. Bhai, A. Bignell, K. Billis, S. Boddu, L. Brooks, M. Charkhchi, C. Cummins, L. Da Rin Fioretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. Garcia Giron, T. Genez, J. Gonzalez Martinez, C. Guijarro-Clarke, A. Gymer, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, T. Juettemann, V. Kaikala, M. Kay, I. Lavidas, T. Le, D. Lemos, J. C. Marugán, S. Mohanan, A. Mushtaq, M. Naven, D. N. Ogeh, A. Parker, A. Parton, M. Perry, I. Pilizota, I. Prosovetskaia, M. Pandian Sakthivel, A. I. A. Salam, B. M. Schmitt, H. Schuilenburg, D. Sheppard, J. G. Pérez-Silva, W. Stark, E. Steed, K. Sutinen, R. Sukumaran, D. Sumathipala, M.-M. Suner, M. Szpak, A. Thormann, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, B. Walts, N. Willhoft, A. Winterbottom, E. Wass, M. Chakiachvili, B. Flint, A. Frankish, S. Giorgetti, L. Haggerty, S. E. Hunt, G. R. IIsley, J. E. Loveland, F. J. Martin, B. Moore, J. M. Mudge, M. Muffato, E. Perry, M. Ruffier, J. Tate, D. Thybert, S. J. Trevanion, S. Dyer, P. W. Harrison, K. L. Howe, A. D. Yates, D. R. Zerbino, P. Flicek, *Nucleic Acids Res*. **2021**, *50*, D988, https://doi.org/10.1093/nar/gkab1049.
[46] M. Chadeau-Hyam, G. Campanella, T. Jombart, L. Bottolo, L. Portengen, P. Vineis, B. Liquet, R. C. H. Vermeulen, *Environ. Mol. Mutagenesis* **2013**, *54*, 542, https://doi.org/10.1002/em.21797.
[47] E. A. Houseman, W. P. Accoamndo, D. C. Koestler, B. C. Christensen, C. J. Marsit, H. H. Nelson, J. K. Wiencke, K. T. Kelsey, *BMC Bioinform*. **2012**, *13*, 1, https://doi.org/10.1186/1471-2105-13-86.
[48] R. J. Pruim, R. P. Welch, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, C. J. Willer, *Bioinformatics* **2010**, *26*, 2336, https://doi.org/10.1093/bioinformatics/btq419.
[49] A. J. Cox, M. C.-Y. Ng, J. Xu, C. D. Langefeld, K. L. Koch, P. A. Dawson, J. J. Carr, B. I. Freedman, F.-C. Hsu, D. W. Bowden, *Atherosclerosis* **2013**, *229*, 155, https://doi.org/10.1016/j.atherosclerosis.2013.04.008.
[50] S. Luo, A. Surapaneni, Z. Zheng, E. P. Rhee, J. Coresh, A. M. Hung, G. N. Nadkami, B. Yu, E. Boerwinkle, A. Tin, D. E. Arking, I. Steinbrenner, P. Schlosser, A, Köttgen, M. E. Grams, *Clin. J. Am. Soc. Nephrol*. **2021**, *16*, 37, https://doi.org/10.2215/CJN.08600520.
[51] G. H. Bruun, T. K. Doktor, B. S. Andresen, *Mol. Gen. Metabol*. **2013**, *110*, 122, https://doi.org/10.1016/j.ymgme.2013.06.005.
[52] T. Lyngdoh, M. Bochud, J. Glaus, E. Castelao, G. Waeber, P. Vollenweider, M. Preisig, *PloS one* **2013**, *8*, e76336, https://doi.org/10.1371/journal.pone.0076336.
[53] M. Kolz, T. Johnson, S. Sanna, A. Teumer, V. Vitart, M. Perola, M. Mangino, E. Albrecht, C. Wallace, M. Farrall, A. Johansson, D. R. Nyholt, Y. Aulchenko, J. S. Beckmann, S. Bergmann, M. Bochud, M. Brown, H. Campbell, for the EUROSPAN Consortium, J. Connell, A. Dominiczak, G. Homuth, C. Lamina, M. I. McCarthy, for the ENGAGE Consortium, T. Meitinger, V. Mooser, P. Munroe, M. Nauck, J. Peden, H. Prokisch, P. Salo, V. Salomaa, N. J. Samani, D. Schlessinger, M. Uda, U. Völker, G. Waeber, D. Waterworth, R. Wang-Sattler, A. F. Wright, J. Adamski, J. B. Whitfield, U. Gyllensten, J. F. Wilson, I. Rudan, P. Pramstaller, H. Watkins, for the PROCARDIS Consortium, A. Doering, H.-E. Wichmann, for the KORA Study, T. D. Spector, L. Peltonen, H. Völzke, R. Nagaraja, P. Vollenweider, M. Caulfield, for the WTCCC, T. Illig, C. Gieger, *PLoS Genetics* **2009**, *5*, e1000504, https://doi.org/10.1371/journal.pgen.1000504.
[54] T. Huan, R. Joehanes, C. Schurmann, K. Schramm, L. C. Pilling, M. J. Peters, R. Mägi, D. DeMeo, G. T. O'Connor, L. Ferrucci, A. Teumer, G. Homuth, R. Biffar, U. Völker, C. Herder, M. Waldenberger, A. Peters, S. Zeilinger, A. Metspalu, A. Hofman, A. G. Uitterlinden, D. G. Hernandez, A. B. Singleton, S. Bandinelli, P. J. Munson, H. Lin, E. J. Benjamin, T. Esko, H. J. Grabe, H. Prokisch, J. B. J. van Meurs, D. Melzer, D. Levy, *Human Mol. Genetics* **2016**, *25*, 4611, https://doi.org/10.1093/hmg/ddw288.
[55] F. Guida, T. M. Sandanger, R. Castagné, G. Campanella, S. Polidoro, D. Palli, V. Krogh, R. Tumino, C. Sacerdote, S. Panico, G. Severi, S. A. Krytopoulos, P. Georgiadis, R. C. H. Vermeulen, E. Lund, P. Vineis, M. Chadeau-Hyam, *Human Mol. Genetics* **2015**, *24*, 2349, https://doi.org/10.1093/hmg/ddu751.
[56] Y. Bergman, H. Cedar, *Nat. Struct. Mol. Biol.* **2013**, *20*, 274, https://doi.org/10.1038/nsmb.2518.

License and Terms

The definitive version of this article is the electronic one that can be found at https://doi.org/10.2533/chimia.2022.1052