Common genetic variants associated with urinary phthalate levels in children: a genome-wide study

Mariona Bustamante [a,b,c,*±], Laura Balagué [a *], Zsanett Buko[d], Amrit Kaur Sakhi[e], Maribel Casas[a,b,c], Lea Maitre[a,b,c], Sandra Andrusaityte[f], Regina Grazuleviciene[f], Kristine B Gützkow[e], Anne Lise Brantsæter[e], Barbara Heude[g], Claire Philippat[h], Leda Chatzi[i], Marina Vafeiadi [j], Tiffany C Yang[k], John Wright[k], Amy Hough[k], Carlos Ruiz-Arenas[m], Ramil Nurtdinov[n], Geòrgia Escaramís[l,c], Juan Ramon Gonzalez[a,b,c], Cathrine Thomsen[e], Martine Vrijheid[a,b,c]

[a]Environment and Health over the Lifecourse, ISGlobal, Barcelona, Spain
[b]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[c]CIBER Epidemiología y Salud Pública (CIBERESP), Madrid, Spain
[d]Department of Oncological Science, Huntsman Cancer Institute, Salt Lake City, United States of America
[e]Division of Climate and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway
[f]Department of Environmental Science, Vytautas Magnus University, Kaunas, Lithuania
[g]Université Paris Cité and Université Sorbonne Paris Nord, Inserm, INRAE, Center for Research in Epidemiology and StatisticS (CRESS), F-75004 Paris, France
[h]University Grenoble Alpes, Inserm U-1209, CNRS-UMR-5309, Environmental Epidemiology Applied to Reproduction and Respiratory Health Team, Institute for Advanced Biosciences, 38000, Grenoble, France
[i]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA
[j]Department of Social Medicine, Faculty of Medicine, University of Crete, Heraklion, Greece
[k]Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK
[l]Departament de Biomedicina, Institut de Neurociències, Universitat de Barcelona (UB), Barcelona, Spain
[m]Computational Biology Program, CIMA University of Navarra, Pamplona, 31008, Spain
[n]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona 08003, Catalonia, Spain

[*]Equal contribution

[±]**Corresponding author:**
Mariona Bustamante
mariona.bustamante@isglobal.org
ISGlobal
Dr. Aiguader 88
08003

Barcelona
Spain

Highlights

- The genetic variation involved in phthalate detoxification in humans is partially known.
- We identified four loci at genome-wide significance, and 113 at suggestive significance, some of them being novel.
- Two copy number variants were also identified.
- Functional annotation highlighted genes in phase I and II detoxification and renal excretion.

Abstract

Introduction: Phthalates, or dieters of phthalic acid, are a ubiquitous type of plasticizer used in a variety of common consumer and industrial products. They act as endocrine disruptors and are associated with increased risk for several diseases. Once in the body, phthalates are metabolized through partially known mechanisms, involving phase I and phase II enzymes.

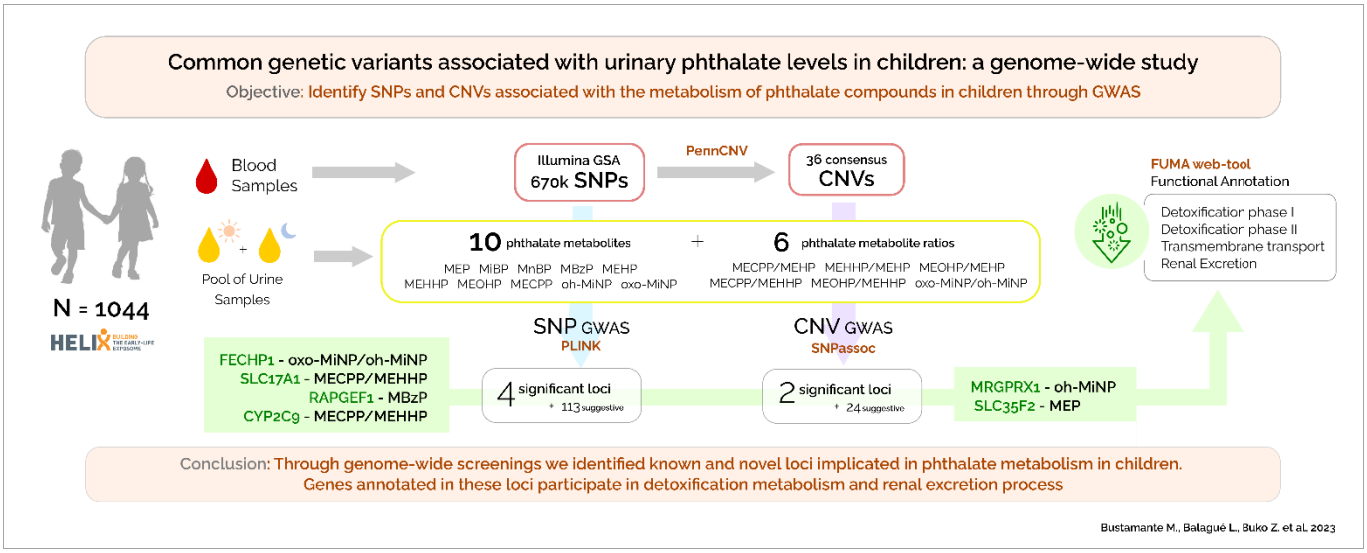
Objective: In this study we aimed to identify common single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) associated with the metabolism of phthalate compounds in children through genome-wide association studies (GWAS).

Methods: The study used data from 1,044 children with European ancestry from the Human Early Life Exposome (HELIX) cohort. Ten phthalate metabolites were assessed in a two-void urine pool collected at the mean age of 8 years. Six ratios between secondary and primary phthalate metabolites were calculated. Genome-wide genotyping was done with the Infinium Global Screening Array (GSA) and imputation with the Haplotype Reference Consortium (HRC) panel. PennCNV was used to estimate copy number variants (CNVs) and CNVRanger to identify consensus regions. GWAS of SNPs and CNVs were conducted using PLINK and SNPassoc, respectively. Subsequently, functional annotation of suggestive SNPs (p-value <1E-05) was done with the FUMA web-tool.

Results: We identified four genome-wide significant (p-value <5E-08) loci at chromosome (chr) 3 (*FECHP1* for oxo-MiNP_oh-MiNP ratio), chr6 (*SLC17A1* for MECPP_MEHPP ratio), chr9 (*RAPGEF1* for MBzP), and chr10 (*CYP2C9* for MECPP_MEHPP ratio). Moreover, 113 additional loci were found at suggestive significance (p-value <1E-05). Two CNVs located at chr11 (*MRGPRX1* for oh-MiNP and *SLC35F2* for MEP) were also identified. Functional annotation pointed to genes involved in phase I and phase II detoxification, molecular transfer across membranes, and renal excretion.

Conclusion: Through genome-wide screenings we identified known and novel loci implicated in phthalate metabolism in children. Genes annotated to these loci participate in detoxification and renal excretion.

Graphical abstract:



Keywords: phthalate, genome-wide association study (GWAS), genetic variant, single nucleotide polymorphism (SNP), copy number variant (CNV), metabolism, toxicity, phase I and II enzymes, renal excretion

Abbreviations

ABC: ATP-binding cassette
BBP: butylbenzyl phthalate
chr: chromosome
CNV: copy number variant
CYP: Cytochrome P450 monooxygenase
DBzP: dibenzyl phthalate
DEP: diethyl phthalate
DEHP: bis(2-ethylhexyl) phthalate
DiBP: diisobutyl phthalate
DINCH: 2-cyclohexane dicarboxylic acid, diisononyl ester
DnBP: di-n-butyl-phthalate
EDCs: endocrine disrupting chemicals
GSA: Infinium Global Screening Array
GWAS: genome-wide association study
HELIX: Human Early Life Exposome project
HMW: high-molecular-weight

HRC: haplotype reference consortium
LD: linkage disequilibrium
LMW: low-molecular-weight
MBzP: mono benzyl phthalate
MECPP: mono-2-ethyl-5-carboxypentyl phthalate
MECPP_MEHP: ratio between MECPP and MEHP
MECPP_MEHHP: ratio between MECPP and MEHHP
MEHHP: mono-2-ethyl-5-hydroxyhexyl phthalate
MEHHP_MEHP: ratio between MEHHP and MEHP
MEHP: mono-2-ethylhexyl phthalate
MEOHP: mono-2-ethyl-5-oxohexyl phthalate
MEOHP_MEHP: ratio between MEOHP and MEHP
MEOPP_MEHHP: ratio between MEOPP and MEHHP
MEP: monoethyl phthalate
MiBP: mono-iso-butyl phthalate
MnBP: mono-n-butyl phthalate
oh-MiNP: mono-4-methyl-7-hydroxyoctyl phthalate
oxo-MiNP: mono-4-methyl-7-oxooctyl phthalate
oxo-MiNP_oh-MiNP: ratio between oxo-MiNP and oh-MiNP
PVC: polyvinyl chloride
QQ: quantile-quantile
SCL: Solute carrier
SNP: single nucleotide polymorphism
SULT: Sulfotransferase
UGT: UDP-glucuronosyltransferase

1. Introduction

Phthalates, or dieters of phthalic acid, are a ubiquitous type of plasticizer used in a variety of common consumer and industrial products, whose exposure is widespread and ever-changing due to our constantly evolving environment and habits (Praveena et al., 2018; Wang et al., 2019). Phthalates are known to be endocrine disrupting chemicals (EDCs), and epidemiological research has suggested that exposure to phthalates is associated with increased risk for several diseases including infertility, allergy, obesity, diabetes, and behavioral problems (Praveena et al., 2018; Wang et al., 2019). Since children are especially vulnerable to contaminants as they are still developing, there is a great concern over the potential for phthalate exposure to disturb normal growth and development (Braun, 2017; Casale and Rice, 2023; Lee et al., 2022, 2023).

Phthalates can be separated into low-molecular-weight (LMW, 3-6 carbon atoms) and high-molecular-weight (HMW, 7-13 carbon atoms) compounds (Praveena et al., 2018; Wang et al., 2019). LMW phthalates are used as solvents and are usually found in medications and personal care items such as deodorants, lotions, and shampoos. HMW phthalates are used in the manufacturing of flexible plastics for purpose of vinyl flooring, adhesives, medical devices, and food packaging. The main intake of LMW phthalates is through skin and inhalation, while HMW phthalates are incorporated into the body through ingestion (Kim and Park, 2014). In the European Union, some phthalates are now banned from use in cosmetics and regulated in material intended to come into contact with food.

After exposure, phthalates are rapidly metabolized and excreted mainly in urine as a result of phase I and phase II enzymes (Domínguez-Romero and Scheringer, 2019; Praveena et al., 2018). First, diester phthalates are hydrolyzed to monoester (primary metabolites) by phase I esterase and lipase enzymes (Bhattacharyya et al., 2022). Subsequently, LMW phthalates are primarily excreted in urine and feces as monoesters, without further metabolism. In contrast, HMW phthalates need further metabolism through hydroxylation and oxidation by phase I enzymes thus producing a number of oxidative metabolites (secondary metabolites). These oxidated metabolites can be directly excreted in the urine, or alternatively, can be additionally metabolized through phase II enzymes. Phase II detoxification consists of conjugation reactions where a compound is added to the parental compound to generate hydrophilic conjugates which can then be easily excreted in the urine. Phthalates and their metabolites can be measured in diverse biological specimens, such as blood, urine, breast milk, and feces. However, urinary phthalate metabolites are the most frequently used biomarkers to track exposure to phthalates (Sakhi et al., 2017; Wang et al., 2019).

Phase I and phase II enzymes are highly polymorphic in humans, with slow and fast metabolic phenotypes determined by single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) (Pinto and Eileen Dolan, 2011). This differential detoxification capacity might modify phthalate effects in the body and is rarely considered in epidemiological studies. In vitro studies demonstrated the importance for phthalate metabolism of genetic polymorphisms in *Cytochrome P450 (CYP) monooxygenases*, a family of phase I detoxification enzymes (Choi et al., 2012).

Moreover, in humans, urinary phthalate metabolite levels were reported to be associated with genetic variants in *CYP monooxygenases* (*CYP2C9* and *CYP2C19)* and in *UDP-glucuronosyltransferases (UGT1A7)* in young adults (Stajnko et al., 2022), and with variants in *GSTP1* and *SOD2* in children (Wang and Karmaus, 2017). However, these studies were based on a priori knowledge of candidate genes and SNPs, likely missing additional loci important for phthalate metabolism. Genome-wide association studies (GWAS), which allow the interrogation of millions of common genetic variants, both SNPs and CNVs, can help to overcome this limitation and reveal new phthalate detoxification pathways (Visscher et al., 2017).

Here, we aimed to identify genetic variants related to phthalate metabolism in children. For this, we analyzed the association of genome-wide SNPs and CNVs with urinary levels of ten phthalate metabolites and six phthalate ratios in 1,044 European ancestry children from the Human Early Life Exposome (HELIX) project and performed functional annotation of the identified variants.

2. Materials and methods

2.1. Study population

HELIX project is an ongoing population-based birth control study in six birth cohorts from different European countries: (1) EDEN - Étude des Déterminants pré et postnatals du développement et de la santé de l'enfant, France; (2) Rhea - the Rhea Mother-Child Study in Crete, Greece; (3) KANC - Kaunus Cohort, Lithuania; (4) MoBa – the Norwegian Mother, Father and Child Cohort Study, Norway; (5) INMA - - INfancia y Medio Ambiente, Spain; (6) BiB - Born in Bradford, United Kingdom (UK) (Maitre et al., 2018). The HELIX project aims to implement novel exposure assessment and biomarker methods to measure the early-life exposure to multiple environmental factors and associate these with omics biomarkers and child health outcomes, thus characterizing the "early-life exposome". The entire study population is 31,472 mother-child pairs. The subcohort study includes 1,304 children with exposure, phenotypes, and molecular data measured at age of 6-12 years. The current study selected 1,044 children of European ancestry with genome-wide genetic and phthalate metabolite levels available from the HELIX subcohort (Appendix A - Supplementary Figure 1). Child's ancestry was predicted from genome-wide genetic data. The following variables were used to describe the population: child's sex (male/female), child's age (in years), child's obesity in 3 categories (normal, overweight, obese) based on the World Health Organization (WHO) body mass index (BMI) classification, and self-reported maternal education (primary, secondary and university or higher).

2.2. Ethics approval and consent to participate

All studies were approved by the national research ethics committees and informed consent to participate was obtained for all participants.

2.3. Phthalate biomarkers measurement

Metabolites for 6 different phthalates were measured in a pool of two urine samples collected in the morning and at bedtime at the mean age of 8 years (Figure 1).

An aliquot of the urine pool was analyzed for the determination of four LMW primary phthalate metabolites [monoethyl phthalate (MEP), mono-n-butyl phthalate (MnBP), mono-isobutyl phthalate (MiBP), mono benzyl phthalate (MBzP)], one HMW primary phthalate metabolite [mono-2-ethylhexyl phthalate (MEHP)], and five HMW secondary metabolites [mono-2-ethyl-5-hydroxyphenyl phthalate (MEHHP), mono-2-ethyl-5-oxohexyl phthalate (MEOHP), mono-2-ethyl-5-oxyhexyl phthalate (MECPP), mono-4-methyl-7-hydroxyoctyl phthalate (oh-MiNP), and mono-4-methyl-7-oxooctyl phthalate (oxo-MiNP)], using high performance liquid chromatography coupled to mass spectrometry (Haug et al., 2018). Both external control samples, in-house control samples and blank samples were analyzed in each batch of about fifty samples. The limits of detection (LOD) of the method ranged from 0.067 to 0.67 ng/mL. All the samples had values >LOD. MEHP, MEHHP and MEOHP had 34, 3 and 1 missing values, due to underperformance of sample quantification.

To account for urine dilution, a second aliquot of urine was analyzed for creatinine concentration, and phthalate metabolite concentrations were divided by urinary creatinine levels (Haug et al., 2018). All creatinine-adjusted concentrations were log2 transformed to obtain normal distributions, as the original distributions were right-skewed. Finally, as a proxy of enzymatic activity, we calculated six ratios between product and substrate, in our case, between secondary and primary phthalate metabolites or between two secondary phthalate metabolites, as follows: MECPP_MEHP, MEHHP_MEHP, MEOHP_MEHP, MECPP_MEHHP, MEOHP_MEHHP and oxo-MiNP_oh-MiNP. The ratios were calculated using untransformed phthalate values unadjusted for creatinine, and then they were log2 transformed. Urine dilution was not taken into account as the ratios were computed using two compounds within the same urine sample.

2.4. Genetic data

DNA was obtained from buffy coats collected in EDTA tubes at the mean age of 8 years. DNA was extracted by cohort using a Chemagen kit in batched of 12 samples. Two techniques were used to determine the DNA concentration 1) NanoDrop 1000 UV-Vis Spectrophotometer (ThermoScientific) and 2) Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies).

Infinium Global Screening Array (GSA) MD version 1 (Illumina) was used for genome-wide genotyping at the Human Genomics Facility (HuGe-F), Erasmus MC (www.glimdna.org). Genotype calling and annotation were done using the GenTrain2.8 algorithm based on a custom cluster file implemented in the GenomeStudio software and the GSAMD-24v1-0_20011747-A4 manifest. SNP coordinates were reported on human reference GRCh37 and on the source strand.

PLINK program was used for the quality control of the genetic data (Purcell et al., 2007). Briefly, samples were filtered out if they had a call rate <97%, had sex inconsistencies, the heterozygosity was >3 standard deviations, if they were related (sharing >18.5% of alleles) or duplicated. For ancestry prediction from GWAS data, we used the Peddy program (Pedersen and Quinlan, 2017). Then genetic ancestry was contrasted with self-reported ethnicity and discordant samples were excluded. Genetic variants were filtered out if they had a call rate <95%, if they were in the non-canonical pseudo-autosomal region (PAR), if they had a minor allele frequency (MAF) <1%, and if they were not in Hardy-Weinberg equilibrium (HWE) at a p-value <1E -06. PLINK was also used to compute the first 20 principal components from the GWAS data of European ancestry children using the linkage disequilibrium (LD) clumping option. Ten PCs explained up to 19.2% of the genetic variance and the five first ones correlated with cohort.

Genome-wide imputation was performed with the Imputation Michigan server using the Haplotype Reference Consortium (HRC) panel, Version r1.1 2016. Before the imputation PLINK data was converted into VCF format and the variants were aligned with the reference genome. Eagle v2.4 was used for the phasing of the haplotypes and minimac4 for the imputation. In the end, we retrieved 40,405,505 variants after imputation. Then we filtered out the genetic variants according to the following parameters: imputation accuracy (R2) <0.9, MAF<5%, and HWE p-value <1E-06. After this the post-imputation dataset had 4,614,947 variants.

2.5. GWAS of SNPs

GWAS of SNPs were conducted with the PLINK program by applying a linear regression model for each phthalate metabolite and each SNP from chromosome 1 to 22, adjusting for sex, age and ten GWAS principal components as a proxy of ancestry (PCs) (Purcell et al., 2007). We run a total of 16 GWAS (five primary metabolites, five secondary metabolites, and six ratios). Genome-wide statistical significance was established at a p-value <5E-08, and suggestive statistical significance at a p-value <1E-5. The lambda genomic inflation factor was estimated as the median of the resulting chi-squared test statistics divided by the expected median of the chi-squared distribution. Quantile-quantile (QQ) plots were conducted with the ggplot2 package and Manhattan plots were conducted with the qqman package in R (D. Turner, 2018; Wickham, 2016). The lead SNP is defined as the SNP with the lowest p-value in a locus, and a locus is defined as a region of 1Mb.

In order to filter out associations between the SNPs and creatinine levels, we run sensitivity models adjusting for log2 transformed creatinine levels, instead of controlling (dividing) the phthalate concentrations by creatinine. Finally, we also run a GWAS for log2 transformed creatinine levels (g per L of urine).

2.6. GWAS of CNVs

Copy number variant (CNV) calls were obtained with PennCNV using BAF/LRR signal files from Illumina BeadStudio (Wang et al., 2007). A total of 22 samples from the initial 1,044 had to be

removed due to poor quality in BAF or LRR values. After that, CNV calls overlapping with the centromeres were filtered out. After the quality control, consensus CNV regions were obtained with the CNVRanger R package following the reciprocal overlap procedure, that merges calls with sufficient mutual overlap, in our case 50% (Conrad et al., 2010; Da Silva et al., 2020). From the resulting consensus regions, only common CNVs with an alternative minor allele frequency (either gain or loss) >3% were kept for the analysis. Consensus CNV regions were classified into: (i) only loss, when we only detected deletions (CN=0 and CN=1); (ii) only gain, when we only detected duplications (CN=3; note that more than 3 copies were never detected); and (iii) loss and gain, when we detected both deletions and duplications (CN=0, CN=1 and CN=3).

Association analyses between each of the resulting regions and each of the 16 metabolites and ratios were performed with SNPassoc R package (González et al., 2007). Distinct association models were applied to each region based on the prior CNV classification. For only-loss regions we tested loss-codominant (0 vs 1 vs 2), loss-dominant (0+1 vs 2), loss-recessive (0 vs 1+2), loss-overdominant (0+2 vs 1) and loss-log-additive models; and for only-gain regions we tested gain-codominant (2 vs 3) model. Finally, for loss-and-gain regions we tested all loss-specific and gain-specific models above, as well as an additive lineal model taking into account loss and gain. All analyses were adjusted for the same covariates as the GWAS of SNPs (sex, age and ten GWAS principal components (PCs)). Multiple-testing correction was conducted by dividing the nominal p-value of 0.05 by the number of detected consensus CNVs (0.05/36 = 0.0014). As was done with SNPs, sensitivity models adjusting for log2 transformed creatinine levels were investigated in order to filter out associations between the CNVs and creatinine levels. Finally, we also run a model for log2 transformed creatinine levels (g per L of urine).

Annotation of genes in each CNV region was done with the biomaRt R package (Durinck et al., 2009).

2.7. Functional annotation of genetic variants, gene-wide association analysis and enrichment

For the annotation, prioritization, and biological interpretation of the genetic variants identified in the GWAS, we used the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) tool (Watanabe et al., 2017). The SNP2GENE module computes LD structure, annotates SNPs using information from different databases, and prioritizes candidate genes. FUMA was run using default parameters except that the p-value threshold was set to 1E-05 instead of 1E-08. For the SNP annotation we used: ANNOVAR, CADD, RegulomeDB, expression quantitative trait locus (eQTL) mapping based on Genotype-Tissue Expression (GTEx) v8 and other databases, comparison with SNPs reported in the GWAS catalog and gene mapping through chromatin interaction maps. Then, with the GENE2FUNC module from FUMA prioritized genes from SNP2GENE module were used to identify shared molecular functions using different databases: Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Reactome, among others. GENE2FUNC was also used to assess enrichment for 54 tissues included in the GTEXv8 dataset.

For gene mapping through chromatin interaction we also used the EPIraction tool (Nurtdinov and Guigó, n.d.). Briefly, the tool contains a catalogue of tissue specific gene-enhancer interactions identified by applying linear models between gene expression and activities of nearby enhancers across 1,529 samples from 77 different tissues. In this study, we annotated SNPs to enhancers and then linked enhancers to genes in two candidate tissues for phthalate detoxification: liver and kidney.

3. Results

3.1. Description of the population

The description of the 1,044 HELIX participants is shown in Table 1. All participants were of European ancestry and from six different countries: Greece, Norway, Spain, Lithuania, France, and the UK. Forty-six percent of the participants were female, and the mean age of phthalate measurements was 8 years (standard deviation (SD): 1.6). Around half of the children were born from highly educated mothers. Urinary phthalate levels (micrograms (μg) per gram (g) of creatinine) were in the range or higher than levels reported in children and adolescents from 12 European countries for the period 2014–2021 (Vogel et al., 2023). After log2 transformation, levels tended to be normally distributed and the correlation among them ranged from 0.18 to 0.98 (Appendix A – Supplementary Figure 2). Highest correlations (r>0.79) were observed among MECPP, MEHHP, MEOHP, MEHP, which are all primary or secondary metabolites of the same parental compound, DEHP. The correlation of oh-MiNP and oxo-MiNP was also high (R=0.73).

3.2. GWAS of SNPs

To identify SNPs associated with the detoxification metabolism of phthalates in children, we run linear regressions for each of the ten phthalate metabolites and the six metabolic ratios adjusting for sex, age, and 10 GWAS principal components.

Genomic inflation factors (lambdas) ranged from 0.989 to 1.021 (Appendix A – Supplementary Figure 3). Four loci reached genome-wide significance (p-value <5E-08) as shown in the Miami plot (Figure 2). They involved two SNPs at chr3 associated with the oxo-MiNP_oh-MiNP ratio (lead SNP rs80064213_C, effect = 0.331, p-value = 4.15E-08, near *FECHP1* gene), 116 SNPs at chr6 associated with the MECPP_MEHPP ratio (rs1359232_A, effect = -0.134, p-value = 4.58E-17, *SLC17A1*), one SNP at chr9 associated with MBzP (rs138702233_T, effect = 0.394, p-value = 4.29E-08, *RAPGEF1*), and 658, 186 and 369 SNPs at chr10 associated with the MECPP_MEHPP, MEOHP_MEHHP and MECPP_MEHP ratios, respectively (rs74494115_T, effect = -0.309, p-value = 1.45E-63, *CYP2C9*). The locus zoom plots of these four loci are shown in Figure 3.

Moreover, 2,140 SNPs representing 158 loci (113 unique) were identified at suggestive significance (p-value <1E-05) for at least one phthalate metabolite or ratio (Table 2, Appendix B - Supplementary Table B1). The number of loci with suggestive statistical significance per phthalate metabolite or ratio was between five and 15. Thirty of the 117 suggestive or genome-wide significant unique loci were shared between more than one phthalate compound or ratio (Appendix A – Supplementary Figure 4). The genome-wide significant loci at chr3 (*FECHP1)*, chr6 (*SLC17A1)* and chr10 (*CYP2C9)*, also had suggestive associations with other phthalate compounds or ratios (Table 3, Appendix B – Supplementary Table B1).

Sensitivity analysis adjusting for creatinine instead that using it to divide phthalate metabolite levels gave similar results with minor differences (Appendix B - Supplementary Table B2). Three of the four genome-wide significant loci had still a p-value <1E-08, and the other hit (chr9 loci associated with MBzP) was marginally significant (p-value = 9.95E-07). A plot comparing p-values and effect sizes between the two models for all the suggestive SNPs can be seen at Appendix A – Supplementary Figure 5A and 5B. Moreover, the GWAS of creatinine levels did not retrieve any genome-wide significant SNP. Only two of the creatinine suggestive SNPs were also suggestively associated with MEOHP and oh-MiNP (Appendinx B Supplementary Table B19).

3.3. GWAS of CNVs

Using the PennCNV tool, we identified a median of seven (interquartile range (IQR): 6-9) CNV regions per child with a median length of 43 (IQR: 29-64) kb. Only two children did not present any CNV event. After applying a reciprocal overlap of 50%, 2,916 distinct CNV regions were detected, including 1,785 that were unique, meaning that they were detected in only one child. Thirty-six of the 2,916 CNVs were identified in > 3% of the children and thus labeled as consensus CNV regions and tested in relation to phthalate levels and ratios (Appendix C – Supplementary Table C0).

In 12 of the 36 consensus regions, we detected single or double deletions (loss CNVs, CN=0 and CN=1), in seven of them we detected single duplications (gain CNVs, CN=3) and in 17 of them we detected single and double deletions as well as single duplications (loss and gain CNVs, CN=0, CN=1 and CN=3). We did not detect double duplications (CN=4) in any CNV consensus region. Different models were tested in each type of region (see 2.6.) and the model with the highest significance was reported for each region. The results of the association of these common CNVs with phthalate levels and ratios can be found in Appendix C – Supplementary Tables C3-18.

Fifty-one CNVs presented nominal significance and two of them passed multiple-testing correction (Figure 2, Appendix C - Supplementary Table C1). First, a single-copy gain in CNV region 11A was associated with higher oh-MiNP levels (chr11:18.95Mb-18.96Mb_gain, effect = 1.231, p-value = 1.77E-06, *MRGPRX1* gene), and second, a single-copy gain in CNV region 11B was associated with higher MEP levels (chr11:107.65Mb-107.67Mb_gain, effect = 0.792, p-value = 1.36E-03, *SLC35F2*). The locus zoom plots of these two CNVs are shown in Figure 4.

In the sensitivity analysis adjusting for creatinine, the 88% of the nominally significant associations were reproduced (45 of 51). The associations between CNV 11A and oh-MiNP and between CNV 11B and MEP were very similar (Appendix C - Supplementary Table C2), however the later did not reach the multiple-testing threshold (chr11:18.95Mb-18.96Mb_gain, effect = 1.120, p-value = 2.51E-06; chr11:107.65Mb-107.67Mb_gain, effect = 0.695, p-value =1.48E-03). A plot comparing p-values and effect sizes between the two models for all the suggestive CNVs can be seen at Appendix A – Supplementary Figure 5C and 5D. Results for creatinine GWAS did not show any significant association, with only four CNV regions, not including 11A and 11B, showing a p-value <0.05 (Appendix C - Supplementary Table C19).

3.4. Functional annotation of genetic variants and identification of candidate genes and pathways

Next, SNPs with suggestive significance were explored with the FUMA tool providing information on SNP annotation (Appendix D), identification of eQTLs (Appendix E), and comparison with SNPs in the GWAS catalog (Appendix F). Appendix G shows the list of candidate genes in each locus based on FUMA results and gene-enhancer interactions from the EPIraction tool. A selection of 12 interesting loci in terms of statistical significance, annotated genes and/or overlap with GWAS traits is listed in Table 3. These loci are annotated to phase I and phase II detoxification genes [*CYP monooxygenases* (*CYP2C* cluster), *Sulfotransferases* (*SULT1B1*, *SULT1E1*, and *SULT1C3)*, *UDP glucuronosyltransferases* (*UGT2A1*, *UGT2A2* and *UGT2B11*), and glutathione related genes (*GSR*, *GPX5* and *GPX6*)]; and genes related to molecule transport across membranes [*Solute carriers* (*SLC17A* cluster, *SLC8A2*, and *SLC35G1*) and others such as *ABCC3*, *ATP9B* and *STAC*].

After the FUMA annotation, we took suggestive prioritized genes to run gene-set enrichment analyses (Appendix H). Among others, we identified gene-sets related to xenobiotic detoxification, drug detoxification, CYP, oxidoreductase reactions, liver, phase I and phase II enzymes. These gene-sets were found basically for MEP, MECPP and its ratios, and MEOHP ratios.

Finally, we searched for enriched GTEXv8 tissues (Appendix I). At adjusted and/or nominal statistical significance, we found enrichment for candidate tissues involved in phthalate metabolism: liver (MEOHP_MEHP and MEOPH_MEHHP), kidney cortex (MBzP, MECPP_MEHP and MEOHP_MEHHP), and bladder (MECPP_MEHP). Other significant enrichments were found for skin and vagina (oxo-MiNP_oh-MiNP), artery aorta (MiBP), and nerve tibial (MEHP).

4. Discussion

In this study we conducted genome-wide screenings of SNPs and CNVs in order to identify genetic variation associated with phthalate metabolism in children. We identified four genome-wide significant loci at chr3 (*FECHP1* – oxo-MiNP_oh-MiNP), chr6 (*SLC17A1* – MECPP_MEHPP), chr9 (*RAPGEF1* – MBzP) and chr10 (*CYP2C9* – MECPP_MEHPP) and two significant CNVs located at chr11 (*MRGPRX1* – oh-MiNP and *SLC35F2* – MEP). Functional annotation of suggestive SNPs highlighted pathways related to xenobiotic and drug detoxification. Also some of the genes annotated

to the SNPs associated with phthalates showed significant enrichment for key tissues as liver and kidney, which is indicative of high gene expression levels in these tissues. Furthermore, previous GWAS have reported significant associations of SNPs located in these loci with drug biotransformation and urinary metabolite levels of several compounds. Below we describe in detail the most relevant loci and genes identified, which are related to phase I and II metabolism, transport of molecules across membranes and kidney function.

4.1. Phase I metabolism

We identified three loci annotated to *CYP* genes, one of them achieving the smallest p-value of the study. This top locus at chr10 was associated with four DEHP metabolite ratios (MECPP_MEHPP; MECPP_MEHP; MEOHP_MEHHP; MEOHP_MEHP) and contains *CYP2C8*, *CYP2C9*, *CYP2C18*, and *CYP2C19* genes. According to the EPIraction tool, the locus can act as an enhancer of these genes in liver and/or kidney. *CYP* genes code for phase I monooxygenases that are key in multitude of metabolic processes of endogenous and exogenous compounds, and, indeed, SNPs in this locus have previously been reported to be associated with serum or plasma levels of several metabolites or proteins (Feofanova et al., 2020; Pazoki et al., 2021; Schlosser et al., 2020) and drug response (ie. warfarin) (Takeuchi et al., 2009). Moreover, CYP2C enzymes are known to participate in phthalate metabolism, and *CYP2C9* variants have been associated with the biotransformation of MEHP into secondary metabolites in young adults (Stajnko et al., 2022). Our study confirms that similar detoxification mechanisms take place in children. The two additional *CYP* loci, one at chr7 harboring *CYP2W1* and the other at chr19 including *CYP2A6*, *CYP2B6*, *CYP2A13* and *CYP2S1*, showed suggestive associations with urinary MEP levels (metabolite of diethyl phthalate (DEP)). Therefore, according to our results, DEP (LMW phthalate present in many personal care products, particularly those containing fragrances) and DEHP (HMW phthalate mainly used for food packaging) seem to be metabolized through different CYP genes.

4.2. Phase II metabolism

We also identified several loci annotated to phase II metabolism genes, including *UDP-glucuronosyltransferases* (*UGTs*), *Sulfotransferases* (*SULTs*) and *Glutathione S-transferases* (*GSTs*). First, we found a locus at chr4, annotated to *UGT2B11*, *UGT2A1* and *UGT2A2*, to be suggestively associated with levels of MnBP, a metabolite of di-n-butyl-phthalate (DnBP), which is found in paints, adhesives, personal care products and deodorants. SNPs in another family member, *UGT1A7*, were previously related to urinary levels of 2-cyclohexane dicarboxylic acid, diisononyl ester (DINCH) metabolite in men and of diisobutyl phthalate (DiBP) and dibenzyl phthalate (DBzP) metabolites in women (Stajnko et al., 2022). Notably, DiBP and DnBP are structural isomers. Second, we found three loci annotated to *Sulfotransferases*: the same locus at chr4 associated with MnBP also contained the *SULT1B1* and *SULT1E1* genes, and two additional suggestive loci at chr2 and chr19 harboring SNPs near *SULT1C3* and *SULT2B1* genes, respectively, that showed associations with DEHP metabolites. Previous studies have described that multiple phthalates have

strong inhibition potential towards *SULT* enzymes, potentially causing reduced body detoxification and liver injury (Ceauranu et al., 2023; Huang et al., 2022). Finally, we identified two loci that contain genes that participate in the glutathione metabolism: a genome-wide significant locus at chr6 annotated to *GPX5*, *GPX6* and *SLC* genes associated with MEHP and oh-MiNP levels; and a suggestive locus at chr8 near *GSR*, a central enzyme of cellular antioxidant defence, associated with MEHP ratios. This is consistent with previous research reporting that in zebrafish, MEHP exposure alters the expression of *gsr* (Kwan et al., 2021).

4.3. Molecular transport across membranes and kidney function

Besides detoxification, we identified several loci annotated to transmembrane transporter genes which could potentially play a role in the excretion of phthalates in the intestine or kidney. These encompass several *Solute carrier* (*SLC*) genes that code for transporters of a wide array of substrates across membranes (Liu, 2019). A locus at chr6 harboring *SLC17A1*, *SLC17A2*, *SLC17A3* and *SLC17A4* was associated with MEHP ratios at genome-wide significance and showed suggestive associations with oh-MiNP. Both gene-enhancer and eQTL information, suggested that these SNPs regulate *SLC* expression in liver and/or kidney. Moreover, it is known that *SLC17A1* and *SLC17A3* participate in urate regulation by facilitating the excretion of intracellular urate from the bloodstream into renal tubule cells. Consistently, previous GWAS reported associations between this locus and serum uric acid levels (Hollis-Moffatt et al., 2012), urinary metabolite levels (Schlosser et al., 2020), estimated glomerular filtration rate (creatinine) (Stanzick et al., 2021), and chronic kidney disease (Torres et al., 2021), among others. Besides, other associations involving the *SLC* family were identified, including the association between a single-copy gain in a CNV that overlaps the final exon and 3'UTR of the *SLC35F2* gene, predicted to enable transmembrane transporter activity, and MEP urinary levels.

ABCC3 encodes a member of the ATP-binding cassette (ABC) transporters that play a role in the transport of biliary and intestinal excretion of organic anions. Moreover, it is reported to be involved in multi-drug resistance (Deeley et al., 2006; Ramírez-Cosmes et al., 2021) and variants in the gene are related to serum/plasma metabolite levels (Bruhn and Cascorbi, 2014). In our study, SNPs in the *ABCC3* intronic region (chr17) were found to be suggestively associated with the MEOHP_MEHP ratio. Notably, another study showed that flies exposed to dibutyl-phthalate (DBP) have increased expression of several genes, including an homologous of the *ABCC3* gene (Williams et al., 2016). Two more transporters were identified. On the one hand, SNPs in a liver and kidney enhancer for *ATP9B* gene (chr18), which is predicted to enable ATPase-coupled intramembrane lipid transport, were suggestively associated with DEHP metabolites and ratios. On the other hand, SNPs near *STAC* (chr3) were associated with the oxo-MiNP_oh-MiNP ratio at genome wide significance and with the MEOHP_MEHHP ratio at suggestive significance. *STAC* is predicted to enable transmembrane transporter binding activity and participate in positive regulation of voltage-gated calcium channels and skeletal muscle contraction.

The last group of results is comprised by two genes potentially implicated in kidney function: *RAPGEF1* and *MRGPRX1*. First, SNPs located in a liver and kidney enhancer for *RAPGEF1* at chr9 were genome-wide associated with MBzP, a metabolite of butylbenzyl phthalate (BBP) used as plasticizer for polyvinyl chloride (PVC). *RAPGEF1* encodes a guanine nucleotide exchange factor, and mutations in genes of the same family have been found in patients with nephrotic syndrome type 7 and membranoproliferative glomerulonephritis (Maywald et al., 2022; Zhu et al., 2019). Second, a loss-gain CNV of 10.8 kb overlapping the whole gene *MRGPRX1* was associated with oh-MiNP, a metabolite of DINP which is used primarily in the manufacture of PVC. *MRGPRX1* encodes a Mas-related G-protein-coupled receptor reported to be responsible for itch sensation, pain transmission and inflammatory reactions, including those to the antimalarial drug chloroquine (Gan et al., 2023). Additionally, mutations in other G-protein-coupled receptors have been implicated in nephrogenic diabetes insipidus, a disease characterized by polyuria and polydipsia (Wang et al., 2018). We found that individuals with three copies of the gene had increased levels of urinary levels of oh-MiNP, whereas individuals with zero or one copy had no difference versus the reference group with two copies. Given the low numbers presenting the alternative allele, further investigation in a larger sample would be required.

Our study found several associations between phthalates and SNPs known to regulate urinary metabolite levels, glomerular filtration rate, and kidney function. These associations might indicate a real participation of these SNPs in phthalate excretion. Nevertheless, as our phthalate measurements were normalized by creatinine to account for urine dilution, some of our results could also just reflect renal excretion capacity. Although both explanations are possible, there is some evidence supporting the former. First, in vitro studies suggest that phthalates can be a substrate for the *SLC* transporters (Klaassen and Aleksunes, 2010); second, the associations found with the phthalate ratios are unaffected by the creatinine normalization; third the sensitivity model adjusting for creatinine instead of normalizing by it provided similar results; and four the GWAS of creatinine did not found many suggestive loci overlapping with loci found for phthalates.

4.4. Strengths of the study

The study has several strengths. First, we measured a large number of phthalate metabolites using a robust analytical method in an equal volume daily pool of two urine samples (morning and night). Second, instead of analyzing individual SNPs in candidate genes, we performed genome-wide screenings of SNPs and CNVs and several downstream annotation methods that allowed us to identify novel loci implicated in phthalate metabolism. Although some genes seemed to be involved in detoxification of more than one phthalate compound (ie. *SULT1C3* for MnBP and MEHP, or *SLC17* for oh-MiNP and MEHP), others appeared to be more specific (ie. *CYP2C* for MEHP while *CYP2A* for MEP). Third, the study was conducted in a large sample of children from the HELIX project, higher than previous similar studies but still limited in comparison with GWAS of other traits. This study, besides producing novel knowledge about the genes participating in phthalate detoxification and renal excretion, also provides a list of SNPs that can be used to stratify individuals in epidemiological research. Bearing the “alternative” alleles of these SNPs can confer different

susceptibility to phthalate's effects depending on the SNP function (i.e., fast biotransformation to a neutral compound and then excretion versus fast biotransformation to a more toxic compound).

4.5. Limitations of the study

However, the study also has some limitations. First, we have restricted our discussion on few genes with interesting functions for phthalate metabolism. However, since SNPs can regulate long distance genes, other genes not highlighted here could also be relevant. Functional analyses in in vitro studies or animal models should be conducted in order to identify the causal variants and genes. Second, we analyzed phthalate metabolites in a pool of two urines and then standardized for creatinine. This has some limitations, especially for pooled samples as discussed elsewhere (O'Brien et al., 2017; Philippat and Calafat, 2021), but currently there is no consensus on how to account for urine dilution in equal volume pools. As discussed above, this could have resulted in the identification of genetic variants that regulate glomerular filtration in general, rather than phthalate excretion in particular. Finally, the analyses were restricted to European ancestry children, thus the translation to other ancestries is unknown.

5. Conclusions

In summary, through genome-wide screenings we identified known and novel loci potentially implicated in phthalate metabolism in children. Genes annotated to these loci participate in phase I and phase II metabolism and renal excretion process. Replication of the findings in independent studies and validation in in vitro systems is required.

6. Competing Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

7. Author Contributions

MB conceptualized the study. MB, LB, ZB, and RN conducted the statistical analysis or functional enrichment analysis with the support of JRG. The HELIX project was coordinated by MV with the support of LM. SA, ALB, MAC, MC, LC, RG, KBG, BH, JL, JS, TCY, JW, are the PIs of the cohorts or participated in sample or data acquisition. AKS was responsible for measurement of phthalate metabolites. MB, GE and CRA participated in genetic data acquisition. LB, ZB and MB wrote the original draft of the paper and all the other co-authors contributed to reviewing and editing the manuscript.

8. Acknowledgements

We would like to thank all the families for their generous contribution.

9. Funding

The study has received funding from the European Community's Seventh Framework Programme (FP7/2007-206) under grant agreement no 308333 (HELIX project) and the H2020-EU.3.1.2. - Preventing Disease Programme under grant agreement no 874583 (ATHLETE project). The genotyping was supported by the project PI17/01225 and PI17/01935, funded by the Instituto de Salud Carlos III and co-funded by European Union (ERDF, "A way to make Europe") and the Centro Nacional de Genotipado-CEGEN (PRB2-ISCIII).

BiB received core infrastructure funding from the Wellcome Trust (WT101597MA) and a joint grant from the UK Medical Research Council (MRC) and Economic and Social Science Research Council (ESRC) (MR/N024397/1), and National Institute for Health Research Applied Research Collaboration Yorkshire and Humber (NIHR200166). The views expressed are those of the author(s), and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. INMA data collections were supported by grants from the Instituto de Salud Carlos III, CIBERESP, and the Generalitat de Catalunya-CIRIT. KANC was funded by the grant of the Lithuanian Agency for Science Innovation and Technology (6-04-2014_31V-66). The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. The Rhea project was financially supported by European projects (EU FP6-2003-Food-3-NewGeneris, EU FP6. STREP Hiwate, EU FP7 ENV.2007.1.2.2.2. Project No 211250 Escape, EU FP7-2008-ENV-1.2.1.4 Envirogenomarkers, EU FP7-HEALTH-2009- single stage CHICOS, EU FP7 ENV.2008.1.2.1.6. Proposal No 226285 ENRIECO, EU- FP7- HEALTH-2012 Proposal No 308333 HELIX), and the Greek Ministry of Health (Program of Prevention of obesity and neurodevelopmental disorders in preschool children, in Heraklion district, Crete, Greece: 2011-2014; "Rhea Plus": Primary Prevention Program of Environmental Risk Factors for Reproductive Health, and Child Health: 2012-15). ISGlobal acknowledges support from the Spanish Ministry of Science and Innovation through the "Centro de Excelencia Severo Ochoa 2019-2023" Program (CEX2018-000806-S), and support from the Generalitat de Catalunya through the CERCA Program. CRG acknowledge the support of the Spanish Ministry of Science, Innovation, and Universities to the EMBL partnership, the Centro de Excelencia Severo Ochoa, and the CERCA Programme/Generalitat de Catalunya.

LM is funded by a Juan de la Cierva-Incorporación fellowship (IJC2018-035394-I) awarded by the Spanish Ministerio de Economía, Industria y Competitividad. MC holds a Miguel Servet fellowship (MS16/00128) funded by Instituto de Salud Carlos III and co-funded by European Social Fund "Investing in your future".

10. Data Availability Statement

Summarized results can be found at https://helixomics.isglobal.org/ **(THIS WILL BE DONE UPON ACCEPTANCE OF THE MANUSCRIPT).** The raw data supporting the current study are

682 available from the corresponding author on request subject to ethical and legislative review. The
683 "HELIX Data External Data Request Procedures" are available with the data inventory in this
684 website: http://www.projecthelix.eu/data-inventory.
685
686
687 **11. References**
688
689 Bhattacharyya, M., Basu, S., Dhar, R., Dutta, T.K., 2022. Phthalate hydrolase: distribution,
690 diversity and molecular evolution. Environ. Microbiol. Rep. 14, 333–346.
691 https://doi.org/10.1111/1758-2229.13028
692 Braun, J.M., 2017. Early-life exposure to EDCs: role in childhood obesity and
693 neurodevelopment. Nat. Rev. Endocrinol. 13, 161–173.
694 https://doi.org/10.1038/nrendo.2016.186
695 Bruhn, O., Cascorbi, I., 2014. Polymorphisms of the drug transporters ABCB1, ABCG2, ABCC2
696 and ABCC3 and their impact on drug bioavailability and clinical relevance. Expert Opin.
697 Drug Metab. Toxicol. 10, 1337–1354. https://doi.org/10.1517/17425255.2014.952630
698 Casale, J., Rice, A.S., 2023. Phthalates Toxicity, StatPearls Treasure Island (FL). StatPearls
699 Publishing.
700 Ceauranu, S., Ciorsac, A., Ostafe, V., Isvoran, A., 2023. Evaluation of the Toxicity Potential of
701 the Metabolites of Di-Isononyl Phthalate and of Their Interactions with Members of Family
702 1 of Sulfotransferases - A Computational Study. Molecules 28, 6748–65.
703 https://doi.org/10.3390/molecules28186748/S1
704 Choi, K., Joo, H., Campbell, J.L., Clewell, R.A., Andersen, M.E., Clewell, H.J., 2012. In vitro
705 metabolism of di(2-ethylhexyl) phthalate (DEHP) by various tissues and cytochrome P450s
706 of human and rat. Toxicol. Vitr. 26, 315–22. https://doi.org/10.1016/j.tiv.2011.12.002
707 Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews,
708 T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K.,
709 MacArthur, D.G., MacDonald, J.R., Onyiah, I., Pang, A.W.C., Robson, S., Stirrups, K.,
710 Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W.,
711 Hurles, M.E., 2010. Origins and functional impact of copy number variation in the human
712 genome. Nature 464, 704–712. https://doi.org/10.1038/nature08516
713 D. Turner, S., 2018. qqman: an R package for visualizing GWAS results using Q-Q and
714 manhattan plots. J. Open Source Softw. 3, 731–732. https://doi.org/10.21105/joss.00731
715 Da Silva, V., Ramos, M., Groenen, M., Crooijmans, R., Johansson, A., Regitano, L., Coutinho,
716 L., Zimmer, R., Waldron, L., Geistlinger, L., 2020. CNVRanger: association analysis of
717 CNVs with gene expression and quantitative phenotypes. Bioinformatics 36, 972–973.
718 https://doi.org/10.1093/bioinformatics/btz632
719 Deeley, R.G., Westlake, C., Cole, S.P.C., 2006. Transmembrane transport of endo- and
720 xenobiotics by mammalian ATP-binding cassette multidrug resistance proteins. Physiol.
721 Rev. 86, 849–899. https://doi.org/10.1152/PHYSREV.00035.2005
722 Domínguez-Romero, E., Scheringer, M., 2019. A review of phthalate pharmacokinetics in
723 human and rat: what factors drive phthalate distribution and partitioning? Drug Metab. Rev.
724 51, 314–329. https://doi.org/10.1080/03602532.2019.1620762
725 Durinck, S., Spellman, P.T., Birney, E., Huber, W., 2009. Mapping identifiers for the integration
726 of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. 4, 1184–1191.
727 https://doi.org/10.1038/NPROT.2009.97

728 Feofanova, E. V., Chen, H., Dai, Y., Jia, P., Grove, M.L., Morrison, A.C., Qi, Q., Daviglus, M.,
729 Cai, J., North, K.E., Laurie, C.C., Kaplan, R.C., Boerwinkle, E., Yu, B., 2020. A Genome-
730 wide Association Study Discovers 46 Loci of the Human Metabolome in the Hispanic
731 Community Health Study/Study of Latinos. Am. J. Hum. Genet. 107, 849–863.
732 https://doi.org/10.1016/J.AJHG.2020.09.003
733 Frederiksen, H., Skakkebæk, N.E., Andersson, A.M., 2007. Metabolism of phthalates in humans.
734 Mol. Nutr. Food Res. 51, 899–911. https://doi.org/10.1002/mnfr.200600243
735 Gan, B., Yu, L., Yang, H., Jiao, H., Pang, B., Chen, Y., Wang, C., Lv, R., Hu, H., Cao, Z., Ren,
736 R., 2023. Mechanism of agonist-induced activation of the human itch receptor MRGPRX1.
737 PLOS Biol. 21, e3001975. https://doi.org/10.1371/JOURNAL.PBIO.3001975
738 González, J.R., Armengol, L., Solé, X., Guinó, E., Mercader, J.M., Estivill, X., Moreno, V.,
739 2007. SNPassoc: An R package to perform whole genome association studies.
740 Bioinformatics 23, 644–645. https://doi.org/10.1093/bioinformatics/btm025
741 Haug, L.S., Sakhi, A.K., Cequier, E., Casas, M., Maitre, L., Basagana, X., Andrusaityte, S.,
742 Chalkiadaki, G., Chatzi, L., Coen, M., de Bont, J., Dedele, A., Ferrand, J., Grazuleviciene,
743 R., Gonzalez, J.R., Gutzkow, K.B., Keun, H., McEachan, R., Meltzer, H.M., Petraviciene,
744 I., Robinson, O., Saulnier, P.-J., Slama, R., Sunyer, J., Urquiza, J., Vafeiadi, M., Wright, J.,
745 Vrijheid, M., Thomsen, C., 2018. In-utero and childhood chemical exposome in six
746 European mother-child cohorts. Environ. Int. 121, 751–763.
747 https://doi.org/10.1016/J.ENVINT.2018.09.056
748 Hollis-Moffatt, J.E., Phipps-Green, A.J., Chapman, B., Jones, G.T., van Rij, A., Gow, P.J.,
749 Harrison, A.A., Highton, J., Jones, P.B., Montgomery, G.W., Stamp, L.K., Dalbeth, N.,
750 Merriman, T.R., 2012. The renal urate transporter SLC17A1 locus: Confirmation of
751 association with gout. Arthritis Res. Ther. 14, R92. https://doi.org/10.1186/ar3816
752 Huang, H., Lan, B. Di, Zhang, Y.J., Fan, X.J., Hu, M.C., Qin, G.Q., Wang, F.G., Wu, Y., Zheng,
753 T., Liu, J.H., 2022. Inhibition of Human Sulfotransferases by Phthalate Monoesters. Front.
754 Endocrinol. (Lausanne). 13, 1–9. https://doi.org/10.3389/fendo.2022.868105
755 Kim, S.H., Park, M.J., 2014. Phthalate exposure and childhood obesity. Ann. Pediatr.
756 Endocrinol. Metab. 19, 69–75. https://doi.org/10.6065/APEM.2014.19.2.69
757 Klaassen, C.D., Aleksunes, L.M., 2010. Xenobiotic, bile acid, and cholesterol transporters:
758 Function and regulation. Pharmacol. Rev. 62, 1–96. https://doi.org/10.1124/pr.109.002014
759 Kwan, W.S., Roy, V.A.L., Yu, K.N., 2021. Review on Toxic Effects of Di(2-ethylhexyl)
760 Phthalate on Zebrafish Embryos. Toxics 9, 193–215.
761 https://doi.org/10.3390/TOXICS9080193
762 Lee, D.W., Lim, H.M., Lee, J.Y., Min, K.B., Shin, C.H., Lee, Y.A., Hong, Y.C., 2022. Prenatal
763 exposure to phthalate and decreased body mass index of children: a systematic review and
764 meta-analysis. Sci. Reports 2022 121 12, 1–18. https://doi.org/10.1038/s41598-022-13154-9
765 Lee, S., Lee, H.A., Park, B., Han, H., Hong, Y.S., Ha, E.H., Park, H., 2023. Prospective
766 association between phthalate exposure in childhood and liver function in adolescence: the
767 Ewha Birth and Growth Cohort Study. Environ. Heal. A Glob. Access Sci. Source 22, 1–12.
768 https://doi.org/10.1186/S12940-022-00953-W/FIGURES/2
769 Liu, X., 2019. SLC Family Transporters. Adv. Exp. Med. Biol. 1141, 101–202.
770 https://doi.org/10.1007/978-981-13-7647-4_3
771 Maitre, L., de Bont, J., Casas, M., Robinson, O., Aasvang, G.M., Agier, L., Andrušaitytė, S.,
772 Ballester, F., Basagaña, X., Borràs, E., Brochot, C., Bustamante, M., Carracedo, A., de
773 Castro, M., Dedele, A., Donaire-Gonzalez, D., Estivill, X., Evandt, J., Fossati, S., Giorgis-

Allemand, L., R Gonzalez, J., Granum, B., Grazuleviciene, R., Bjerve Gützkow, K., Småstuen Haug, L., Hernandez-Ferrer, C., Heude, B., Ibarluzea, J., Julvez, J., Karachaliou, M., Keun, H.C., Hjertager Krog, N., Lau, C.-H.E., Leventakou, V., Lyon-Caen, S., Manzano, C., Mason, D., McEachan, R., Meltzer, H.M., Petraviciene, I., Quentin, J., Roumeliotaki, T., Sabido, E., Saulnier, P.-J., Siskos, A.P., Siroux, V., Sunyer, J., Tamayo, I., Urquiza, J., Vafeiadi, M., van Gent, D., Vives-Usano, M., Waiblinger, D., Warembourg, C., Chatzi, L., Coen, M., van den Hazel, P., Nieuwenhuijsen, M.J., Slama, R., Thomsen, C., Wright, J., Vrijheid, M., 2018. Human Early Life Exposome (HELIX) study: a European population-based exposome cohort. BMJ Open 8, e021311. https://doi.org/10.1136/bmjopen-2017-021311

Maywald, M.L., Picciotto, C., Lepa, C., Bertgen, L., Yousaf, F.S., Ricker, A., Klingauf, J., Krahn, M.P., Pavenstädt, H., George, B., 2022. Rap1 Activity Is Essential for Focal Adhesion and Slit Diaphragm Integrity. Front. Cell Dev. Biol. 10, 790365, 1–14. https://doi.org/10.3389/FCELL.2022.790365/BIBTEX

Nurtdinov, R., Guigó, R., n.d. EPIraction - an atlas of candidate enhancer-gene interactions in human tissues and cell lines. [Manuscript in preparation].

O'Brien, K.M., Upson, K., Buckley, J.P., 2017. Lipid and Creatinine Adjustment to Evaluate Health Effects of Environmental Exposures. Curr. Environ. Heal. reports 4, 44–50. https://doi.org/10.1007/S40572-017-0122-7/FIGURES/2

Pazoki, R., Vujkovic, M., Elliott, J., Evangelou, E., Gill, D., Ghanbari, M., van der Most, P.J., Pinto, R.C., Wielscher, M., Farlik, M., Zuber, V., de Knegt, R.J., Snieder, H., Uitterlinden, A.G., Boezen, H.M., Franke, L., van der Harst, P., Navis, G., Rots, M., Swertz, M., Wolffenbuttel, B.H.R., Wijmenga, C., Lynch, J.A., Jiang, X., Said, S., Kaplan, D.E., Lee, K.M., Serper, M., Carr, R.M., Tsao, P.S., Atkinson, S.R., Dehghan, A., Tzoulaki, I., Ikram, M.A., Herzig, K.H., Järvelin, M.R., Alizadeh, B.Z., O'Donnell, C.J., Saleheen, D., Voight, B.F., Chang, K.M., Thursz, M.R., Elliott, P., Ballas, Z.K., Bhushan, S., Boyko, E.J., Cohen, D.M., Concato, J., Aslan, M., Zhao, H., Constans, J.I., Dellitalia, L.J., Fayad, J.M., Fernando, R.S., Florez, H.J., Gaddy, M.A., Gappy, S.S., Gibson, G., Godschalk, M., Greco, J.A., Gupta, S., Gutierrez, S., Hammer, K.D., Hamner, M.B., Harley, J.B., Hung, A.M., Huq, M., Hurley, R.A., Iruvanti, P.R., Ivins, D.J., Jacono, F.J., Jhala, D.N., Kaminsky, L.S., Klein, J.B., Liangpunsakul, S., Lichy, J.H., Moser, J., Huang, G.D., Muralidhar, S., Mastorides, S.M., Mathew, R.O., Mattocks, K.M., McArdle, R., Meyer, P.N., Meyer, L.J., Moorman, J.P., Morgan, T.R., Murdoch, M., Okusaga, O.O., Oursler, K.A.K., Ratcliffe, N.R., Rauchman, M.I., Robey, R.B., Ross, G.W., Servatius, R.J., Sharma, S.C., Sherman, S.E., Sonel, E., Sriram, P., Stapley, T., Striker, R.T., Tandon, N., Villareal, G., Wallbom, A.S., Wells, J.M., Whittle, J.C., Whooley, M.A., Wilson, P.W., Sun, Y. V., Xu, J., Yeh, S.S., Connor, T., Argyres, D.P., Hauser, E.R., Beckham, J.C., Stephens, B., Aguayo, S.M., Ahuja, S.K., Pyarajan, S., Cho, K., Gaziano, J.M., Kinlay, S., Nguyen, X.M.T., Brewer, J. V., Brophy, M.T., Do, N. V., Humphries, D.E., Selva, L.E., Shayan, S., Whitbourne, S.B., Breeling, J.L., Romero, J.P.C., Ramoni, R.B., 2021. Genetic analysis in European ancestry individuals identifies 517 loci associated with liver enzymes. Nat. Commun. 12, 2579, 1–12. https://doi.org/10.1038/S41467-021-22338-2

Pedersen, B.S., Quinlan, A.R., 2017. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am. J. Hum. Genet. 100, 406–413. https://doi.org/10.1016/j.ajhg.2017.01.017

Philippat, C., Calafat, A.M., 2021. Comparison of strategies to efficiently combine repeated

urine samples in biomarker-based studies. Environ. Res. 192, 110275, 1–9. https://doi.org/10.1016/J.ENVRES.2020.110275

Pinto, N., Eileen Dolan, M., 2011. Clinically Relevant Genetic Variations in Drug Metabolizing Enzymes. Curr. Drug Metab. 12, 487–497. https://doi.org/10.2174/138920011795495321

Praveena, S.M., Teh, S.W., Rajendran, R.K., Kannan, N., Lin, C.C., Abdullah, R., Kumar, S., 2018. Recent updates on phthalate exposure and human health: a special focus on liver toxicity and stem cell regeneration. Environ. Sci. Pollut. Res. 25, 11333–11342. https://doi.org/10.1007/s11356-018-1652-8

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81, 559–575.

Ramírez-Cosmes, A., Reyes-Jiménez, E., Zertuche-Martínez, C., Hernández-Hernández, C.A., García-Román, R., Romero-Díaz, R.I., Manuel-Martínez, A.E., Elizarrarás-Rivas, J., Vásquez-Garzón, V.R., 2021. The implications of ABCC3 in cancer drug resistance: can we use it as a therapeutic target? Am. J. Cancer Res. 11, 4127–40.

Sakhi, A.K., Sabaredzovic, A., Cequier, E., Thomsen, C., 2017. Phthalate metabolites in Norwegian mothers and children: Levels, diurnal variation and use of personal care products. Sci. Total Environ. 599–600, 1984–1992. https://doi.org/10.1016/J.SCITOTENV.2017.05.109

Schlosser, P., Li, Y., Sekula, P., Raffler, J., Grundner-Culemann, F., Pietzner, M., Cheng, Y., Wuttke, M., Steinbrenner, I., Schultheiss, U.T., Kotsis, F., Kacprowski, T., Forer, L., Hausknecht, B., Ekici, A.B., Nauck, M., Völker, U., Walz, G., Oefner, P.J., Kronenberg, F., Mohney, R.P., Köttgen, M., Suhre, K., Eckardt, K.U., Kastenmüller, G., Köttgen, A., 2020. Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. Nat. Genet. 52, 167–176. https://doi.org/10.1038/S41588-019-0567-8

Stajnko, A., Runkel, A.A., Kosjek, T., Snoj Tratnik, J., Mazej, D., Falnoga, I., Horvat, M., 2022. Assessment of susceptibility to phthalate and DINCH exposure through CYP and UGT single nucleotide polymorphisms. Environ. Int. 159, 107046, 1–13. https://doi.org/10.1016/j.envint.2021.107046

Stanzick, K.J., Li, Y., Schlosser, P., Gorski, M., Wuttke, M., Thomas, L.F., Rasheed, H., Rowan, B.X., Graham, S.E., Vanderweff, B.R., Patil, S.B., Robinson-Cohen, C., Gaziano, J.M., O'Donnell, C.J., Willer, C.J., Hallan, S., Åsvold, B.O., Gessner, A., Hung, A.M., Pattaro, C., Köttgen, A., Stark, K.J., Heid, I.M., Winkler, T.W., 2021. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. Nat. Commun. 12, 4350, 1–17. https://doi.org/10.1038/S41467-021-24491-0

Takeuchi, F., McGinnis, R., Bourgeois, S., Barnes, C., Eriksson, N., Soranzo, N., Whittaker, P., Ranganath, V., Kumanduri, V., McLaren, W., Holm, L., Lindh, J., Rane, A., Wadelius, M., Deloukas, P., 2009. A genome-wide association study confirms VKORC1, CYP2C9, and CYP4F2 as principal genetic determinants of warfarin dose. PLoS Genet. 5, e1000433, 1–9. https://doi.org/10.1371/JOURNAL.PGEN.1000433

Torres, A.M., Dnyanmote, A. V., Granados, J.C., Nigam, S.K., 2021. Renal and non-renal response of ABC and SLC transporters in chronic kidney disease. Expert Opin. Drug Metab. Toxicol. 17, 515–542. https://doi.org/10.1080/17425255.2021.1899159

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., Yang, J., 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. 101, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Vogel, N., Schmidt, P., Lange, R., Gerofke, A., Sakhi, A.K., Haug, L.S., Jensen, T.K., Frederiksen, H., Szigeti, T., Csákó, Z., Murinova, L.P., Sidlovska, M., Janasik, B., Wasowicz, W., Tratnik, J.S., Mazej, D., Gabriel, C., Karakitsios, S., Barbone, F., Rosolen, V., Rambaud, L., Riou, M., Murawski, A., Leseman, D., Koppen, G., Covaci, A., Lignell, S., Lindroos, A.K., Zvonar, M., Andryskova, L., Fabelova, L., Richterova, D., Horvat, M., Kosjek, T., Sarigiannis, D., Maroulis, M., Pedraza-Diaz, S., Cañas, A., Verheyen, V.J., Bastiaensen, M., Gilles, L., Schoeters, G., Esteban-López, M., Castaño, A., Govarts, E., Koch, H.M., Kolossa-Gehring, M., 2023. Current exposure to phthalates and DINCH in European children and adolescents – Results from the HBM4EU Aligned Studies 2014 to 2021. Int. J. Hyg. Environ. Health 249, 114101, 1–11. https://doi.org/10.1016/J.IJHEH.2022.114101

Wang, I.J., Karmaus, W.J.J., 2017. Oxidative Stress-Related Genetic Variants May Modify Associations of Phthalate Exposures with Asthma. Int. J. Environ. Res. Public Health 14, 162–75. https://doi.org/10.3390/IJERPH14020162

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H., Bucan, M., 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17, 1665–1674. https://doi.org/10.1101/GR.6861907

Wang, X.X., Wang, D., Luo, Y., Myakala, K., Dobrinskikh, E., Rosenberg, A.Z., Levi, J., Kopp, J.B., Field, A., Hill, A., Lucia, S., Qiu, L., Jiang, T., Peng, Y., Orlicky, D., Garcia, G., Herman-Edelstein, M., D'Agati, V., Henriksen, K., Adorini, L., Pruzanski, M., Xie, C., Krausz, K.W., Gonzalez, F.J., Ranjit, S., Dvornikov, A., Gratton, E., Levi, M., 2018. FXR/TGR5 Dual Agonist Prevents Progression of Nephropathy in Diabetes and Obesity. J. Am. Soc. Nephrol. 29, 118–137. https://doi.org/10.1681/ASN.2017020222

Wang, Y., Zhu, H., Kannan, K., 2019. A Review of Biomonitoring of Phthalate Exposures. Toxics 7, 1–28. https://doi.org/10.3390/TOXICS7020021

Watanabe, K., Taskesen, E., Van Bochoven, A., Posthuma, D., 2017. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8, 1826–36. https://doi.org/10.1038/s41467-017-01261-5

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis, Use R! Springer Cham. https://doi.org/10.1007/978-3-319-24277-4

Williams, M.J., Wiemerslage, L., Gohel, P., Kheder, S., Kothegala, L. V., Schiöth, H.B., 2016. Dibutyl Phthalate Exposure Disrupts Evolutionarily Conserved Insulin and Glucagon-Like Signaling in Drosophila Males. Endocrinology 157, 2309–2321. https://doi.org/10.1210/EN.2015-2006

Zhu, B., Cao, A., Li, J., Young, J., Wong, J., Ashraf, S., Bierzynska, A., Menon, M.C., Hou, S., Sawyers, C., Campbell, K.N., Saleem, M.A., He, J.C., Hildebrandt, F., D'Agati, V.D., Peng, W., Kaufman, L., 2019. Disruption of MAGI2-RapGEF2-Rap1 signaling contributes to podocyte dysfunction in congenital nephrotic syndrome caused by mutations in MAGI2. Kidney Int. 96, 642–55. https://doi.org/10.1016/J.KINT.2019.03.016

Tables

Table 1. Descriptive of the study population

For continuous variables mean and SD are shown, while for categorical variables the sample size and its percentage % are reported. LMW: low molecular weight. HMW: high molecular weight. N=1,044.

Variable	N (%), Mean (SD)	N missing
Cohort		
BiB (UK)	90 (8.62%)	0
EDEN (France)	135 (12.93%)	0
KAUNAS (Lithuania)	196 (18.77%)	0
MoBa (Norway)	239 (22.89%)	0
RHEA (Greece)	186 (17.82%)	0
INMA (Spain)	198 (18.97%)	0
Child's sex		
Females	473 (45.31%)	0
Males	571 (54.69%)	0
Child's age (years)	7.96 (1.54)	0
Child's obesity		
Normal	830 (79.50%)	0
Overweight	155 (14.85%)	0
Obese	59 (5.66%)	0
Maternal education		
Primary	115 (11.02%)	0
Secondary	353 (33.81%)	0
University or higher	550 (52.68%)	0
Urinary creatinine levels (g per ml of urine)	1.004 (0.350)	0
Urinary LMW phthalate levels (μg per g of creatinine)		
MBzP	8.85 (16.14)	0
MEP	77.32 (184.98)	0
MiBP	57.48 (60.29)	0
MnBP	33.78 (37.75)	0
Urinary HMW phthalate levels (μg per g of creatinine)		
MEHP	4.67 (11.09)	34
MECPP*	53.56 (131.78)	0
MEHHP*	31.20 (84.53)	3
MEOHP*	18.80 (46.17)	1
oh-MiNP	9.09 (16.20)	0
oxo-MiNP**	5.62 (15.82)	0

*Metabolites of MEHP, **Metabolite of oh-MiNP

915

Table 2. Summary of the results of the single-trait GWAS of phthalate levels and their ratios

A locus is defined as a region of 1 Mb with SNPs with similar effect size on the trait
Suggestive statistical significance is set at p-value <1E-05

Class	Phthalate	N	Lambda	N genome-wide sig. SNPs	N genome-wide sig. loci	N suggestive sig. SNPs	N suggestive sig. loci	Minimum p-value
LMW	MBzP	1044	0.996	1	1	30	6	4.28E-08
LMW	MEP	1044	0.997	0	0	21	8	9.37E-07
LMW	MiBP	1044	1.007	0	0	62	7	1.80E-07
LMW	MnBP	1044	1.008	0	0	58	10	4.98E-07
HMW	MEHP	1010	1.000	0	0	25	11	1.39E-06
HMW	MEHHP	1041	1.018	0	0	128	15	6.17E-08
HMW	MECPP	1044	1.021	0	0	505	15	1.36E-07
HMW	MEOHP	1043	1.018	0	0	67	12	2.66E-07
HMW	oh-MiNP	1044	1.013	0	0	32	12	2.23E-06
HMW	oxo-MiNP	1044	0.987	0	0	17	5	1.33E-06
HMW ratio	MEHHP_MEHP	1009	0.986	0	0	83	11	3.62E-07
HMW ratio	MECPP_MEHP	1010	0.987	369	1	563	11	2.10E-20
HMW ratio	MECPP_MEHH P	1041	1.017	774	2	1091	11	1.45E-63
HMW ratio	MEOHP_MEHP	1009	0.986	0	0	417	12	5.72E-07
HMW ratio	MEOHP_MEHH P	1040	1.006	186	1	307	11	2.81E-14
HMW ratio	oxo-MiNP_oh-MiNP	1044	0.999	2	1	66	7	4.15E-08
	Total*			1332	6	3472	164	
	Total**					2140	158	
	Total unique*				4		117	
	Total unique**						113	

*including genome-wide significant SNPs
**without including genome-wide significant SNPs

916
917

Table 3. Selected loci associated with phthalate levels and/or ratios, ordered by p-value

A locus is defined as a region of 1Mb with SNPs with similar R2 on the trait.

The lead SNPs is the SNP with the lowest p-value in the locus. If more than one, then one is arbitrary selected and shown.

* According to the notation for the lead SNPs of the single-trait GWAS

** In brachets suggestive associations

*** After annotation with FUMA, thus including SNPs not genotyped but in linkage disequilibrium

Locus*	Phthalate**	Chr	Pos_hg19	Lead SNP	Non-effect allele	Effect allele	Beta	Standard error	P-value	N SNPs in locus***	Number independent SNPs	Nearest gene to lead SNP	Relative gene position	Other genes in locus	eQTLs	Gene-enhancer interactions	GWAS catalog	Candidate gene function
1	MECPP_MEHPP; MECPP_MEHP; MEOHP_MEHHP; (MECPP); (MEOHP_MEHP)	10	96719845	rs74494115	G	T	-0.3087	0.0171	1.45E-63	1132	26	*CYP2C9*	Intronic	*MYOF; FRA19AC1; LGI1; SLC35G1; PLCE1; NOC3L; TBC1D12; HELLS; CYP2C18; CYP2C19; CYP2C8; C10orf129; PDLIM1; SORBS1; ALDH18A1; TCTN3; ENTPD1; CCNJ; M2; others*	Several but not liver or kidney (MYOF; FRA19AC1; LGI1; SLC35G1; NOC3L; TBC1D12; HELLS; CYP2C18; CYP2C19; CYP2C8; C10orf129; PDLIM1; ALDH18A1; TCTN3; ENTPD1; M2 (in trans))	Liver *(LGI1; SLC35G1; PLCE1; NOC3L; TBC1D12; HELLS; CYP2C18; CYP2C19; CYP2C8; C10orf129; PDLIM1; SORBS1); (NOC3L; TBC1D12; HELLS; CYP2C18; CYP2C9; CYP2C8; PDLIM1; SORSB1; ALDH18A1; TCTN3; ENTPD1; CCNJ)* kidney	Serum/plasma metabolite levels; liver enzyme levels (alkaline phosphatase); drug response/metabolization (ex. warfarin); gout	***Cytochrome P450 monooxygenases:*** Catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids: ***CYP2C8*** (xenobiotics); ***CYP2C9*** (xenobiotics); ***CYP2C18*** (unknown); and ***CYP2C19*** (xenobiotics) // ***SLC35G1 (Solute Carrier Family 35 Member G1)*** : Drug/metabolite transporter protein.
2	MECPP_MEHPP; (MECPP); (MEOPH_MEHP); (MEHHP_MEHP); (oh-MiNP)	6	25809716	rs1359232	C	A	-0.1335	0.0156	4.58E-17	462	12	*SCL17A1*	Intronic	*SCGN; HIST1H cluster; SCL17A2; SLC17A3; SLC17A4; TRIM38; GPX5; GPX6 and others*	Liver *(SLC17A1; SLC17A3; SLC17A4)* ; kidney *(SLC17A3);* others	Liver *(SCGN; SLC17A1; SLC17A2; SLC17A3; SCL17A4; HIST1H cluster and others)* ; kidney *(SCL17A1; SCL17A3; SCL17A3)*	Urate levels; urinary metabolite levels; serum/plasma metabolite levels; gout; diastolic blood pressure; hematocrit; estimated glomerular filtration rate (creatinine)	***Solute Carrier Family 17 Members:*** Facilitate transport of molecules across membranes: ***SLC17A4*** (phospate excretion); ***SLC17A1*** and ***SLC17A3*** (urate excretion); ***SLC17A2*** (sialic acid transport) // ***Glutathione Peroxidase Members : GPX5*** and ***GPX6*** (reduction of hydrogen peroxide, organic hydroperoxides and lipid hydroperoxides)
3	oxo-MiNP_oh-MiNP; (MEOHP_MEHHP)	3	34882408	rs80064213	T	C	0.3307	0.0598	4.15E-08	59	2	*FECHP1*	Intergenic	*PDCD6IP; ARPP21; STAC; DCLK3*	None	None	Body mass index; breast cancer	***STAC (SH3 And Cysteine Rich Domain):*** Predicted to enable transmembrane transporter binding activity.
4	MBzP	9	134523638	rs138702233	C	T	0.3938	0.0713	4.29E-08	40	2	*RAPGEF1*	Intronic	*AIF1L; NUP214; FAM78A; PPAPDC3; PRRC2B; POMT1; UCK1; MED27; NTNG2; SETX; TTF1, C9orf171; DDX31; GTF3C4; AK8; C9orf9; TSC1; GFI1B*	Several but not liver or kidney *(AIF1L; PPAPDC3; PRRC2B; POMT1, RAPGEF1)*	Liver *(AIF1L; NUP214; POMT1; UCK1; RAPGEF1)* ; kidney *(FAM78A; POMT1; UCK1; RAPGEF1; MED27; NTNG2)*	None	***RAPGEF1 (Rap Guanine Nucleotide Exchange Factor 1):*** Human guanine nucleotide exchange factor. It is related to Nephrotic Syndrome, Type 7 and Membranoproliferative Glomerulonephritis.
5	(MEHHP); (MEOHP); (MECPP)	2	107637715	rs77376475	A	G	0.3092	0.0567	6.17E-08	2	1	*AC005040.3*	Intergenic	*SULT1C3; ST6GAL2*	Dorsolateral Prefrontal Cortex *(ST6GAL2)*	None	None	***SULT1C3 (Sulfotransferase Family 1C Member 3):*** Sulfate conjugation of endogenous compounds and xenobiotics.
6	(MEOHP_MEHHP)	1	242361140	rs10926644	T	C	-0.0522	0.0097	8.43E-08	31	2	PLD5	Intronic	*CHRM3; FH; KMO; CHML; WDR64; EXO1; MAP1LC3C; CEP170; SDCCAG8; ZBTB18; ADSS; C1orf101*	Brain *(PLD5)*	None	Response to cytadine analogues (cytosine arabinoside); total PHF-tau	***FH (Fumarate Hydratase)*** : Enzyme of the Krebs cycle that catalyzes the formation of L-malate from fumarate. // ***KMO (Kynurenine 3-Monoozygenase):*** Catalyzes the hydroxylation of L-tryptophan metabolite, L-kynurenine, to form L-3-hydroxykynurenine. // ***PLD5 (Phospholipase D Family Member 5):*** Predicted to enable catalytic activity. // ***ADSS (Adenylosuccinate Synthase 2):*** Catalyzes the first committed step in the conversion of inosine monophosphate to adenosine monophosphate.
7	(MEOHP); (MEHHP); (MECPP); (MECPP_MEHP); (MEHHP_MEHP)	18	76807781	18:76807781:G:A	A	G	0.1458	0.0288	5.00E-07	116	2	*SALL3*	Intergenic	*GALR1; HSBP1L1; PARD6G-AS1; NFATC1; ATP9B*	Several but not liver or kidney (*GALR1; HSBP1L1; PARD6G-AS1; NFATC1; ATP9B)*	Kidney and liver *(ATP9B)*	Frontotemporal dementia	***ATP9B (ATPase Phospholipid Transporting 9B - Putative) :*** Predicted to enable ATPase-coupled intramembrane lipid transporter activity.
8	(MECPP_MEHPP)	19	49250239	19:49250239:C:T	T	C	-0.0765	0.0161	2.42E-06	104	1	*IZUMO1*	Intergenic	*SLC8A2; PLA2G4C; CARD8, SULT2B1; FAM83E; SPAC4; RPL18; SPHK2; DBP; CA11; NTN5; FUT2; MAMSTR; RASIP1; FUT1; FGF21, BCAT2; HSD17B14; BAX; FTL; GYS1; RUVBL2; LHB; others*	Several tissues (*PLA2G4C; CARD8, SULT2B1; FAM83E; RPL18; SPHK2; DBP; CA11; NTN5; FUT2; MAMSTR; RASIP1; IZUMO1 - kidney; FUT1; BCAT2; HSD17B14; RUVBL2)*	Liver *(RPL18; SPHK2; DBP; CA11; MAMSTR; RASIP1; FUT1; FGF21; HSD17B14; BCAT2)* ; kidney *(FAM38E; RPL18; SPHK2; DBP; CA11; NTN5; FUT2; MAMSTR; RASIP1; FUT1; FGF21; HSD17B14; FTL; GYS1; RUVBL2; LHB)*	Paired metabolites in plasma and urine; plasma proteins; intestinal-type alkaline phosphatase levels; liver enzyme levels (gamma-glutamyl transferase); urinary albumin-to-creatinine ratio; urinary metabolites, aspartate aminotransferase levels; others	***HSD17B14 (Hydroxysteroid 17-Beta Dehydrogenase 14):*** Metabolism of steroids and also of other substrates, such as fatty acids, prostaglandins, and xenobiotics. // ***SLC8A2 (Solute Carrier Family 8 Member A2):*** Predicted to enable calcium:cation antiporter activity, involved in regulation of postsynaptic cytosolic calcium ion concentration and calcium:sodium antiporter activity. // ***SULT2B1 (Sulfotransferase Family 2B Member 1):*** Sulfate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds.
9	(MEOHP_MEHP)	17	48751114	rs72837555	C	T	-0.1472	0.0319	4.54E-06	12	1	*ABCC3*	Intronic	*CHAD; RSAD1; MYCBPAP; EPN3; SPATA20; CACNA1G; ANKRD40; LUC7L3; WFIKKN2; TOB1; NME1-NME2*	Several tissues but not liver or kidney *(SPAT20; ABCC3; ANKRD40; WFIKKN2; NME1-NME2)*	Liver *(ABCC3; ANKRD40; LUC7L3)*	Serum/plasma metabolite levels	***ABCC3 (ATP Binding Cassette Subfamily C Member 3):*** Member of the superfamily of ATP-binding cassette (ABC) transporters of molecules across extra- and intra-cellular membranes. Involved in multi-drug resistance.
10	(MEP)	19	41148101	rs2561564	T	G	-0.3124	0.0684	5.61E-06	2	1	*LTBP4*	Intergenic	*AKT2, C19orf47, PLD3, SERTAD1, SERTAD3, SHKBP1, NUMBL, ADCK4, ITPKC, C19orf54, SNRPA, MIA, MIA-RAB4B, RAB4B, RAB4B-EGLN2, EGLN2, CYP2A6, CYP2B6, CYP2A13, CYP2S1, others*	Several but not liver or kidney *(SERTAD1; SHKBP1; LTBP4; NUMBL; ADCK4; ITPKC; C19orf54, RAB4B; RAB4B-EGLN2; EGLN2)*	Liver *(CYP2A13)*	None	***Cytochrome P450 monooxygenases:*** Catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids: ***CYP2A6*** (xenobiotics); ***CYP2B6*** (xenobiotics); ***CYP2A13*** (a major nitrosamine specific to tobacco); and ***CYP2S1*** (carcinogens).
11	(MnBP)	4	70619038	rs1394032	G	C	0.18421	0.041	7.6E-06	46	1	*SULT1B1*	Intronic	*TMPRSS11BNL; TPRSS11B; UGT1B11; UGT2A1; UGT2A2; UGT2A2; SULT1B1; SULT1E1; CSN1S1; ODAM; FDCSP; CSN3; PROL1; AMTN*	Several tissues (*UGT2B11; SULT1B1 - liver; SULT1E1; CSN1S1)*	Liver *(FDCSP; PROL1; AMTN)*	Uterine fibroids	***UDP Glucuronosyltransferase Family 2 Members:*** UDP conjugation and subsequent elimination of potentially toxic xenobiotics and endogenous compounds: ***UGT2B11*** (estrogen metabolic process and xenobiotic glucuronidation); ***UGT2A1*** and ***UGT2A2*** (clearing lipophilic odorant molecules from the nasal sensory epithelium). // **Sulfotransferase Family Members:** Sulfate conjugation of many hormones, neurotransmitters, drugs, and xenobiotics: ***SULT1B1*** ; ***SULT1E1*** (transfers a sulfo moiety to and from estrone, which may control levels of estrogen receptors).
12	(MECPP_MEHPP); (MEHHP_MEHP)	8	30424915	8:30424915:A:G	G	A	-0.0851	0.0189	7.74E-06	161	1	*RBPMS*	Intronic	*INTS9; HMBOX1; DUSP4; TMEM66; LEPROTL1; MBOAT4; DCTN6; GTF2E2; SMIM18; GSR; PPP2CB; PURG; WRN*	Several tissues but not liver or kidney *(TMEM66; RBPMS; GTF2F2; SMIM18; GSR; PPP2CB)*	Liver (TMEM66; LEPROTL1; GTF2E2; SMIM18; PPP2CB; PURG); Bladder *(MBOAT4; DCTN6)*	None	***GSR (Glutathione-Disulfide Reductase):*** Central enzyme of cellular antioxidant defence, and reduces oxidized glutathione disulfide (GSSG) to the sulfhydryl form GSH, which is an important cellular antioxidant.

Figures

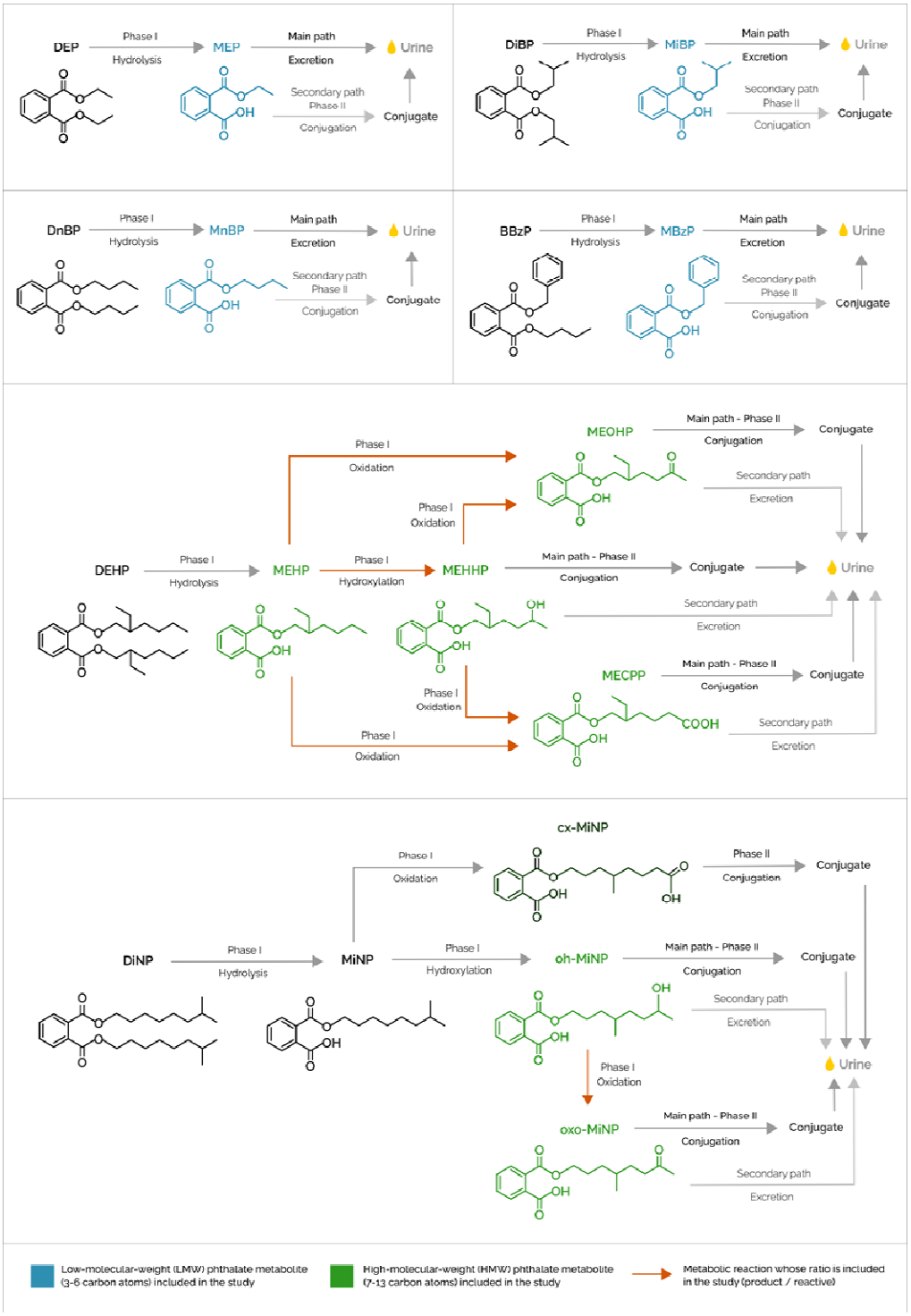


Figure 1. Summary of the phthalate metabolites and ratios included in the study.
Diagram showing the phthalate metabolites included in the study, along with their precursor phthalates and intermediate metabolites. Low-molecular-weight metabolites (LWM) in the study are coloured in blue and high-molecular weight metabolites (HMW) in the study are coloured in green. Orange arrows represent the metabolic reactions whose ratio is included in the study (product/substrate).

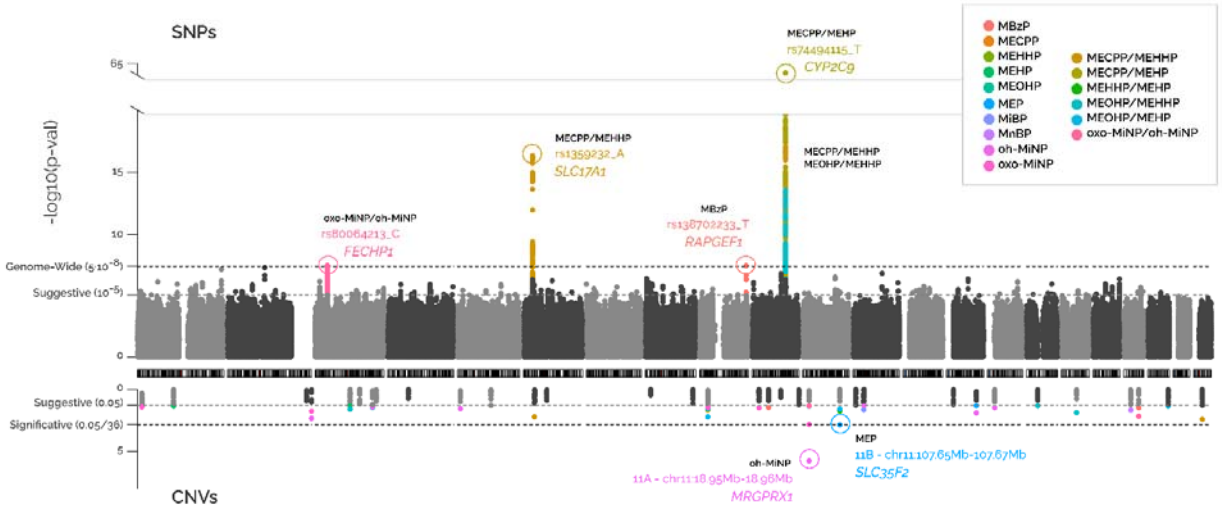


Figure 2. Miami plot of the association between common SNPs (top panel) and CNVs (bottom panel) vs phthalate levels or ratios

Each dot represents the association of a SNP or CNV. The x-axis indicates the position of the SNP or CNV in the genome. The y-axis shows the statistical significance, the –log10(p-value). Colours indicate the phthalate compound or ratio the SNPs or CNVs are associated with. The SNPs/CNVs passing multiple-testing correction (5E-08 for SNPs and 1.39E-03 for CNVs) are annotated to the closest gene.

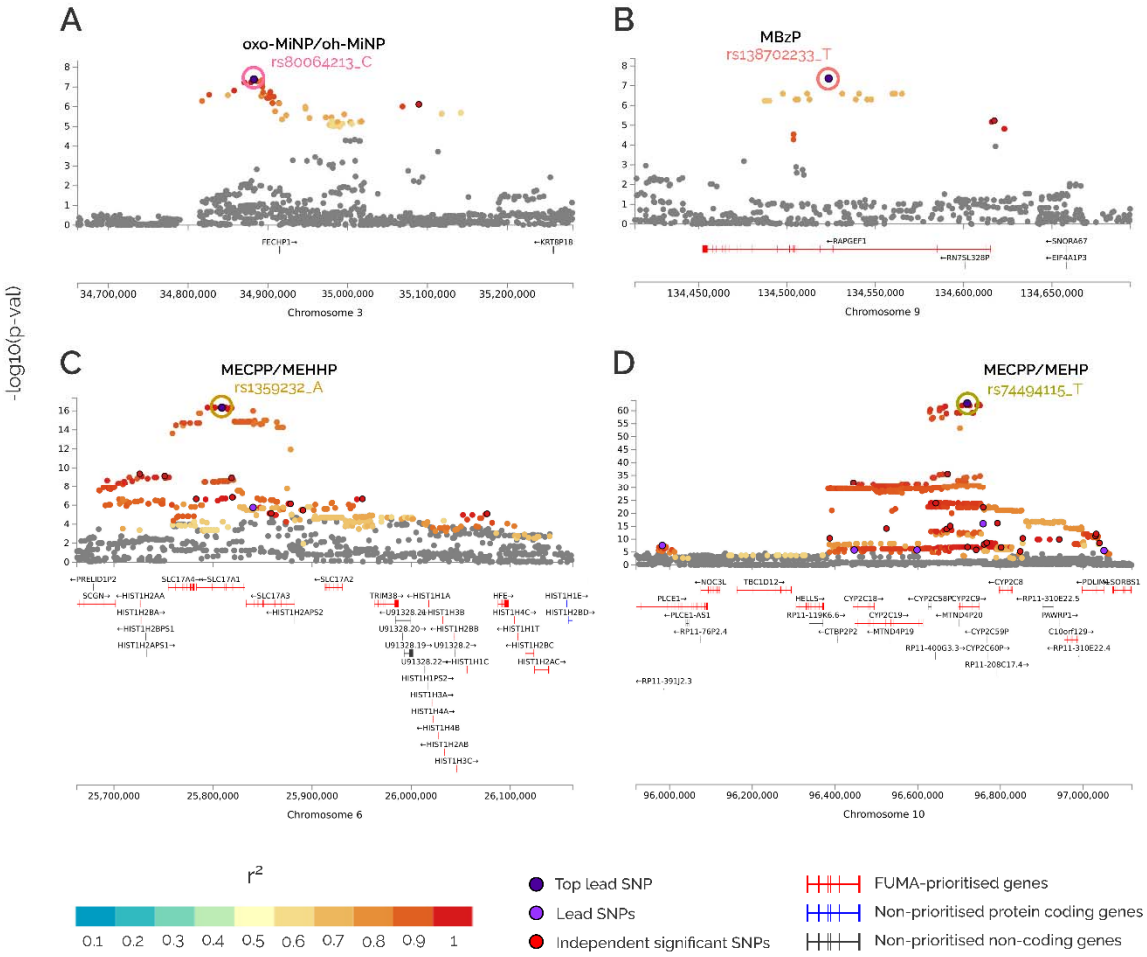


Figure 3. Locus zoom of the SNP-associated regions identified after multiple-testing correction
Each dot represents the association of a SNP. The lead SNPs are coloured in dark purple, other lead SNPs (SNPs in the region with LD <0.1 with top lead SNP) in light purple and independent significant SNPs in red. Other SNPs are coloured according to their linkage disequilibrium (r^2) with the lead SNP. The y-axis shows the statistical significance, the –log10(p-value). The x-axis indicates the position of the SNP in the genome. FUMA prioritized coding genes and FUMA non-prioritized coding and non-coding genes are shown.

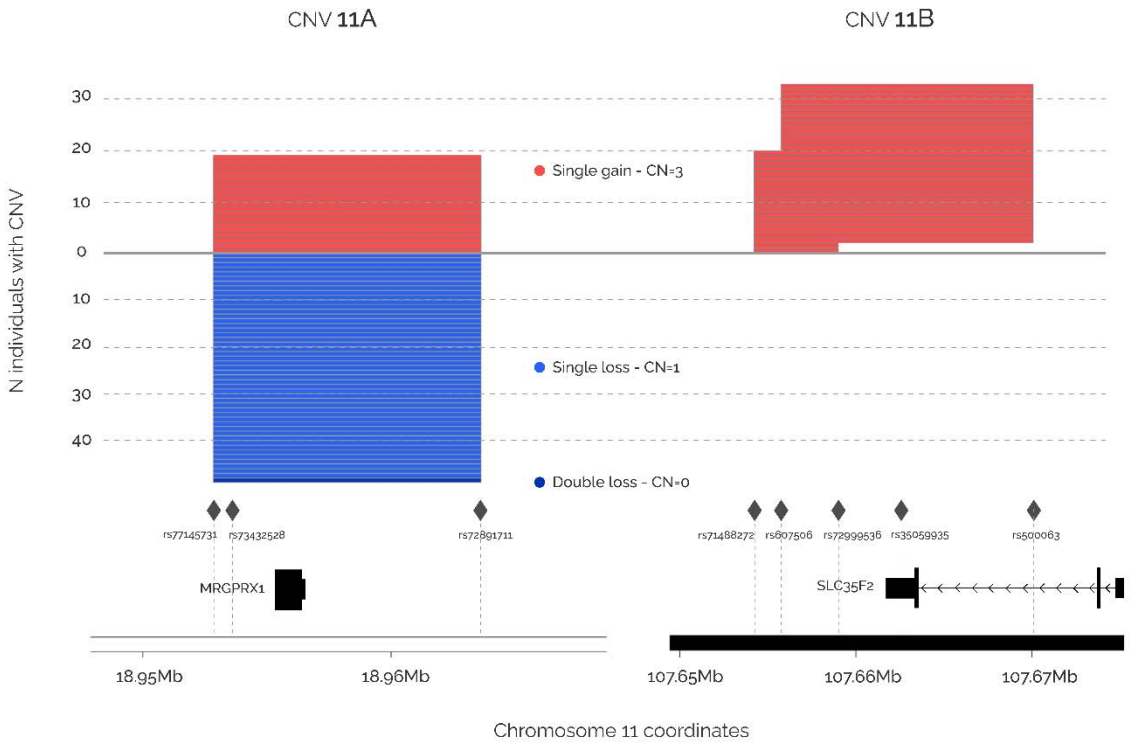


Figure 4. Locus zoom of the consensus CNV-associated regions identified after multiple-testing correction
Each line represents an CNV region in a single sample. Single gain CNVs are coloured in red, single loss CNVs in blue and double loss CNVs in dark blue. The y-axis shows the number of individuals presenting a CNV for the specific region. The x-axis shows the genomic coordinates. Rhombus represent the SNPs used to detect the CNVs, and genes in the regions are also shown.